ISSN 1831-9424

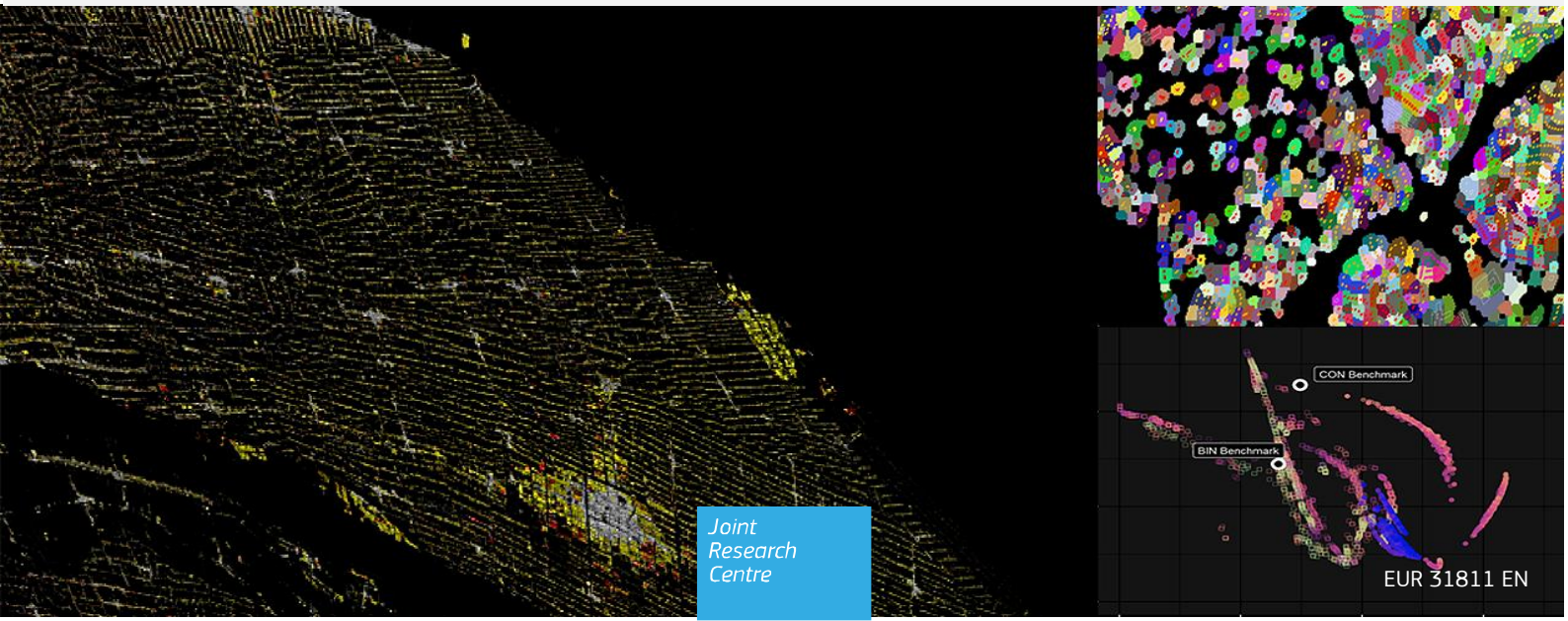
European Commission

JRC TECHNICAL REPORT

Operational procedure for multi-temporal assessment of built-up surfaces and volumes in the Global Human Settlement Layer R2023A

Pesaresi, M., Politis, P., Goch, K., Kemper, T.

2024



Joint Research Centre

EUR 31811 EN

Contact information
Name: Thomas Kemper
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
Email: thomas.kemper@ec.europa.eu
Tel.: +39 0332 78 5576

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC135315

EUR 31811 EN

PDF ISBN 978-92-68-11541-1 ISSN 1831-9424 doi:10.2760/664949 KJ-NA-31-811-EN-N

Luxembourg: Publications Office of the European Union, 2024

How to cite this report: Pesaresi, M., Politis, P., Goch, K. and Kemper, T., *Operational procedure for multi-temporal assessment of built-up surfaces and volumes in the Global Human Settlement Layer R2023A*, Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/664949, JRC135315.

Contents

Abstract

Long-term monitoring of the built-up area is essential for a number of applications. The analysis of growth trends over time allows an ex-ante assessment of policy decisions and can inform the implementation of policies like the monitoring of progress towards achieving the Sustainable Development Goals (SDG's). At the same time, they can form the basis for future projections and public discussion of sustainable development paths. So far, the joint use of Landsat and Sentinel satellite sensors for long-term built-up surface monitoring was an unsolved task in the state-of-the-art applications of remote sensing data analytics. This study introduces an integrated solution for inferring changes on built-up surfaces from Sentinel-2 MSI images (at 10 m resolution), combined with historical Landsat scenes (at 30 m and 60 m resolution), organized into four epochs 1975, 1990, 2000 and 2014. The objective of this study is two-fold. First, we aim to develop a methodology for estimating multi-temporal global built-up surfaces and volumes that allow for controlled estimates of built-up change in time in rural and urban areas. Secondly, we aim to deliver the multi-temporal assessment of global built-up surfaces and volumes with greater accuracy than in the previous Global Human Settlement Layer (GHSL) products. Our approach relies on stratified multiple-quantization associative rule learning applied to Earth Observation data, object-oriented image processing, and multiple decision support ensemble modelling. Initial assessment of our model show that the built-up surface change predictions of the proposed solution are more accurate than those reported in the previous GHSL data package (GHS-BU R2022A), as well as in other current multi-temporal estimates of built-up surface with worldwide coverage.

Acknowledgements

We would like to thank Lewis Dijkstra and the GHSL team members that contributed to the discussion on the current solution for multi-temporal estimation of built-up surfaces. The processing of the Landsat and Sentinel-2 data was carried out on the JRC Big Data Analytics Platform (BDAP).

Authors

Martino Pesaresi, Panagiotis Politis, Katarzyna Goch, Thomas Kemper

1 Introduction

The Global Human Settlement Layer (GHSL) project produces global spatial information, evidence-based analytics, and knowledge describing the human presence on the planet. The GHSL relies on the design and implementation of new spatial data mining technologies that allow automatic processing, data analytics and knowledge extraction from large amounts of heterogeneous data including global, fine-scale satellite image data streams, census data, and crowd sourced or volunteered geographic information sources. Since its first release in 2016, the GHSL data were continuously improved in quality and coverage.

The work presented here addresses for the first time a global fine-scale representation of the built-up surface changes jointly using the data collected from Landsat sensors (historical) and more recent Sentinel-2 (S2) multi-spectral instrument data. This fundamental point of joint use of Landsat and Sentinel input sensors facilitates keeping an open link between the information extracted from the new Copernicus sensor data and from the patrimony of historical remote sensing data collected since the beginning of the Earth Observation (EO) era in '70s, allowing the monitoring and the future projection of the sustainable development indicators. Such long-term monitoring of the built-up environment is essential for a number of applications including regional and urban planning, sustainable development or disaster risk reduction. The analysis of growth trends over time allows an ex-ante assessment of policy decisions and can inform the implementation of policies like the monitoring of progress towards achieving the Sustainable Development Goals (SDG's). In particular indicators of SDG 11 on sustainable cities and communities can be monitored directly with the GHSL data (Melchiorri et al. 2019; Schiavina et al. 2019). With the combined Landsat and Sentinel-2 time series and the future updates of the built-up and population layers through the exposure component of the Copernicus Emergency Management Service (Melchiorri and Kemper 2023), it will be possible to cover the full period of the 2030 Agenda for Sustainable Development from 2015 to 2030. In addition, the data can form the basis for future projections and public discussion of shared socio-economic pathways used in the context of climate scenarios (Gao and O'Neill 2020; Gao and Pesaresi 2021).

1.1 Question addressed

Commonly known challenges in the construction of global fine-scale multi-temporal land cover classification products are related to seasonal/illumination reflectance changes that must be discriminated from thematic information changes. Moreover, if "built-up surface" is the target abstraction class, and the spatial resolution of the available sensor is at the decametric scale approximating the average size of the target built-up structures, then additional challenges should be taken in to account. They are related to the fact that the collected remotely sensed data in a specific data sample is always generated by the reflection/emission of a mixture of component surfaces including different types of roof covers, neighbouring surfaces as paved surfaces or vegetation grass/ trees, plus a strong shadow component cast by the buildings that is variable depending on the illumination angle and on the building height. The JRC GHSL was pioneering the technical possibility of a global fine-scale mapping of built-up areas (M. Pesaresi, Ehrlich, et al. 2016), (Corbane et al. 2019), followed by other research efforts leveraging on the Google Earth Engine (GEE) processing platform and concentrating on using Landsat data input (Gong et al. 2020; Huang et al. 2021; 2022; Liu et al. 2020; Marconcini et al. 2021; Zhang et al. 2022).

The image data managed by the GEE has a high level of pre-processing cost (data cube), and high density of image data records, including all the 10 million of Landsat scenes collected by the Landsat satellite platform. On the contrary, the image data supporting the GHSL is low-cost, scarce and reduced, being the supporting Landsat data constituted by only 35 000 single Landsat scenes collected in arbitrary points in time (sparse set, supplemented with Global Land Survey[1] data), not organized as data cube, and being the supporting Sentinel-2 (S2) image data a radiometric *composite* of original imagery, reducing the quantity of radiometric information and injecting noise derived from not exact spatial alignment between the individual images.

The scarce and reduced characteristic of the image data supporting the GHSL demonstrates the robustness, the computational efficiency, and effectiveness of the inferential engine developed by the JRC (M. Pesaresi, Corbane, et al. 2016; M. Pesaresi, Syrris, and Julea 2016), in much more hostile data environment, as compared to the mainstream machine learning methods made available in the GEE environment. The above facts are

[1] https://www.usgs.gov/landsat-missions/global-land-survey-gls: GLS1975, GLS1990, GLS2000

framed in the overall perspective of *democratization of the information production process* from public remotely sensed data and reduction of the cost and of the environmental impact (energy consumption) of the processing infrastructure as it was set at the origin of the GHSL concept definition.

1.2 Objectives of the research/study

The first global production of decametric-scale BU surface multi-temporal information (from multi-sensor image data) was proposed by (M. Pesaresi, Ehrlich, et al. 2016), followed by subsequent GHSL data package releases (European Commission. Joint Research Centre. 2022; 2023; Florczyk et al. 2019). The GHS-BU R2022A data release was the first attempt for multi-platform multi-temporal built-up surface and volume estimation. It introduced several innovative product components, as compared to the previous release R2019:

- new observed epoch 2018 vs. 2014
- new 10 m-res vs. 30 m-res assessment of the built-up surfaces
- new sub-pixel built-up surface fraction prediction vs. Boolean class prediction
- new differentiation of residential (RES) and non-residential (NRES) built-up classes
- new 100 m-res built-up volume vs. just 250 m-res built-up surface
- new multi-temporal assessment
- multiple-sensor: Landsat MSS, TM, and Sentinel-2
- equal-time-interval grids vs. arbitrary time interval
- extrapolation to the 2025 and 2030 time points

These innovative product components were obtained by a substantial investment in research and methodological development of the JRC in the precedent years that can be summarized in the following points:

- new associative rule learning on quantized EO data based on multiple quantization minimal support schema generalization of the symbolic machine learning (SML) approach (M. Pesaresi, Syrris, and Julea 2016), supporting the S2 sub-pixel built-up fraction estimation; S2 RES/NRES class differentiation, and Landsat image classification
- new low-level textural classification based on multi-scale generalization of PANTEX (M. Pesaresi, Gerhardinger, and Kayitakire 2008), supporting the RES/NRES classification, and the hyper-dense built-up cores identification
- new low-level segmentation based on the watershed of the inverse of the saliency on multi-scale image morphological decomposition by geodesic connected components (Characteristic-Saliency-Leveling frame) (Martino Pesaresi, Ouzounis, and Gueguen 2012), supporting the object-based RES/NRES classification
- new multiple-scene SML-extracted semantic Φ composite based on divergent cumulative mechanism of logically complementary hypothesis (built-up vs. non built-up) , improving the robustness of the multi-temporal classification of historical Landsat imagery
- new built-up volumetric estimation based on the integration of Digital Elevation Models (DEMs) and S2 derived features, supporting the built-up height (ANBH, AGBH) and volume (VOL) prediction (European Commission. Joint Research Centre. 2023)
- new spatial-temporal evolution prediction schema based on rank-optimization of the composite of data-driven empirical land suitability (ELS) and data-driven built-up expansion-inertial dynamical fields (BUDYN)

Despite the high accuracy of GHSL R2022A in single epochs, a positive bias was observed in predicted change rates of built-up surfaces and built-up volumes after the year 2000, especially in rural areas (domain as set by the GHS-SMOD R2022A).

The observed bias stems for the need for both maximal accuracy of the built-up grids and the built-up change rates measured on admin spatial units, required for reliable estimation of policy indicators (SDG11 for example). However, the inner trade-offs between predicting 100 m built-up surface grids predicted at the different epochs, and in predicting the aggregated (by administrative units) built-up change rates, may conflict – due to prediction bias components at different scales. For example, at the given data conditions, the maximization of the change grid accuracy led to extremely conservative aggregated change rates, thus driving the system toward unrealistically static change rates unsuitable for supporting monitoring of the policy indicators. This implies that of a satisfactory solution should be determined by observation of more than single performance measure in the ensemble of model choices. The observed limitation in the prediction of change rates in GHSL R2022A, coupled with the unprecedented availability of the new multi-temporal building footprints (MTBF) reference

data, was the driving force behind research into improving the prediction in built-up surface change, finalized into the new data release (European Commission. Joint Research Centre. 2023). The main objectives of the study are summarized below:

1. Assess the performances of the prediction of built-up surface change as provided by the available prior knowledge, further referred to as the "priors" (Corbane et al. 2019; Gong et al. 2020; Huang et al. 2021; 2022; Liu et al. 2020; Marconcini et al. 2021; M. Pesaresi, Ehrlich, et al. 2016; Zhang et al. 2022), and in comparison with the previous R2022A GHSL data release.
2. Experiment the possibility to improve the capacity to discriminate built-up/non built-up semantic abstraction from historical Landsat satellite data by changing the train sampling mechanism, by stratification of the problem solving in different domains potentially exposed to specific information retrieval biases, by changing the mechanism of downscaling the predicted historical Landsat Φ to the new S2 geometry, and by changing the mechanism of phi maximization across the different Landsat satellite scenes belonging to the same epoch or different epochs.
3. Experiment the possibility to compose the multiple-time multiple-strata multiple-model change predictions in to an ensemble prediction and measure the performances as compared to the previously available options in the URBAN vs. RURAL application domains, as set by the GHS-SMOD R2022A (and referred to as URBAN and RURAL henceforth). In the frame of this objective, two sub-experiment were performed a) a binary (BIN) change detection schema and b) a continuous (CON) change detection schema.
4. Experiment the possibility to select an independent ensemble model for given change epoch with a Pareto multi-objective optimization, combining the distance between the predicted and the observed change grids, the change rate error in the urban application domain and the change rate error in the RURAL application domain.
5. Answer to the pragmatic question if the R2022A observed built-up change bias could be reduced by improvements of the processing methods, and if yes by using which data support for the automatic decision: a) any of the priors already available, b) a composite of the priors, c) the scarce Landsat data supporting in the GHSL infrastructure, or d) the scarce Landsat data supporting in the GHSL infrastructure augmented by the priors.

1.3 Novelty

The study on multi-temporal estimation of built-up surfaces described in this report, resulting in the production of data release R2023A, improves the prediction of the spatial-temporal patterns of change in the built-up surfaces, by extending the processing workflow with the following main packages:

- extension of the reference database with new multi-temporal building footprint (MTBF) reference data
- semantic and spatial harmonization, quality control, filtering, identification of spatial-temporal data valid conditions in the MTBF
- new design, development, and validation of the methods supporting the Φ semantic extraction from Landsat image data (neighbouring spatial equalized sampling, stratified learning, temporal interlaced maximization)
- new design, development, and validation of the methods supporting the composite of the multiple-sensor, multiple-time, multiple-model, multiple-decision support, multiple-strata predictions to the final ensemble model prediction
- introduction of the existing multi-temporal estimates of the built-up surface (priors) to the workflow, new empirical test of the priors, combination of priors, and Φ semantic extraction from Landsat data models
- multiple decision support ensemble approach for model evaluation and selection.

2 Data

2.1 Earth Observation 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.

Landsat Imagery

The Landsat image data used in input include 35 479 individual scenes organized in four epochs 1975, 1990, 2000, and 2014. The image data resume the whole history of civilian Remote Sensing, including all Landsat missions from L1 to L8, and four different sensors: MSS, TM, ETM, and OLI, with substantially different technical specifications[2]. The epochs 1975, 1990, 2000, and 2014 are dominated by image data collected by the MSS, TM and ETM sensors, respectively. The average absolute time tolerance of the image data collection time vs. the nominal time barycentre of the epoch is 2.4, 2.7, 1.8, and 0.9 years for the 1975, 1990, 2000, and 2014 epochs, respectively. The aggregated time precision of all the data in the four epochs is of 2 years. The empirical time barycentre for the epochs 1975, 1990, 2000, and 2014 is the year 1975.1, 1989.4, 2000.8, and 2009.7, respectively (Table 1).

Table 1 – Summary of the Landsat image data

Epoch	Sensor	Mission	Count of scenes	Average year of scenes	Standard deviation of year of scenes
1975	MSS	-	7 355	1975.1	2.4
		L1	3 495	1973.1	1.0
		L2	3 099	1976.5	1.6
		L3	750	1978.8	0.8
		L4	11	1982.0	0.0
1990	-	-	8 011	1989.4	2.7
	MSS	-	242	1983.9	1.1
		L3	3	1983.0	0.0
		L4	132	1983.1	0.3
		L5	107	1985.0	0.8
	TM	-	7 769	1989.6	2.6
		L4	1 312	1989.0	1.3
		L5	6 457	1989.7	2.7
2000	-	-	9 774	2000.8	1.8
	ETM	L7	9 276	2000.6	1.4
	TM	L5	498	2004.1	3.8
2014	-	-	925	2009.7	0.9
	ETM	L7	259	2009.5	0.5
	OLI	L8	28	2013.5	0.5
	TM	L5	638	2009.6	0.6
Total	-	-	26 065		

Sentinel-2 imagery

The epoch 2018 is made by the GHS_composite_S2_L1C_2017-2018_GLOBE_R2020A ([3]) that corresponds to global cloud-free pixel based composite created from the Sentinel-2 data archive (Level L1C) available in Google Earth Engine ([4]) for the period January 2017 - December 2018.

---

[2] For more details see: https://www.usgs.gov/landsat-missions/landsat-satellite-missions

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

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

2.2 Priors

In January 2008 Barbara Ryan, the Associate Director for Geography at the U.S. Geological Survey, and Michael Freilich, NASA's Director of the Earth Science Division, signed off a Landsat Data Distribution Policy that made Landsat images free to the public[5]. The USGS announced the free-and-open data policy on April 21, 2008. Since the pioneering work of the JRC GHSL (M. Pesaresi and Ehrlich 2009) (M. Pesaresi et al. 2013) (M.Pesaresi 2014) some global, fine-scale and multi-temporal assessment of the built-up surface have been produced using the public Landsat image data. They are listed in chronological order below:

GHS_B_P2016 : "Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014" (M. Pesaresi, Ehrlich, et al. 2016)

GHS_B_P2019 : "Automated global delineation of human settlements from 40 years of Landsat satellite data archives" (Corbane et al. 2019)

GAIA : "Annual maps of global artificial impervious area (GAIA) between 1985 and 2018" (Gong et al. 2020)

GAUD : "High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015" (Liu et al. 2020)

GISA : "30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019" (Huang et al. 2021)

WSF_EVO : "World Settlement Footprint Evolution 1985-2015" (Marconcini et al. 2021)

GISA2 : "Toward accurate mapping of 30-m time-series global impervious surface area (GISA2.0)" (Huang et al. 2022)

GISD30 : "Global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform" (Zhang et al. 2022, 30)

Those satellite-data-derived land cover classifications describe the temporal evolution of class abstraction semantics that are not necessarily consistent ([6]) with data classification methodologies that are very different: nevertheless, they may be considered in principle as largely overlapping and all highly correlated with the semantic notion of "built-up surface" as expressed in the new GHSL data (European Commission. Joint Research Centre. 2022). Moreover, all the non-GHSL sources have been produced in 2020-2022 with the support of the whole Landsat archive (~10 million of scenes), in contrast with the poverty of the GHSL historical image data support relying only on a limited archive of ~35 000 Landsat scenes. Therefore, those global multi-temporal land cover products represent prior knowledge that may be potentially valuable in support to various processing tasks of the new GHSL MT data. They include: training set sampling, adaptive learning, data fusion and data gap filling.

2.3 Multi-temporal building footprints (MTBF)

The multi-temporal building footprints (MTBF) were collected from three main sources: the Geographic Information System of the Commission (GISCO) portal, JRC autonomous search in the city administration open data, and the MTBF33 project (Uhl and Leyk 2022). MTBF data are vector datasets of building footprints with date of construction assigned. The original sources have been subdivided by small administrative units (NUT3 in Europe, County in US) in order to increase the number of test cases and to be able to test the variability of the average performance metrics. Figure 1 shows the distribution of available MTBF data tested in the study.

[5] https://www.usgs.gov/media/files/imagery-everyone-timeline-open-landsat-archive

[6] They include "built-up areas" GHS, "artificial impervious surfaces" GAIA, "impervious surface area" GISA & GISA2, "urban area" GAUD, and "settlement footprint" WSF_EVO.

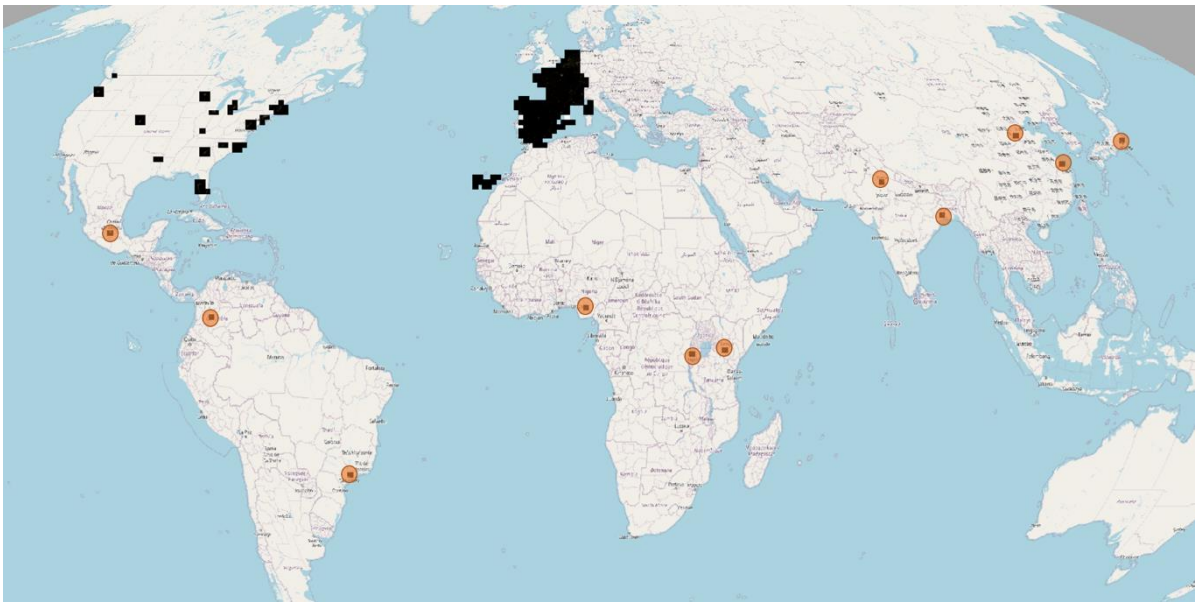


Figure 1 – Test data tiles supporting the study. Black: data tiles where MTBF data is available, Red circle: out-of-the-sample data tiles in worst case scenarios processed for visual inspection.

Due to the uncertainties related to their actuality (up-to-datedness) and completeness, the collected MTBF were not suitable for testing all the epochs considered. For each epoch, we included in testing the MTBF samples (100 m x 100 m grids) that were passing the suitability criteria:

- Minimum construction year greater or equal 1000 – to eliminate samples with implausible construction dates;
- Maximum construction year lesser or equal 2022 – to eliminate samples with implausible construction dates, greater than the date of the MTBF production;
- 99.5 percentile of construction years greater or equal the considered epoch – to ensure the observation of change in the number of buildings in the considered epoch.
- Exclusion of samples including building footprint data with no attribute of the epoch of construction (NODATA exclusion)

Table 2 shows the number of MTBF test cases by source and by Country, classified regarding the suitability to support the assessment of specific epochs. As can be noticed, the availability of test cases is maximal in the epochs 1975, 1990, and 2000, decreases in the epoch 2014, and is minimal in the 2018 epoch. Analogue observation can be done on the number of valid data test samples at 100 m-res in the different epochs (Table 3). The whole MTBF data passing the suitability criteria is representative of all the settlement classes as defined in the degree of urbanization level 2 and stratified using GHS settlement layers GHS-SMOD R2022A (SMOD L2) (Table 4).

Table 2 – Number of MTBF test cases by source and by Country, classified regarding the suitability to support the assessment of specific epochs, depending on the conditions of update and completeness of the MTBF data.

Source	**Country**	**N of MTBF test cases 1975**	**N of MTBF test cases 1990**	**N of MTBF test cases 2000**	**N of MTBF test cases 2014**	**N of MTBF test cases 2018**
GISCO	FRANCE	677	677	677	677	661
	SPAIN	96	96	96	87	9
JRC	Netherlands	225	206	192	192	95
	Switzerland	8	8	8	8	8
	USA	74	74	74	49	4
MTBF33	USA	203	203	203	7	0
Total		**1283**	**1264**	**1250**	**1020**	**777**

Table 3 – Number of 100m-res samples supported by the MTBF test cases by source and by Country, in test cases classified regarding the suitability to support the assessment of specific epochs, depending on the conditions of update and completeness of the MTBF data.

Source	Country	≤1975	≤1990	≤2000	≤2014	≤2018	≤2020	Total
GISCO	FRANCE					10 3876	6 895 379	6 999 255
	SPAIN			4 904	220 828	500 805	18 482	745 019
JRC	Netherlands	86 438	52 854			506 984	538 720	1 184 996
	Switzerland						48 408	48 408
	USA			342 824		441 678	8 141	792 643
MTBF33	USA			1 948 617	112 821			2 061 438
Total		**86 438**	**52 854**	**2 296 345**	**333 649**	**1 553 343**	**7 509 130**	**11 831 759**

Table 4 – Total number of 100m-res valid samples supported by the MTBF data, by SMOD L2 application domain strata

Source	Country	Mostly uninhabited area (11)	Dispersed rural area (12)	Village (13)	Suburbs or peri-urban area (21)	Semi-dense town (22)	Dense towns (23)	City (30)	Total
GISCO	FRANCE	1 043 128	1 547 495	254 748	341 388	131 136	55 534	125 989	3 499 418
	SPAIN	79 814	136 380	21 167	67 428	23 459	12 237	31 988	372 473
JRC	Netherlands	19 228	183 947	41 394	145 394	44 273	51 831	104 717	590 784
	Switzerland	689	4 809	864	8 039	108	2 262	7 433	24 204
	USA	34 369	105 884	5 950	155 930	6 642	19 124	68 167	396 066
MTBF33	USA	98 595	300 094	20 458	175 000	40 133	37 674	158 450	1 030 404
Total		**1 275 823**	**2 278 609**	**344 581**	**1 093 179**	**245 751**	**178 662**	**496 744**	**5 913 349**

As expected, the number of samples supporting the epoch 2018, and consequently the change maps 2018 vs. all the other epochs (2014, 2000, 1990, and 1975) are much smaller than the total number of samples, but still representing all the settlement patters in the L2 SMOD strata, degree of urbanization level 2 (Table 5). As a result, MTBF samples selected for the model development cover mostly the area of France, partially Spain, Netherlands, Switzerland, and an individual county from US (Figure 2). The selected samples cover mostly rural areas, with many samples of the urban centres excluded, as no significant change was observed there since 1975.

Table 5 – Total number of 100m-res valid samples supported by the MTBF data and suitable for the assessment of the epoch 2018, by SMOD L2 application domain strata

Source	Country	Mostly uninhabited area (11)	Dispersed rural area (12)	Village (13)	Suburbs or peri-urban area (21)	Semi-dense town (22)	Dense towns (23)	City (30)	Total
GISCO	FRANCE	1 018 067	1 527 797	251 766	339 790	130 219	54 825	125 016	3 447 480
	SPAIN	1 382	4 040	2 204	245	535	173	652	9 231
JRC	Netherlands	8 641	97 613	18 813	68 811	24 929	21 759	28 794	269 360
	Switzerland	689	4 809	864	8 039	108	2 262	7 433	24 204
	USA		244		2 873		953		4 070
Total		**1 028 779**	**1 634 503**	**273 647**	**419 758**	**155 791**	**79 972**	**161 895**	**3 754 345**

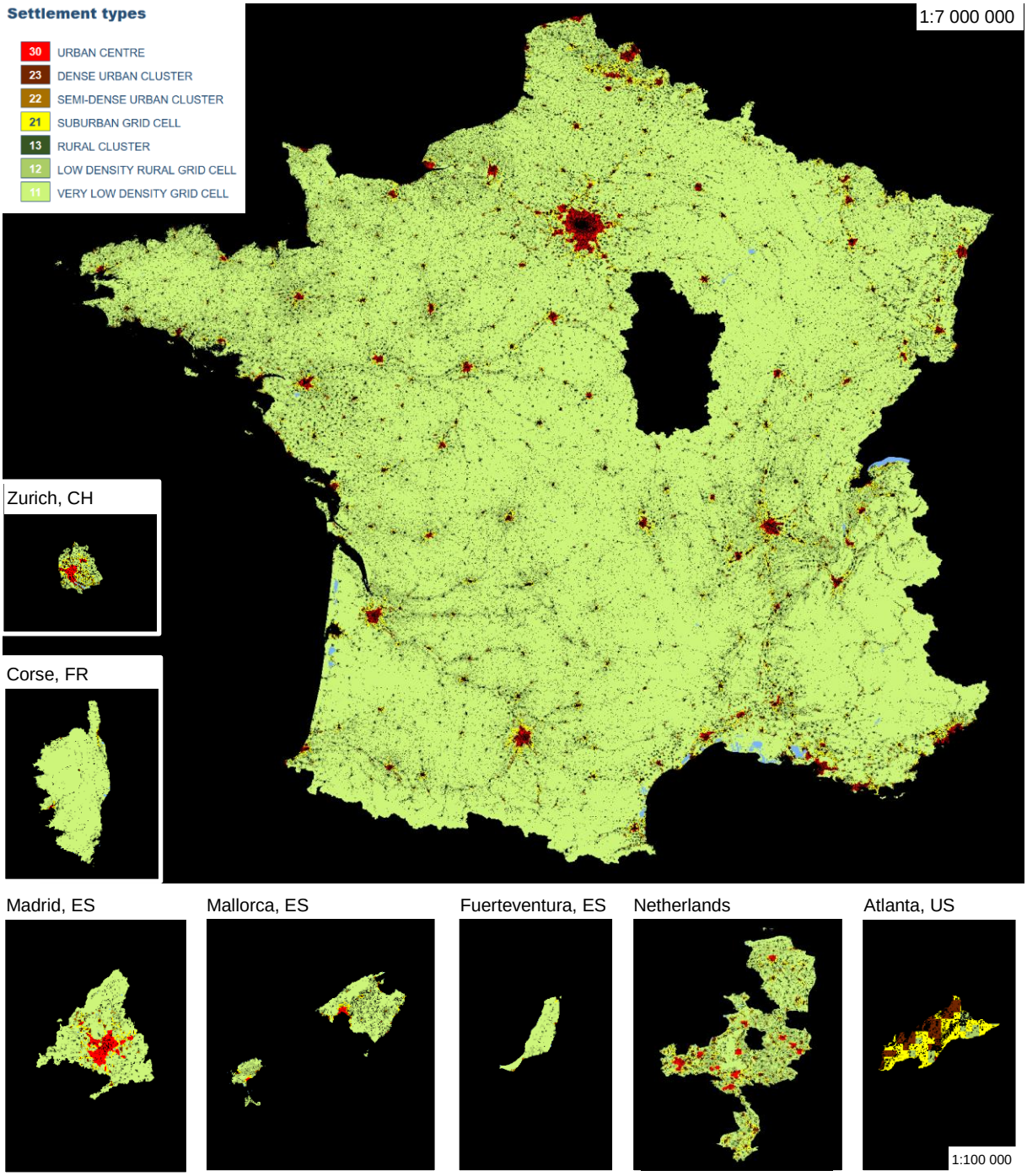


Figure 2 – Valid 100 m x 100 m MTBF samples, suitable for the assessment of the epoch 2018 and the change maps 2018 vs all the other epochs (2014, 2000, 1990, and 1975). The colour background represents SMOD classification. Noteworthy, samples selected do not cover cores of urban centre SMOD class

3 Methods

3.1 Problem setting

Jointly using the data collected from Landsat sensors and more recent S2 multi-spectral instrument data is subject to issues arising from their different technical capabilities, such as completeness, accuracy and resolution. As a result, the unsupervised change detection techniques, used in the earlier GHS multi-temporal estimations, are particularly challenging for scattered built-up patches, where the empirical probability to be associated to the NBU semantic background is systematically higher than the probability to be associated to the BU semantic foreground. This systematic empirical bias is a consequence of the limited spatial/spectral resolution capacity of the historical sensors supporting the MT assessment, as compared to the resolution necessary for supporting the discovery of the built-up surfaces in the year 2018. As noticed in testing which was observed for R2022A product, the unsupervised change detection technique is exposed to the risk of producing unrealistic high change rate of built-up surfaces, especially in rural areas connoted by scattered settlement patterns.

