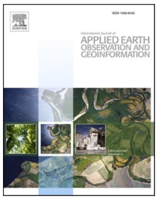
International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105153

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Do-it-yourself built-up mapping tool: A practical cloud-based solution using Sentinel imagery for mapping urban expansion in Africa

Marta Sapena [a,b], Johannes Mast [a], Elisabeth Schoepfer [a,*], Hannes Taubenböck [a,c]

[a] *German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), 82234 Weßling, Germany*
[b] *Image Processing Laboratory, Universitat de València, 46980 Valencia, Spain*
[c] *Julius-Maximilians-Universität Würzburg, Institute of Geography and Geology, 97074 Würzburg, Germany*

ARTICLE INFO

Keywords:
Built-up mapping
Global South
Google Earth Engine
Copernicus
Machine learning

ABSTRACT

Urban areas across Africa have undergone unprecedented growth, posing significant challenges for sustainable development, infrastructure planning, and climate resilience. Existing mapping products often struggle to capture the dynamic and heterogeneous nature of these evolving urban landscapes, highlighting the need for maps that are both up-to-date and locally relevant. This study introduces a practical, cloud-based solution: an online tool for site-specific mapping (i.e., tailored maps for a defined area of interest) that leverages the capabilities of Google Earth Engine. The tool uses Sentinel-1 and Sentinel-2 imagery to derive a wide range of spectral and texture metrics, supplemented by terrain data, and is trained using open building footprint datasets available for 2022. In an experimental setup, nine model configurations were tested under varying data availability conditions across 100 urban sites in Africa. The best-performing model achieved a mean F1-score of 0.59 (recall 0.63; precision 0.51) when validated against building footprints, with higher accuracy observed in dense urban areas. This configuration was integrated into the freely available 'Do-it-yourself built-up mapping tool' (DIY-BU). A quantitative analysis across the 100 test sites showed that the maps generated by our tool for 2022 were substantially more accurate (with an increase of F1-score by 0.18–0.30) than global multi-temporal products analysed for the same period (i.e., Dynamic World, ESRI land cover, GISA, GLC_FCS30D and GISD30). While the quantitative assessment was limited to the 2022 reference year, and the multi-temporal maps rely on a monotonic growth assumption (preventing the detection of demolition), a qualitative analysis highlighted the tool's advantages in capturing detailed urban expansion and small-scale structures. The DIY-BU-mapping tool offers a valuable resource for a variety of applications, including urban planning, infrastructure monitoring, disaster preparedness and climate adaptation. Beyond presenting the tool's functionality, the paper discusses its limitations and potential applications across diverse geographic and data availability contexts.

1. Introduction

Africa's urban landscape is undergoing an unprecedented transformation. Projections suggest the continent's population will double by 2050, with 950 million people expected to move to cities (OECD/SWAC, 2020). This places Africa at the epicentre of future global urban growth and presents significant challenges to urban governance, infrastructure planning, and service delivery (Acuto et al., 2018).

Monitoring urban development deepens our understanding of both historical and current trends and provides essential insights for informed decision making (Patino and Duque, 2013). Yet, despite its importance, Africa still faces significant data deficits. In many areas, existing data is either outdated or inconsistent, or is entirely missing, which makes it difficult to identify priority areas or design sustainable urban policies. This issue is particularly urgent in the context of the Sustainable Development Goals (SDGs) (UN, 2023), which emphasise the importance of creating inclusive, safe, resilient and sustainable cities. As the latest Intergovernmental Panel on Climate Change (IPCC) report highlights, the lack of localised data restricts regional-level research and policy, forcing reliance on global assessments that often fail to reflect local realities (IPCC, 2023).

In recent decades, there has been a surge in the availability of satellite imagery from Earth observation programmes. This has supported the development of systematic mapping approaches and multi-scale

---

[*] Corresponding author.
E-mail address: Elisabeth.Schoepfer@dlr.de (E. Schoepfer).

https://doi.org/10.1016/j.jag.2026.105153
Received 14 August 2025; Received in revised form 14 January 2026; Accepted 2 February 2026

analyses of urban growth (Taubenböck, 2019). However, the transformation of raw imagery into reliable into actionable information remains a technical and resource-intensive task. This has led to growing efforts to develop analysis-ready datasets that broaden the accessibility of Earth observation data for wider audiences (Digital Earth Africa, 2022). In the field of built-up mapping, several globally harmonized datasets now provide spatial and temporal coverage (Sapena, 2020).

In Africa, global datasets can be divided into two broad categories: (1) general-purpose land cover products, in which built-up is one among several classes (e.g., Dynamic World, ESRI Land Cover, Worldcover, Global Dynamic Land Cover, and GLC_FCS30D) and (2) datasets that focus exclusively on built-up areas (e.g., GHSL2023, WSF, GAIA, GISA and GISD30). While these grided products represent major advancements, their performance across African regions remains inconsistent (Chakraborty et al., 2024; Huang et al., 2022; Mu et al., 2022; Zhang et al., 2022). Common issues include the underrepresentation of informal settlements and refugee camps (Mück et al., 2017; Van Den Hoek and Friedrich, 2021), reduced accuracy in vegetated and rural areas (Chakraborty et al., 2024; Zhang et al., 2022) and overall lower availability and higher uncertainty in comparison to other regions (Taubenböck et al., 2024). Recent advances, such as GHSL2023, have started to address some of these gaps, with notable improvements in mapping vulnerable areas and refugee camps (Pesaresi et al., 2024).

Complementing these products, very-high-resolution vector building footprint datasets, such as Google Open Buildings dataset (Sirko et al., 2021), Microsoft's Building Footprints (Microsoft, 2023), Ecopia (Ecopia, 2023) and GlobalBuildingAtlas (Zhu et al., 2025), offer great geometric detail and are increasingly used to train built-up mapping models (Pesaresi et al., 2024). However, these datasets often remain mono-temporal and have uneven spatial coverage. A recent comparison by Chamberlain et al., (2024) found large discrepancies between countries and settlement types across Africa, concluding that in the absence of high-quality reference data, it remains difficult to determine which dataset is most reliable. A promising development is Google's Open Buildings 2.5D Temporal Dataset (Sirko et al., 2023), which integrates building footprints with Sentinel-2 imagery to produce annual maps of built-up areas and building heights across the Global South. While this represents a step towards multi-temporal built-up mapping, its long-term stability and local accuracy remain to be assessed.

Despite the global progress made, most current datasets suffer from significant limitations in the African context. Many are not tailored to local conditions, rely on a single data source or only represent static time points. They rarely capture recent urban developments and often lack consistent updates or transparent processes. This limits their practical utility, particularly in areas with limited technical capacity and high demand for localised, timely and actionable information. Consequently, the absence of high-resolution, frequently updated and site-specific maps continues to hinder planning and policy formulation (Acuto et al., 2018; Dodman et al., 2022; Chen et al., 2023) and contributes to uncertainty in global urban assessments (Taubenböck et al., 2024).

A recent review has also highlighted the lack of reproducibility and documentation in urban land-use models. These models tend to focus on large cities, despite the fact that much of Africa's urban growth is projected to occur in smaller and medium-sized cities (Reba and Seto, 2020). In this context, there is a clear need for open, adaptable and user-friendly tools that enable local technicians, planners and analysts to produce reliable, up-to-date maps of built-up areas in their own regions. The limitations of current raster- and vector-based datasets highlight the urgent need for more context-sensitive, scalable and transparent solutions for mapping urban change in Africa. In this paper, we refer to such locally adapted outputs as site-specific maps, which are tailored to a defined area of interest, such as a city or town.

Existing cloud computing platforms, such as SEPAL and Digital Earth Africa, have greatly improved access to satellite data and supported the development of decision-ready products across various sectors (Digital Earth Africa, 2024; SEPAL, 2025). However, to the best of our knowledge, these platforms do not currently offer tools specifically designed for multi-temporal mapping of built-up areas or tracking urban expansion. These remain critical needs for sustainable urban planning. To address this gap, we introduce a practical, user-centred solution: an online, site-specific mapping tool built in Google Earth Engine (GEE). This transparent, free, and cloud-based tool leverages large, multimodal datasets and machine learning to produce annual, high-resolution (10-m), and temporally consistent up-to-date maps of built-up area expansion for African cities using Sentinel imagery. We adopt the OECD (2024) definition of built-up area as the presence of buildings, and define 'up-to-date' as maps generated using the most recent full year of available Sentinel imagery.

The main objective of this study is to develop and validate a cloud-based tool that can be used to generate tailored, analysis-ready maps of built-up area expansion monitoring. The tool has been designed for practitioners with general geospatial skills (e.g., technicians supporting urban planning, students, educators), as well as advanced users interested in extending or adapting the method. To ensure the tool's reliability and relevance, multiple modelling configurations were tested across 100 urban sites in Africa, and the results were benchmarked against state-of-the-art (SOTA) global mapping products.

The DIY-BU-mapping tool and its underlying code are fully open access. We aim to (1) empower African technicians, researchers and students to produce timely and relevant urban expansion maps; (2) promote the use of Sentinel satellite data for monitoring urban development via a user-centred approach; (3) encourage adaptation and reuse of the tool for new applications beyond built-up area mapping; and (4) contribute to transparent, reproducible science in support of solutions to global challenges such as climate change and urbanization.

The remainder of this paper is structured as follows: we describe the datasets, pre-processing steps, and experimental design used to select the optimal model configuration for the tool; evaluate the results across 100 African test sites; introduce the tool's components; and discuss the strengths and limitations of the resulting maps.

2. Data

This section describes the satellite data and geospatial datasets used for mapping built-up area expansion in Africa (Sect. 2.1). We present the reference data used for training and validating the models (Sect. 2.2), and introduce the SOTA global built-up maps that approximately coincide in time with the reference data (Sect. 2.3). These datasets are used to benchmark the results of the proposed approach.