Increased access to MTBF data has allowed to re-set the problem in a supervised change detection schema allowing to minimize the errors induced by systematic bias and gain of historical Landsat EO data as compared to the new S2 image data injected in the system. Moreover, the increased amount of MTBF allowed for inferring systematic relations between remote sensing data and semantic information in multiple observation contexts or information STRATA. Thus, the new model is based on the composite of stratified minimization of the empirical risk, keeping the inferential engine working by associative rule learning applied to the combination (sequence) of quantized remotely sensed data, but generalizing the previous approach from the single-model-solution to the ensemble-model-solution, potentially allowing to increase both accuracy and robustness of the change detection.

3.2 General overview of the new solution

We build the time-series of the built-up surface and volume information in equal time (5-year) intervals from 1975 to 2030. Spatial-temporal interpolation was applied in order to transform the predictions of the amount of BU surface in the global spatial grids corresponding to arbitrary epochs in the time domain [1975, 1990, 2000, 2014, 2018] in equal-time interval 1975:5:2030. Equal-time interval information grids are supposed to facilitate the use of the GHSL data as input of causal models or future predictive models using explicit temporal variables.

3.2.1 ***Master workflow***

The general logic followed for the production of the multi-temporal GHSL data is the following:

a) Extract high-level semantic from the best image data available for the study: Sentinel2 image composite, 10m-res year 2018
   - Sub-pixel built-up surface fraction TOTAL (sum in the 100m sample): $f_x BU^S_{TOT}$
   - Non-residential (NRES) land use classification (surface share in the 100m sample) : $f_x NRES$
   - Average of the net building height (ANBH) prediction (100m sample): $f_x ANBH$

b) Extract logical derivative of the BU high-level semantic
   - Non-residential built-up surface : $f_x BU^S_{NRES} = f_x BU^S_{TOT} \cdot f_x NRES$
   - Residential built-up surface: $f_x BU^S_{RES} = f_x BU^S_{TOT} \cdot (1 - f_x NRES)$
   - Total built-up volume: $f_x BU^V_{TOT} = f_x BU^S_{TOT} \cdot f_x ANBH$
   - Residential built-up volume: $f_x BU^V_{RES} = f_x BU^S_{RES} \cdot f_x ANBH$
   - Non-residential built-up volume: $f_x BU^V_{NRES} = f_x BU^S_{NRES} \cdot f_x ANBH$

c) Extract the BU surface persistence factor (100m sample) in the arbitrary points in time (GHSL epochs) where historical image data was available : $\beta_{t \in [1975,1990,2000,2014]}$

d) Project the BU high-level semantic in the past epochs, under the assumption BU high-level semantic varying proportionally the change of the supporting BU surface, thus:

- $f_x BU^{S}_{RES\ t\in[1975,1990,2000,2014]} = \beta_{t\in[1975,1990,2000,2014]} \cdot f_x BU^{S}_{RES\ t=2018}$
- $f_x BU^{S}_{NRES\ t\in[1975,1990,2000,2014]} = \beta_{t\in[1975,1990,2000,2014]} \cdot f_x BU^{S}_{NRES\ t=2018}$
- $f_x BU^{V}_{RES\ t\in[1975,1990,2000,2014]} = \beta_{t\in[1975,1990,2000,2014]} \cdot f_x BU^{V}_{RES\ t=2018}$
- $f_x BU^{V}_{NRES\ t\in[1975,1990,2000,2014]} = \beta_{t\in[1975,1990,2000,2014]} \cdot f_x BU^{V}_{NRES\ t=2018}$

e) Build equal-time interval BU high-level semantics in the 1975:5:2030 range by spatial-temporal interpolation and extrapolation of the corresponding predictions in the 1975, 1990, 2000, 2014 points in time.

The general logic of the proposed solution is summarized at Figure 3: assumed solved the prediction of built-up surface in the year 2018 from S2 image composite by other independent model predicting the continuous built-surface function in the spatial grid *x* $f_x BU_{t=2018}$, the historical Landsat imagery is used to predict the persistence factor β_t (inverse of change) to be applied to the function $f_x BU_{t=2018}$ in order to predict the other observed epochs 2014, 2000, 1990, and 1975.

$$f_x BU_{t\in[1975,1990,2000,2014]} = \beta_{t\in[1975,1990,2000,2014]} * f_x BU_{t=2018}$$

Equation 1 - persistence factor β_t: basic mechanism to predict change

β_t = 1 means that $f_x BU_t = f_x BU_{t=2018}$ at the specific spatial sample *x* and point in time *t*, while β_t = 0 means that $f_x BU_t = 0$ at the specific spatial sample *x* and point in time *t*.

In order to mitigate issues related to different image sensor resolution and raster grid misalignment in the input data, the factor β_t is estimated in a multi-step approach involving image segmentation at 10m-res, and subsequent aggregation to the final 100m-grid.

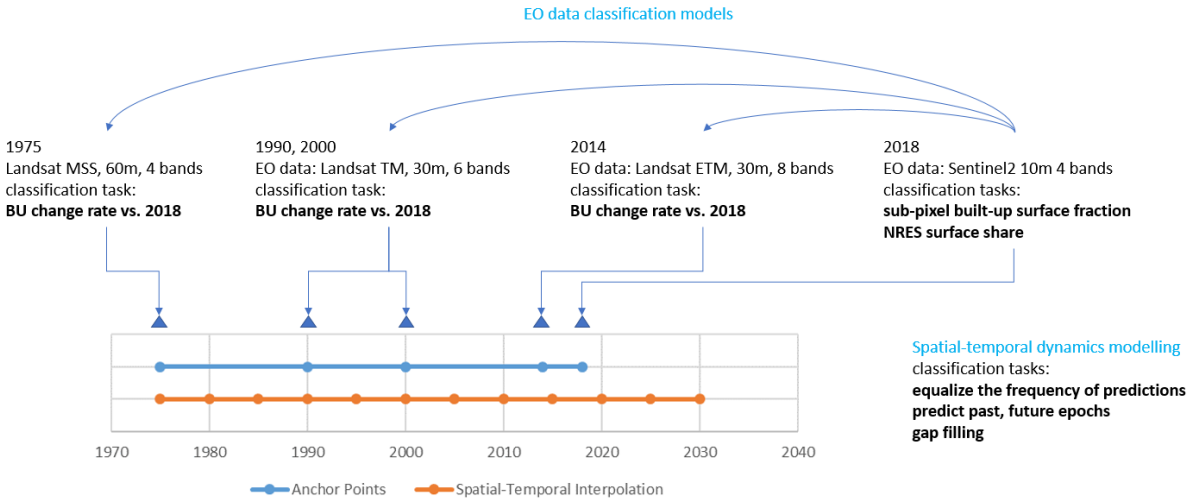


Figure 3 – General Logic of the proposed solution

3.2.2 *Extracting semantic Φ from Landsat data*

The semantic Φ is the measure of the association between a combination of quantized EO data collected in different wavelengths (called a "data sequence") and a given semantic, e.g the BU, the NBU classes. In the approach discussed here, the semantic Φ is extracted from EO data using symbolic machine learning or SML (M. Pesaresi, Syrris, and Julea 2016). In particular, a generalization of the SML to the multiple-quantization (MQ) is adopted. In the MQ SML a hierarchical pyramid of inferences is done by iterative application of single quantization SML in a given list of quantization parameters, solved by increasing order. Thus, the hierarchical inferential pyramid is scanned from the bottom to the top (from small to large quantization parameters), and the first SML inference with a minimal empirical support is selected. In this approach, two basic hyper-parameters are governing the function of the classifier a) the vector of quantization parameters Q^l and b) the minimal support *MinSupp*. In the application discussed here the two hyper-parameters of the MQ SML were set as

$$Q^l = 2^{[0..6]} = [1, 2, 4, 8, 16, 32, 64]$$

$$MinSupp = 10$$

They are considered as constant across the whole volume of image data processed (all sensors, all epochs), and assuming the quantized digital number (*DN*) at the sample *x* of the EO data at the level *l* of the quantization pyramid calculated as

$$\boxed{DN_x^{l\in 1..7}} = round(DN_x / Q^{l\in 1..7})$$

In the application discussed here, the direct digital numbers as collected by the EO sensors are used in input and quantized for the SML encoder, without the necessity to apply external models and assumptions for atmospheric radiance transfer and reflectance calibration. This is possible because continuous machine learning is done through low-computational cost SML processing, thus no model transfer is needed from one satellite scene to another.

Different EO sensors were adopted in the various Landsat missions operational from 1975 to 2020. They can be grouped in four main cases: the Multi Spectral Scanner (MSS), the Thematic Mapper (TM), the Enhanced Thematic Mapper (ETM), and the most recent Operational Land Imager (OLI).

Table 6- Table 9 shows their main technical characteristics and which spectral bands were used in input of the SML for producing the inference of the semantic Φ. As general rule, the maximum amount of spectral data available from the specific sensor was used, with the only exception of the thermal infrared (TIR) and the panchromatic (Pan) data. In particular, Blue, Green, Red, and near infra-red (NIR) bands are always used in input because available since the first MSS sensor. TM, ETM sensors introduced the short-wave infrared (SWIR) bands that were included in the SML input data encoder as well. Finally, OLI sensors added the new coastal aerosol and the cirrus bands that were included in the SML input data encoder as well.

Accordingly, the SML input data encoder was built around 4-elements quantized data sequences for the MSS data, 6-elements quantized data sequences for the TM and ETM data, and 8-elements quantized data sequences for the OLI data, with a constant $Q^l = 2^{[0..6]}$, $MinSupp = 10$ hyper-parameter set. The whole Landsat data by sensor and epoch are listed in **Error! Reference source not found.**.

Multi Spectral Scanner (MSS)

Band # (L1-L2)	Band # (L3)	Band # (L4-L5)	wavelength µm	Resolution*	L4/L5 TM Band Equivalent		SML input
4	4	1	0.5-0.6	68 m X 83 m	~ 2 (0.52–0.60 µm)	Blue	yes
5	5	2	0.6-0.7	68 m X 83 m	~ 3 (0.63–0.69 µm)	Green	yes
6	6	3	0.7-0.8	68 m X 83 m	~ 4 (0.76–0.90 µm)	Red	yes
7	7	4	0.8-1.1	68 m X 83 m	~ 4 (0.76–0.90 µm)	NIR	yes
N/A	8	N/A	10.4-12.6	68 m X 83 m	~ 6 (10.41-12.5 µm)	TIR	no

Table 6 – Spectral bands of the MSS sensor used in input of the SML inference. Source: https://landsat.gsfc.nasa.gov/multispectral-scanner/

Thematic Mapper (TM)

Band #	wavelength µm	Resolution*		SML input
1	0.45-0.52	30 m	Blue	yes
2	0.52-0.60	30 m	Green	yes
3	0.63-0.69	30 m	Red	yes
4	0.76-0.90	30 m	NIR	yes
5	1.55-1.75	30 m	SWIR-1	yes
6	10.41-12.5	120 m	TIR-1	no
7	2.08-2.35	30 m	SWIR-2	yes

Table 7 – Spectral bands of the TM sensor used in input of the SML inference. Source: https://landsat.gsfc.nasa.gov/thematic-mapper/

Enhanced Thematic Mapper (ETM)

Band # (L1-L2)	wavelength µm	Resolution*		SML input
1	0.45-0.515	30m	Blue	yes
2	0.525-0.605	30m	Green	yes
3	0.63-0.69	30m	Red	yes
4	0.775-0.90	30m	NIR	yes
5	1.55-1.75	30m	SWIR-1	yes
6	10.4-12.5	60m	TIR-1	no
7	2.08-2.35	30m	SWIR-2	yes
8	0.52-0.9	15m	Pan	no

Table 8 – Spectral bands of the ETM sensor used in input of the SML inference. Source: https://landsat.gsfc.nasa.gov/etm-plus/

Operational Land Imager (OLI)

Band #	wavelength µm	Resolution*		SML input
1	0.433–0.453	30 m	coastal aerosol	yes
2	0.450–0.515	30 m	blue	yes
3	0.525–0.600	30 m	green	yes
4	0.630–0.680	30 m	red	yes
5	0.845–0.885	30 m	NIR	yes
6	1.560–1.660	30 m	SWIR-1	yes
7	2.100–2.300	30 m	SWIR-2	yes
8	0.500–0.680	15 m	PAN	no
9	1.360–1.390	30 m	cirrus	yes
10	10.6-11.2	100 m	TIR-1	no
11	11.5-12.5	100 m	TIR-2	no

Table 9 – Spectral bands of the OLI sensor used in input of the SML inference. https://landsat.gsfc.nasa.gov/satellites/landsat-8/spacecraft-instruments/operational-land-imager/

3.2.3 *Predicting semantic changes from satellite data*

Conceptual design of the new solution comprises of the three main steps: 1/ extraction of semantic Φ from Landsat imagery, 2/ composition of the decision on the image segment level and 3/ aggregation of the prediction to the raster grid and evaluation (Figure 4). Sematic Φ associated to built-up (BU) and non-built-up (NBU) abstraction classes Φ_{BU} Φ_{NBU} is extracted from Landsat imagery organized in four epochs (1975, 1990, 2000, and 2014). This Φ is encoded as raster data respecting the arbitrary geometry (resolution, origin of the grid, projection) of each specific input image. This fact ensures the minimal degradation of the original satellite

image data, thus the maintenance of the maximum possible signal-to-noise ratio in the subsequent automatic image information retrieval tasks. Successively, the Φ_{BU} Φ_{NBU} is downscaled to data segments (Figure 13) derived from the built-up fraction predicted from the S2 image composite of 2018, at 10 m resolution, Mollweide (MWD) projection that is adopted in the final GHSL products. The downscale from pixel of arbitrary origin and different size to 10m-res image segments is done by the surface-weighted average operator. The Φ_{BU} and Φ_{NBU} are estimated for each satellite scene collected at a specific point in time at the level of the data segments. Subsequently, the Φ_{BU} and Φ_{NBU} are composed by epoch and by spatially overlapping scenes, using different optimization approaches with the objective of minimizing the effect of image changes induced by seasonal changes and/or non-stationary noise in the input image data. The temporal composite-by-optimization of the Φ_{BU} and Φ_{NBU} hypothesis at the temporal point t is noted as $\overline{\circledast \Phi}_t$. Non-stationary noise include cloud obstructions or cloud shadows, other biasing atmospheric conditions or data gaps in the input images (e.g. stripes in the old MSS image data) that are automatically excluded from the $\overline{\circledast \Phi}_t$. inference without the need of a dedicated masking module as is common practice in multiple-scene classification tasks. The $\overline{\circledast \Phi}_t$. by epoch summarized by image segment is used as predictor of the persistence factor β_t by the mean of linear regression (LR) or direct cut-off mechanisms. The image segment-based (also called "object-based") automatic change detection ensure a stable spatial baseline of the analysis across the different epochs observed by the different sensors, minimizing the spurious probabilities possibly originated by the misalignment of the image raster grids, thus minimizing the error of the change information inferred from the $\overline{\circledast \Phi}_t$.. Two main approaches are tested regarding the decision composite at the segment level: a. binary decision mechanism (BIN) and continuous decision mechanism (CON). BIN engine is based on different strategies for automatic finding of a best $\overline{\circledast \Phi}_t$. cut-off based on available prior information, while the CON engine in based on linear regression (LR). In both cases, a multiple-decision support approach is taken: as set of 7 strata define the change detection problem from different machine learning point of views (ML STRATA), and N best models predicting changes are selected for each ML STRATA, thus they are composite in one final ensemble decision. This methodological choice is designed in order to increase the accuracy and the robustness of the automatic change detection, allowing the inferential engine to concentrate in specific simpler tasks or objectives defined by the ML STRATA, and founding the final decision on the robust model ensemble of the best model predictors by strata. Lastly, the model decision is aggregated and evaluated at the raster cell grid of 100m-resolution, MWD projection, that is the GHSL product specification.

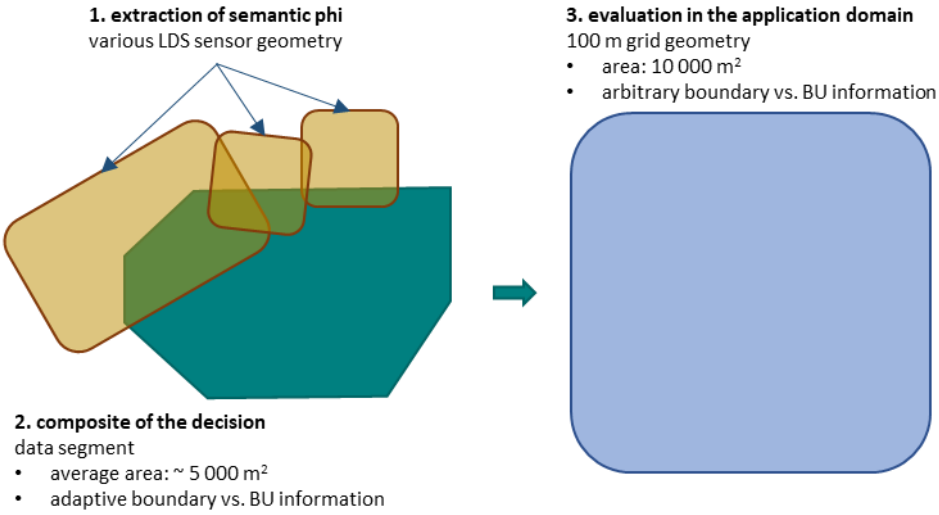


Figure 4 – Conceptual design of the method for semantic PHI extraction, composition and aggregation.

Given the conceptual design of the method, the experimental workflow was organized around two main steps: A) understanding how to best extract sematic Φ from Landsat Imagery and compose in multiple—scene single-epoch optimized summary $\overline{\circledast \Phi}_t$ per strata (Figure 5), and B) understanding how to composite the best semantic $\overline{\circledast \Phi}_t$. from the various strata to the final ensemble decision (Figure 6).

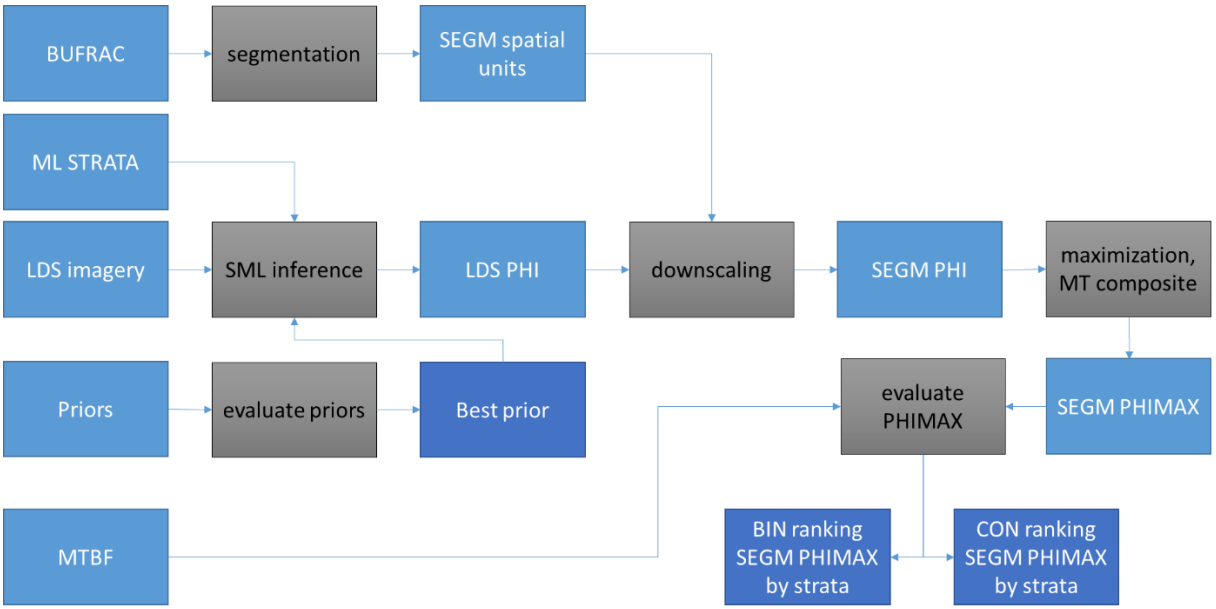


Figure 5 – Main experimental step A: understanding how to best extract sematic PHI from multi-temporal Landsat Imagery

The inputs of the experimental step A are i) the sub-pixel built-up surface estimates from the S2 image data at 10m-resolution (BUFRAC), ii) various strata domains used to support the machine learning (ML STRATA), iii) historical earth observation data collected by the Landsat program (LDS imagery) organized in four epochs (2014, 2000, 1990, and 1975), a set of multi-temporal land cover classification of the earth surface (Priors) done by previous GHSL releases or other studies, that can be considered as prior knowledge regarding global evolution of global built-up surfaces, and finally the multi-temporal building footprints data (MTBF) that are used for evaluation purposes (Figure 7 - Figure 12). The outputs of the experimental step A are the best prior, the BIN ranking of the best change predictor model by strata, and the CON ranking of the best change predictor model by strata. They are used in input of the experimental step B (Figure 6), focused on testing the composite of different ensemble decision (multiple-predictors, multiple-strata/objective) for each epoch.

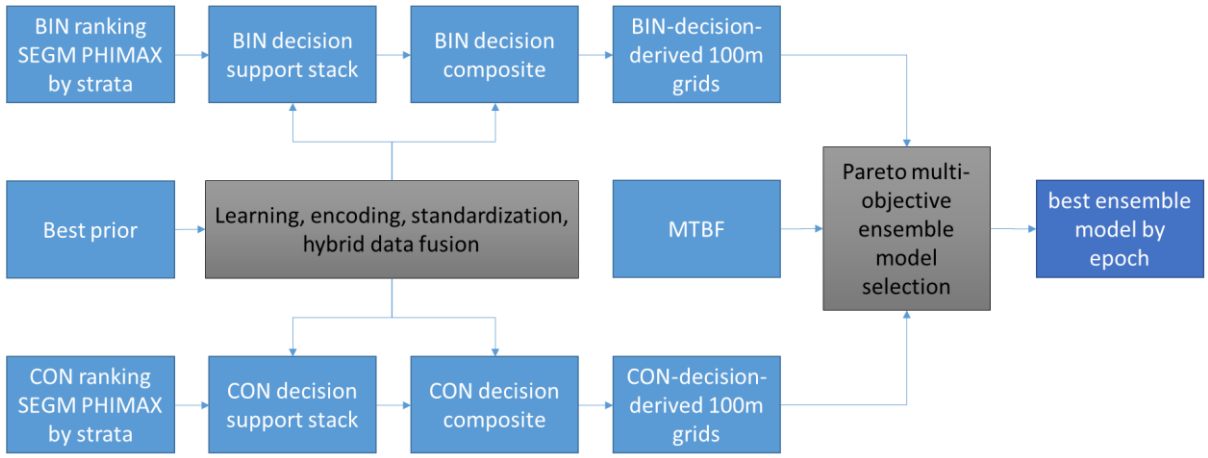


Figure 6 – Main experimental step B: understanding how to composite the best semantic PHI from various strata to the final ensemble decision

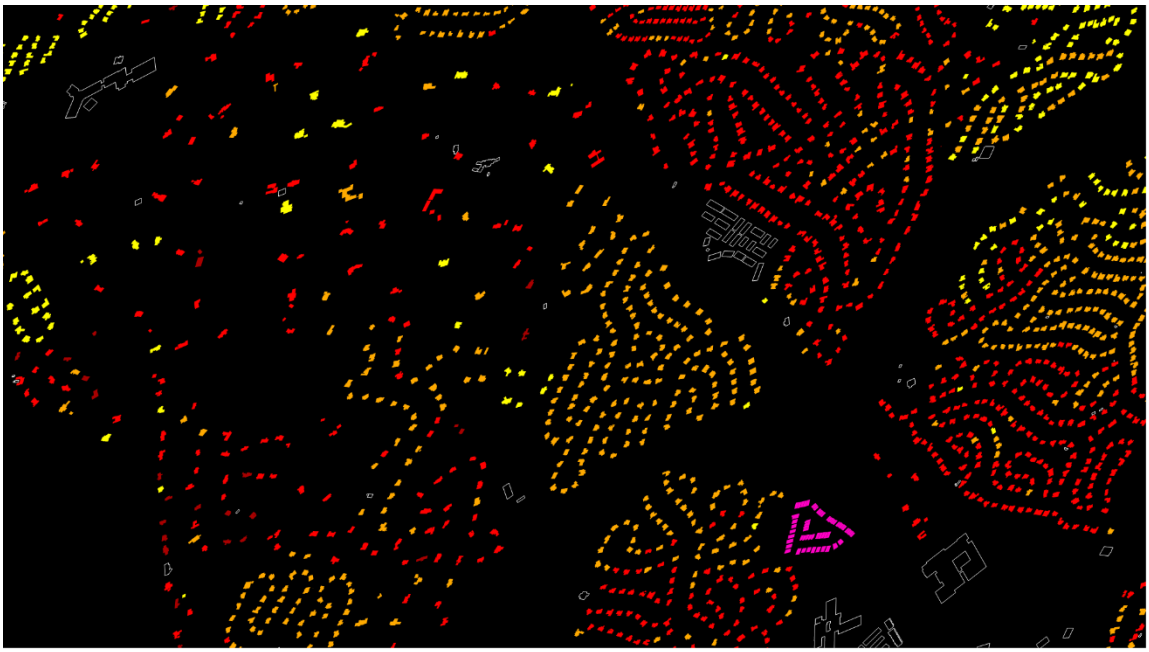


Figure 7 – Example of reference MTBF data test case name US_GA_City_of_Johns_Creek N-E Atlanta, USA. Dark red 1975, red 1990, orange 2000, yellow 2014, magenta 2018. Transparent buildings have no temporal data attached.

Figure 8 – Example Landsat image data from the epoch 2014, MTBF overlay

Figure 9 – Example of Landsat image data from the epoch 2000, MTBF overlay

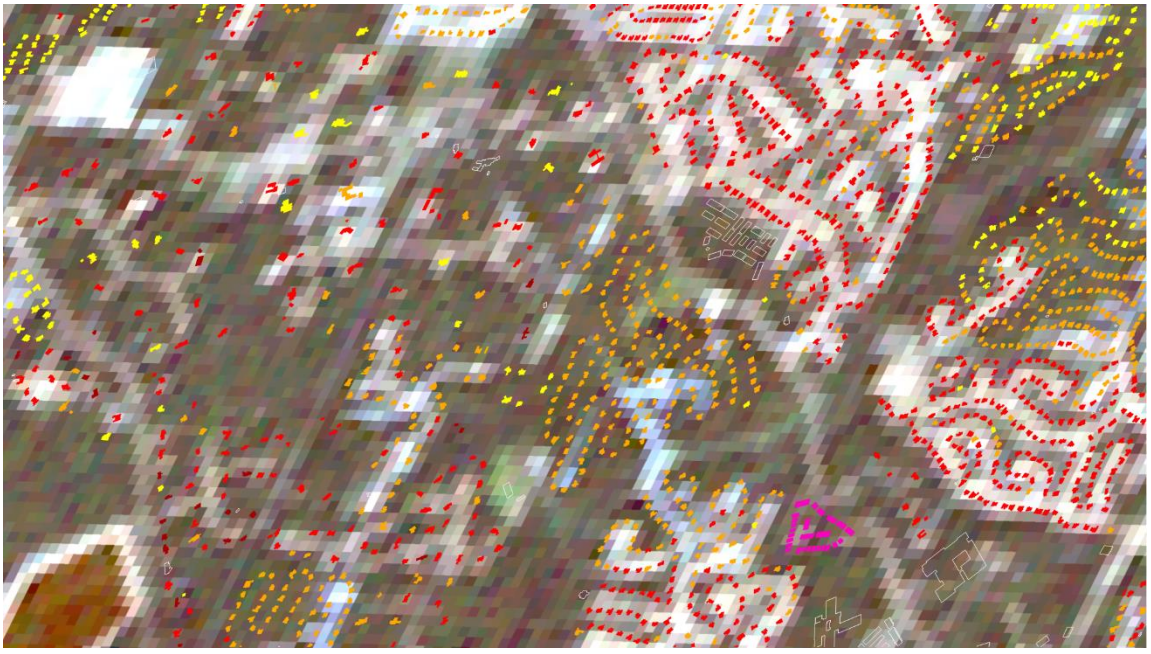


Figure 10 – Example of Landsat image data from the epoch 1990, MTBF overlay

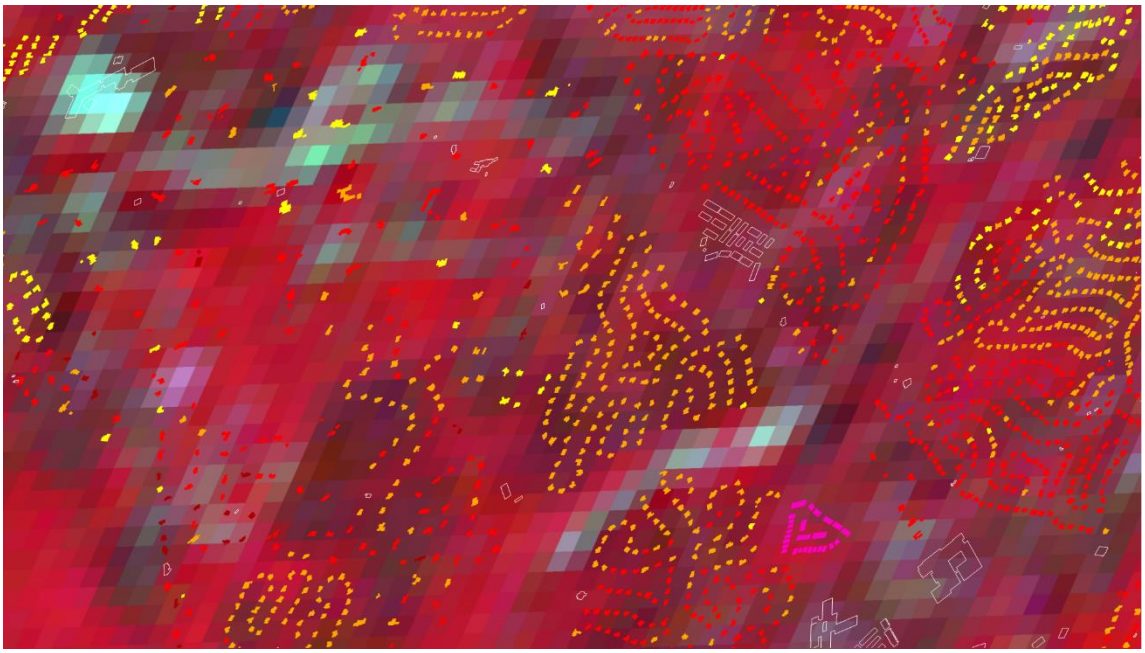


Figure 11 – Example of Landsat image data from the epoch 1975, MTBF overlay

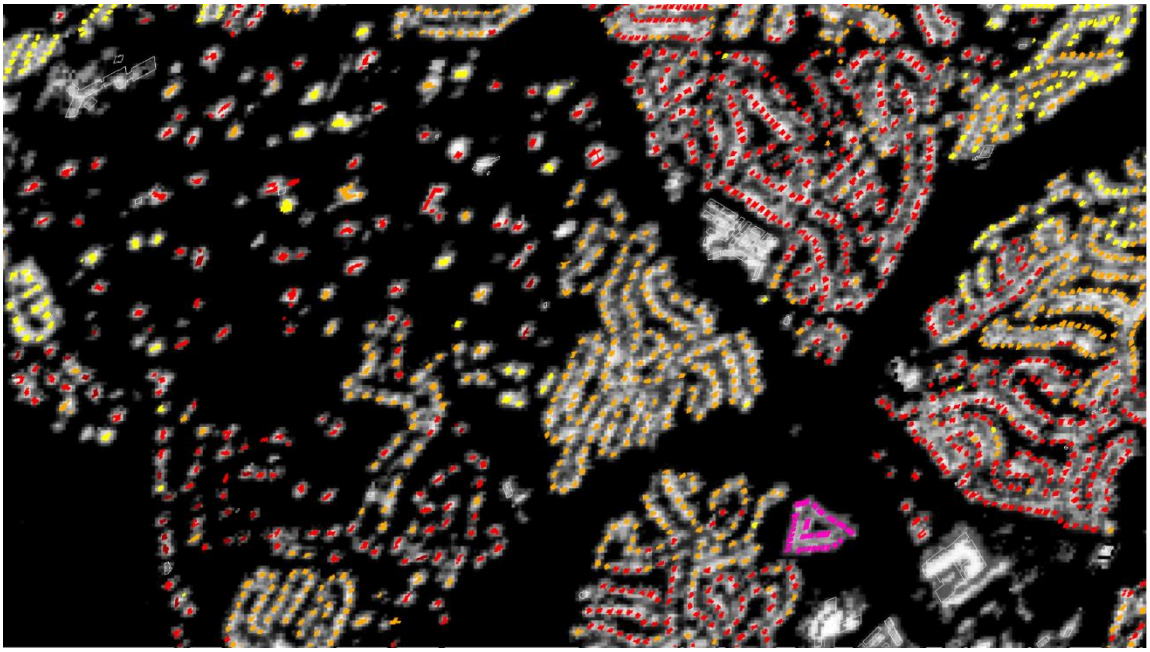


Figure 12 – New sub-pixel built-up surface fraction prediction (BUFRAC) from S2 image data 2018, MTBF overlay

3.3 Baseline spatial data definition

3.3.1 *BUFRAC segmentation*

BUFRAC is the sub-pixel built-up surface prediction made at 10m-resolution from S2 image data composite of the most recent year observed in this study (2018), noted as $f_x BU_{t=2018}$. Watershed segmentation techniques are used to extract image segments (Figure 13) from this raster data. In particular, the watershed segmentation was obtained from the inverse of the $f_x BU_{t=2018}$ at 10m-res, filtered by a local Gaussian low-pass filtering ($\sigma = 0.5$) in order to reduce over-segmentation. Image segments are used as constant spatial baseline for the subsequent multi-temporal classification supported by the historical Landsat image data collections. Generally, the sematic Φ estimated at the grid level of the Landsat imagery, is downscaled to the 10m-res segment by the surface-weighted average operator.