2.1. Satellite data

We employ Sentinel-1 (S-1) and Sentinel-2 (S-2) satellite imagery from the Copernicus programme, to derive spectral and texture indices used as features (i.e., attributes) in the classification models. Previous studies have shown that combining data from these sensors can improve the accuracy of mapping products (e.g., Hafner et al., 2022; Marconcini et al., 2021).

S-1 is a radar mission providing continuous, all-weather, day-and-night imagery (ESA, 2023a). In this study, we use the Interferometric Wide swath (IW) Ground Range Detected (GRD) Level-1 data products, distributed by ESA, from the S-1 Synthetic Aperture Radar (SAR) instrument, which are accessible via GEE. S-2 is a multi-spectral imaging mission (ESA, 2023a). We use data from its MultiSpectral Instrument (MSI), which provides spectral bands at 10-, 20- and 60-m spatial resolutions. In Africa, GEE offers S-2 Level-1C data (orthorectified and Top-Of-Atmosphere (TOA) reflectance) since 2016 onwards, while S-2 Level-2A data (orthorectified and surface reflectance) have only been available since 2018. To maximize temporal coverage, we use Level-1C data in this study. Sentinel data was chosen for its higher spatial resolution (10-m), which is substantially finer than that of earlier missions such as Landsat (30-m), and is therefore critical for detecting the small buildings commonly found in urban areas in Africa. The time frame available (2016–present) complements and extends the WSF Evolution dataset (1985–2015), enabling analysis of long-term built-up change.

2.2. Reference data

In addition to the satellite imagery, we use ancillary geospatial data for model training and validation. Reference data were derived from two vector building footprint datasets: 'Open Buildings' and 'Building Footprints'. As the exact date of the building footprints is not provided by the providers, we conducted a manual verification process.

The 'Open Buildings' version 3 (OB3) dataset was released is May 2023. Following the provider's recommendations, we manually checked the historical image dates in Google Earth Pro for each of the 100 test sites described in Sect. 3.3. Most reference images were captured between May 2022 and May 2023, although a few sites had older imagery (see Table A1 for details). For Microsoft's 'Building Footprints' (MBF) dataset (Samapriya Roy et al., 2024), the provider indicates the use of Bing Maps imagery collected between 2014 and 2021. Due to limited metadata, it was not possible to confirm the acquisition date for each site. To minimize temporal uncertainty, we set the reference period to the full collection of Sentinel imagery for 2022, and limited MBF usage to five out of the 100 sites (where OB3 was not available). Additionally, to ensure high-quality reference data, only building footprints with a confidence score of 85% or higher were selected, although such filtering was not possible for MBF. Then, to ensure robust and independent map evaluation, we incorporated additional building footprint datasets (see Sect. 3.4 and Table 1). These include: (1) the 'Overture Maps Foundation' (OMF) dataset, which aggregates building footprints from 'OpenStreetMap', Esri, Google, and Microsoft; (2) the 'Open Buildings 2.5D Temporal' (OBT) dataset for the year 2022, which provides information on building presence, counts, and heights; (3) the 'GlobalBuildingAtlas' (GBA), which fills in the gaps left by other datasets; (4) the OB3 dataset and (5) MBF datasets, which were partly used for training; and (6) a fusion of all the above datasets, which provide the highest level of completeness since no single existing dataset is complete or globally consistent. Fig. A1 shows an overview of the differences in the percentage of built-up areas for 100 test sites per dataset.

Table 1
Reference and validation data: datasets description, resolution, date, and source.

Name	Description	Resolution / Year	Source
Open Buildings v3 (OB3)	Google building footprints version 3	– / mostly 2022–2023 (Table A1)	(Sirko et al., 2021)
Building Footprints (MBF)	Microsoft building footprints	– / 2014–2021	(Microsoft, 2023)
Overture Maps Foundation (OMF)	Building footprints from multiple sources	– / –	(Overture Maps Foundation, 2024)
Open Buildings 2.5D Temporal (OBT)	Annual maps of estimated building presence, counts and heights	0.5 m / 2022	(Sirko et al., 2023)
GlobalBuildingAtlas (GBA)	Building polygons with estimated heights	3 m / 2019	(Zhu et al., 2025)
Fused Building (FB)	A combination of OB3, MBF, OMF, OBT and GBA	– / 2014–2023	–
WorldCover v200	Land-use/land-cover map	10 m / 2021	(Zanaga et al., 2022)
Africapolis	Urban settlements with population estimates	– / 2015	(Denis et al., 2008)
Agroecological zones	Agroecological zones for Africa (reclassified)	– / 2015	(RCMRD, 2015)

Additionally, we used the 'WorldCover' land-use/land-cover map to extract stratified reference samples for non-built-up areas (Table 1) (cf. Sect. 3.2). To account for Africa's environmental diversity and explore location-specific model configurations (cf. Sect. 3.3), we selected calibration and test sites based on two additional datasets: the Africapolis dataset, which includes the location, extent, and estimated population of urban settlements in Africa, and an African Agroecological Zones (AEZ) map, reclassified into four main classes: arid, semiarid, subhumid, and humid (Table 1). These datasets supported the selection of geographically diverse test regions with different sizes under varying environmental conditions.

2.3. State-of-the-art built-up datasets available for the study period

As introduced, both mono-temporal (WorldCover v200, WSF2019, as well as diverse Building Footprint datasets) and multi-temporal (Dynamic Word, ESRI land cover, GLC_FCS30D, GISA, GISD30, GHSL2023) global land cover and built-up products are available for Africa. To assess the performance of our DIY-BU approach, we compare our results (see Sect. 3.3.2) against SOTA datasets. For multi-temporal comparison, we selected datasets covering the period 2016–2023, while mono-temporal comparison focuses on datasets aligned with the 2022 reference period (Table 2). The comparison methodology is described in Sect. 3.4.

Table 2
Global built-up and land cover maps representing built-up areas used for comparison with SOTA datasets. The years shown refer to those used in the comparison, not the full temporal coverage.

Name	Description	Type / Resolution / Years	Source
Dynamic World (DW)	Land cover (LC) probability derived from S-2 including 9 land covers, one being built-up areas. The annual mode was used to extract built-up areas from 2016 to 2022.	LC / 10 m / 2016–2023	(Brown et al., 2022)
ESRI land cover (ESRI)	Multi-temporal land cover classification derived from S-2, including 9 land covers, one being built-up areas.	LC / 10 m / 2017–2023	(Karra et al., 2021)
WorldCover v200 (ESA)	Land cover classification for 2021 derived from S-1 and S-2, including 11 land covers, one being urban areas.	LC / 10 m / 2021	(Zanaga et al., 2022)
WSF2019 (WSF)	Built-up (BU) map for 2019 derived from S-1 and S-2, representing built-up structures excluding roads.	BU / 10 m / 2019	(Marconcini et al., 2021)
GLC_FCS30D (GLC)	Land cover classification from 1985 to 2022 derived from Landsat imagery, including 35 land covers, one of which is impervious surfaces.	LC / 30 m / 2016–2022	(Zhang et al., 2021)
GISA	Multi-temporal global impervious areas derived from Landsat imagery, version 2.0.	BU / 30 m / 2016–2019	(Huang et al., 2022)
GISD30 (GISD)	Global impervious-surface dynamics derived from Landsat imagery.	BU / 30 m / 2020	(Zhang et al., 2022)
GHSL2023 (GHSL)	Total area of built-up within 1-hectare grid cells, derived from S-2 and Landsat imagery.	BU / 100 m / 2020	(Pesaresi and Politis, 2023)

3. Methods

This section outlines the methodological framework of the study. We first describe the data pre-processing steps and the derivation of spectral and texture indices (Sect. 3.1), followed by the sampling strategy used for model training and validation (Sect. 3.2). We then introduce the machine learning models applied for built-up area classification and detail the experimental setup with which we test the models across various regions and scenarios reflecting different levels of data availability (Sect. 3.3). Finally, we present the accuracy metrics used to assess the model performance (Sect. 3.4).

All these processes were implemented in GEE, unless otherwise stated. GEE is a cloud-based geospatial analysis platform that requires no specialized local hardware or software, only internet access to run the web application and download results (Gorelick et al., 2017). However, GEE imposes certain limitations on machine learning workflows, including: (1) the absence of widely used model architectures, such as convolutional neural networks; (2) a limit on the size of individual queries (10 MB); (3) constraints on the size of training datasets, where the product of the number of features and samples must not exceed 20 million (equivalent to ~ 100 MB); (4) restrictions on the size of model parameters, which must not exceed 100 MB (Google for developers, 2023); and (5) a limitation in the task quota, restricting the number of concurrent tasks and volume of data that can be processed at a time per user. Accordingly, our methodology was designed with these limitations in mind. Prior to the experimental phase, we conducted a model calibration step to reduce the feature set and thus the computational load. This calibration involved selecting a subset of features, including descriptive statistics of image bands, spectral and texture indices for annual image collections, and slope, to balance model performance and computational efficiency. An overview of the workflow is provided in Fig. 1.

3.1. Pre-processing and indices

The pre-processing workflow focuses on generating annual cloud-free and cloud-shadow-free mosaics from S-2 imagery. For each year from 2016 to 2023, we select all available images within the Area of Interest (AOI) that have a cloud cover percentage below 75%. We use the S-2 cloud probability dataset developed by the Sentinel Hub's cloud detector (*s2cloudless*), as it has been previously validated (Skakun et al., 2022) and is available in GEE. Based on *s2cloudless,* we mask out areas covered with clouds based by following a set of steps, adapted from Braaten (2023): (1) pixels with a cloud probability greater than 40% are