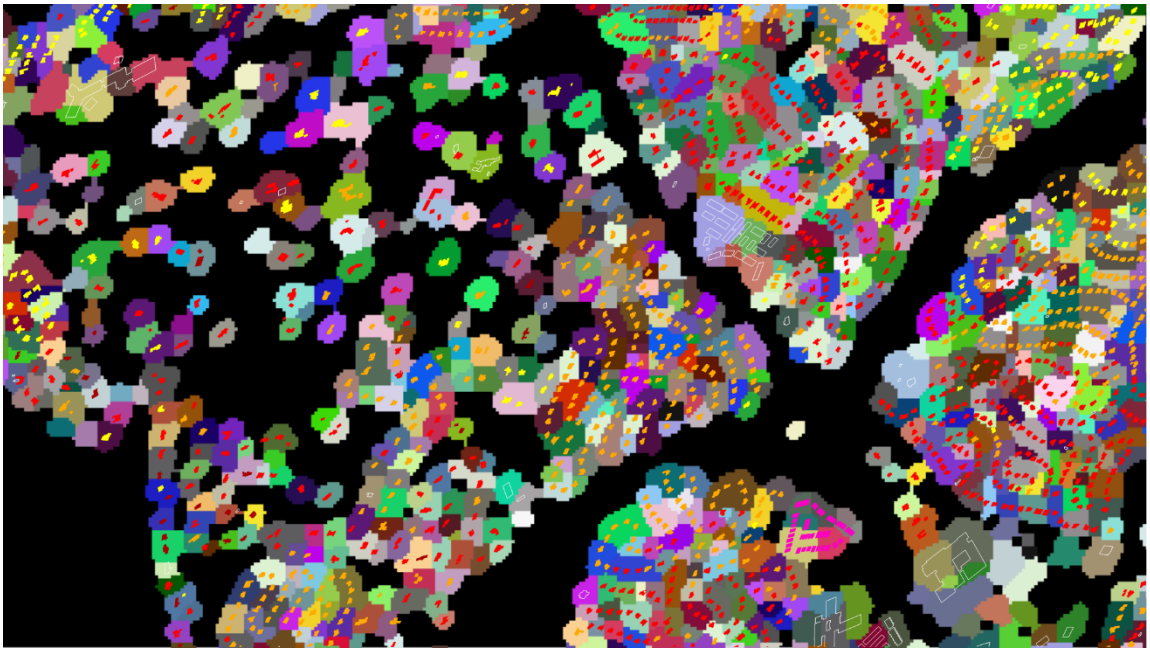


Figure 13 – Watershed segmentation of the BUFRAC 10m supporting the multi-temporal assessment from the historical Landsat imageries, MTBF overlay

3.3.2 *ML STRATA*

The machine learning strata are defined in order to provide more specific tasks or objectives to the automatic change detection prediction, taking in to account the know technical characteristics of the EO sensors and the known physical characteristics of the semantic targets (buildings, settlement spatial-temporal patterns). They are seven non-mutually-exclusive strata defined as follows

1. COMPACT: the target built-up structure (BU) observed by the specific sensor is collected as a patch of adjacent samples (pixels), thus largely influenced by the radiometric reflection of the foreground built-up neighbouring surfaces (BU). The samples belonging to this stratum are potentially exposed to underestimate of the change due to the similarity to the foreground information associated with the BU abstraction class.
2. SPARSE: the target built-up structure (BU) observed by the specific sensor is collected as isolated sample (pixel), thus largely influenced by the radiometric reflection of the background non-built-up (NBU) surfaces. The samples belonging to this stratum are potentially exposed to overestimate of the change due to the similarity to the background information associated with the NBU abstraction class.
3. RES: the target built-up structure is classified as residential, according to the GHSL specs (European Commission. Joint Research Centre. 2023). Thus, having expected average size (scale) in the order of 10 meters producing specific radiometric mixture in the EO raster data composing the reflection of the roof material, the reflection of the surrounding surfaces (gardens, roads, soil), and – important – a relatively large shadow component.

4. NRES: the target built-up structure is classified as non-residential, according to the GHSL specs (European Commission. Joint Research Centre. 2023). Thus, having expected average size (scale) in the order of 100 meters producing specific radiometric mixture in the EO raster data composing the reflection of the roof material, the reflection of the surrounding surfaces (parks, roads, soil), and – important – a relatively small or absent shadow component.
5. IN_PRIOR: the target built-up structure is belonging to the spatial domain already solved by existing semantic multi-temporal global priors having their specific limitations and methodological constraints. Thus, available global MT priors can be used to support the automatic change decision process.
6. OUT_PRIOR: the target built-up structure is outside the spatial domain already solved by existing semantic multi-temporal global priors having their specific limitations and methodological constraints. It is new built-up surface domain discovered by the S2 image data classification in 2018. Thus, available global MT priors cannot be used to support the automatic change decision process.
7. ALL: The logical union of all the above strata, ensuring a coherent average response across the various specific strata.

Examples of the above mentioned strata in the N-E of Atlanta, US are displayed in the Figure 14, Figure 15, Figure 16, Figure 17, Figure 18, and Figure 19.

Figure 14 - Membership to the COMPACT strata of the image segments supporting the MT classification. N-E Atlanta, US.

Figure 15 - Membership to the SPARSE strata of the image segments supporting the MT classification. N-E Atlanta, US.

Figure 16 - Membership to the RES strata of the image segments supporting the MT classification. N-E Atlanta, US.

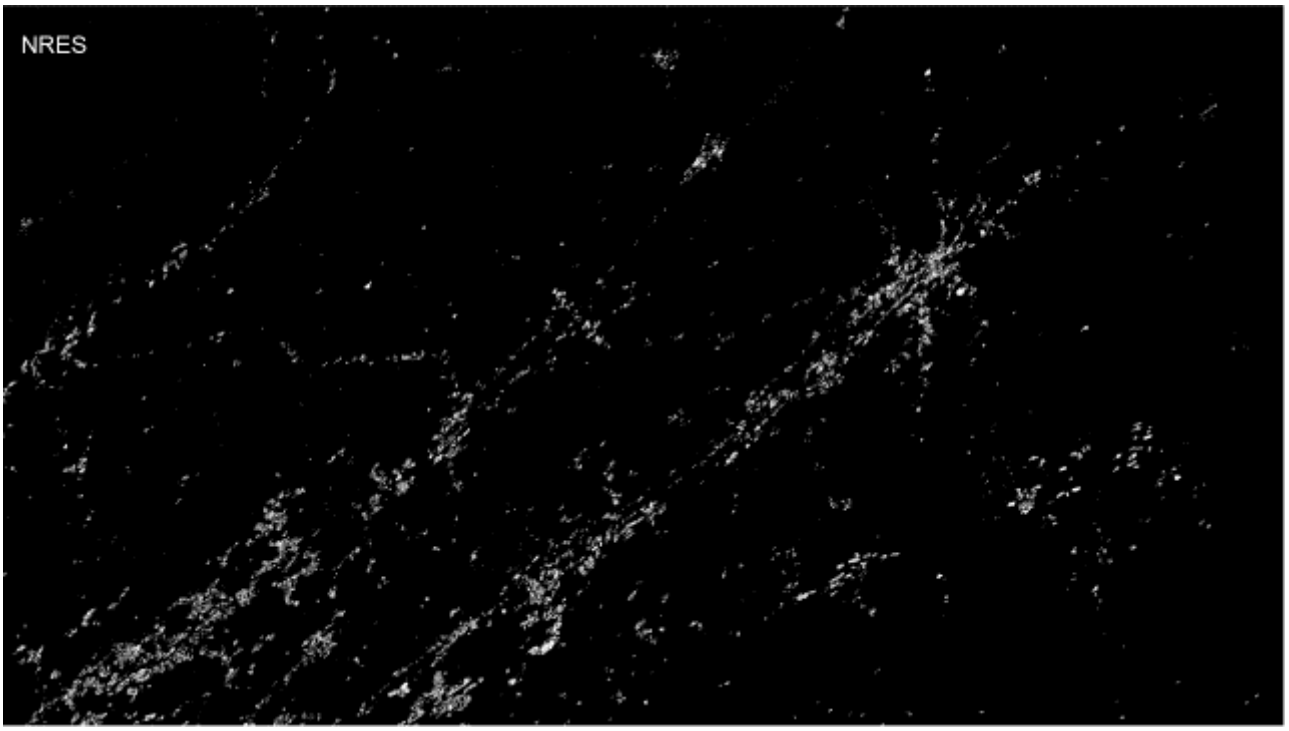


Figure 17 - Membership to the NRES strata of the image segments supporting the MT classification. N-E Atlanta, US.

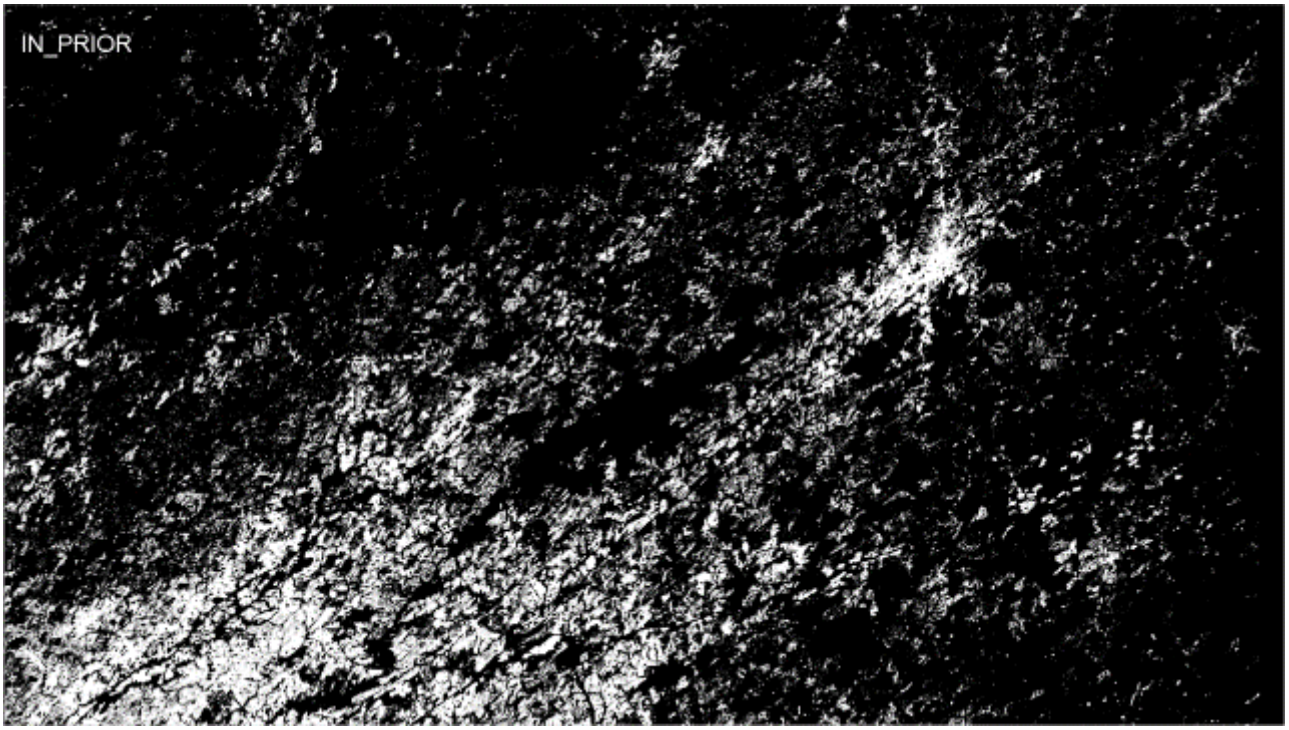


Figure 18 - Membership to the IN_PRIOR strata of the image segments supporting the MT classification. N-E Atlanta, US.

Figure 19 - Membership to the OUT_PRIOR strata of the image segments supporting the MT classification. N-E Atlanta, US.

3.3.3 ***Evaluation of the priors***

The PRIORS are used in several steps of the GHSL MT classification: i) training set for machine learning, ii) rescaling or standardization of the decision support, and iii) data fusion in hybrid predictions including gap filling.

GAIA, GAUD, GISA, GISA2.0, GISD30, WSFEVO, GHS_P2016, GHS_P2019, priors are evaluated and compared with the GHS_P2022A benchmark for their capacity to predict the BU surface changes of the anchor point 2018 vs. other GHSL epochs 2014, 2000, 1990, and 1975.

Additionally, composites of union of the priors (UCOMPO) are evaluated and compared with the GHS_P2022A benchmark: be BU_2018 the BU surface predicted in the epoch 2018, and be U_OFF(t) the union of all the prior signals to switch off the BU_2018 in a specific epoch t (2014, 2000, 1990, 1975), and U_ON(t) the union of all the prior signals to maintain switched on the BU surface in the specific epoch t (2014, 2000, 1990, 1975). Four UCOMPO options (A, B, C, D; Figure 20) are evaluated: In the "all the change" option (B and D) the whole U_OFF(t) is taken, while in the "conservative change" option (A and C) only the domain of U_OFF(t) not contradicting the BU_ON(t) its retained, thus the intersection of the U_OFF with the negation of the U_ON is taken (exclusive or logic, also called XOR). Moreover, two supporting sets for the priors are evaluated: "all the priors" (A and B) and "selected priors" subset (C and D).

Figure 21 shows the accuracy of the single priors in predicting the stock of BU surface in the epochs 1975, 1990, 2000, 2014, and 2018. According to these findings WSF_EVO, GISA, and GHS_B_P2019 are selected to support the UCOMPO_C and UCOMPO_D options. WSF_EVO has the best performances in mid epochs but not including the extrema, GISA is the prior including the extrema with the best performances, and the GHS_B_P2019 in the UCOMPO would improve consistency with previous GHSL releases (user requirements).

Figure 22 shows the accuracy of the single priors compared with the UCOMPO priors in predicting the stock of BU surface in the epochs 1975, 1990, 2000, 2014, and 2018. Average of the URBAN, RURAL application domain strata. These findings support the choice of the UCOMPO_D option (WSF_EVO, GISA, and GHS_B_P2019, all the change) as the best option to predict changes vs. the target epoch 2018. This choice is confirmed by the observation of the accuracy in predicting the change map from 2018 to the other epochs (2014, 2000, 1990, and 1975), where the UCOMPO_D rank the best in all the epochs (Figure 23).

The capacity of the UCOMPO_D to predict continuous changes from the anchor point 2018 to the other epochs (2014, 2000, 1990, and 1975) is compared with the other available options using the Pearson linear correlation coefficient, and the UCOMPO_D solution is confirmed as the best performing in both URBAN and RURAL application domain strata (Figure 24, Figure 25, Figure 26).

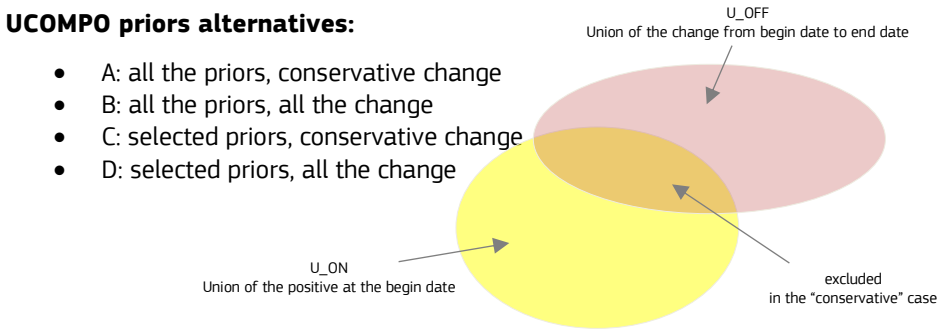


Figure 20 - Logical schema of the union of the prior composite (UCOMPO) applied in the study

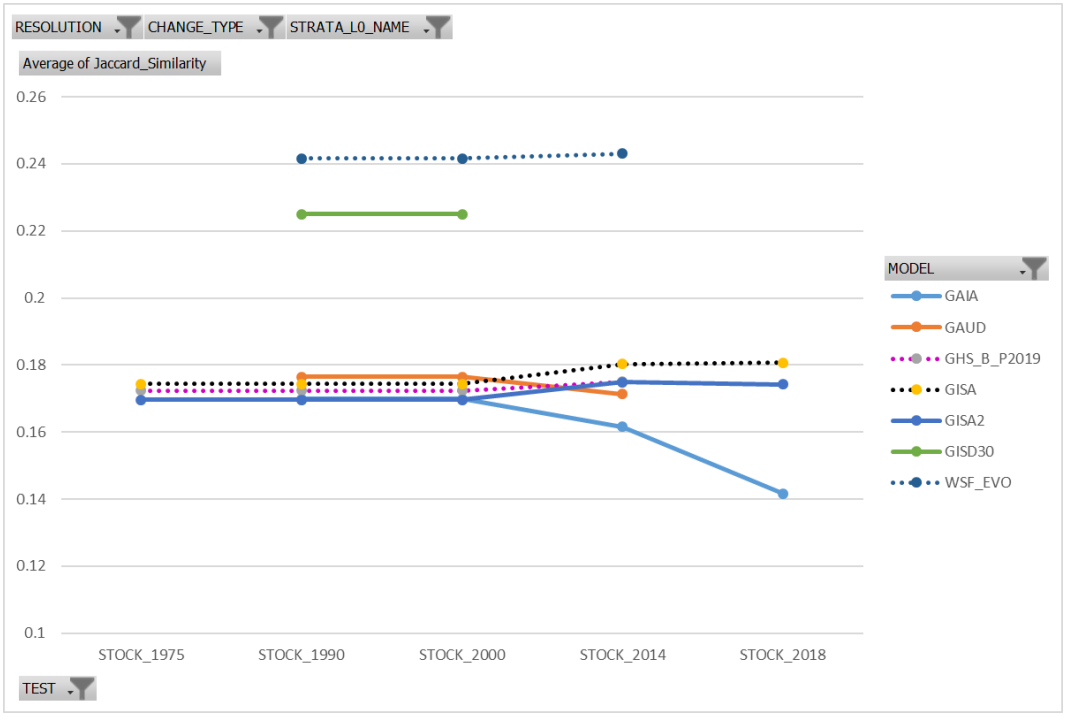


Figure 21 – Accuracy of the single priors in predicting the stock of BU surface in the epochs 1975, 1990, 2000, 2014, and 2018. Average of the URBAN, RURAL application domain strata.

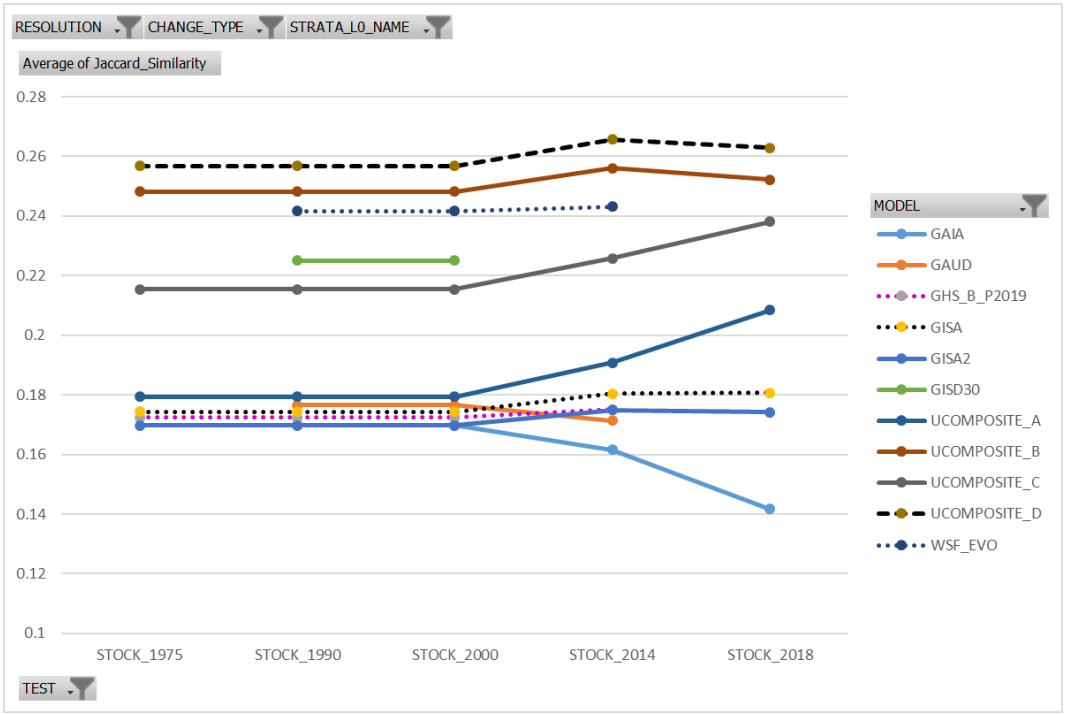


Figure 22 - Accuracy of the single priors compared with the UCOMPO priors in predicting the stock of BU surface in the epochs 1975, 1990, 2000, 2014, and 2018. Average of the URBAN, RURAL application domain strata

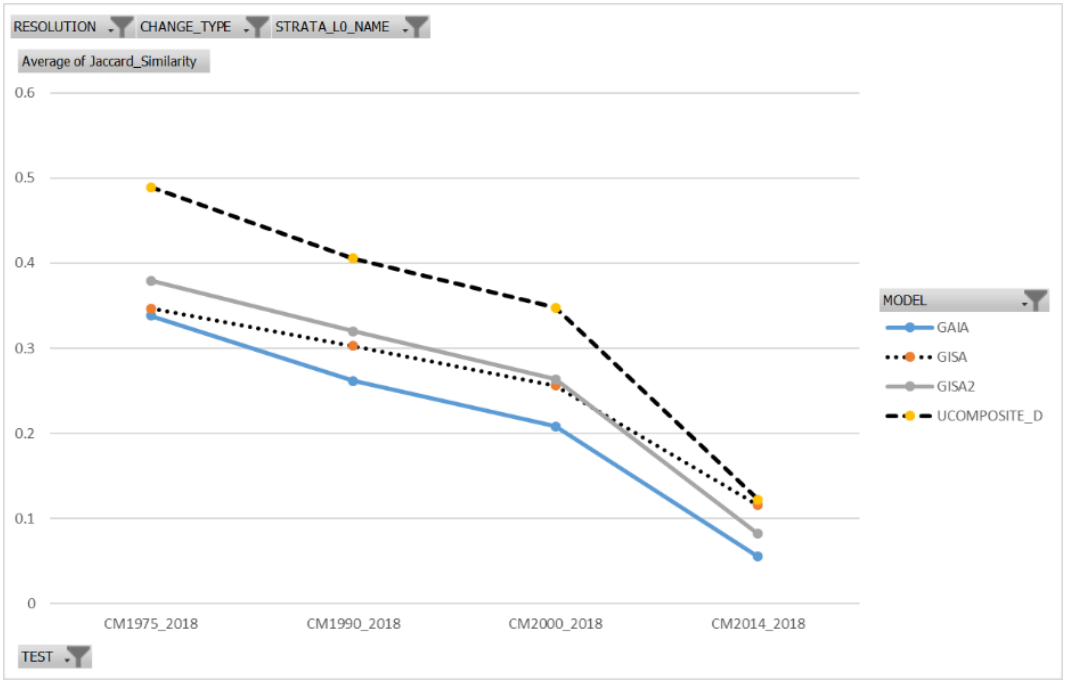


Figure 23 - Accuracy of the priors and the UCOMPO_D in predicting the change from the epoch 2018 to relevant points in time 2014, 2000, 1990, and 1975. Average of the URBAN, RURAL application domain strata.

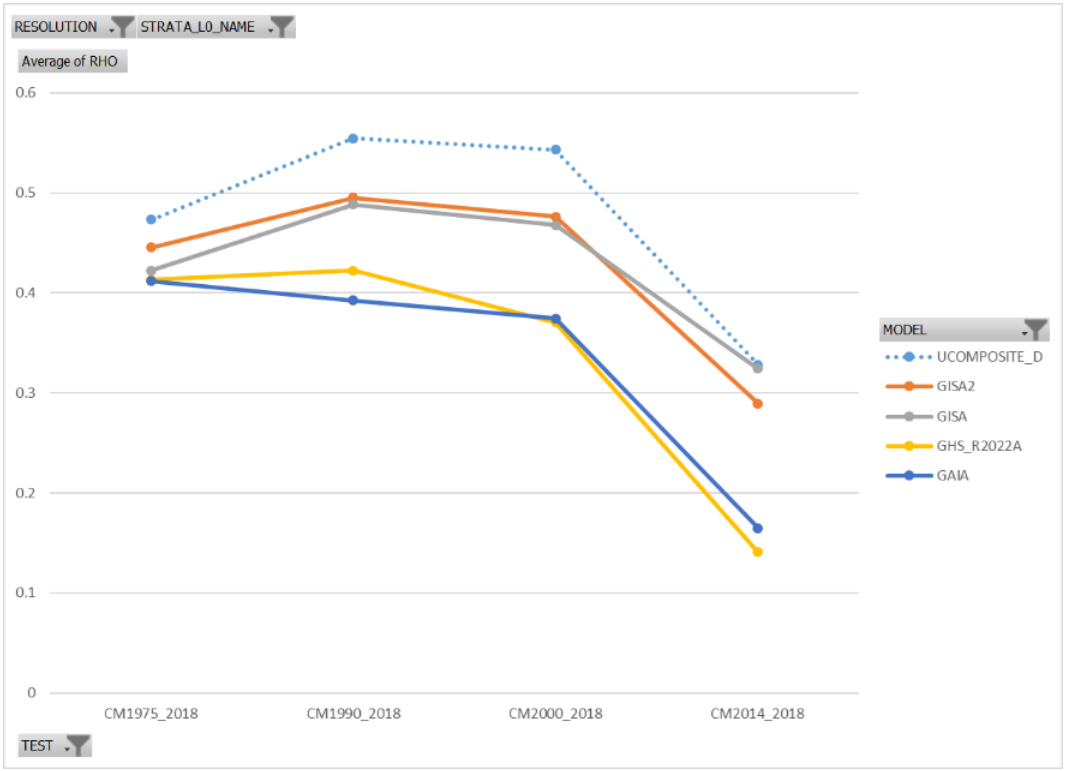


Figure 24 – Linear correlation (Pearson rho) of the change map multiplied with BU surface 2018, with observed changes in MTBF between the 2018 anchor point and the epochs 2014, 2000, 1990, and 1975. Average of the URBAN, RURAL application domain strata

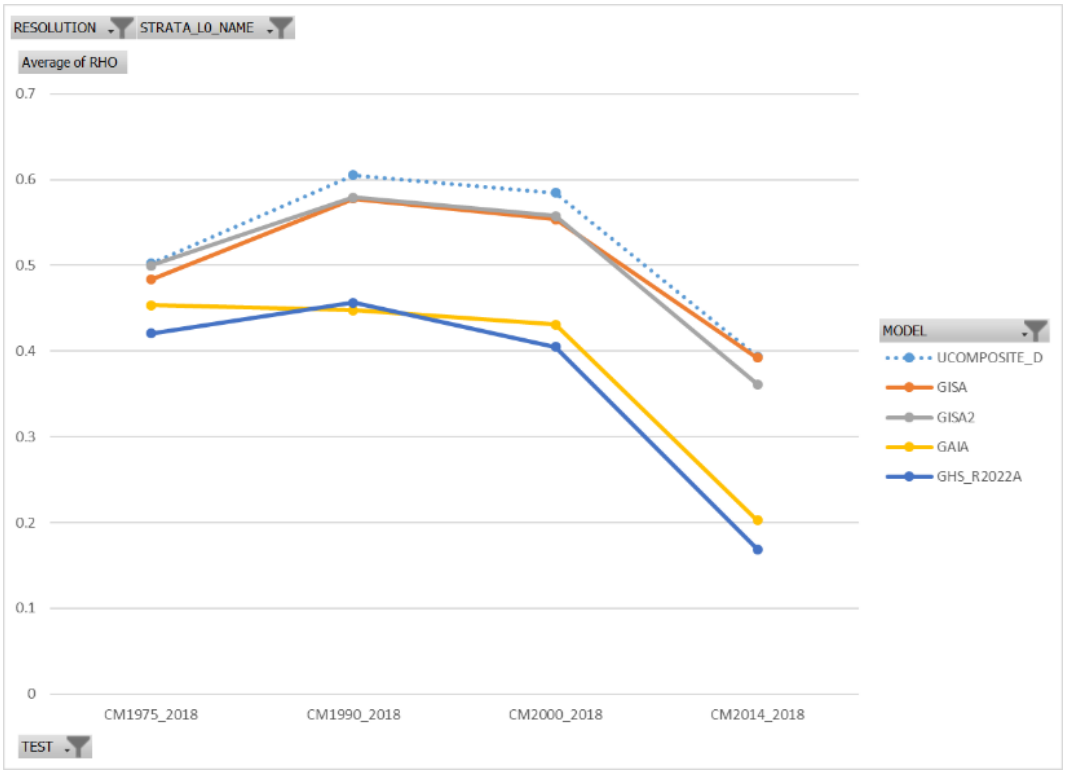


Figure 25 - Linear correlation (Pearson rho) of the change map multiplied with BU surface 2018, with observed changes in MTBF between the 2018 anchor point and the epochs 2014, 2000, 1990, and 1975. URBAN application domain strata