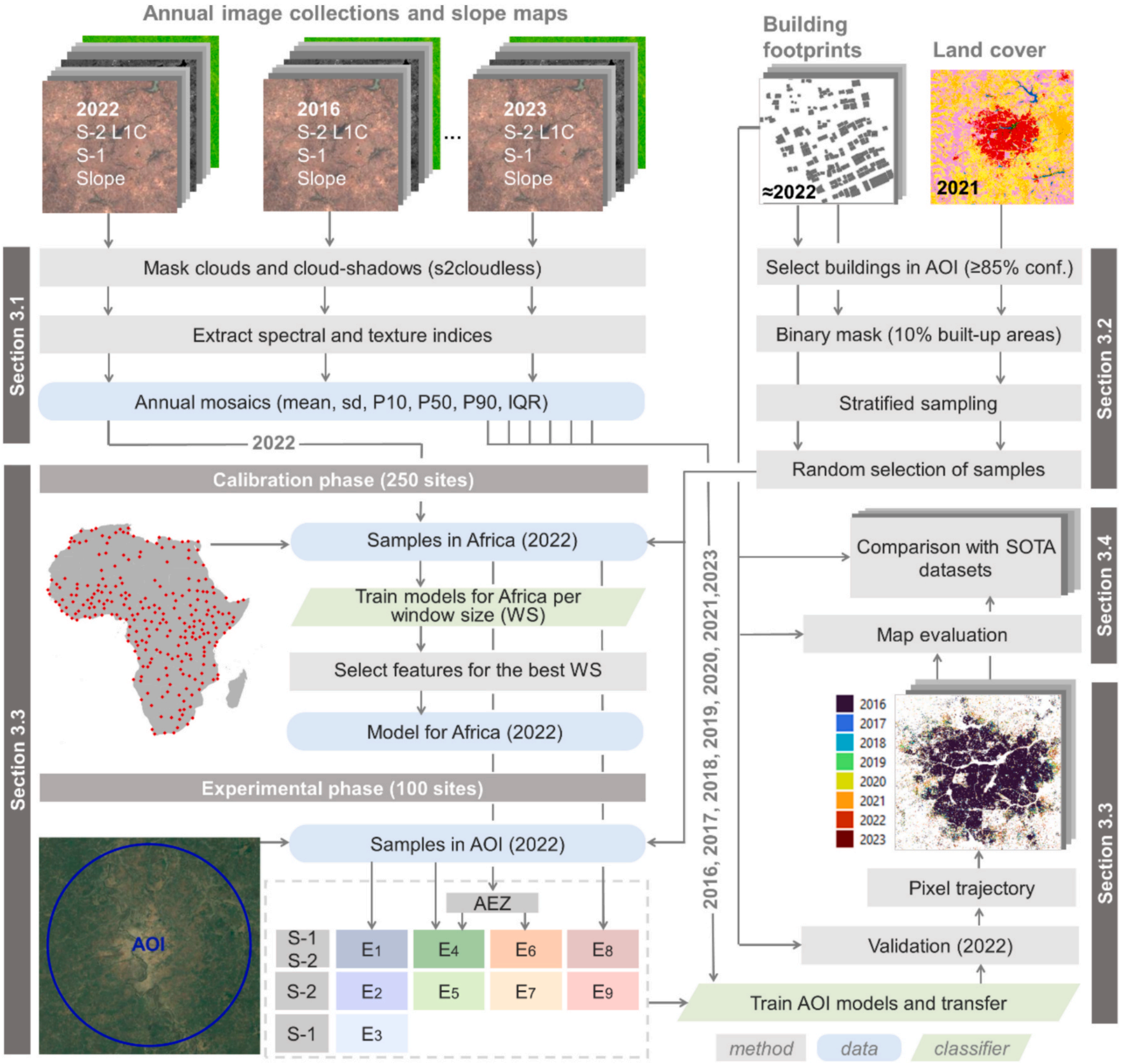


Fig. 1. Workflow of the study. The main steps are grouped according to the sections of the paper. The abbreviations used in the figure are: sd, standard deviation; P10, P50 and P90, percentiles 10, 50, and 90; IQR, interquartile range; AOI, area of interest or local test site; AEZ, agroecological zones; E, experiments one to nine as explained in Table 5; SOTA, state-of-the-art.

classified as 'clouds' (Aybar et al., 2022; Skakun et al., 2022); to identify cloud-shadow pixels, (2) water pixels are excluded from the set of potential shadow candidates using the Global Surface Water Explorer (GSW) dataset from the JRC (Pekel et al., 2016), specifically the 'Yearly Water Classification History' version 1.4, which provides annual classifications of water seasonality based on occurrence values detected throughout the year. Pixels classified as 'seasonal water' or 'permanent water' for the respective year are masked out. This step is critical because water absorbs almost all light and results in low digital numbers (DN) in the near infrared (NIR) band, similar to shadows. By filtering these pixels out of the shadow detection process, we ensure that they are not erroneously removed from the image stack. For the remaining non-water pixels, (3) those with DNs lower than 1,500 in the NIR band are classified as 'dark pixels' and are considered potentially cloud-shadow candidates as proposed by Braaten (2023). (4) A mode filter with a radius of 2 pixels is applied to the 'clouds' mask to reduce noise, generating a 'cleaned clouds' mask. (5) A maximum filter then extends the cleaned mask by 4 pixels to include semi-transparent cloud pixels surrounding the detected clouds that might not have been initially classified. Lastly, (6) the average sun azimuth angle is used to determine the shadow direction, and the cloud-shadows are projected from the 'cleaned clouds' mask accordingly. Pixels identified as 'dark' and lying along this projected direction are classified as 'shadows'. The final 'cloud-shadow' mask is obtained by combining the 'cleaned clouds' and 'shadow' masks, and small patches are removed using focal filters.