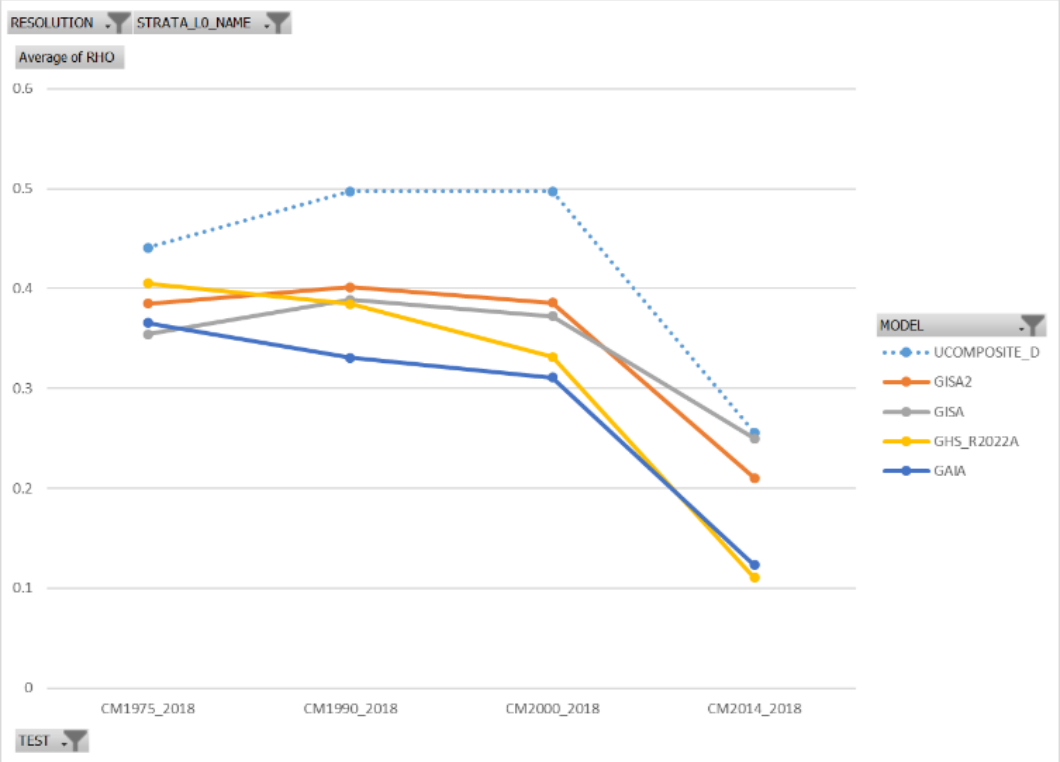


Figure 26 - Linear correlation (Pearson rho) of the change map multiplied with BU surface 2018, with observed changes in MTBF between the 2018 anchor point and the epochs 2014, 2000, 1990, and 1975. RURAL application domain strata

3.4 Multiple-scene semantic composite

In the model discussed here, an innovation is introduced regarding the resuming of the Φ inferences collected from multiple input images (or "scenes" in the Landsat glossary) when assessing the same spatial sample and falling in the same nominal epoch interval. Satellite imagery are collected in a precise point in time (hour, day), falling in determined conditions of illumination, changing atmospheric conditions (clouds, cloud shadows, haze, humidity, and other turbidity components in atmosphere) and seasonal changes (vegetation, water, soil reflectance changing), determining various conditions of separability between the semantic foreground (the target class) vs. the background (the logical complement). The SML classification approach allows to make continuous learning, discovering the best rules associating image data to a specific semantic at low computational cost, therefore allowing to avoid model transfer from spatial and temporal domains different than the one processed in the specific image under classification. This fact augment the precision of the inference at the specific point in time of each different satellite imagery, by providing the best discriminant function at the precise conditions of the image data collection. Still, different satellite imagery in different points in time will report about different Φ assessing the same spatial samples and the same semantic, because of the changing semantic foreground-vs-background separability conditions in each input image. In the model solution discussed here a new composite mechanism called "cumulative divergent Φ composite" is adopted. The basic idea of this approach is that for each spatial sample we can cumulate the semantic Φ available from different imageries in different point in time by independent maximization of the hypothesis that the sample belongs to the foreground vs. the background semantic: e.g. Φ_{BU} vs. Φ_{NBU} in the application discussed here. By maximization of the foreground vs. the background semantic hypothesis, we ensure that we pick up the best-discriminant point in time assessed by the available images, where the specific sample *x* was associated to a specific semantic and his semantic competitor. The final decision on which semantic should be associated to the sample x is done after the maximization of the divergent hypothesis, ensuring they all have the same chance of success in the available supporting image data. The maximized Φ over multiple image data collected in the given epoch t is noted as $\overline{\circledast \Phi_t}$. By definition, the statistical distribution of the composed $\overline{\circledast \Phi_t}$ by maximization of divergent hypothesis will be more polarized toward the extrema as compared to the single-

scene Φ : this fact decreases the domain of incertitude, where the SML Φ inference is around the intermediate zero domain. Moreover, also by mathematical construction, the samples biased by non-stationary noise in a given image, under the condition that are randomly placed in the spatial domain vs. the learning set supporting the SML, will receive a semantic Φ score around the zero value (the intermediate value in [-1..1] range) in any given image. Therefore, the influence of the non-stationary noise in the input images is minimized by construction in the composed $\overline{\circledast \Phi_t}$. Relevant non-stationary noise in the satellite data processed in this work include cloud obstructions, cloud shadows, or anomalies in the sensor (e.g stripes of anomalous values in the old MSS data). The minimization-by-construction of non-stationary noise in the composed $\overline{\circledast \Phi_t}$ inference, allowed us to avoid relying on specific models for detection of clouds or other obstructions/anomalies in the images, on the contrary to common practices allocating such models in the pre-processing phase. Therefore, in the solution designed here we decrease the model dependency to external assumptions. Figure 27 shows the dominant paradigm in multiple-scene classification composite: image data is transformed in measurements of absolute amount of energy (reflectance calibration), clouds, shadows and other obstructions are masked out from valid data using additional hypothesis and parameters, thus the inference is made and the final output is mosaicked. A simplified version of this masking-composite approach (radiometric calibration was excluded) was also applied in the previous releases of the GHSL (M. Pesaresi, Ehrlich, et al. 2016) (Corbane et al. 2019). In alternative, images can be composed upstream e.g. with the objective to minimize cloud cover and thus the inference is extracted from the "cloud-free" image composite (Figure 28): this is the approach taken for the extraction of the BUFRAC. Figure 29 shows the paradigm applied in the work discussed here. The multiple-class inference is extracted from single scenes, and subsequently composed by maximization of all the semantic hypothesis, thus the decision is taken after the maximization.

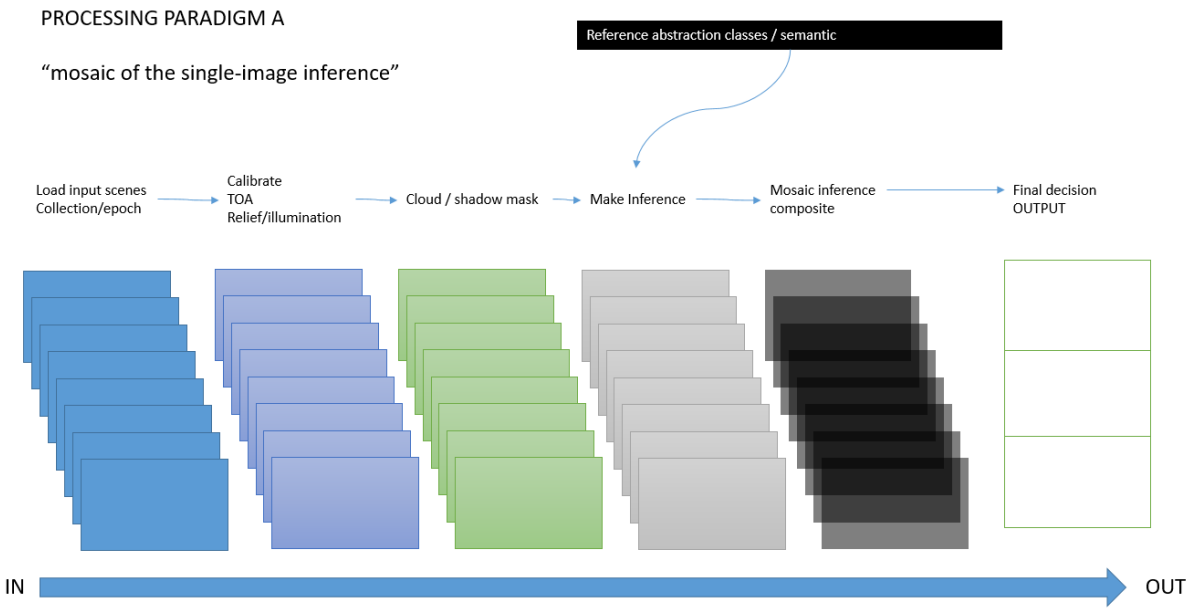


Figure 27- Multiple-image classification paradigm A: "mosaic of the single-image inference"

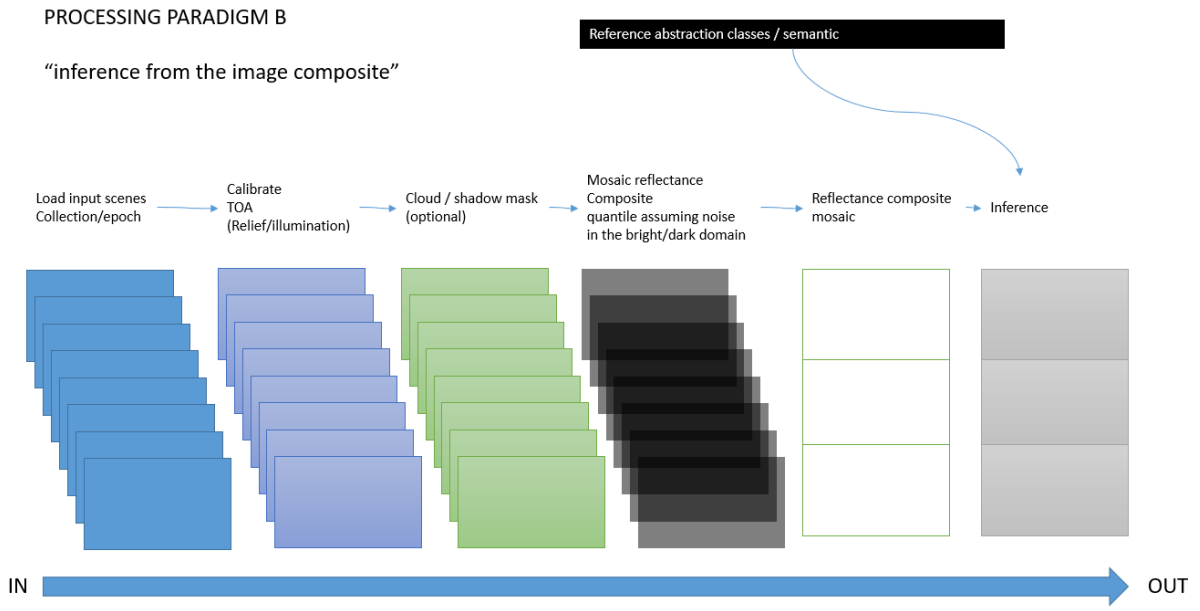


Figure 28 - Multiple-image classification paradigm B: inference from the image composite"

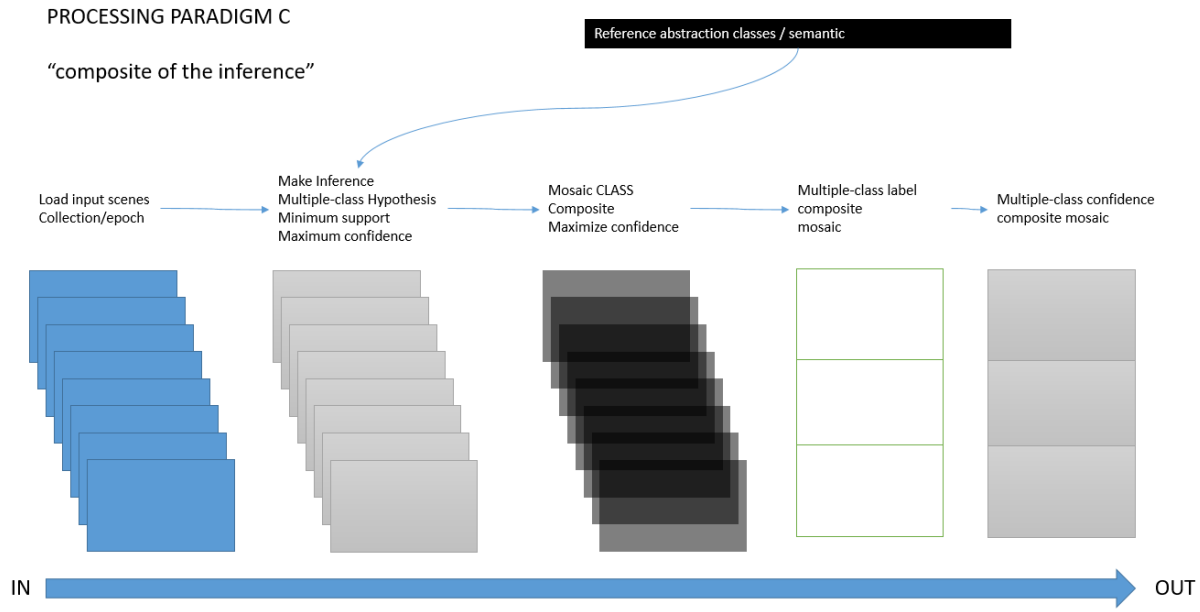


Figure 29 – Multiple-image classification paradigm C: "composite of the inference"

3.5 Model architecture

The prediction of the image segment changes from BU to NBU (switch off) is based on the application of the SML generalized to the multiple-quantization case, using in input the quantized values recorded in the different bands of the Landsat satellite imageries. Two main approaches are tested: a binary (BIN) change detection schema and b) a continuous (CON) change detection schema.

The model architecture for image segment changes is defined by the combination of 7 model architectural choices listed below:

1. TRAIN_STRATA: which strata is used for training set sampling.
    a. 'ALL': the whole data tile under processing,
    b. 'COMPACT': the compact BU patches,
    c. 'SPARSE': the sparse BU samples,
    d. 'COMPO': a linear composite of COMPACT and SPARSE.
2. SAMPLING: which strategy is used for random train sampling. In the case of target BU surface, being the positive examples rare they are all sampled, while a sampling strategy is need for the negative examples belonging to the semantic background:
    a. 'EQNEAR': equalized random sampling of the negative examples in the neighbouring of the positive samples.
    b. 'EQFULL': equalized random sampling of the negative examples in the whole semantic background of valid data,
    c. 'FULL': all the available negative samples are used.
3. DECISION: which logical combination (data fusion) of semantic Φx and prior Px is used for supporting the decision Dx at the spatial sample *x*:
    a. 'PRIOR': only the change rates as extrapolated from available priors $Dx = Px$
    b. 'PHI' : only the Φ from SML is used $Dx = \Phi x$
    c. 'PHI_OR_PRIOR' : the fuzzy union is considered $Dx = \max(Px, \Phi x)$
    d. 'PHI_AND_PRIOR', the fuzzy intersection is considered $Dx = \min(Px, \Phi x)$
    e. 'PHI_MED_PRIOR'. The median decision is considered $Dx = Px + \Phi x/2$
4. PHITYPE: which PHI type (M. Pesaresi, Syrris, and Julea 2016) is used for supporting the decision:
    a. 'PHI_A', $\Phi_E^a(X, Y^+, Y^-) = f^+ - f^-/f^+ + f^-$ using frequency of occurrences: maximization of the overall accuracy
    b. 'PHI_B', $\Phi_E^b(X, Y^+, Y^-) = p^+ - p^-/p^+ + p^-$ using empirical probabilities: maximization of the balanced accuracy
    c. 'PHI_AB'. $\Phi_E^{ab} = \Phi_E^a + \Phi_E^b/2$: median solution
5. PHISEMANTIC: which maximized $\overline{\circledast\, \Phi}$ is used to support the change decision:
    a. 'BU', $\overline{\circledast\, \Phi}_t = \max(1 - \Phi_{BU}^{s \in t})$
    b. 'NBU', $\overline{\circledast\, \Phi}_t = \max(\Phi_{NBU}^{s \in t})$
    c. 'NBU_BU_D': NBU-BU difference, $\overline{\circledast\, \Phi}_t = \max(\Phi_{NBU}^{s \in t} - \Phi_{BU}^{s \in t})$
    d. 'NBU_BU_ND': normalized NBU-BU difference, $\overline{\circledast\, \Phi}_t = \max(\Phi_{NBU}^{s \in t} - \Phi_{BU}^{s \in t}/\Phi_{NBU}^{s \in t} + \Phi_{BU}^{s \in t})$
    e. 'OFF_factor': $\overline{\circledast\, \Phi}_t = \max(\Phi_{NBU}^{s \in t}/\Phi_{BU}^{s \in t})$
6. PHITEMPORAL: which strategy is used to maximize the semantic $\overline{\circledast\, \Phi}$ extracted from the Landsat imagery (Figure 30):
    a. 'BYEPOCH': each epoch collection is considered independently for the maximization purpose,
    b. 'INTERLACED': image data of the various epochs are interlaced in the maximization of the BU, NBU semantic PHI.
7. PHIFUN: which function is used to downscale the sensor-derived raster semantic Φ_{BU}, Φ_{NBU} to the spatial baseline image segments (Figure 4):
    a. 'fun_meanGT0': the average of the valid Φ_{BU}, Φ_{NBU} data,

b. 'fun_max': the maximal value of the valid Φ_{BU}, Φ_{NBU} data.

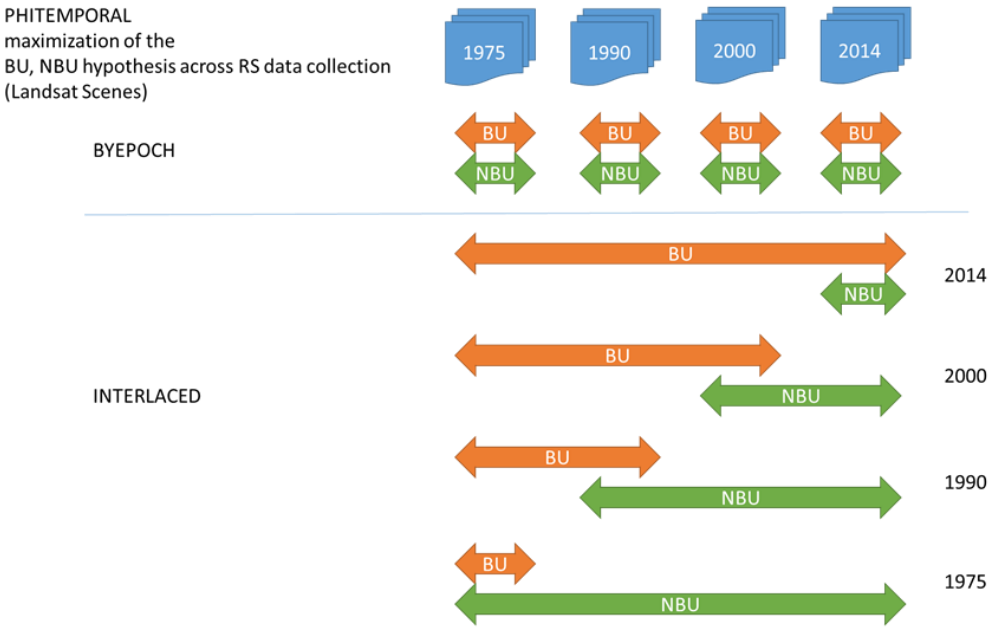


Figure 30 - PHITEMPORAL maximization of the BU, NBU class abstraction hypothesis across the remotely sensed data collections

3.5.1 ***Model architecture ranking***

A total of 3 600 combinations of model architecture were tested for each of seven ML STRATA and each of four change epochs (2014-2018, 2000-2018, 1990-2018, and 1975-2018), for a total of 100 800 experimental tests. Binary (BIN) and continuous (CON) model predictions were evaluated using MTBF as the reference data. Only the most restrictive set of the REF data domain was retained in this experimental design, which was passing the test for the 2014-2018 change detection.

The binary (BIN) model is obtainable by a direct cut-off of the SML PHI model prediction in the binary change set: higher values of PHI indicating NBU than BU result in the binary decision to "switch-off" given segment.

The BIN choice evaluation is done by the receiver operating characteristic curve (ROC) technique, the used metric is the maximal Jaccard Similarity (Jaccard 1901), a metric commonly used in information retrieval, machine learning and image processing for estimation of similarity of objects (Kosub 2019). The Jaccard index *J* measures the relative size of the overlap of two finite sets:

$$J(X_{BIN}, Y_{BIN}) \stackrel{\text{def}}{=} \frac{|X_{BIN} \cap Y_{BIN}|}{|X_{BIN} \cup Y_{BIN}|}$$

Where X_{BIN} is the predicted semantic on segment level, Y_{BIN} is the observed semantic, and $|\cdot|$ denote the cardinal of the set, that is the number of segments that either intersect or overlap.

The continuous (CON) choice evaluation is solved by observation of the Pearson linear correlation coefficient for predicted vs. observed amount of built-up surface changes relatively to the 2018 anchor point:

$$\rho(X_{CON}, Y_{CON}) = \frac{cov(X_{CON}, Y_{CON})}{\sigma_{X_{CON}}, \sigma_{Y_{CON}}}$$

Where X_{CON} is the predicted change of built-up surface change between given epoch and 2018 in each segment, Y_{CON} is the observed change, cov is the covariance and σ is the standard deviation of predicted and observed changes on segment level.

For each ML STRATA, we ranked the combinations of model architecture per change epoch, by decreasing Jaccard Similarity index for BIN predictions, and by decreasing Pearson coefficient for CON predictions.

3.6 Ensemble decision model design

The ensemble decision is made in order to increase the robustness and the accuracy of the final decision to switch off a specific spatial sample containing BU surface in 2018. As the decision can be either binary (BIN) or continuous (CON), thus the architecture of the ensemble is tested across different options regrouped under the BIN and CON branches.

3.6.1 *BIN branch*

Under this hypothesis, the following options are tested

a) FINAL_DECISION_TYPE: whether is better to rely on a single model prediction over the ALL strata vs. the composite of various strata

b) NBESTMODEL: number of best models supporting the decision at each strata. Option of 1:5 was tested.

c) CUTOFFCASE: which strategy can be used to cut-off the base model response to a binary decision (support). 11 options including the application of the average best cut-off by model by strata learned from the ROC analysis, and the adaptive learning of good cut-off by observation of the model prediction in the prior change domain.

d) FINALDECISION: how we compose the multiple decision support stack to the ensemble decision. 20 options tested including voting schema (additive approach) and associative rule learning (data pattern recognition)

3.6.2 *CON branch*

Under this hypothesis, the following options are tested

a) FINAL_DECISION_TYPE: whether is better to rely on a single model prediction over the ALL strata vs. the composite of various strata

b) NBESTMODEL: number of best models supporting the decision at each strata. Option of 1:5 was tested.

c) FINALDECISION: how we compose the multiple decision support stack to the ensemble decision. 'MEDIAN', 'MEAN', 'MIN', 'MAX', and quantile 'Q1', 'Q2', 'Q3', 'Q4', 'Q5', 'Q6', 'Q7', 'Q8', 'Q9' options were tested.

3.7 Multi-objective model ensemble selection

For each change epoch (2014, 2000, 1990, and 1975 vs the 2018 anchor point), three agreement measures are calculated: the Ruzicka distance of the predicted vs. the observed change grids, and the mean absolute error (MAE) of the built-up surface change rate in the URBAN and RURAL application domains; all using the MTBF as the reference.

The Ruzicka distance measures are calculated for model predictions aggregated to 100 m grids,. Only the most restrictive set of the REF data domain was retained in this experimental design, which was passing the test for the 2014-2018 change detection. The Ruzicka distance is formulated as the inverse of the Ruzicka similarity:

$$d_{Ruz} = 1 - \frac{\sum_x \min(f_x\ ,\ r_x)}{\sum_x \max(f_x\ ,\ r_x)}$$

Where f_x and r_x are vectors of the x 100m-res spatially explicit grid cells with non-negative real numbers reporting about, respectively, the predicted and observed persistence factor $\beta_{t\in[1975,1990,2000,2014]}$, with:

$$\beta_{t\in[1975,1990,2000,2014]} = \ f_x BU_{t=2018}\ / f_x BU_{t\in[1975,1990,2000,2014]}$$

The change rate of built-up surface area informs about the increase of built-up surface in-between the observed epochs, and is required by spatial policy indicators (SDG11 for example). We calculate change rate of built-up surface in the admin units used as test cases in the experiment, stratified per URBAN and RURAL domain as determined by the GHS-SMOD R2022A predictions at 1km of spatial resolution.

$$BC_t = \frac{B_{t+1} - B_t}{B_t} \times 100\%$$

Where B_t is the built-up surface predicted in time *t* in a given admin unit used as test case in the experiment, and in the specific URBAN vs. RURAL strata. The MAE of the built-up surface persistence factor $\beta_{t\in[1975,1990,2000,2014]}$ was computed for each model ensemble, in each observed change epoch, for *N* number of admin units used as test cases in the experiment (samples):

$$MAE_t = \frac{\sum\left|BC_{t\_obs} - BC_{t\_pred}\right|}{N}$$

We apply Pareto multi-objective optimization (Pareto 1912) for selecting the optimal solution from the model ensemble for each change epoch. The ensemble predictions are evaluated under three objectives:

1. minimize the Ruzicka distance of the predicted vs. the observed change grids
2. minimize the MAE of the change rate error in the URBAN application domain stratum
3. minimize the MAE of the change rate error in the RURAL application domain stratum

For each change epoch, a set of non-dominated solutions (Pareto front) is ranked by minimization of the distance to the optimal (Utopian) solution, with the highest performing model selected as the final solution for the given observed epoch (1975, 1990, 2000, 2014).

3.8 Building 5-year interval time-series of built-up surfaces and volumes

3.8.1 *General workflow for 5-year interval time-series*

The general logic of the spatial-temporal interpolation and extrapolation solution applied here is summarized below:

1. Analyse the multi-temporal evolution of the *BU support domain* defined as the spatial domain where predicted built-up surface is greater than zero.

$$BU^{x>0}_{t\in[1975,1990,2000,2014,2018]} = x: \; f_x BU^{S}_{TOT\ t\in[1975,1990,2000,2014,2018]} > 0$$

2. Predict the sum of the BU support domain $\sum_x BU^{x>0}_{t\epsilon 1975:5:2030}$ per data tile of 100x100km
    a. in the years 1975:5:2020: solved by piecewise linear interpolation based on nearest observed temporal anchor points
    b. in the years 2025,2030: solved by extrapolation of the curve fit determined at point 1.
3. Predict the $f_x BU_{t\in 1975:2030}$ quantitative grids per data tile of 100x100km
    a. In the years 1975:5:2020: solved by linear interpolate the f_x predictions based on nearest observed anchor points.
    b. In the years 2025, 2030: solved by f_x spatial pattern generative algorithm
4. Spatially allocate the predicted $BU^{x>0}_{t\epsilon 1975:5:2030}$ support domain samples in the $t\epsilon 1975:5:2030$ equal-time interval by spatial rank optimization, knowing the expected $\sum_x BU^{x>0}$ per data tile (100x100km).
5. Calculate the final spatial-temporal interpolated $f_x \ddot{BU}$ quantitative grids as product of the $f_x BU$ quantitative grids by the $BU^{x>0}$ Boolean support domain predicted in the $t\epsilon 1975:5:2030$:

$$f_x \ddot{BU}_{t\in 1975:5:2030} = f_x BU_{t\in 1975:2030} \cdot BU^{x>0}_{t\epsilon 1975:5:2030}$$

Step 1 is solved by polynomial curve fitting of the $\sum_x BU^{x>0}$ sum per data tile (100x100km) in function of the time. Two scenarios are developed: i) first order (linear) polynomial, ii) second-order polynomial. Final prediction of $\sum_x BU^{x>0}$ in function of the time, based on the median solution between the two scenarios.

Step 4 of the above schema is solved by spatial-temporal interpolation or extrapolation of five observed epochs (1975, 1990, 2000, 2014, 2018), based on a rank-optimal spatial allocation method. This supporting spatial optimization function combines static and dynamical components: the static component, Empirical Land

Suitability (ELS), is determined by the observation of the empirical association between the occurrence of specific land form combinations (slope, elevation, water and high vegetation) and the occurrence of human settlement development from remotely sensed data. The empirical association is measured using symbolic machine learning (M. Pesaresi, Syrris, and Julea 2016). The dynamical component (BUDYN) 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.

3.8.2 *Empirical Land Suitability*

The "Empirical Land Suitability" (ELS) is a spatial grid representing the probability of a given grid cell of becoming developed based on the empirically and locally estimated association between the increase in built-up surface in period 1975-2020 and the land form combination. The built-up increase domain is defined by 100 m resolution grid cells where surface of built-up area in 2020 is greater than surface in 1975. The land form combination is evaluated at the spatial resolution of 100m. It includes elevation, slope, distance to water, and distance to high vegetation. The elevation information is derived from global open Copernicus DEM (GLO-30), at 30 m spatial resolution (European Space Agency and Airbus 2022) generalized to the 100m-res by surface-weighted average. Slope is generated by application of morphological grey-tone gradient defined as dilation minus erosion (Serra 1982) with a flat structuring element of 3x3 pixels to the 100m-res generalized elevation layer. The presence of water is derived from GHSL land fraction grid (Martino Pesaresi and Politis 2022), for pixels with land fraction lower than 0.9. The presence of high vegetation is derived from NDVI composite (Corbane et al. 2018), using pixels with values greater than 0.5. Distance to water and distance to high vegetation information layer is derived by application of Euclidean distance function. We apply associative rule learning algorithm on quantized land form combination per 100x100 km data tile, thus the output Φ_{ELS} expresses the empirical association (range [-1..1]) between observed land forms and the presence of new built-up surfaces in period 1975-2020, in the specific 100x100km data tile where the inference was evaluated (Figure 31).

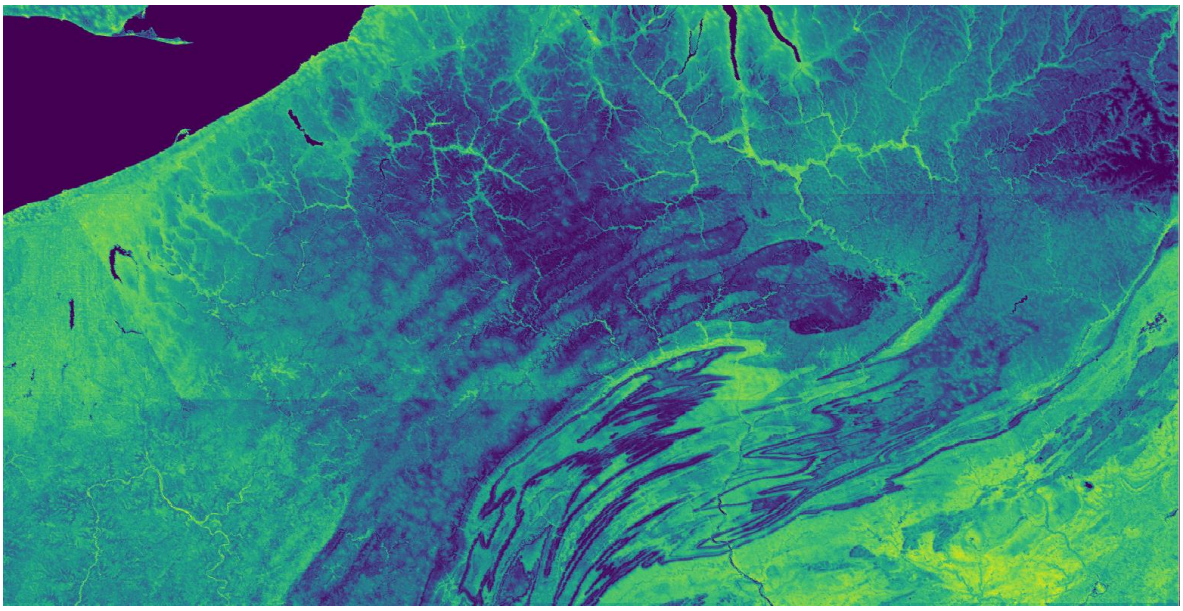


Figure 31 – ELS spatial grid for area of the city of Pittsburgh, Pennsylvania (US). ELS values grow from dark blue (low suitability) to light green (high suitability).

3.8.3 *Dynamic trends in built-up surface (BUDYN)*

Built-up dynamics (BUDYN) is a spatial estimation of the expansion potential of the built-up domain in time, decomposed in two factors

- "expansion" factor f^+ : the tendency to expand the settlement, as determined by human need interlinked with local conditions/opportunities Ω
- "inertial" factor f^- : the tendency to maintain the status quo, as determined by human need interlinked with local conditions/opportunities Ω

Both f^+, f^- are expression of the same implicit drivers: local human needs interlinked with local conditions/opportunities Ω. The effect of those drivers can be measured empirically from the observed data on the built-up surfaces over time, without the need to collect additional spatial information about forbidden expansion areas (as protected areas, religious or cultural heritage respect areas and similar), plans, regulations, local practices, market preferences etc. as it is common practice in spatial temporal extrapolation models.

The factors f^+, f^- are formalized in the spatial grid of 100m-res by the application of two spatial operators to the spatial variable f_x:

a) The probabilistic potential field of f_x at the scale *s*: $P_s(f_x)$. Solved by the application of a Gaussian low-band-pass filter with a kernel of size *s*.
b) The spatial expansion of f_x at the scale *s*: $D_s(f_x)$. Solved by the application of a morphological dilation of f_x using a non-flat structuring element of spherical shape and size *s* (Haralick and Shapiro 1992).

The measured change and inertia assuming two points in time *t1>t0* are formalized as

a) Measured positive change (expansion): $C1_x = x : f_xBU_{t1} > f_xBU_{t0}$
    - *Definition: "measured change from t0 to t1 is greater than zero at the sample x point"*
b) Measured zero change (inertia): $C0_x = x : \begin{cases} D_s(f_xBU_{t0}) > 0 \\ f_xBU_{t1} - f_xBU_{t0} = 0 \end{cases}$
    - *Definition: "the sample x could have changed because spatially close to positive evidence at t0, but it was observed as stable at t1"*

The potential fields of the measured changes and inertia are defined as

a) The potential of expansion: $PC1_x = P_s(C1_x)$
b) The potential of inertia: $PC0_x = P_s(C0_x)$

Finally, the spatial expansion of the potential fields f^+, f^- are defined as

a) Spatial expansion of the expansion potential $f^+ = D_s(PC1_x)$
b) Spatial expansion of the inertial potential $f^- = D_s(PC0_x)$

The BUDYN model is substantially deductive, thus designed to work with a limited number of assumptions in unsupervised way. The only inductive parameter is the scale *s* informing the size of the spatial potential $P_s(f_x)$ and the spatial expansion $D_s(f_x)$ operators. The scale *s* was estimated by observing the average size of built-up domain expansion defined as $f_xBU > 0$ in the maximum temporal period considered in the study (1975-2020). The *s* estimate is done by each processing data tile of 100x100km. Figure 32 shows the output of the BUDYN in the territory of the plain around the city of Troyes (FR). The factors f^+, f^- are composed in the green and red+blue channels, respectively, of the RGB colour composition. Samples that are built in 2020 are masked out with a grey tone. In this representation, samples with magenta tone have a dominant f^- component, samples with green tone have a dominant f^+ component, and samples with white tone have both f^+, f^- factors equally active. E.g. in predicting which spatial samples will be switched on in the future 2025 point in time, the samples with a green tone will have a precedence respect to the ones with a magenta tone. It is worth to notice that the BUDYN model provides spatially anisotropic predictions informed by the spatial-temporal trends measured from data. Similar size settlements, in the same geographical conditions (alluvial plane, no visible physical obstacles) can be predicted to behavior very differently in the temporal domain: case A more dynamical expansion, vs. case B where the dominant component is static inertial. The reasons why those two settlements are performing differently in the temporal domain remain implicit in the model without the need to introduce new supporting data and explanatory hypothesis, as could be the introduction of propriety data or local spatial regulation data that may contribute to explain the observed spatial-temporal pattern. The similar observation can be done for the case C: this small town exhibits a clear preference of anisotropic expansion on the right side, while the left side has a stronger inertial component. In the decision where to allocate new BU support samples the right side will have priority based on past spatial-temporal trends.

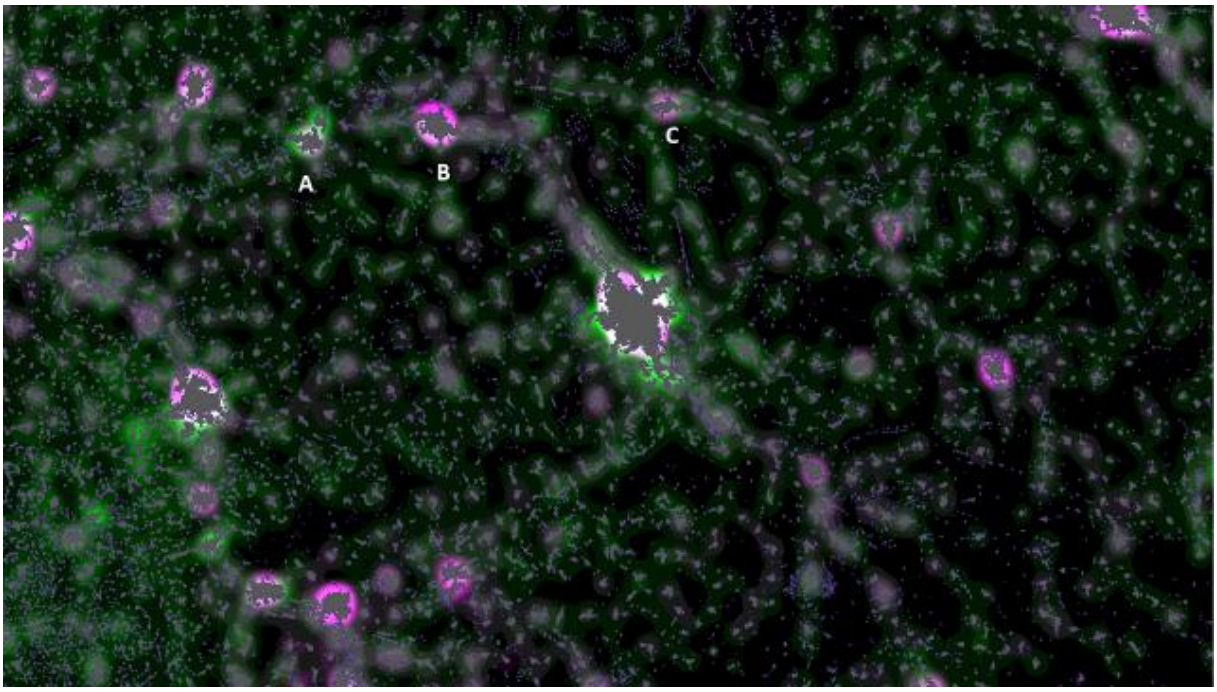


Figure 32 - BUDYN composite for city Troyes in France, showing expansion factor related to the built-up area change between time steps t_1 and t_0; and inertial factor related to the built-up area in time step t_0. Colors indicate the values of each band: expansion – green; inertial – magenta; both – white. Built-up areas in 2020 are marked in grey.

3.8.4 ***Rank-optimal spatial allocation of the BU support domain***

The logical schema of the rank-optimal allocation of the BU domain is as follows:

Given it is known the sum of the samples *n* that should change from the state α: $f_x BU^S_{TOT\ t} = 0$ to the state β: $f_x BU^S_{TOT\ t+1} > 0$, in the change from the time *t* to the time *t+1*, in a given spatial domain *D*.

The solution is provided by ranking all the samples $\alpha : f_x BU^S_{TOT\ t} = 0, x \in D$ by a function objective f_{obj} expressing the probability to change state → β , and thus by selection of the *n* top candidates.

The above process is done piecewise iteratively for all the time points solving the 1975:5:2030 series (see logical schema in chapter 3.8). The scale parameter for calculating the f^+, f^- factors is estimated in the local data tile of 100x100km based on average size of BU expansion patches in the maximum observed time domain (1975-2020).

The time points in the set 1975:5:2020 are solved by the application of allocation criteria AC_i, defined as a volume independent normalized factor of expansion f^+ over factor of inertia f^-:

$$AC_i = \frac{1 + f^+}{1 + f^-} + \varepsilon$$

Where ε is uniformly distributed random variable.

The time points in the set 2025,2030 are solved by the application of allocation criteria AC_e, defined as a volume dependent difference between the expansion factor f^+ and the inertia factor f^-:

$$AC_e = f^+ - f^-$$

Both AC_i , AC_e are bounded in the spatial domain by the condition B defined as the intersection between three criteria:

$$B = x : f_x \Phi_{ELS} > 0 \cap f_x LAND > 0.1 \cap f_x HWAYS < 0.5$$

With $f_x \Phi_{ELS}$ the empirical land suitability encoded in the [-1..+1] range, $f_x LAND$ the land surface share in the sample (Martino Pesaresi and Politis 2022), and $f_x HWAYS$ is the share of the surface of the sample covered

by high ways from OSM data (Martino Pesaresi and Politis 2023). The first criteria is adaptive to the empirical conditions observed in the data tile domain of 100x100km, while the two others are universal and valid globally.

3.8.5 ***Spatial pattern generative algorithm***

The spatial pattern generative algorithm (SPGA) is applied for solving the Step3.b of the general workflow (3.8.1), regarding the prediction of the $f_x BU_{t\in 2025,2030}$ quantitative grids per data tile of 100x100km in the time points 2025, 2030. The SPGA output is used in input together with the Boolean BU domain solved by spatial rank-optimization, in the multiplicative model supporting the overall spatial-temporal interpolation process (see point 5 of the general workflow 3.8.1). Accordingly, the objective of the SPGA is to fill the whole spatial domain with the patterns detected in the $f_x BU$ quantitative spatial grids. Leaving to the Boolean mask determined by the spatial rank-optimization process the task to select which samples generated by the SPGA will be retained in the final prediction.

The SPGA workflow is defined as follows:

1. Select the salient pattern seeds (SPS)
2. Determine the influence zone of each SPS
3. Replicate the pattern inside each specific SPS influence zone

Step1 is solved by morphological opening (Serra 1983) with a structuring element of 3x3 pixels, that corresponds to 300x300m on the ground. The assumption here is that a spatial "pattern" must exhibit at least three consecutive non-zero samples in order to be considered as salient. Step2 is solved by the morphological watershed segmentation of the Euclidean distance function from the SPS binary set (Vincent and Soille 1991). The assumption here is that the decision on which pattern will be replicated is governed by the distance function from the seeds. For each sample to be filled, the nearest SPS is taken in to account as example. Step3 is solved by a traversal recursive replication spatial operator (Rosenthal et al. 1986). The assumption here is that salient patterns will replicate as-it-is in their influence zone.

Figure 33 to Figure 40 show the SPGA process for an example of GHSL data taken from the region around the city of Milano (Italy).

Figure 33 shows the continuous $f_x BU$ surface predicted in 2020 at 100m-res. The $f_x BU$ functions exhibits values in the range 0:10,000 expressing the amount of m^2 of predicted built-up surface in the year 2020. Figure 34 shows the $f_x BU$ filtered by the morphological opening operator, used to select the salient pattern seeds (SPS). Figure 35 shows the influence zones determined by the watershed segmentation on the Euclidean distance function from the SPS. Figure 36 shows the output of the pattern replication process, replicating the pattern inside each specific influence zone. Figure 37 shows (red) the samples that are selected to change state in the future year 2025 according to the rank-optimal spatial allocation of the BU support domain (see 3.8.4), vs. the samples already in the BU support domain in 2020 (white). Figure 38 shows the final prediction for the $f_x \ddot{BU}$ quantitative grids in 2025 as product of the $f_x BU$ quantitative grids generated by the pattern replication process by the $BU^{x>0}$ Boolean support domain predicted for 2025 in the rank-optimal spatial allocation process. Figure 39 and Figure 40 show the same logic applied to the prediction of the year 2030.

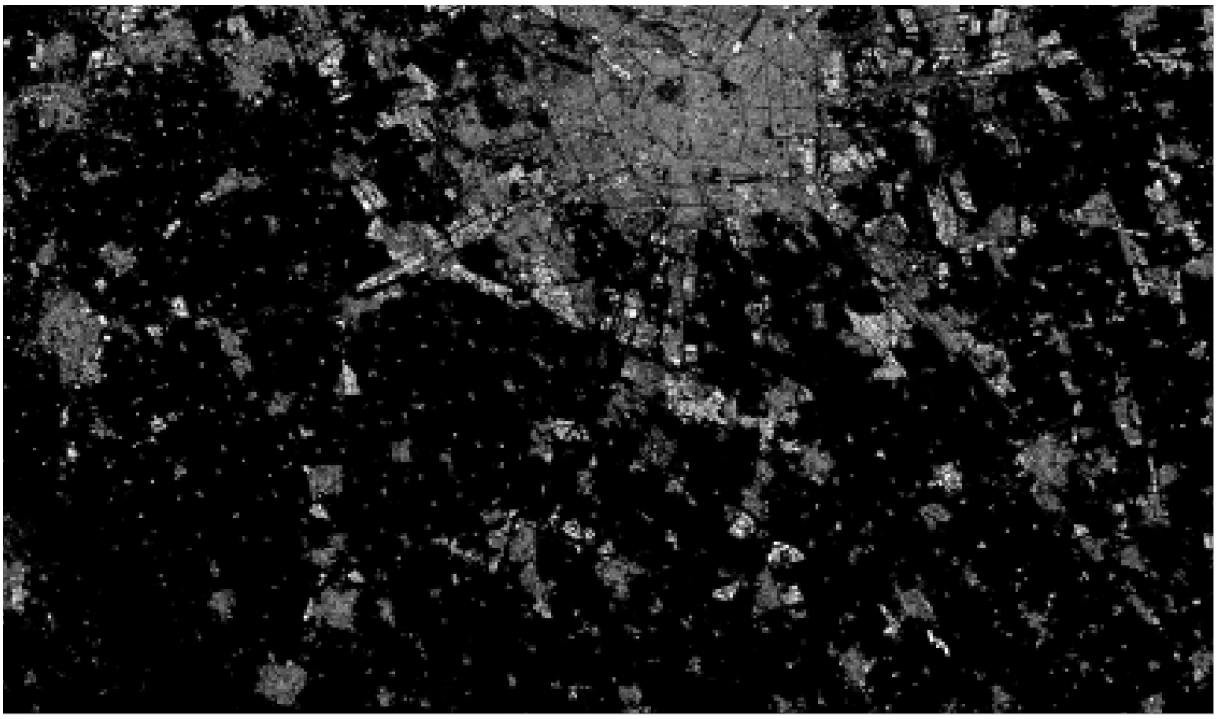


Figure 33 – Continuous f_xBU surface predicted in 2020 at 100m-res

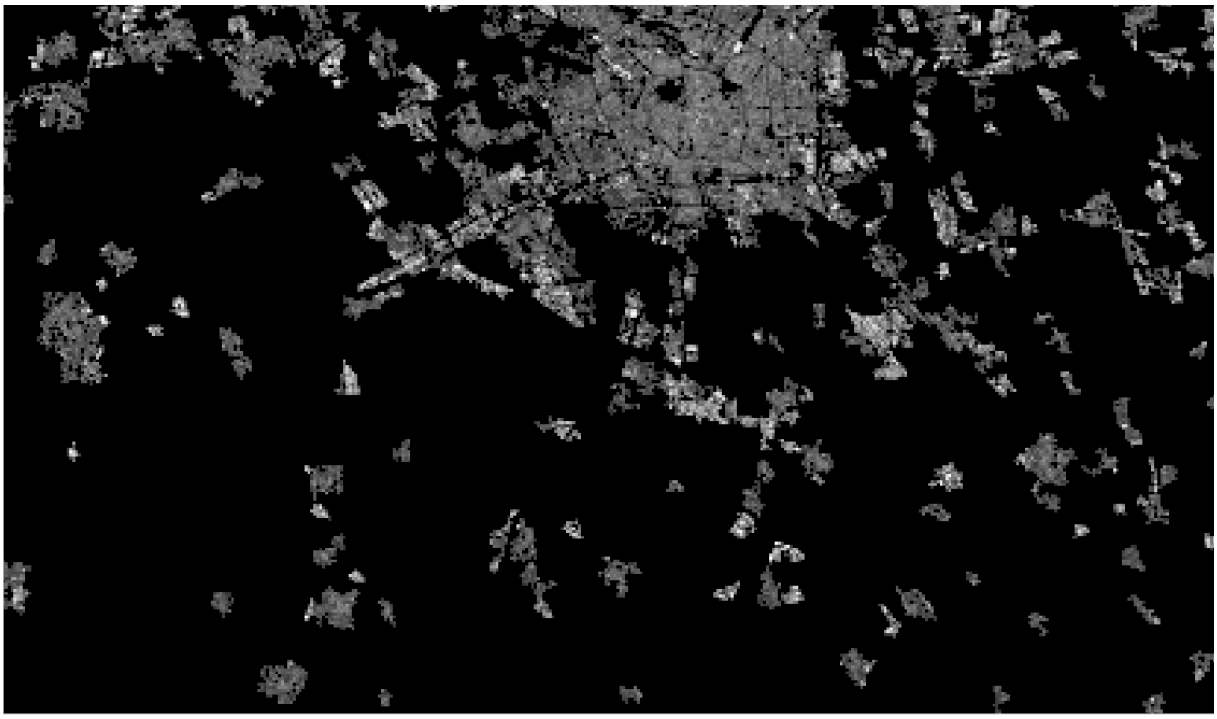


Figure 34 – Selection of the salient patterns seeds (SPS)

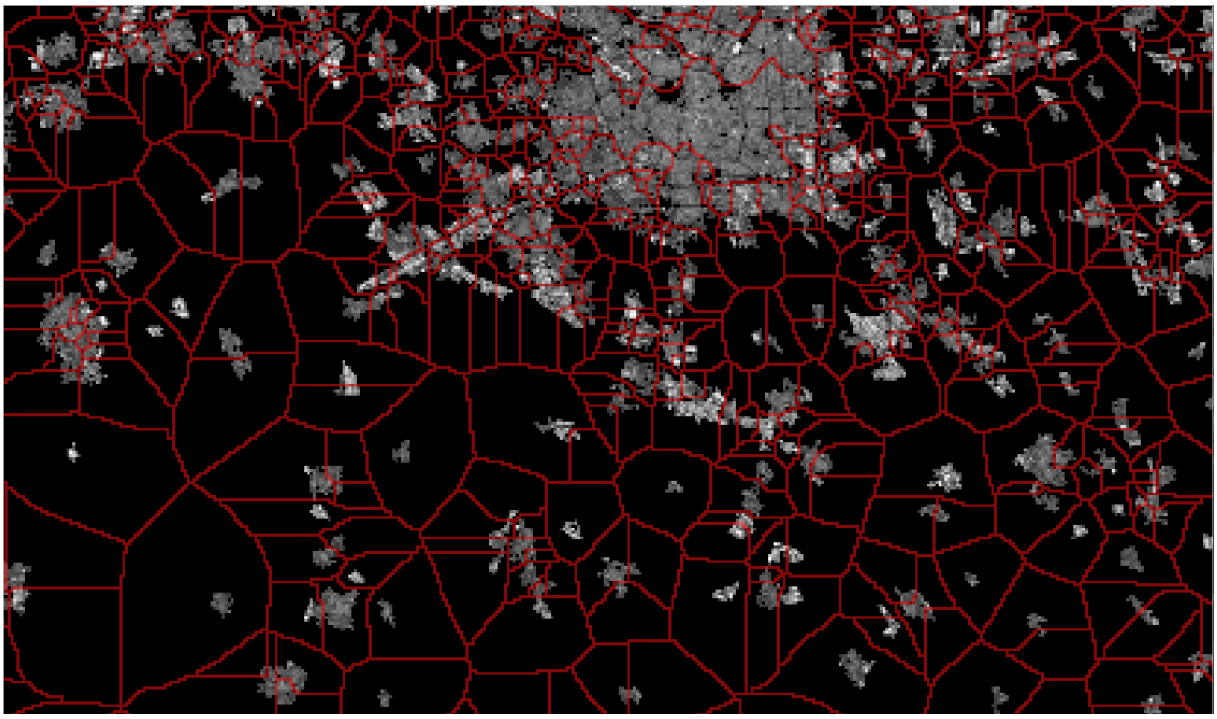


Figure 35 – Influence zones of the SPS

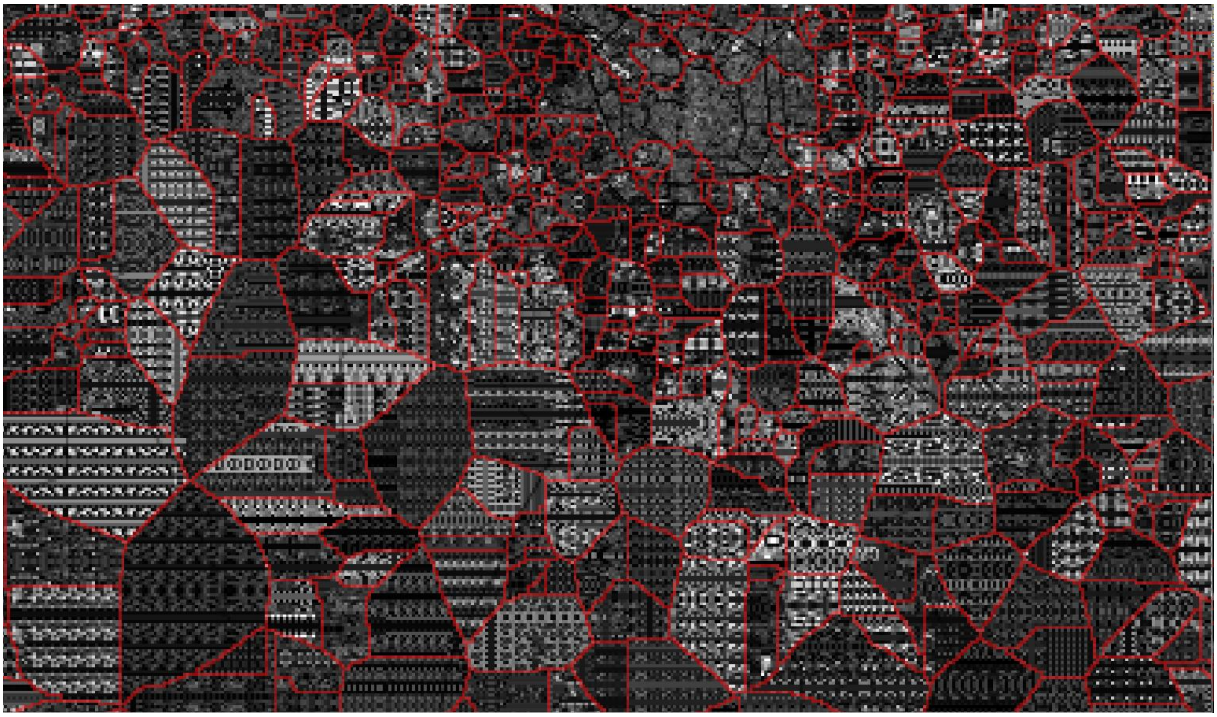


Figure 36 – Replication of the SPS pattern in the influence zones

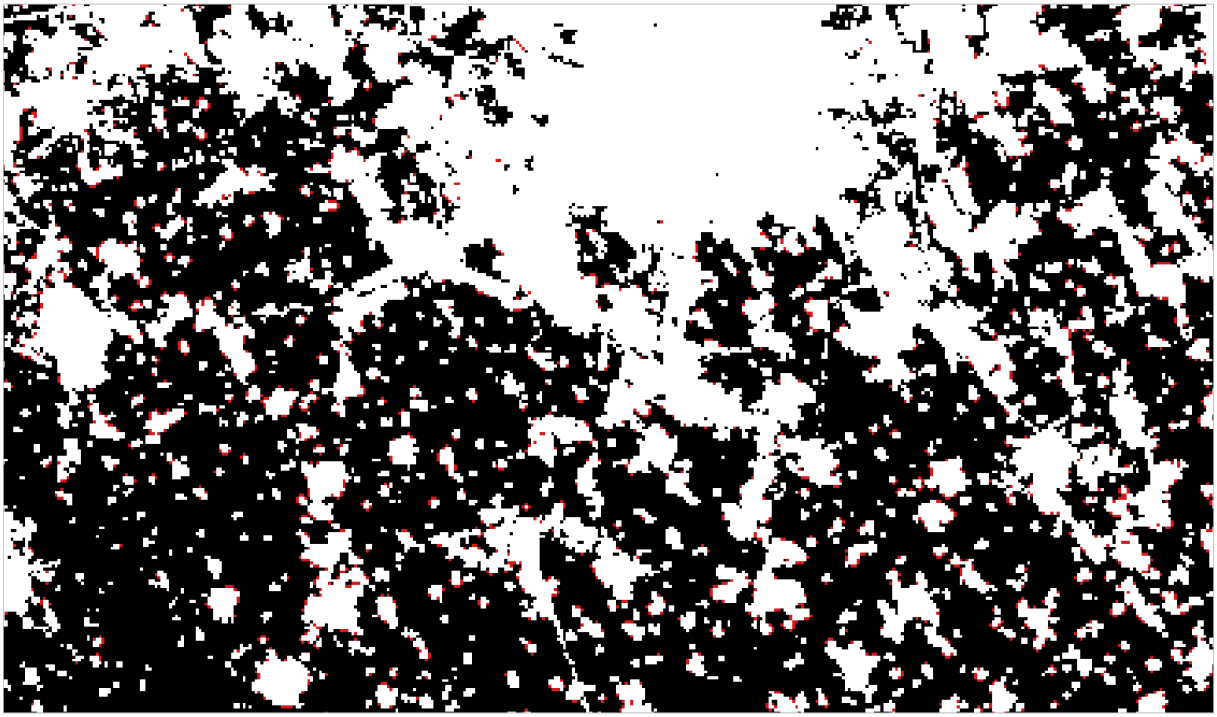


Figure 37 – BU domain in 2020 (white) and BU domain increase 2025 (red) predicted by the spatial-temporal rank optimization process

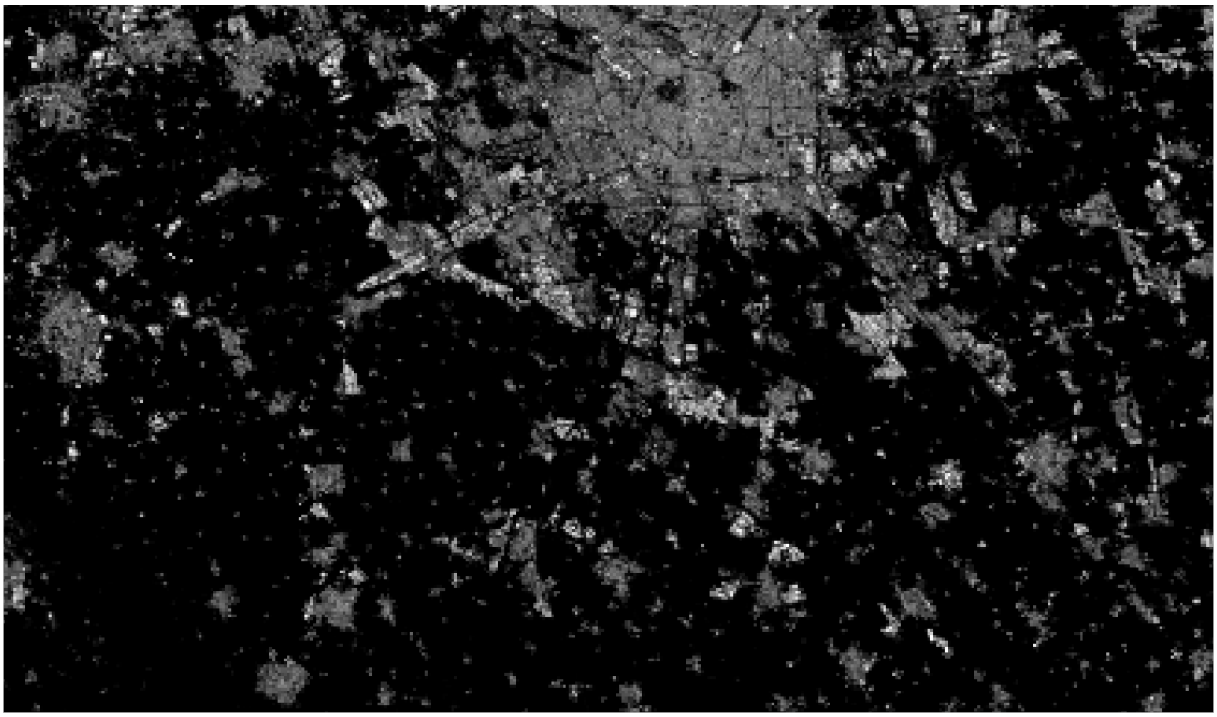


Figure 38 – BU continuous surface predicted in 2025

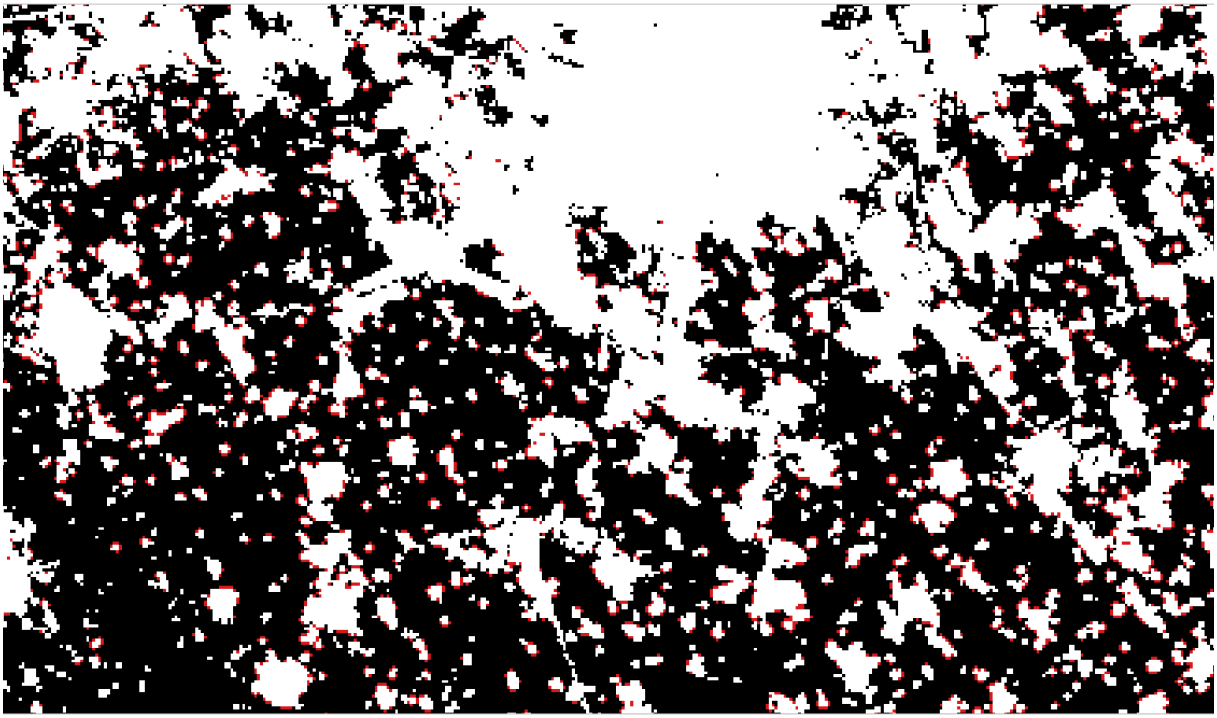


Figure 39 – BU domain in 2020 (white) and BU domain increase 2025 (red) predicted by the spatial-temporal rank optimization process

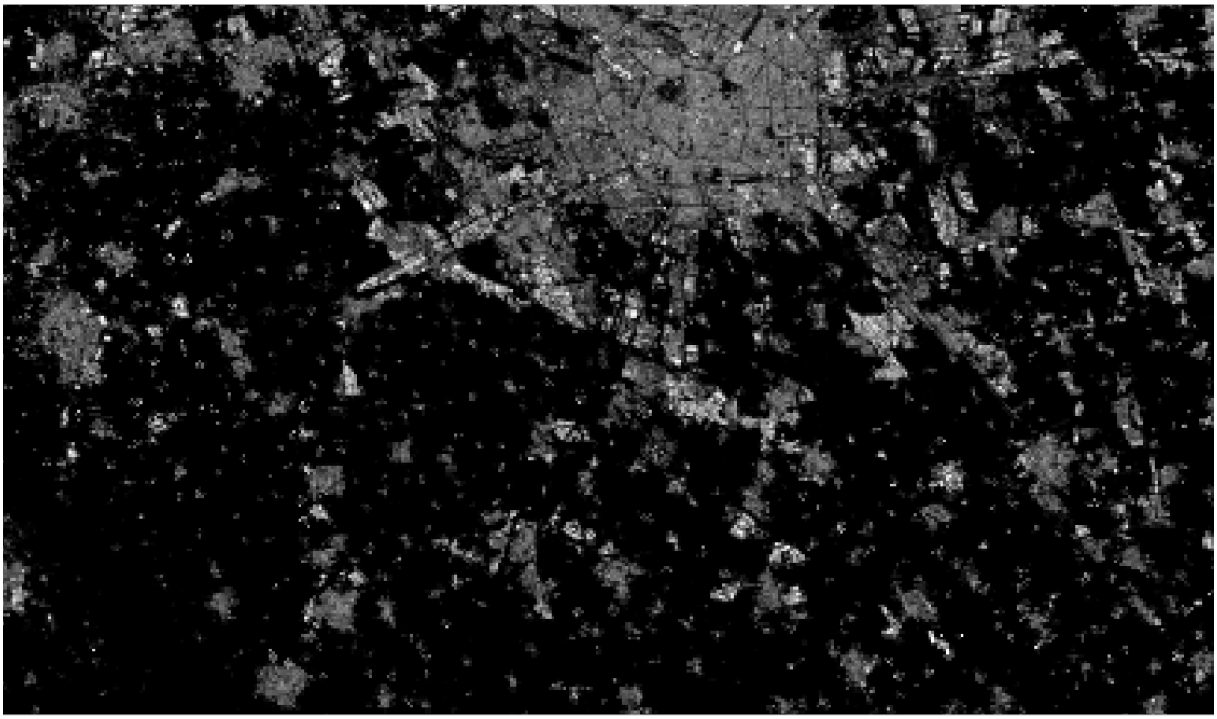


Figure 40 – BU continuous surface predicted in 2030

4 Results

4.1 Ranking of SEGM PHIMAX by strata results

We examined the architecture of the ensemble of the five best performing models of the binary and continuous change detection scheme (Figure 41 - Figure 44). Within the five best performing models, tested in seven strata (COMPACT, SPARSE, RES, NRES, IN_PRIOR, OUT_PRIOR, TOTAL), for all four change epochs combined or separately, we examined which parameters occurred most often. The parameters examined were related to: the stratification (TRAIN_STRATA) and sampling strategy (SAMPLING) used for training set preparation; the training data used to support the change decision (DECISION); the strategy used to maximizes the multi-temporal change decision on the segment level (PHITEMPORAL); the method and semantics for estimating PHI in the change decision (PHITYPE, PHISEMANTIC); and the method for downscaling of change decision to the segment geometry (PHIFUN).