Once the cloud-shadow masks are generated for each individual image of the annual collections, the masks are applied to the corresponding bands. Subsequently, we calculate multiple spectral and texture indices for the cloud-shadow-free pixels. These indices are described in Table S1 and have been widely used to distinguish between built-up and non-built-up areas, with a particular focus on separating built-up from bare soils (Ettehadi Osgouei et al., 2019; He et al., 2010; Rasul et al., 2018; Waqar et al., 2012). As an additional feature, the slope of the terrain is extracted from the Copernicus digital elevation model (DEM) 'GLO-30 DEM' dataset, which provides a spatial resolution of 30-m (Copernicus, 2024). The slope is used in combination with the indices, as previously done by 'WorldCover' (ESA, 2020). With this additional feature we aim to constrain the classification of built-up areas based on the slope gradient.

Finally, to generate the annual mosaics, we calculate several statistics from the annual image collections. The mean is calculated for the S-2 and S-1 bands (i.e., B2, B3, B4, B8, VH, and VV), and the derived spectral indices. Additionally, for the S-1 bands and the spectral indices, we compute the percentiles 10, 50, and 90, the Interquartile Range (IQR), and the standard deviation. Finally, we calculate texture indices for the annual means of B2, B3, B4, and B8 using the Gray-Level Co-occurrence Matrix (GLCM) with various windows sizes (from 3x3 to 17x17 pixels), whose performance is tested in the calibration phase (cf. Sect. 3.3.1).

3.2. Sampling approach

We sample training and validation data from the reference data products available in GEE. For each AOI, we first select building footprints from the OB3 dataset with a confidence level ≥ 0.85 (i.e., a high likelihood of representing true buildings). This threshold prioritizes high-quality reference data and reduces the inclusion of false positives. If no OB3 buildings are present in an AOI, we use the MBF dataset instead. To ensure consistency with the S-2 resolution, we rasterize the building footprints to create a binary 'built-up mask' at 10-m resolution. In this step, pixels are considered built-up if at least 25% of their area is covered by buildings. This mask is used both to randomly select built-up samples and to exclude potentially missing built-up areas from the ESA land cover product.

To remain within the training data size limits imposed by GEE, the number of samples is limited to 50,000 per AOI. From the 'built-up mask', up to 25,000 pixels are randomly selected and assigned to class 1 (built-up). For non-built-up samples, we applied a stratified sampling approach using the ESA land cover map ensuring representation across land cover type, and excluding built-up areas from the mask to avoid overlap. For each land cover class, 5% of the available pixels are selected, with a minimum of 500 pixels and a maximum of 10,000 pixels. This allows for a more balanced representation. A final random sample is then drawn from this pool to complete the non-built-up portion of the 50,000 total points. Non-built-up samples are assigned to class 0 (non-built-up).

Although the thresholds used in this sampling procedure, such as the OB3 confidence threshold, the 25% pixel coverage criterion, and the sample size of 50,000 points, have been tested and are recommended, they are not hard-coded into the tool. Users can easily adjust these parameters within the GEE script to accommodate different contexts or data availability. For example, increasing the sample size could help to capture diverse spectral signatures in highly heterogeneous urban landscapes, whereas adjusting the OB3 confidence threshold could facilitate the detection of less visible or obvious informal buildings while also avoiding false positives. Additionally, adjusting the rasterization percentage could facilitate the identification of smaller buildings by defining the minimum pixel coverage required during training. This flexibility makes the tool more applicable in a broader range of situations. Detailed instructions for modifying these settings can be found in the step-by-step user guide in the Supplementary Material. Additionally, a sensitivity analysis of the pixel coverage threshold (Fig. A2) and sample size (Fig. A3) is included in the Appendix to illustrate the potential effects of these adjustments.

3.3. Supervised classification of built-up areas

We aim to identify the most efficient and accurate classification model for the DIY-BU-mapping tool. To this end, we first fit a series of preliminary models, which are structured in two stages: a *calibration phase* and an *experimental phase*.

In the *calibration phase*, we train models for Africa using sample data extracted from a diverse set of 250 AOIs. These models are fitted using all available features, including spectral and texture indices, as well as slope information, while varying the window sizes for the texture metrics. Based on the accuracy assessment of these models, we identify the best window size. Once this is established, we apply correlation analysis to detect and remove highly correlated features.