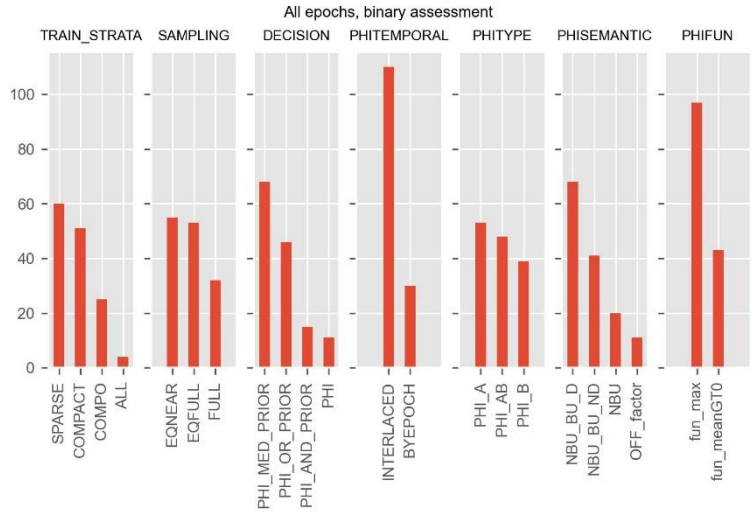


Figure 41 - Architecture of five best performing models of binary change detection schema, tested in seven strata (COMPACT, SPARSE, RES, NRES, IN_PRIOR, OUT_PRIOR, TOTAL), for all four change epochs (n=140). Bars represent the frequency of parameters used in model ensemble architecture.

For all change epochs combined in the binary change detection schema (Figure 41), the majority of models in the selected ensemble was trained using sparse BU samples. Interestingly, examination of model ensemble per change epoch showed the models selected for 1990 and 2014 change epoch performed better trained on sparse samples, while the models selected for the change epochs 1975 and 2000 performed better using compact training data (Figure 42). Equalized random sampling strategy for selecting examples belonging to the semantic background had the highest performance for all epochs combined, and for each change epoch individually. In most cases method of selecting negative examples in the neighbourhood of the positive samples performed the best (EQNEAR) apart from the change epoch 1975, where the EQFULL strategy of sampling whole semantic background of valid data was more effective.

In every case, the interlaced approach, combining various epochs in the maximization of the PHI semantic, had the highest performance. PHI_A was the most common PHI type used to support the change decision, apart from the change epoch 1990, where the PHI_AB type was selected. In every case, the NBU-BU difference semantic (NBU_BU_D) downscaled by the maximal value to a segment level (PHI_FUN) had the highest performance. In the investigated model ensemble, the binary change decision was supported by the combination of the PHI and prior, combined by "OR" or by "MED" operator defined as, respectively, the point-wise maxima and the point-wise median operators on the decision spatial vectors.

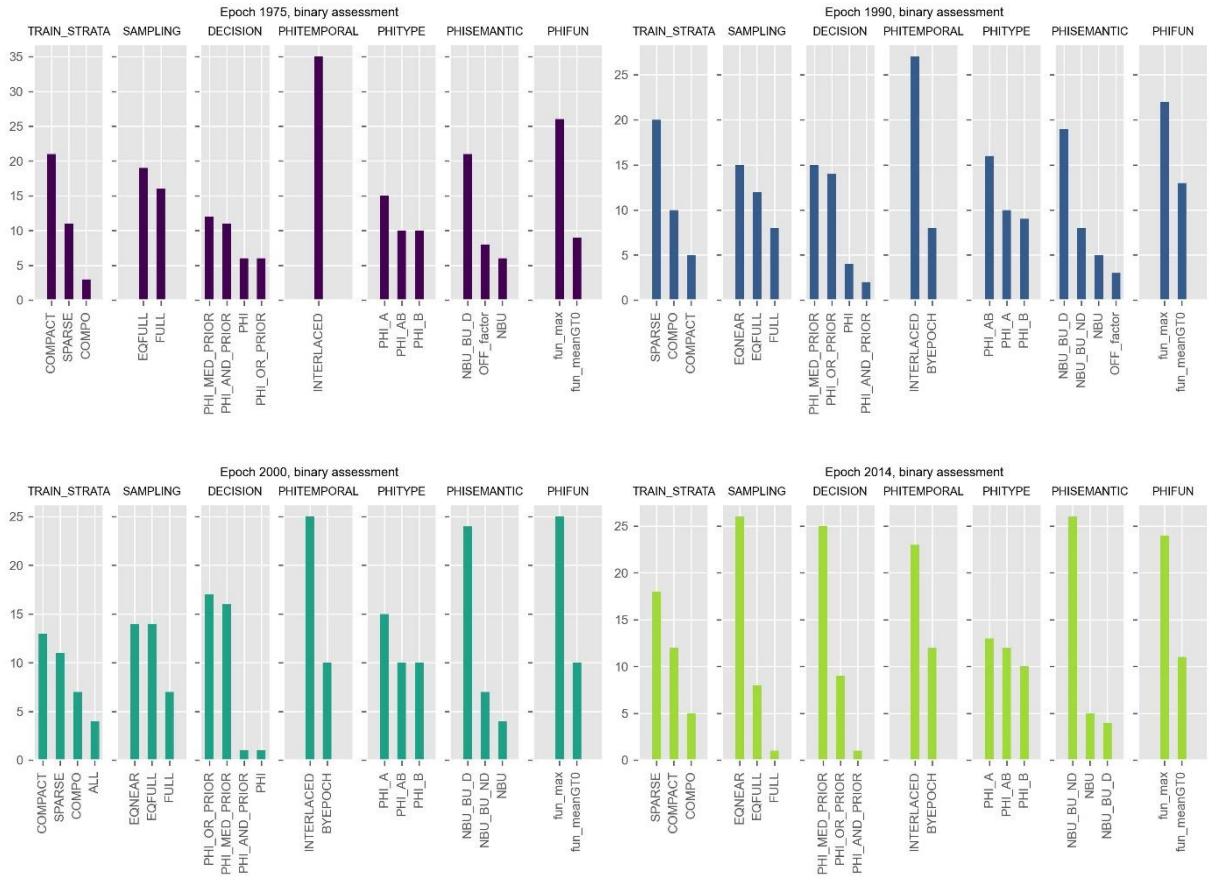


Figure 42 - Architecture of five best performing models of binary change detection schema, tested in seven strata (COMPACT, SPARSE, RES, NRES, IN_PRIOR, OUT_PRIOR, TOTAL), for each change epoch (n=35). Bars represent the frequency of parameters used in model ensemble architecture.

In the continuous change detection schema, the selected ensemble was trained using a composite of the compact and sparse BU samples (COMPO strata) (Figure 43). Noteworthy, only for the change epoch 2014 the model ensemble performed better using the compact training data (Figure 44). Similarly to the binary change detection schema, the EQNEAR equalized random sampling strategy had the highest performance – for every case except for change epoch 2000, where in majority all negative samples were used (FULL).

A different strategy was observed for combining various epochs in maximizing PHI semantic than in the binary approach. The interlaced approach had the highest performance in the change epochs 1975 and 2014 that are placed at the extrema of the temporal window solved by the method, including the most difficult task of detecting thematic changes while changing the supporting image sensors. The independent consideration of each epoch collection (BYEPOCH) performed better in the change epochs 1990 and 2000, connoted by more homogenous sensor specifications. For all epochs combined, both strategies had similar performance. There was no clear advantage in the change decision support strategy, with the PHI_B type performing best across all epochs combined. Comparably to the binary approach, the direct NBU-BU difference semantic (NBU_BU_D) had the highest performance, with the exception of the 2014 change epoch, when the normalized difference (NBU_BU_ND) had a higher rank. Unlike the binary approach, the downscaling of the change decision to the segment level (PHI_FUN) was usually done by averaging of the valid PHI data, apart from the change epoch 2014, when the maximum operator occurred in most model. In every case, the continuous change decision was supported by a combination of the PHI and prior, combined by “MED” operator, expressing the point-wise median operators on the decision spatial vectors.

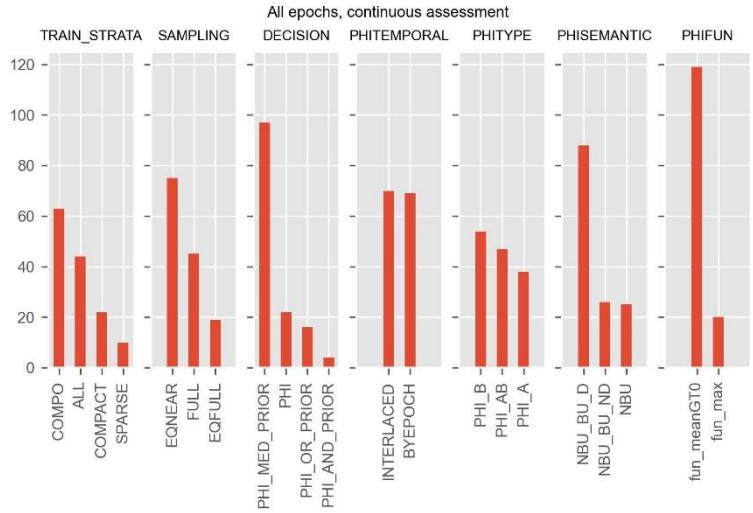


Figure 43 - Architecture of five best performing models of continuous change detection schema, tested in seven strata (COMPACT, SPARSE, RES, NRES, IN_PRIOR, OUT_PRIOR, TOTAL), for all four change epochs (n=140). Bars represent the frequency of parameters used in model ensemble architecture.

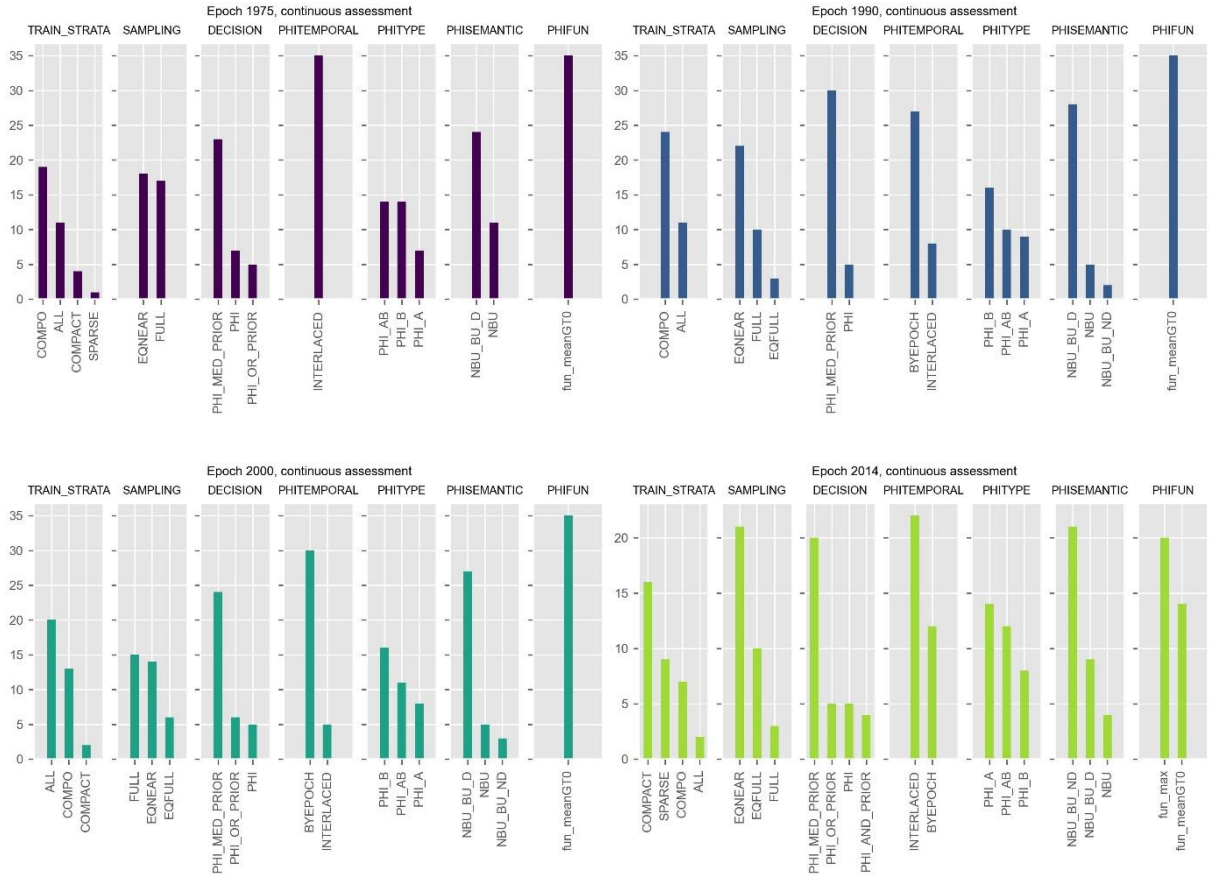


Figure 44 - Architecture of five best performing models of continuous change detection schema, tested in seven strata (COMPACT, SPARSE, RES, NRES, IN_PRIOR, OUT_PRIOR, TOTAL), for each change epoch (n=35). Bars represent the frequency of parameters used in model ensemble architecture.

The results of examining the architecture of the ensemble of best-performing models of the binary and continuous change detection scheme supports the modelling approach used in the study. Firstly, the selection of best-performing models confirms the usefulness of adding supplementary predictions (priors) in the model learning phase. Secondly, the equalized random strategy for sampling the semantic background proves to be

useful, especially the EQNEAR strategy of sampling negative examples in the close neighbourhood of positive samples.

4.2 Evaluation of the ensemble decisions

Figure 45, Figure 46, Figure 47, and Figure 48 show the performances of the ensemble decision options for solving the change between the anchor point 2018 and the epochs 2014, 2000, 1990, and 1975, respectively. Empty squares represent BIN ensembles. Filled round represent CON ensembles. Pattern/colours represent different strategies for solving the stratified multiple decision support. The benchmark is the R2022A release, represented as a bold disk point. The performances are summarized as J-accuracy of the 100m-res change grids ("grid accuracy") on the horizontal axis, and the change rate error (MAE) on the vertical axis, related to both the RURAL and URBAN application domains.

As can be noticed, many ensemble models may provide better solutions than the benchmark R2022A showing there is a high potential for improvement. In particular, BIN ensemble models can reach a lower average URBAN-RURAL aggregated change rate error, but at the price of a lower grid accuracy as compared to the CON ensemble solutions. In general, and for all the considered change epochs, CON ensemble models show a better joint performance of grid accuracy maximization and change rate error minimization. This fact is confirmed by the results of the Pareto model selection.

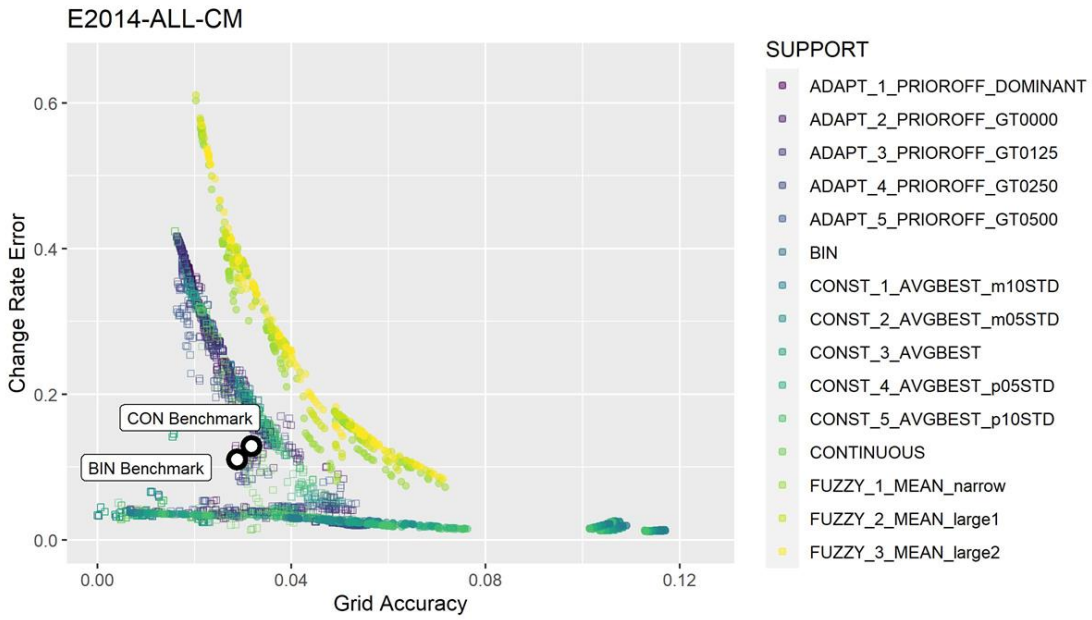


Figure 45 – Ensemble decision options for solving the change 2018 – 2014. Empty squares: BIN ensembles. Filled round: CON ensembles. Pattern/colours different strategies for solving the stratified multiple decision support. The benchmark is the R2022A release.

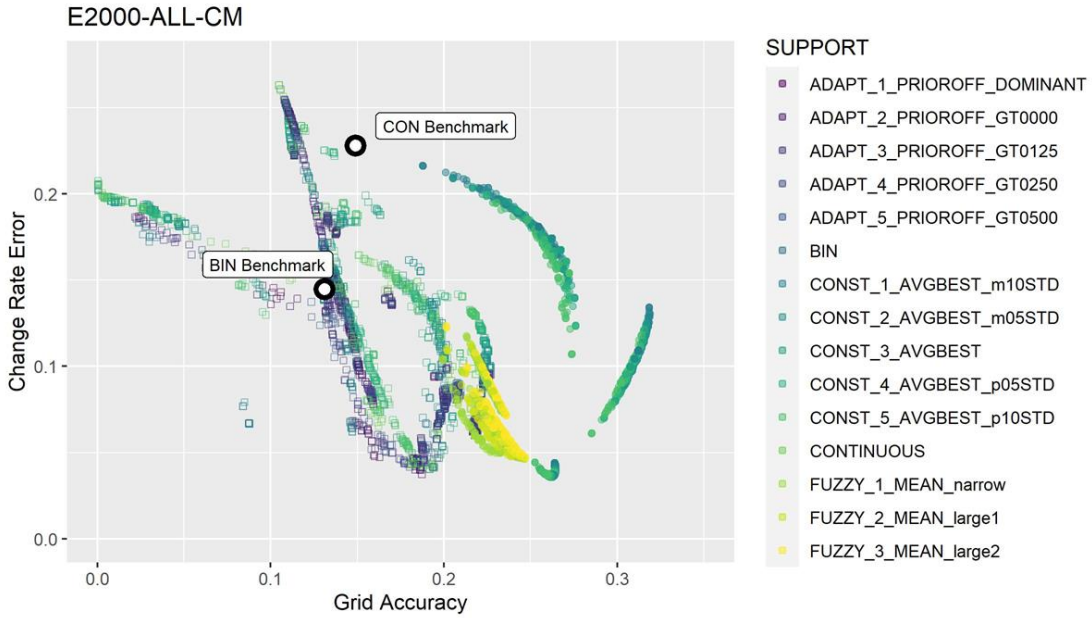


Figure 46 - Ensemble decision options for solving the change 2018 – 2000. Empty squares: BIN ensembles. Filled round: CON ensembles. Pattern/colours different strategies for solving the stratified multiple decision support. The benchmark is the R2022A release.

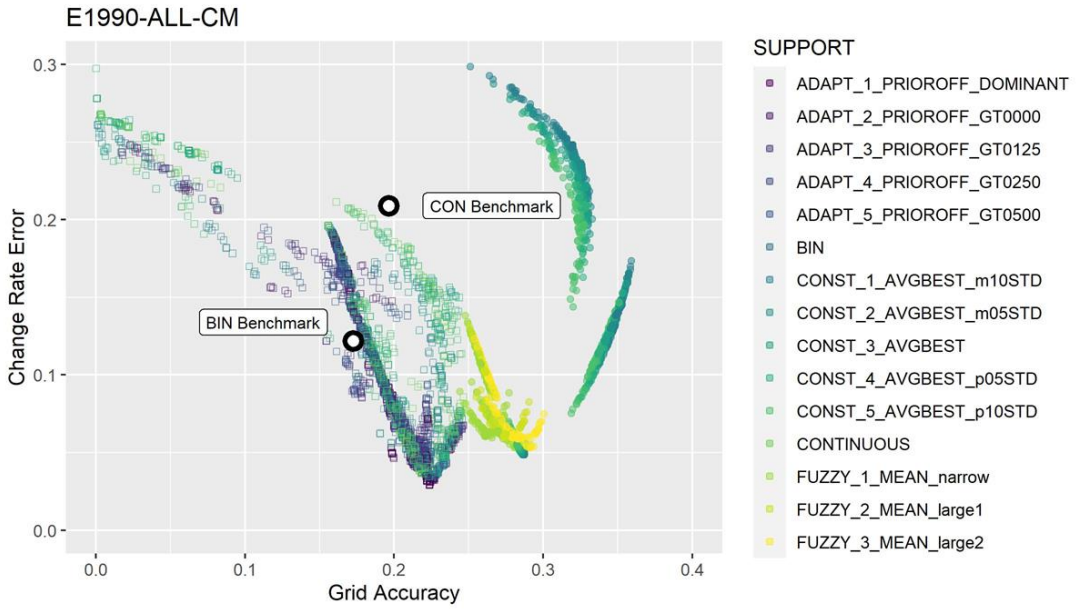


Figure 47 - Ensemble decision options for solving the change 2018 – 1990. Empty squares: BIN ensembles. Filled round: CON ensembles. Pattern/colours different strategies for solving the stratified multiple decision support. The benchmark is the R2022A release.

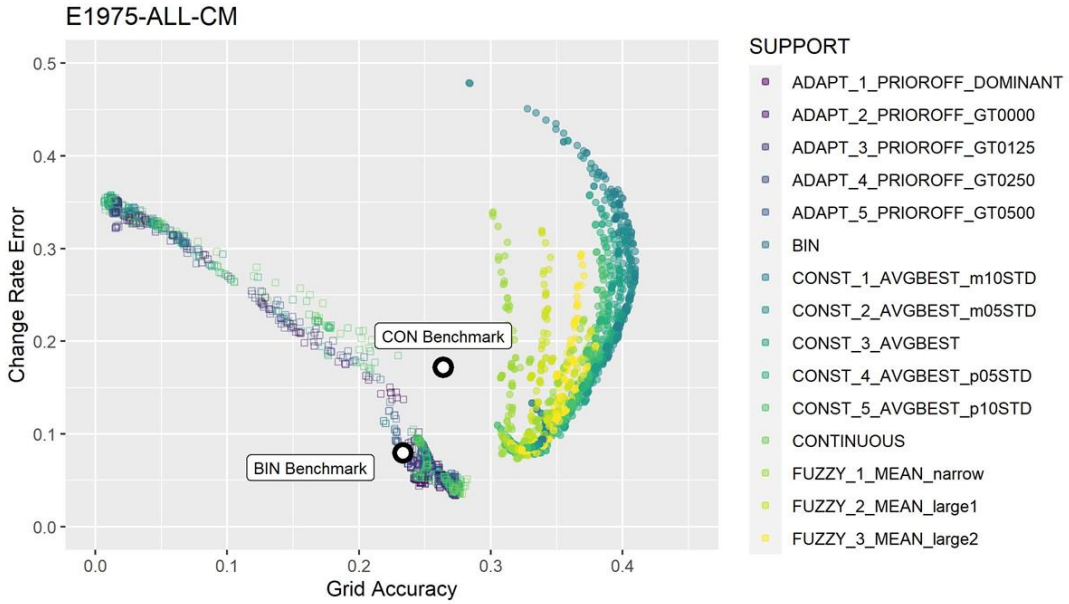


Figure 48 - Ensemble decision options for solving the change 2018 – 1975. Empty squares: BIN ensembles. Filled round: CON ensembles. Pattern/colours different strategies for solving the stratified multiple decision support. The benchmark is the R2022A release.

4.3 Final ensemble selection

Table 10 shows the final ensemble selection by the PARETO multi-objective optimization. Important findings are:

1. LOGIC: The CON branch outperform the BIN branch in all the change epochs: thus mechanism based on linear regression are better than solutions based on hard decisions of on/off. This is true for all the change epochs considered.
2. DECISION-TYPE: The stratified approach (COMPOSITE_BEST_STRATA) always outperform the single-solution approach (SINGLE_BEST_TOTAL), in all considered change epochs. Thus confirming the choice to focus the machine learning on specific challenges as regarding the contextual conditions is a good choice (*dividet et impera*).
3. NBESTMODEL: The best number of supporting models for each stratum to be included in the ensemble decision is not constant across predicted time of changes. As general tendency it seems the optimal number of supporting models is grossly proportional to the time gap to be solved. This may be explained by the fact that classification uncertainty increases by increasing time gap because of decreasing quality of the supporting imagery, thus the inclusion of a larger number of best models stabilize the results and increases robustness of the predictions.
4. SUPPORT: The best approach for setting the support for the regression coefficients or other inferential parameters is to rely on the parameters collected during the model calibration assuming them as constant across the various data tiles solving the global data. This is the “CONST” option as opposed to the “ADAPT” option that was tested as well, including adaptation of the parameters, based on the observation of the statistical distribution of priors in the specific data tile. In the CONST option the parameter tested in the Pareto optimization is a function of the mean and standard deviation of the best parameters discovered during the model calibration. Interestingly, the winning parameters have a standard deviation from the average that is decreasing by increasing the time gap of the predicted change.
5. FINALDECISION: testing different approaches for composing the decisions collected from different strata, various number of supporting models each. The winning approach is to make the sensitivity to change inverse proportional to temporal window to be solved. In short-term predictions better to rely on maximal change mechanism (2018-2014), becoming an average (2018-2000), then a 0.1 quantile (2018-1990) and finally a more conservative 0.4 quantile in the 2018-1975 prediction.

6. POSTPROECSSING: testing which approach was preferable for gap filling: interpolating vs. injecting the priors ("HYBRID" design) in an inclusive vs. and exclusive way, respectively taking in to account interpolated values or disregarding them and inject just the prior predictions. The HYBRIDEXCLUSIVE approach was considered preferable in the Pareto optimization process.

Figure 49 shows the performance of tested model design during the multi-objective optimization, as regarding the composed three criteria in the Utopia accuracy metric.

Table 11 shows the comparison of the performances expected from the ensemble models selected by the Pareto optimization vs. the benchmark R2022A, relatively to the four change epochs 1975, 1990, 2000, and 2014. It is worth noting the general increase of performances both in terms of decrease of the MAE of the predicted BU change grids and increased realistic aggregated change rates, with a off change rate gain closer to one. In particular, it is important to notice the improvement related to the predicted change in the period 2014-2018, connoted by a short time interval (little change information) and a change of sensor from Landsat to S2 (large background noise), thus expressing the worst case as signal-to-noise ratio conditions. In the RURAL application domain the new ensemble model is largely reducing the MAE of the predicted BU surface change at the 100m grid cell (40 m^2 vs. 239 m^2 of the benchmark). Even more importantly, the new ensemble model produces an aggregated change rate more realistic, then more suitable to be used in input of policy indicators. The off change rate gain of the new ensemble solution is estimated as 0.951 (slighly underestimatig change), while the benchmark was 4.917 (largely overestimatig change). Consequently, there is an empricial indication that the the new model development significantly overperforms R2022A benchmark model.

Table 10 – Final Ensemble decision model design selected by the PARETO multi-objective optimization. Row descriptions listed in the model design column corresponds to the terms used in the model implementation.

Model design	1975	1990	2000	2014
LOGIC (BIN/CON)	CON	CON	CON	CON
DECISION-TYPE	COMPOSITE_BEST_STRATA	COMPOSITE_BEST_STRATA	COMPOSITE_BEST_STRATA	COMPOSITE_BEST_STRATA
NBESTMODEL	5	3	4	1
SUPPORT	CONST_3_AVGBEST	CONST_2_AVGBEST_m05STD	CONST_1_AVGBEST_m10STD	CONST_1_AVGBEST_m10STD
FINALDECISION	COMPOSITE_Q4	COMPOSITE_Q1	COMPOSITE_MEAN	COMPOSITE_MAX
POSTPROCESSING	HYBRIDEXCLUSIVE	HYBRIDEXCLUSIVE	HYBRIDEXCLUSIVE	HYBRIDEXCLUSIVE

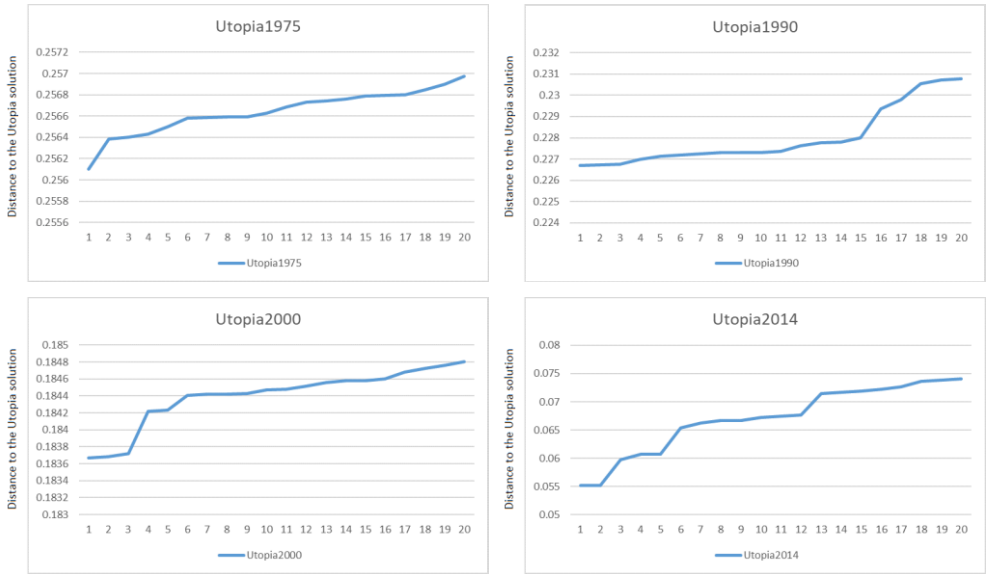


Figure 49 – Distance of the composite of the Ruzicka distance and of the MAE of the change rate error in RURAL and URBAN domain to the optimal (Utopia) solution, for the 20-best ensemble models ranked by the distance, by each change epoch 1975, 1990, 2000 and 2014.

Table 11 – Comparison of the performances expected from the ensemble models selected by the Pareto optimization vs. the benchmark R2022A, for the four change epochs 1975, 1990, 2000, and 2014.

EPOCH	MODEL_ID	MODEL	MAE_ALL	MAE_URBAN	MAE_RURAL	OFFCHANGERATE_gain_ALL	OFFCHANGERATE_gain_URBAN	OFFCHANGERATE_gain_RURAL
1975	**3883**	**PARETO_ENSEMBLE_selected**	**483.54**	**775.05**	**393.29**	**0.962**	**0.961**	**0.986**
1975	1	BENCHMARK_R2022A	640.19	721.09	600.62	1.029	0.751	1.187
1990	**3211**	**PARETO_ENSEMBLE_selected**	**348.97**	**538.96**	**293.21**	**1.013**	**1.020**	**1.041**
1990	1	BENCHMARK_R2022A	615.72	619.17	616.17	1.359	0.954	1.574
2000	**3475**	**PARETO_ENSEMBLE_selected**	**235.80**	**348.51**	**205.43**	**0.959**	**0.920**	**1.015**
2000	1	BENCHMARK_R2022A	544.57	495.19	570.84	1.637	1.057	1.918
2014	**2536**	**PARETO_ENSEMBLE_selected**	**44.90**	**68.24**	**40.11**	**0.893**	**0.786**	**0.951**
2014	1	BENCHMARK_R2022A	206.59	126.06	239.27	3.750	1.293	4.917
total		**PARETO_ENSEMBLE_selected**	**278.30**	**432.69**	**233.01**	**0.96**	**0.92**	**1.00**
total		BENCHMARK_R2022A	501.77	490.38	506.72	1.94	1.01	2.40

4.4 Post-processing options

As already introduced, the prediction of changes in the 1975, 1990, 2000, and 2014 epochs vs. the 2018 anchor point are made by independent ensemble models. As corollary by construction this fact means that i) each sample x at any epoch will have less or equal BU surface than the 2018, and ii) a sample x at the subsequent time t+1 will not have a predicted BU surface necessarily greater or equal than in the precedent time t (Figure 50). Thus in principle the new ensemble model could be used to model both the *increase* or the *decrease* of the BU surface from the past to the recent epochs, being the decrease induced for example by the phenomena of demolishing of built-up structures and re-set the use of the area as parks or other non-built-up uses.

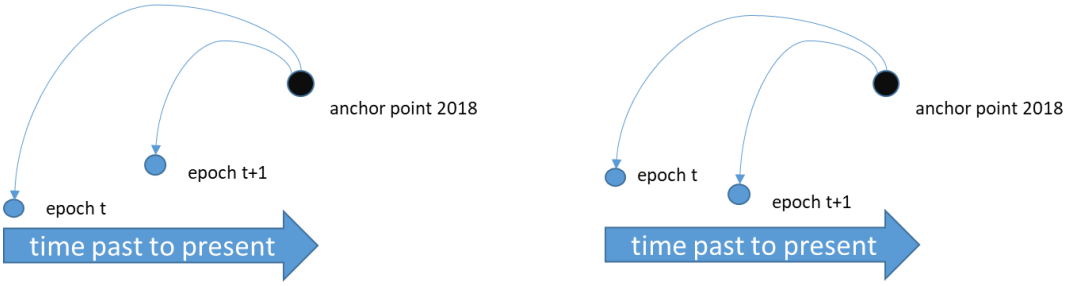


Figure 50 – Independent temporal estimates of the epochs t and t+1, precedent of the anchor point 2018

In practice, the potential claim of the model to predict the *decrease* of built-up surfaces from past to recent epochs cannot be empirical tested with the available MTBF, that unfortunately *are assuming a growing built-up surface by construction of the supporting vector cartographic data*. Moreover, the potentially predicted decrease of BU surface quantity is significantly smaller than the predicted increase of BU surface, thus more exposed to low signal-to-noise ratio generating less reliability of these estimates (Table 12).

Epoch change	Average t->t+1 BU stable	Average t->t+1 BU increase year	Average t->t+1 BU decrease year
1975-1990	200.73 m^2	25.28 m^2	-9.21 m^2
1990-2000	962.52 m^2	25.63 m^2	-3.79 m^2
2000-2014	1 331.80 m^2	20.31 m^2	-4.64 m^2
2014-2018	1 388.61 m^2	18.80 m^2	n.a.

Table 12 – Empirical assessment of the predicted average yearly increase and decrease of BU surface from past to recent epochs in the data tiles included in the test.

Taking in to account the above known constraints, few post-processing options have been evaluated in order to consolidate the ensemble model predictions. In particular, three "temporal constraint" options were evaluated i) "Free" or not constrained temporal output, ii) "Backward constraint" imposing at each sample x less or equal predicted BU surface from recent to past epochs, and iii) "Forward constraint" imposing at each sample x more or equal predicted BU surface from past to recent epochs. Additionally, an "acceleration of change" option (A, B) was tested by imposition of the maximal change rate as inferred by the ensemble model or deducible from the available priors.

4.5 Basic quantitative characteristics of the predicted grids

Figure 51 shows the absolute BU surfaces estimates in the epochs 1975, 1990, 2000, 2014, and 2018, in all the data tile considered. The new ensemble model, in any of the post-processing options, produces a temporal pattern of the BU surfaces that is closer to the REF observed data from the MTBF. Consequently, can be claimed that the post-processing options are expected to produce negligible effects in terms of absolute amount of predicted total BU surfaces. To be noticed also the evident acceleration of change reported by the benchmark R2022A (dash red line, Boolean change map sum), largely overestimating the change rates in the most recent years.

Figure 52 shows the BU surface estimates relative to the 1975, by the different post-processing options and the reference observed data. All the 'A' post-processing options are substantially predicting the observed temporal pattern, while the 'B' post-processing options show an evident acceleration of changes vs. the observed temporal pattern, thus they are rejected.

Figure 53 shows the error (MAE) of the prediction of the relative change rates to 1975, in all the considered epochs (1975, 1990, 2000, 2014, and 2018), for the ensemble model in the 'A' post-processing path and various already existing data/model, including the benchmark R2022A. To be noticed that the three ensemble model options "Free_A", "BackwardConstraint_A", and "ForwardConstraint_A" have substantially the same error rate, that is the best as compared to the other already existing data/models, including the benchmark R2022A that shows the worst performance. The least ensemble model error option is the "Free_A" as it was expected being the one winning the Pareto optimization process. The best option with a form of temporal constraint is the "BackwardConstraint_A".

Figure 54 shows the J-accuracy (Ruzicka similarity) of the BU surface estimates in the 100m-cell samples of the five considered epochs. Comparison of the previous model benchmark R2022A (binary change map version), the new ensemble model development, and all the other already available models. All the new model ensample options show an accuracy greater than three times of all the other alternatives, in all the epochs. The "BackwardConstraint_A" shows an acceptable compromise in terms of accuracy vs. the analogue "B" that scores slightly better as grid cell accuracy but accelerates the aggregated change rate (Figure 52).

Figure 55 shows the error (MAE) of the BU surface estimates in the 100m-cell samples of the five considered epochs. Comparison of the previous model benchmark R2022A (binary change map version), the new ensemble model development, and all the other already available models. The new ensemble model leads to a noticeable lower error of the 100m-cell estimates that is substantially constant across the various epochs.

Finally, Table 13 and Table 14 show the ranking of the prediction/models by the average grid accuracy and the average grid error, respectively, in all the epochs stratified by the degree of urbanization (SMOD) level 2. The

"BackwardConstraint_A" ensemble model option is the most accurate and the least error, thus is the one selected for the production purposes.

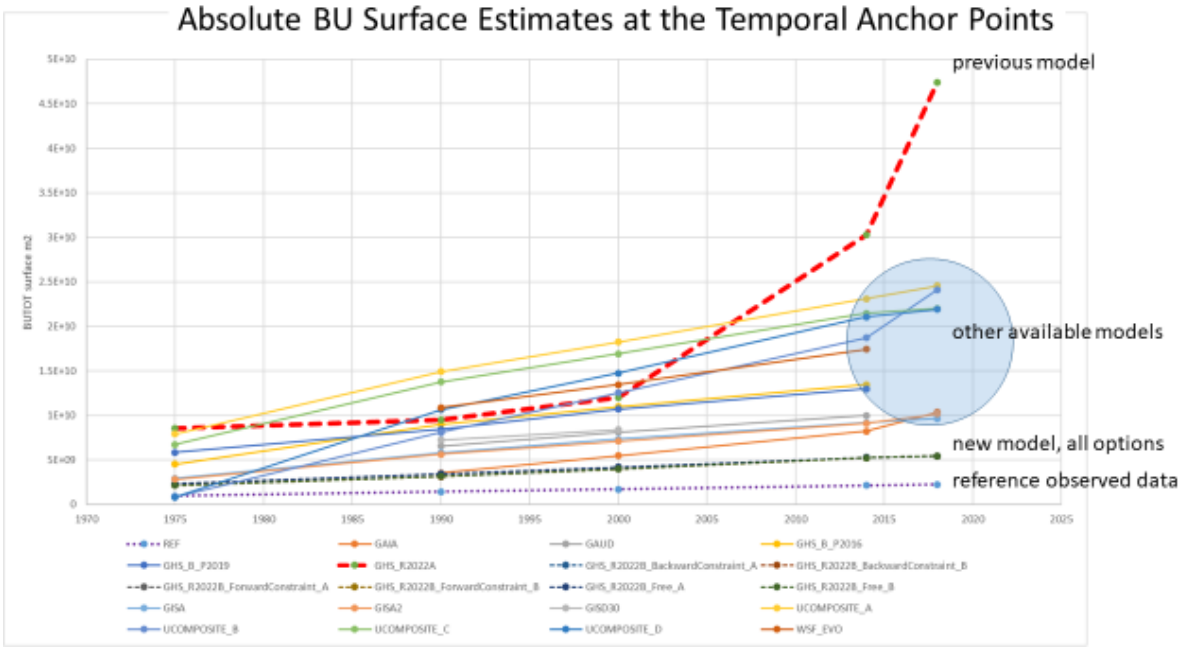


Figure 51 - Absolute BU surfaces estimates in the epochs 1975, 1990, 2000, 2014, and 2018.

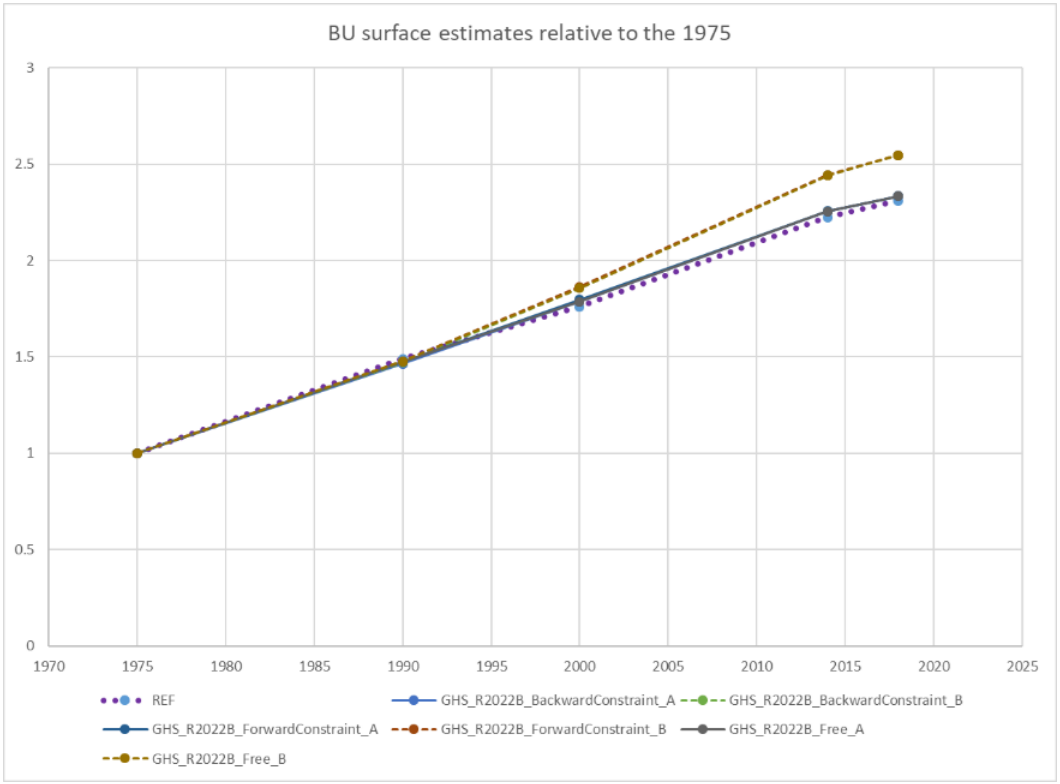


Figure 52 – BU surface estimates relative to the 1975, by the different post-processing options and the reference observed data

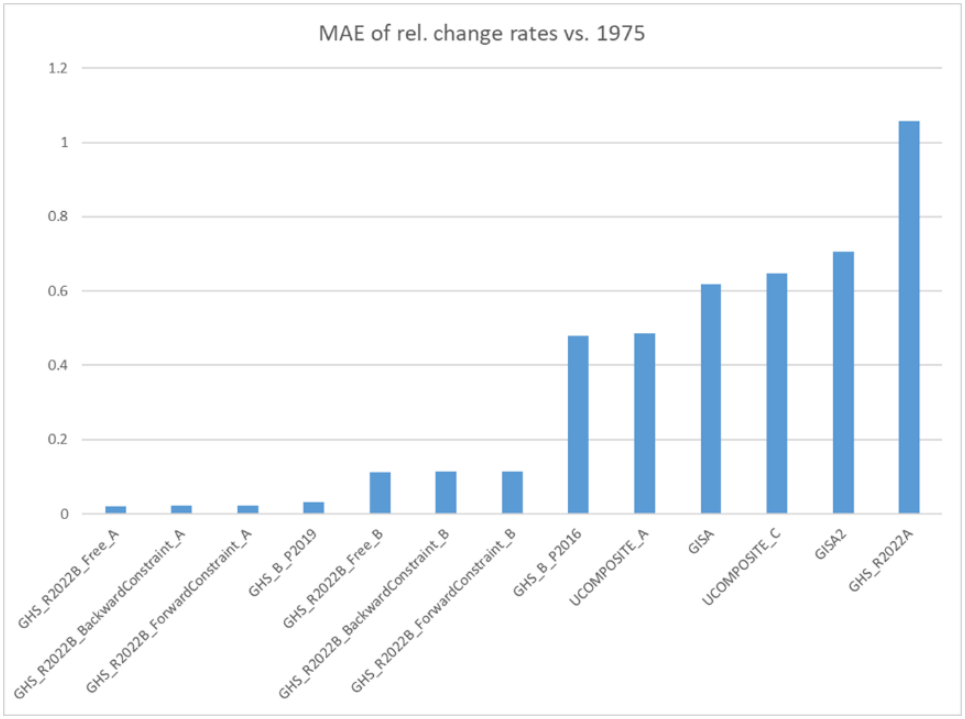


Figure 53 – Error (MAE) of the prediction of the relative change rates to 1975, in all the considered epochs (1975, 1990, 2000, 2014, and 2018)

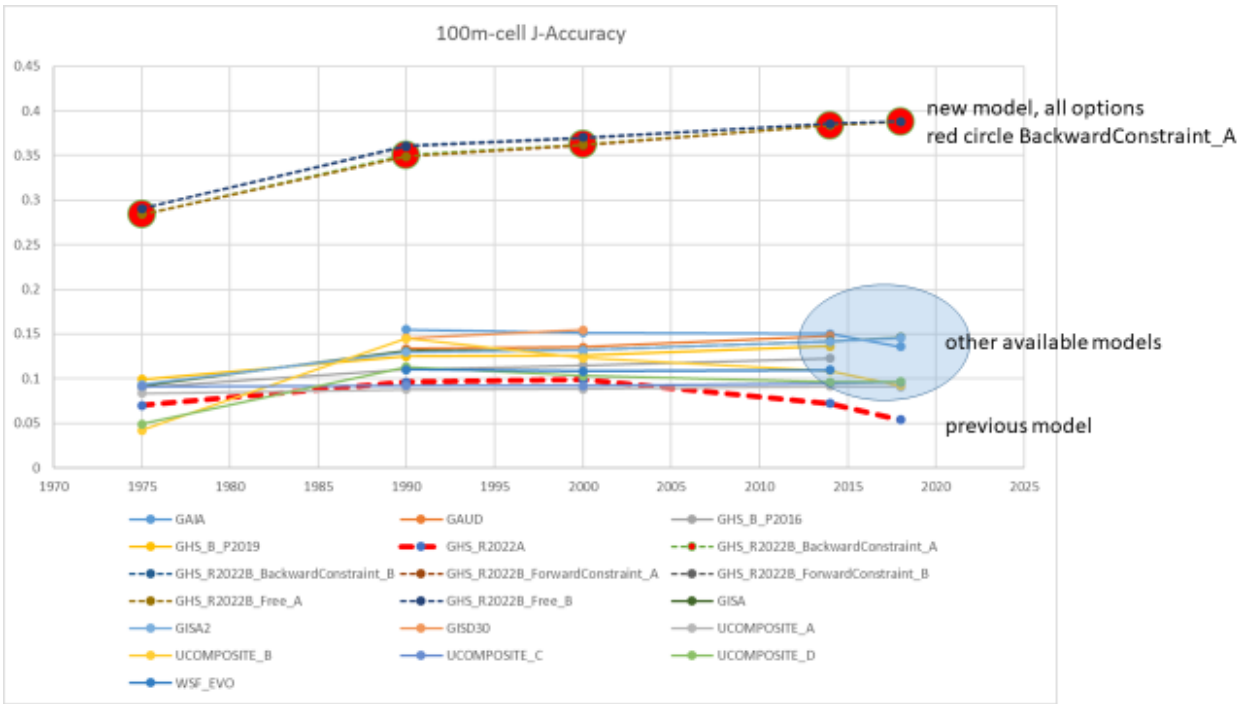


Figure 54 – J-Accuracy (Ruzicka similarity) of the BU surface estimates in the 100m-cell samples of the five considered epochs. Comparison of the previous model benchmark R2022A, the new ensemble model development, and all the other already available models

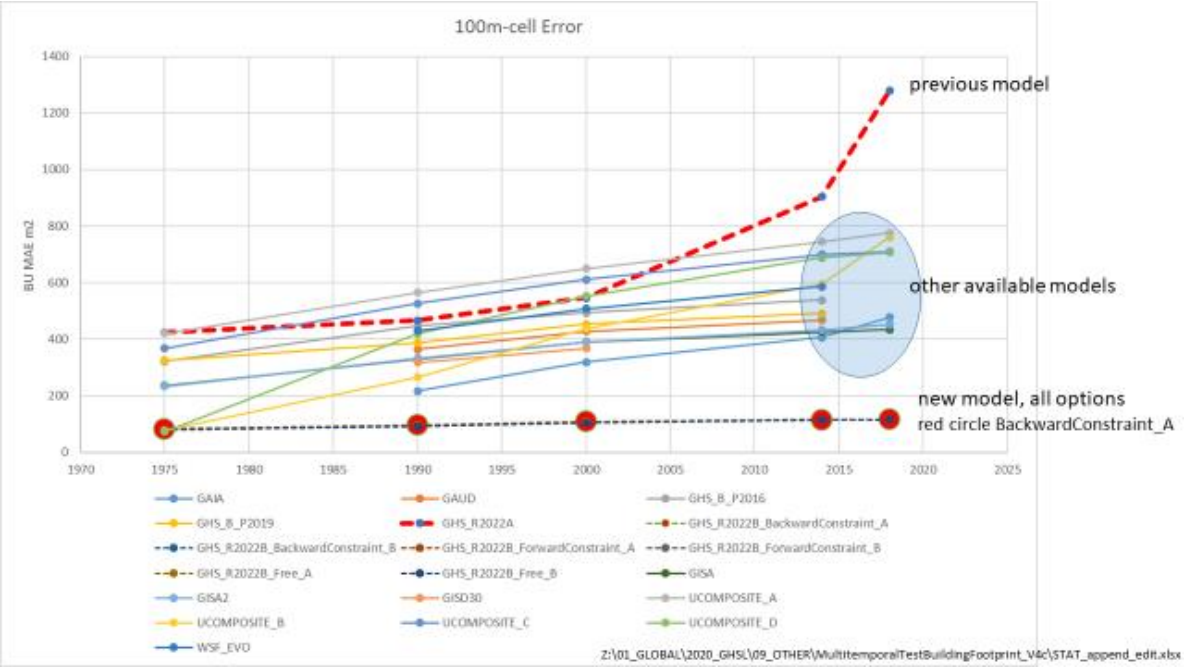


Figure 55 – Error (MAE) of the BU surface estimates in the 100m-cell samples of the five considered epochs. Comparison of the previous model benchmark R2022A, the new ensemble model development, and all the other already available models

Table 13 – Ranking of the prediction/models by the average grid accuracy (Ruzicka similarity) in all the epochs, stratified by the degree of urbanization (SMOD) level 2.

Prediction / model	**Rural area (1)**	**Towns & semi-dense area (2)**	**City (3)**	**Total**
GHS_R2022B_BackwardConstraint_A	0.340	0.399	0.427	**0.374**
GHS_R2022B_Free_A	0.340	0.399	0.427	**0.374**
GHS_R2022B_ForwardConstraint_A	0.339	0.399	0.427	**03.74**
GISD30	0.140	0.166	0.161	**0.153**
GAIA	0.109	0.187	0.171	**0.149**
GAUD	0.122	0.147	0.140	**0.135**
GISA	0.107	0.151	0.150	**0.130**
GISA2	0.107	0.148	0.145	**0.128**
GHS_B_p2019	0.109	0.134	0.135	**0.122**
WSF_EVO	0.101	0.122	0.134	**0.113**
UCOMPOSITE_B	0.089	0.131	0.140	**0.112**
GHS_B_P2016	0.094	0.126	0.127	**0.111**
UCOMPOSITE_C	0.082	0.107	0.119	**0.096**
UCOMPOSITE_D	0.081	0.108	0.117	**0.096**
UCOMPOSITE_A	0.078	0.131	0.114	**0.091**
GHS_R2022A	0.062	0.096	0.115	**0.082**
All	**0.148**	**0.188**	**0.197**	**0.170**

Table 14 – Ranking of the prediction/models by the average grid error in all the epochs, stratified by the degree of urbanization (SMOD) level 2.

Prediction / model	Rural area (1)	Towns & semi-dense area (2)	City (3)	Total
GHS_R2022B_BackwardConstraint_A	139.96	529.63	882.91	**378.44**
GHS_R2022B_Free_A	140.12	529.81	883.00	**378.60**
GHS_R2022B_ForwardConstraint_A	140.28	529.98	883.08	**378.76**
GISD30	378.87	1 921.69	3 629.80	**1 351.20**
GAIA	333.46	1 946.50	3 917.88	**1 366.91**
GAUD	360.84	1 930.79	3 932.59	**1 374.24**
GISA	344.79	2 040.64	4 103.88	**1 430.89**
GISA2	635.94	2 157.79	3 459.08	**1 560.18**
GHS_B_p2019	480.10	2 296.46	4 252.54	**1 619.84**
WSF_EVO	427.73	2 351.55	4 629.09	**1 653.60**
UCOMPOSITE_B	462.82	2 367.50	4 522.97	**1 667.45**
GHS_B_P2016	717.99	2 655.61	4 244.10	**1 888.53**
UCOMPOSITE_C	768.99	3 010.85	4 860.33	**2 124.37**
UCOMPOSITE_D	846.77	3 249.73	5 247.74	**2 300.97**
UCOMPOSITE_A	920.41	6 447.13	5 517.07	**2 446.69**
GHS_R2022A	1 120.53	3 678.67	5 449.36	**2 636.24**
All	**518.06**	**2 144.89**	**3 707.15**	**1 521.62**

4.6 Visual inspection of the ensemble model predictions for observed epochs

The ensemble model predictions of BU surfaces in the different observed epochs (1975, 1990, 2000, 2014, 2018) at 100 m-res were visually inspected in order to i) detect unexpected qualitative anomalies not evident in the quantitative error assessment as for example data tiling effects, spatial-temporal biased patterns induced by noise or gaps in the supporting historical satellite data, ii) check the behaviour of the ensemble model predictions in the off-the-sample and eccentric geographical regions not represented by the MTBF test data. They were selected by prioritizing the different geographical background and settlement patterns as respect the tested regions, thus maximizing the expected difficulty of the ensemble model transfer, potentially leading to the worst-case error scenarios.

The MT ensemble model predictions at 100 m-res are continuous in the range 0-10 000, expressing the amount of predicted BU square meter surface in each cell, in each epoch. It is represented here as a RGB composite image by associating to the predicted BU surface of the 2018, 2000, and 1975 epochs, respectively, to the RGB colour channels. In this visual representation, grey indicates a stable BU component since 1975, yellow a BU component produced 1975-2000 time window, and red the BU component produced in the 2000-2018 time window. Figure 56, Figure 57, and Figure 58 are extracted from the MTBF test domain, thus in the comfort zone of the ensemble model automatic inferential mechanism. Figure 59, Figure 60, Figure 61, and Figure 62 are extracted form data tiles outside the MTBF test domain and far from the comfort zone of the ensemble model automatic inferential mechanism. In particular, Figure 59 shows the case of a complete absence of the Landsat imagery supporting the 1975 epoch. In the previous R2019 and R2016 GHSL releases this case was solved by assuming constant (no change) pattern from recent to past epochs, because no image data evidences were supporting the off-change inference. In the current model development, the zones where no Landsat image data are available are solved by the injection of the best composite of available priors, which are used as semantic proxies. Figure 60, Figure 61, and Figure 62 show the case of rural scattered regions in Africa and Asia that were discovered in the S2 image data of the epoch 2018, and are mostly in the omission domain of the available MT priors (including the R2019 and R2016 precedent GHSL releases), thus forcing the model to transfer the inference to an unknown environment. In absence of MTBF allowing quantitative error assessment, the visual inspection shows a plausible spatial-temporal evolution pattern in those cases.

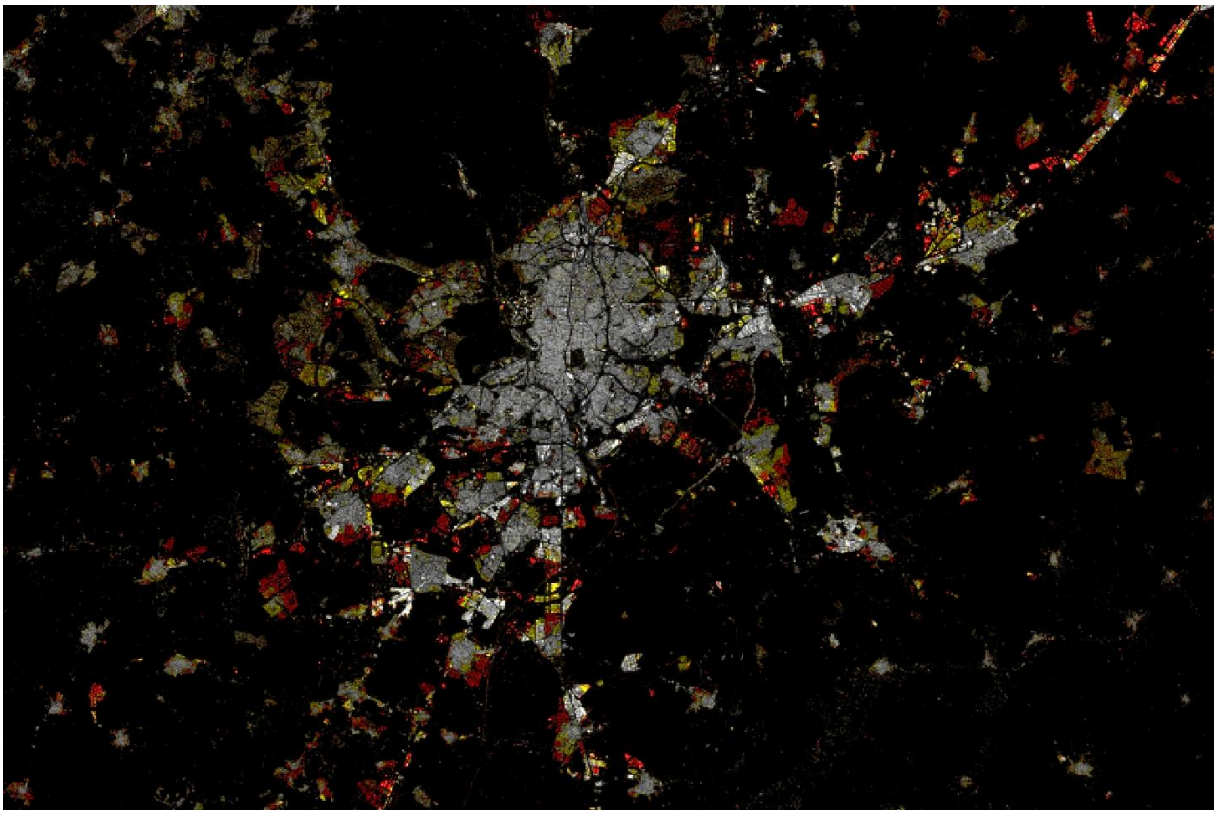


Figure 56 – Ensemble model predictions for the city of Madrid, Spain. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10 000 m^2 of BU surface.

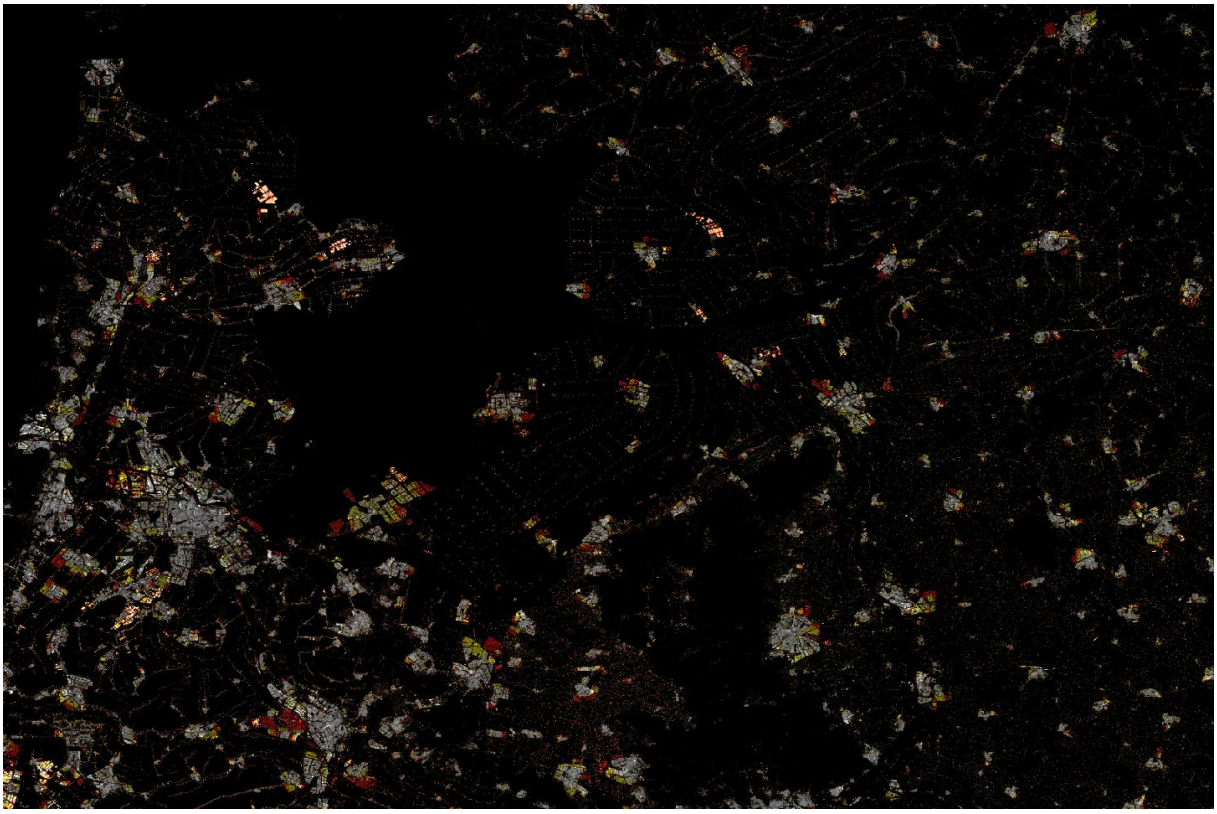


Figure 57 - Ensemble model predictions for the region N-E of the city of Amsterdam, Netherland. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10 000 m^2 of BU surface.