In the second stage, we use the calibration results as a baseline for the models applied in the *experimental phase*. Here, we test the designed experimental setup with a distinct set of 100 AOIs and evaluate the performance of different model configurations. Fig. 2 illustrates the distribution of sampled AOIs for both the calibration and the experimental phases.

3.3.1. Model calibration

As previously outlined, the model calibration phase serves to optimize computing time and prevent memory errors in GEE, as well as to avoid the use of redundant features. In this phase, we train models for Africa with the objective of determining the optimal window size for the texture indices in combination with the spectral features. Then, we reduce the dimensionality by removing highly correlated features, and quantify the accuracy loss. The final model configuration is then used as a baseline for the experimental setup.

To this end, we gathered sample data for representative sites across Africa for the year 2022. In total, we selected 250 AOIs using a random sampling approach stratified by geographic characteristics. Specifically, we considered geographical location, size (total built-up area), total population, and AEZ (Fig. 2). These characteristics are derived from the Africapolis dataset (Table 1). For each of the selected 250 urban areas, we defined an influence zone of 5 km to include the urban–rural border. Within each AOI, we selected 1,000 sample pixels of built-up and 1,000 sample pixels of non-built-up, following the sampling approach

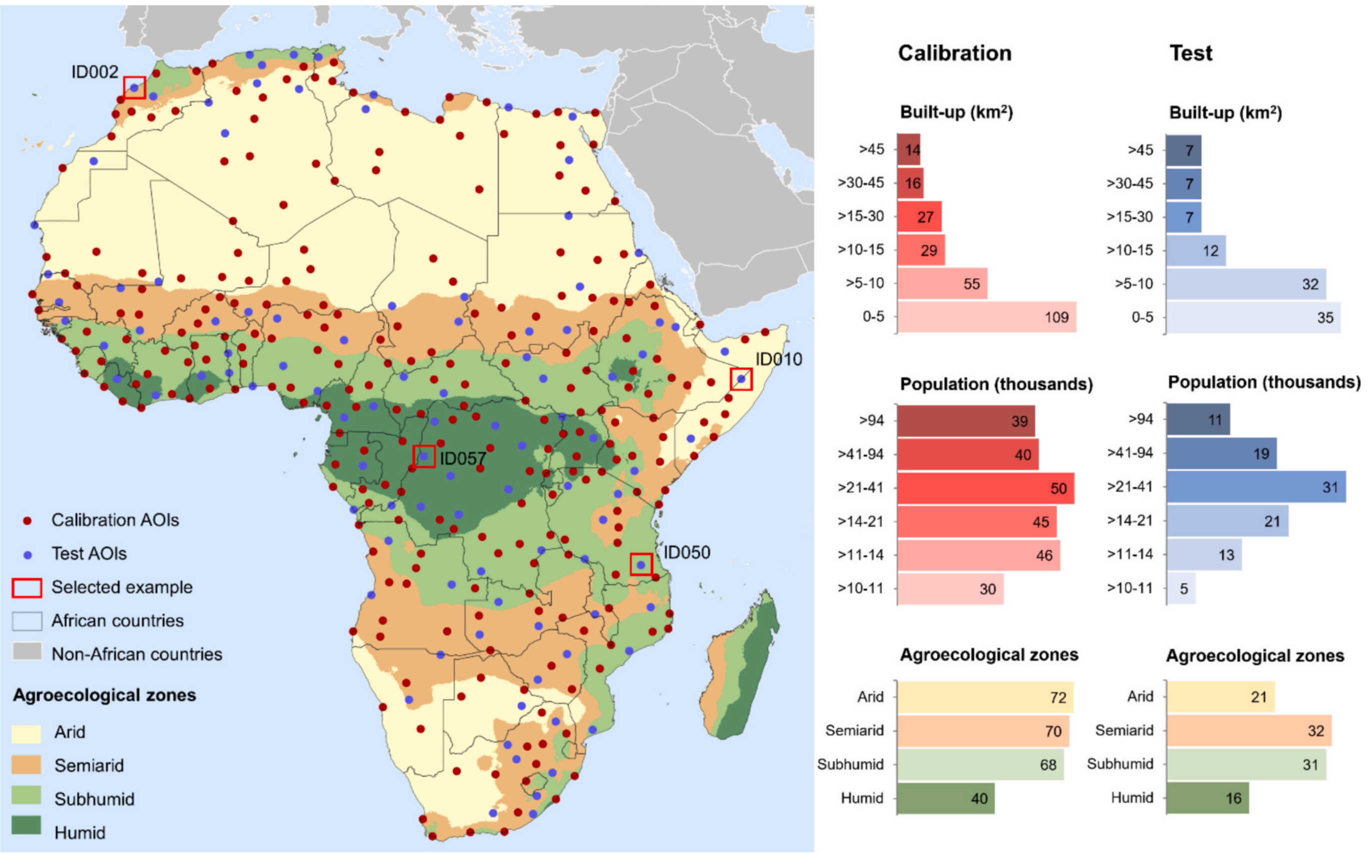


Fig. 2. On the left, the location map of sampled Areas of Interest (AOIs) is displayed. In red 250 AOIs used in the calibration phase are shown. In blue 100 AOIs used to test the experimental setup are displayed. The red squares show the selected AOIs for detailed qualitative and quantitate evaluations. In the background of the map, the agroecological zones (AEZ) can be seen. On the right, a set of statistics summarize the diversity of total built-up area, population and AEZ for the selected AOIs for the calibration and test phases.

described in Sect. 3.2.