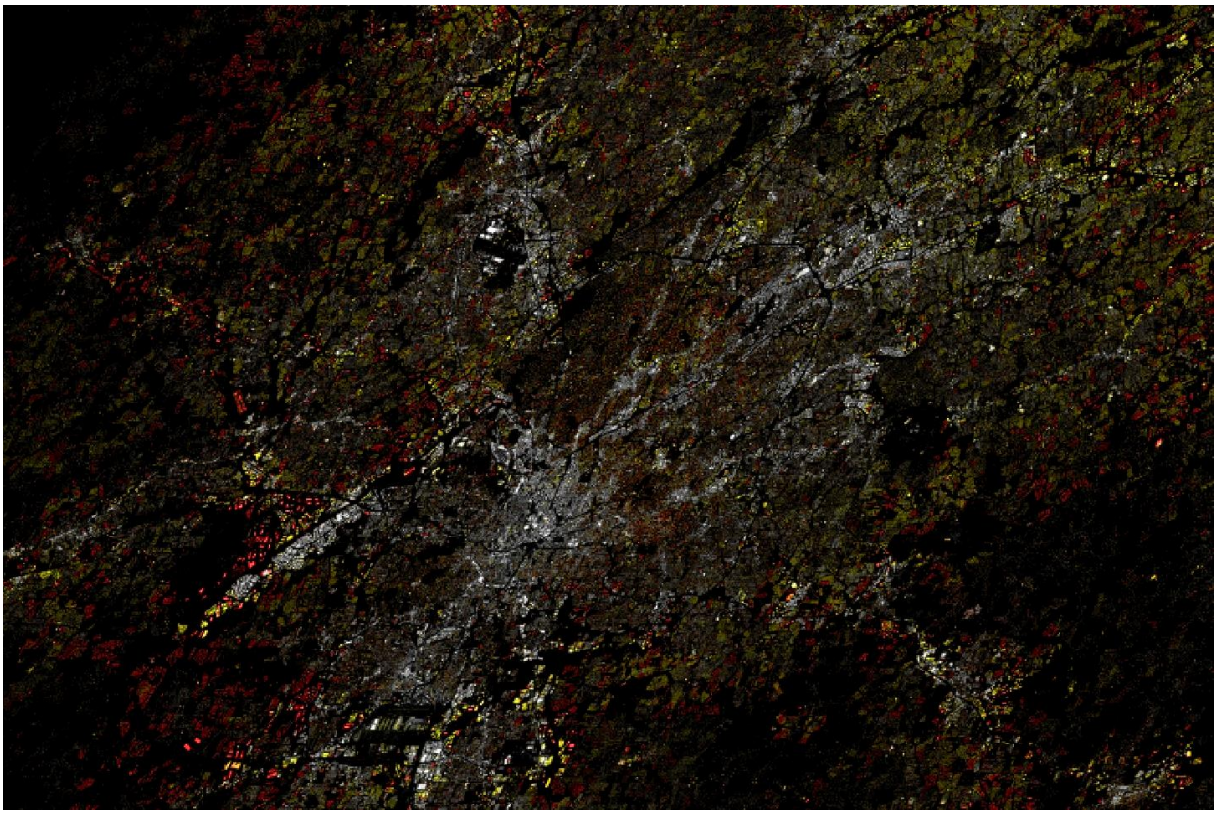


Figure 58 – Ensemble model predictions for the city of Atlanta, Unites States of America. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10000 m2 of BU surface.

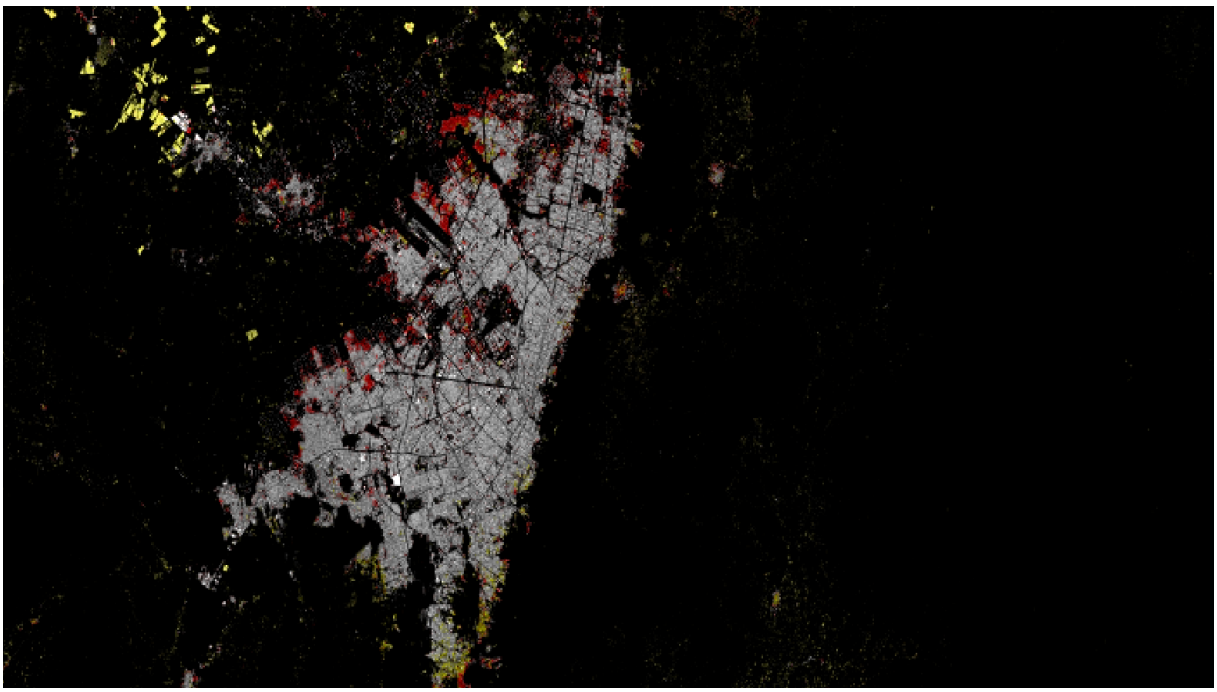


Figure 59 – Ensemble model predictions for the city of Bogotá, Colombia. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10000 m2 of BU surface.

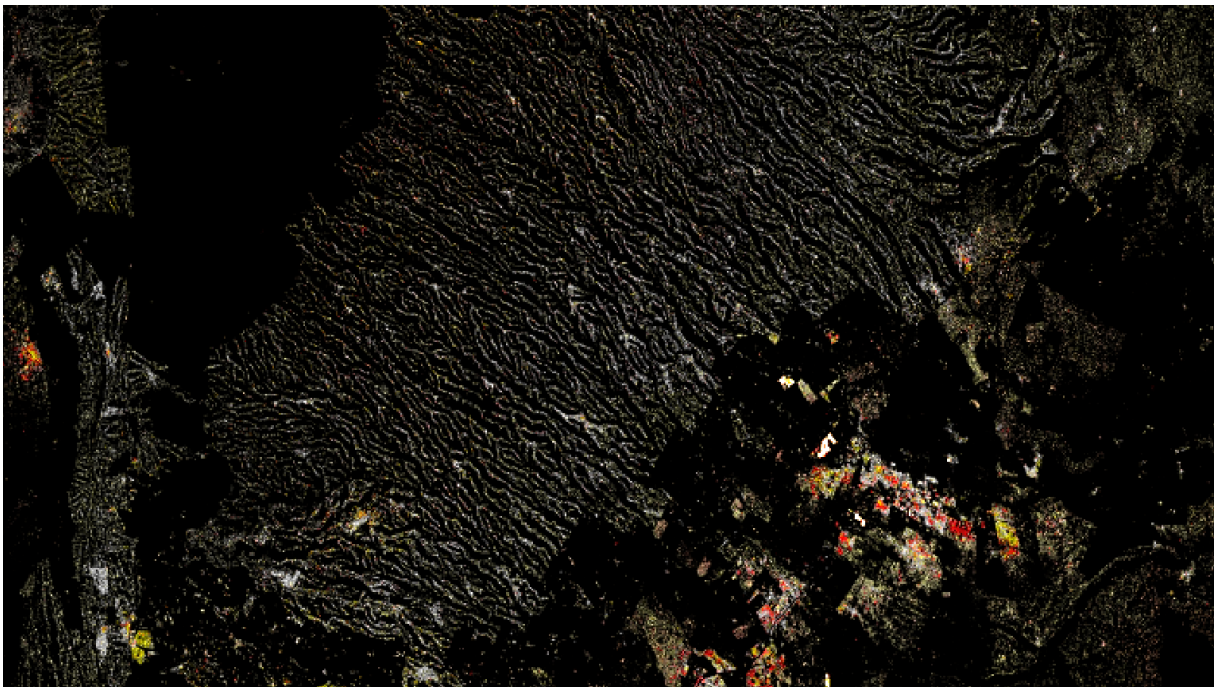


Figure 60 - Ensemble model predictions for the region N of Nairobi, Kenya. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10 000 m^2 of BU surface

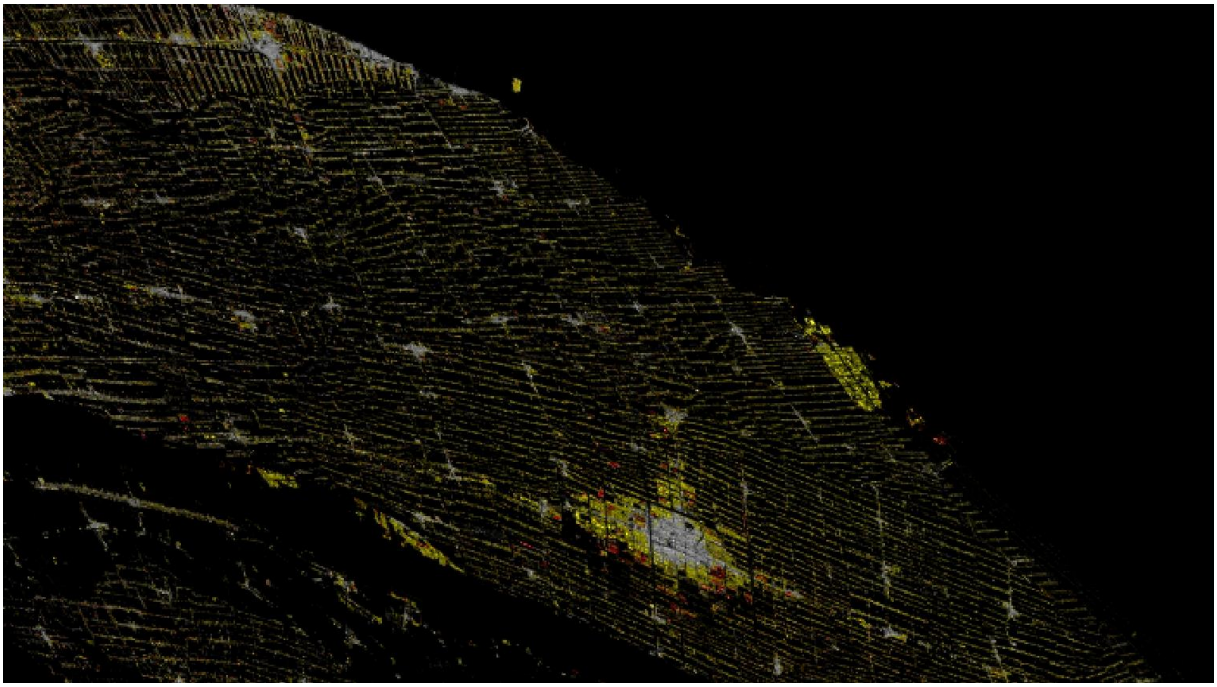


Figure 61 - Ensemble model predictions for the region N of Shanghai, China. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10 000 m^2 of BU surface

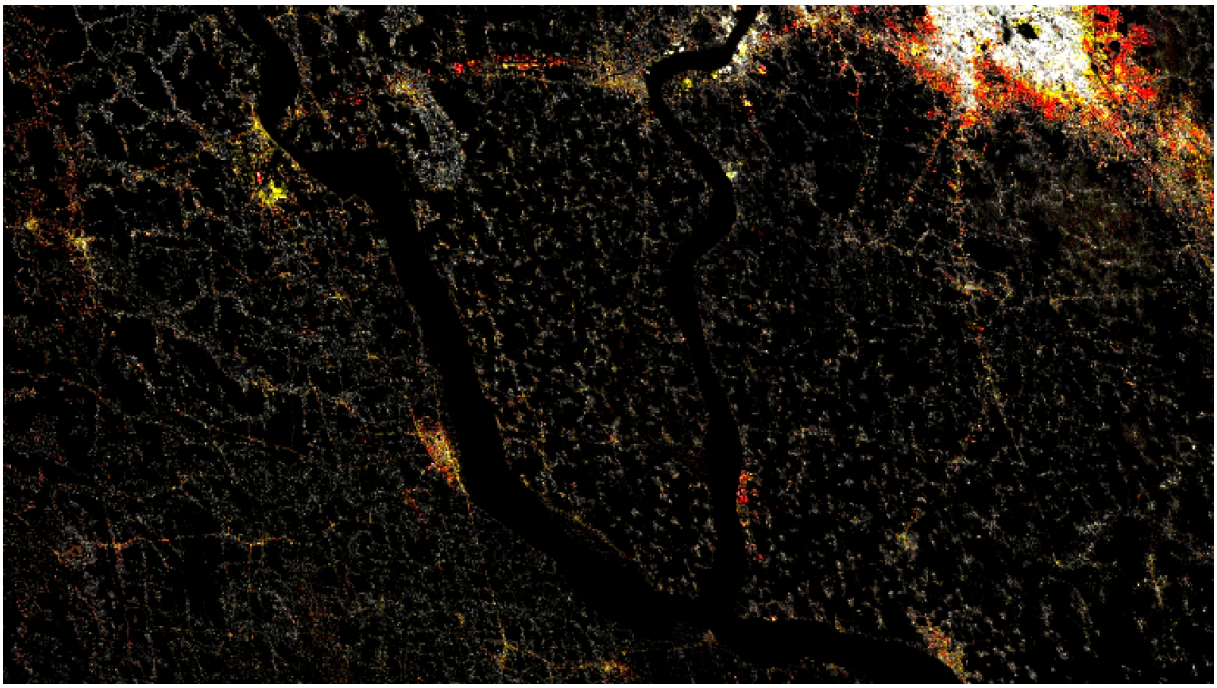


Figure 62 - Ensemble model predictions for the region S-W of Kolkata, India. RGB composite R:2018, G:2000, B:1975, grey indicates a stable BU component since 1975, linear histogram stretching min:0, max:10 000 m^2 of BU surface

4.7 Visualisation of the ensemble model predictions interpolation

Interpolation and extrapolation of the ensemble model predictions into 5-year interval layers is presented on Figure 63. In this visual representation, grey indicates a stable BU component since 1975, and colours from red to magenta denote BU components produced for 1980 – 2030 5-year intervals.

Validation of the multi-temporal model predictions will be delivered in dedicated peer-reviewed publications.

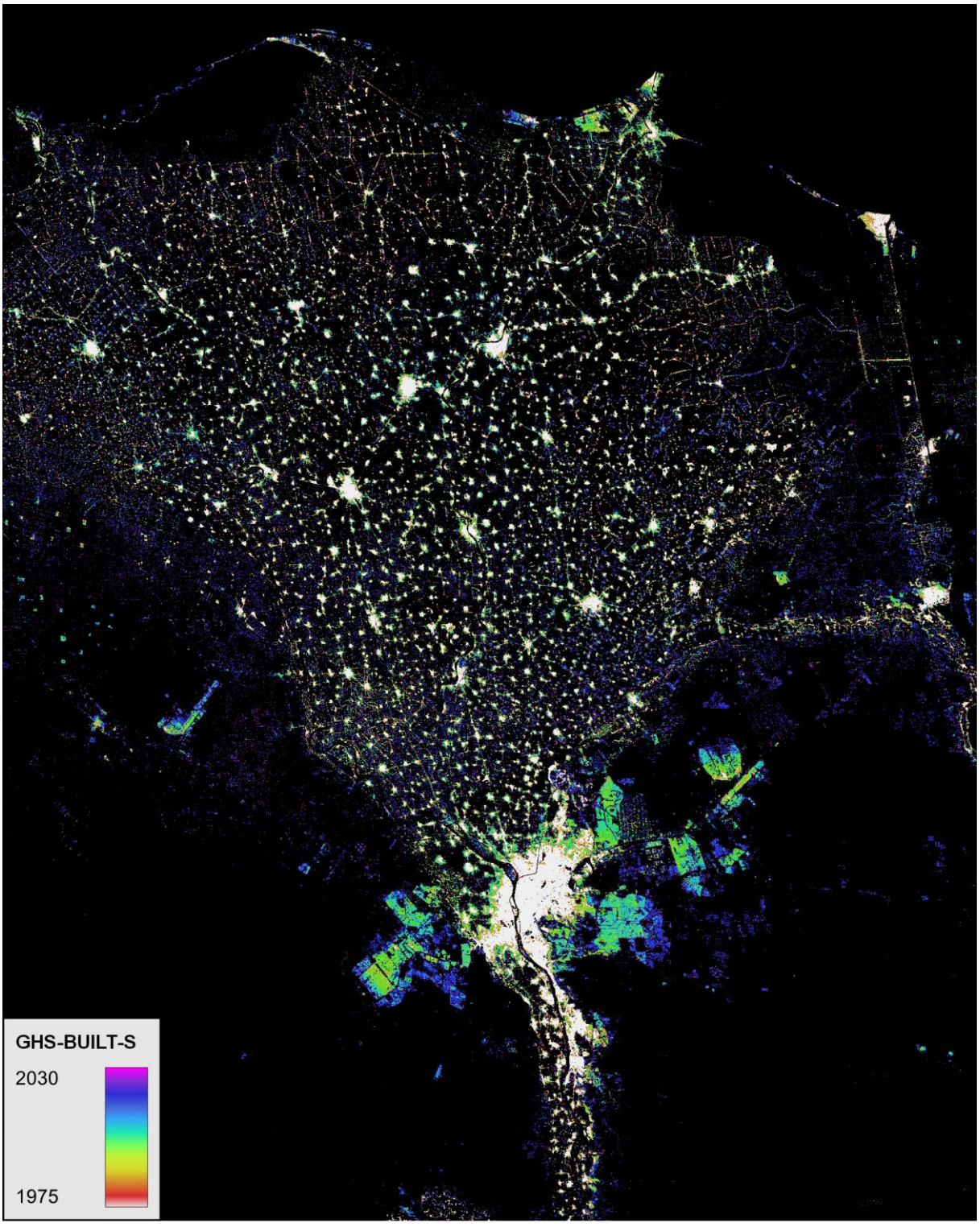


Figure 63 – Interpolation layers of the ensemble model predictions for the region Cairo, Egypt. Colours represent the presence of built-up surface in each interpolated year, with grey colour marking indicating a stable BU component since 1975.

5 Conclusions

The long-term spatial distributions of global built-up density and of resident people produced by the GHSL provide baseline data for monitoring the progress towards achieving the SDG's. In particular, spatial policy indicators monitoring SDG 11 on sustainable cities and communities require reliable information about the increase of built-up surface in-between the observed epochs. Therefore, a new approach for solving the prediction of the multi-temporal change of built-up surfaces and volumes in the GHSL data ecosystem from the anchor point 2018 and the epochs 2014, 2000, 1990, and 1975 was developed.

The new approach relies on stratified multiple-quantization associative rule learning applied to EO data, object-oriented image processing, and multiple decision support ensemble modelling. For each predicted change epoch independently, the final ensemble model was selected trough Pareto multiple objective optimization, minimizing the error of the BU surface change maps and the errors in aggregated BU surface change rates in the URBAN, RURAL application domain strata. Moreover, some post-processing options have been discussed and evaluated.

According to the collected empirical records in the available multi-temporal building footprints reference data, the change predictions of the new ensemble model are substantially more accurate than the ones claimed in the GHS-BU R2022A, and in any other prior alternative or combination of priors available in the literature. In particular, the expected error (MAE) of the predicted BU change grids by the new ensemble model in all the considered epochs is 432 m^2 and 233 m^2, respectively for the URBAN and RURAL domains, vs. the 490 m^2 and 506 m^2, respectively, for the benchmark R2022A. Thus, an average decrease of half of the error in the prediction of BU surface changes in rural domain is expected, as compared to the R2022A benchmark in all the considered epochs. The decrease of the error is even more important in the prediction of the change 2018-2014 that is the most challenging case involving the change of the supporting EO sensors from Landsat TM (30m-res) to Sentinel2 MSI (10m-res). In this case, the estimated error of the new ensemble model is 68 $m^2$and 40 m^2, respectively, in the URBAN and RURAL domains, vs. the benchmark R2022A scoring 126 m^2 and 239 m^2, correspondingly. This means that in the new ensemble model predictions the expected error in this challenging case will decrease of a factor close to 2 in the urban domain and decrease of a factor close to 6 in rural domain. Coherently, in the change 2018-2014 the aggregated change rate gain factor of the new ensemble model in the rural domain is close to 1, vs. the benchmark R2022A that is 4.9, thus largely overestimating the change rate in rural application domain.

In order to understand the value added of the research and development achievements discussed, it is worth noting that the new multiple-sensor (S2-Landsat) built-up surface change detection challenge is performed in the same GHSL scarce and reduced data environment as in the previous GHSL data releases. This implies that the proposed object-oriented, ensemble-decision framework contribute to the robustness and stability of the presented model. Such solution, capable of handling scarce and spatially and temporally varying multiple-sensor input data facilitates continuity and repeatability of the modelling of the global multi-temporal built-up surfaces and volumes.

References

Corbane, C., M. Pesaresi, T. Kemper, P. Politis, Aneta J. Florczyk, V. Syrris, M. Melchiorri, F. Sabo, and P. Soille. 2019. "Automated Global Delineation of Human Settlements from 40 Years of Landsat Satellite Data Archives." *Big Earth Data* 3 (2): 140–69. https://doi.org/10.1080/20964471.2019.1625528.

Corbane, C., M. Pesaresi, P. Politis, A.J. Florczyk, M. Melchiorri, S. Freire, M. Schiavina, D. Ehrlich, G. Naumann, and T. Kemper. 2018. "The Grey-Green Divide: Multi-Temporal Analysis of Greenness across 10,000 Urban Centres Derived from the Global Human Settlement Layer (GHSL)." *International Journal of Digital Earth*, October, 1–18. https://doi.org/10.1080/17538947.2018.1530311.

European Commission. Joint Research Centre. 2022. *GHSL Data Package 2022: Public Release GHS P2022.* LU: Publications Office. https://data.europa.eu/doi/10.2760/19817.

———. 2023. *GHSL Data Package 2023.* LU: Publications Office. https://data.europa.eu/doi/10.2760/098587.

European Space Agency and Airbus. 2022. "Copernicus DEM." European Space Agency. https://doi.org/10.5270/ESA-c5d3d65.

Florczyk, A. J., C. Corbane, D. Ehrlich, S. Freire, T. Kemper, L. Maffenini, M. Melchiorri, et al. 2019. *GHSL Data Package 2019*. JRC117104. Publications Office of the European Union. 10.2760/729240.

Gao, Jing, and Brian C O'Neill. 2020. "Mapping Global Urban Land for the 21st Century with Data-Driven Simulations and Shared Socioeconomic Pathways." *Nature Communications* 11 (1): 2302.

Gao, Jing, and Martino Pesaresi. 2021. "Downscaling SSP-Consistent Global Spatial Urban Land Projections from 1/8-Degree to 1-Km Resolution 2000–2100." *Scientific Data* 8 (1): 281.

Gong, Peng, Xuecao Li, Jie Wang, Yuqi Bai, Bin Chen, Tengyun Hu, Xiaoping Liu, et al. 2020. "Annual Maps of Global Artificial Impervious Area (GAIA) between 1985 and 2018." *Remote Sensing of Environment* 236 (January): 111510. https://doi.org/10.1016/j.rse.2019.111510.

Haralick, Robert M, and Linda G Shapiro. 1992. *Computer and Robot Vision*. Vol. 1. Addison-wesley Reading, MA.

Huang, Xin, Jiayi Li, Jie Yang, Zhen Zhang, Dongrui Li, and Xiaoping Liu. 2021. "30 m Global Impervious Surface Area Dynamics and Urban Expansion Pattern Observed by Landsat Satellites: From 1972 to 2019." *Science China Earth Sciences* 64 (11): 1922–33. https://doi.org/10.1007/s11430-020-9797-9.

Huang, Xin, Yihong Song, Jie Yang, Wenrui Wang, Huiqun Ren, Mengjie Dong, Yujin Feng, Haidan Yin, and Jiayi Li. 2022. "Toward Accurate Mapping of 30-m Time-Series Global Impervious Surface Area (GISA)." *International Journal of Applied Earth Observation and Geoinformation* 109: 102787.

Jaccard, Paul. 1901. "Étude Comparative de La Distribution Florale Dans Une Portion Des Alpes et Des Jura." *Bull Soc Vaudoise Sci Nat* 37: 547–79.

Kosub, Sven. 2019. "A Note on the Triangle Inequality for the Jaccard Distance." *Pattern Recognition Letters* 120 (April): 36–38. https://doi.org/10.1016/j.patrec.2018.12.007.

Liu, Xiaoping, Yinghuai Huang, Xiaocong Xu, Xuecao Li, Xia Li, Philippe Ciais, Peirong Lin, et al. 2020. "High-Spatiotemporal-Resolution Mapping of Global Urban Change from 1985 to 2015." *Nature Sustainability* 3 (7): 564–70. https://doi.org/10.1038/s41893-020-0521-x.

Marconcini, Mattia, Annekatrin Metz-Marconcini, Thomas Esch, and Noel Gorelick. 2021. "Understanding Current Trends in Global Urbanisation-the World Settlement Footprint Suite." *GI_Forum* 9 (1): 33–38.

Melchiorri, Michele, and Thomas Kemper. 2023. "Establishing an Operational and Continuous Monitoring of Global Built-up Surfaces with the Copernicus Global Human Settlement Layer." In *2023 Joint Urban Remote Sensing Event (JURSE)*, 1–4. Heraklion, Greece: IEEE. https://doi.org/10.1109/JURSE57346.2023.10144201.

Melchiorri, Michele, Martino Pesaresi, Aneta J. Florczyk, Christina Corbane, and Thomas Kemper. 2019. "Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator—SDG 11.3.1." *ISPRS International Journal of Geo-Information* 8 (2): 96. https://doi.org/10.3390/ijgi8020096.

Pareto, Vilfredo. 1912. “Manuel d’économie Politique.” *Bull. Amer. Math. Soc* 18 (462–474): 3.

Pesaresi, M., C. Corbane, A. Julea, A. Florczyk, V. Syrris, and P. Soille. 2016. “Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas.” *Remote Sensing* 8 (4): 299. https://doi.org/10.3390/rs8040299.

Pesaresi, M., and D. Ehrlich. 2009. “A Methodology to Quantify Built-up Structures from Optical VHR Imagery.” In *Global Mapping of Human Settlement Experiences, Datasets, and Prospects*, edited by P. Gamba and M. Herold, 27–58. CRC Press.

Pesaresi, M., D. Ehrlich, S. Ferri, A. Florczyk, Manuel Carneiro Freire Sergio, S. Halkia, Andreea Julea, T. Kemper, Pierre Soille, and Vasileios Syrris. 2016. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014*. Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/handle/111111111/40182.

Pesaresi, M., A. Gerhardinger, and F. Kayitakire. 2008. “A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1 (3): 180–92. https://doi.org/10.1109/JSTARS.2008.2002869.

Pesaresi, M., G. Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, et al. 2013. “A Global Human Settlement Layer from Optical HR/VHR RS Data: Concept and First Results.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (5): 2102–31. https://doi.org/10.1109/JSTARS.2013.2271445.

Pesaresi, M., V. Syrris, and A. Julea. 2016. “A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning.” *Remote Sensing* 8 (5): 399. https://doi.org/10.3390/rs8050399.

Pesaresi, Martino. 2014. “Global Fine-Scale Information Layers: The Need of a Paradigm Shift.” In *Conference on Big Data from Space (BiDS’14)*, edited by Pierre Soille and Pier Giorgio Marchetti. JRC.

Pesaresi, Martino, Georgios K Ouzounis, and Lionel Gueguen. 2012. “A New Compact Representation of Morphological Profiles: Report on First Massive VHR Image Processing at the JRC.” In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, 8390:839025. International Society for Optics and Photonics.

Pesaresi, Martino, and Panagiotis Politis. 2022. “GHS-LAND R2022A - Land Fraction as Derived from Sentinel2 Image Composite (2018) and OSM Data.” European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/AB7AD451-5ED5-44A6-A4D0-9F7A4E848CEE.

———. 2023. “GHS-BUILT-C R2023A - GHS Settlement Characteristics, Derived from Sentinel2 Composite (2018) and Other GHS R2023A Data.” European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/3C60DDF6-0586-4190-854B-F6AA0EDC2A30.

Rosenthal, Arnon, Sandra Heiler, Umeshwar Dayal, and Frank Manola. 1986. “Traversal Recursion: A Practical Approach to Supporting Recursive Applications.” *ACM SIGMOD Record* 15 (2): 166–76.

Schiavina, Marcello, Michele Melchiorri, Christina Corbane, Aneta J Florczyk, Sergio Freire, Martino Pesaresi, and Thomas Kemper. 2019. “Multi-Scale Estimation of Land Use Efficiency (SDG 11.3. 1) across 25 Years Using Global Open and Free Data.” *Sustainability* 11 (20): 5674.

Serra, Jean. 1982. “Morphology for Grey-Tone Functions.” *Image Analysis and Mathematical Morphology* 1: 424–78.

———. 1983. *Image Analysis and Mathematical Morphology*. Academic Press, Inc.

Uhl, Johannes H, and Stefan Leyk. 2022. “MTBF-33: A Multi-Temporal Building Footprint Dataset for 33 Counties in the United States (1900-2015).” *arXiv Preprint arXiv:2203.11078*.

Vincent, Luc, and Pierre Soille. 1991. “Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations.” *IEEE Transactions on Pattern Analysis & Machine Intelligence* 13 (06): 583–98.

Zhang, Xiao, Liangyun Liu, Tingting Zhao, Yuan Gao, Xidong Chen, and Jun Mi. 2022. “GISD30: Global 30 m Impervious-Surface Dynamic Dataset from 1985 to 2020 Using Time-Series Landsat Imagery on the Google Earth Engine Platform.” *Earth System Science Data* 14 (4): 1831–56.

List of abbreviations and definitions

AGBH	Average of the Gross Building Height
ANBH	Average of the Net Building Height
BIN	binary change detection schema
BU	built-up class
BUDYN	spatial estimation of the expansion potential of the built-up domain in time
BUFRAC	sub-pixel built-up surface prediction made at 10m-resolution from S2 image data composite
CON	continuous change detection schema
DEM	Digital Earth Model
ELS	Empirical Land Suitability layer
EO	Earth Observation
GAIA	"Annual maps of global artificial impervious area (GAIA) between 1985 and 2018" (Gong et al., 2020)
GAUD	"High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015" (Liu et al., 2020)
GEE	Google Earth Engine
GHSL	Global Human Settlement Layer
GHS_B_P2016	"Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014" (Pesaresi, Ehrlich, et al., 2016)
GHS_B_P2019	"Automated global delineation of human settlements from 40 years of Landsat satellite data archives" (Corbane et al., 2019)
GISA	"30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019" (Huang et al., 2021)
GISA2	"Toward accurate mapping of 30-m time-series global impervious surface area (GISA2.0)" (Huang et al., 2022)
GISCO	the Geographic Information System of the Commission
GISD30	"Global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform" (Zhang et al., 2022)
JRC	Joint Research Centre
MT	multi-temporal
MTBF	multi-temporal building footprints
MTBF33	MTBF33 project (Uhl & Leyk, 2022)
MWD	The Mollweide projection ESRI:54009
NBU	non built-up class
NRES	non residential built-up class
PHI	decision probability in SML method
RES	residential built-up class
RURAL	URBAN application domains, as set by the GHS-SMOD R2022A
S2	Sentinel-2
SML	symbolic machine learning
UCOMPO	union of the priors

URBAN URBAN application domains, as set by the GHS-SMOD R2022A

VOL built-up volume

WSF_EVO “World Settlement Footprint Evolution 1985-2015” (Marconcini et al., 2021)

List of figures

List of tables

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

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

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub
EU Science Hub - Joint Research Centre
EU Science, Research and Innovation
EU Science Hub
@eu_science

Publications Office
of the European Union