Out of the 250 AOIs, 187 had reference data from OB3, 11 from MBF, 17 did not have building reference data, and 35 lacked Sentinel-1 data for 2022. In total, we generated nearly 400,000 sample pixels and computed the values for the spectral bands, indices, slope, statistics, and GLCM at various window sizes. Because of the full dataset exceeds GEE's processing limits (Google for developers, 2023), we exported the sample data for each AOI and conducted the calibration phase in R, where we performed the training, validation, and feature selection.

For the classification, we used a random forest model, as it is available in GEE and has been widely demonstrated to reliably perform in built-up area mapping, often outperforming methods such as support vector machines, gradient boosting, and maximum likelihood classifier (Amarsaikhan et al., 2024; Ouma et al., 2023). The model, with 100 trees, was trained with 70% of the sample data and validated with the remaining 30%. The results show that the best-performing model uses a window size of 17x17 pixels for the texture indices, along with a total of 169 features, including the mean, standard deviation, median, P10, P90 and IQR for 13 S2- and 3 S1-based spectral indices, as well as 17 texture indices for bands B2, B3, B4 and B8 (Table 3).

Model accuracy was found to decrease as the window size used for texture indices decreased. The lowest accuracy was observed in the model without texture metrics. To further reduce the number of features and ensure compatibility with GEE, we excluded highly correlated features using a correlation threshold of 0.85, resulting in a final set of 86 uncorrelated features. Although the model trained with all features and the optimal window size showed a slightly better performance, the version using uncorrelated features resulted in only a marginal reduction of 0.48% F1-score (Table 3). This small trade-off is justified by the significant improvements in computational efficiency and stability, which are critical given GEE's memory and processing constraints. While random forest models are generally robust to multicollinearity, reducing feature redundancy allowed us to avoid processing failures and ensure broader usability of the tool. As a result, we selected the model with uncorrelated features and a 17x17 window size for the experimental setup. The final subset of selected metrics is marked with an asterisk in Table S1.

Table 3
Calibration models' accuracy metrics with all features (spectral indices and grey-level co-occurrence matrix, GLCM) and with uncorrelated features (using the best window size). For a description of the accuracy metrics, see Sect. 3.4.

Model	No. of features	OA	Recall	Precision	F1-score
All (–)	101	0.9090	0.9079	0.9046	0.9062
All (GLMC 3x3)	169	0.9240	0.9370	0.9089	0.9227
All (GLMC 5x5)	169	0.9253	0.9368	0.9114	0.9239
All (GLMC 7x7)	169	0.9244	0.9346	0.9115	0.9229
All (GLMC 9x9)	169	0.9248	0.9336	0.9130	0.9232
All (GLMC 11x11)	169	0.9255	0.9335	0.9145	0.9239
All (GLMC 13x13)	169	0.9265	0.9340	0.9160	0.9249
All (GLMC 15x15)	169	0.9273	0.9345	0.9169	0.9257
All (GLMC 17x17)	169	0.9281	0.9352	0.9178	0.9264
Uncorr. (GLMC 17x17)	86	0.9233	0.9303	0.9129	0.9216

3.3.2. *Experimental setup*

The availability of datasets and the geographic specifics differ across Africa. Against this background, we compare the performance of similar models under these varying conditions. In this section, we assess the capabilities of nine different models, defined by input data and regional specifics (Table 4). The experimental setup is designed to investigate several scenarios, including the evaluation of model's performance using different sensors. This is particularly relevant, as many regions in Africa lack S-1 coverage, or the persistence of clouds limits the availability of S-2 imagery.

Moreover, during the calibration phase, it was observed that reference data is unavailable for many regions. As a result, it becomes important to evaluate the performance of models trained for Africa when no reference data tailored to a specific AOI is available. Additionally, given the substantial environmental diversity observed across the continent, the question arises whether models trained specifically for each agroecological zone offer improvements over models trained for Africa as a whole.

To test these scenarios, we select a diverse set of 100 AOIs based on the same criteria as the calibration setup, ensuring they are excluded from the calibration sample (Fig. 2), and generate site-specific sample data for each. Subsequently, for each AOI, we train one model for the year 2022 under all scenarios of data availability levels (E1-E9, Table 4). These models are then applied to infer built-up areas for earlier and later years based on the corresponding feature sets. For each site, this results in eight annual built-up area maps from 2016 to 2023, which are subsequently combined to generate urban expansion maps (Sect 3.3.3). Of the 100 AOIs, 94 had building reference data (89 from OB3 and 5 from MBF). In addition, 15 AOIs lacked S-1 data during the study period, primarily due to the end of the Sentinel-1B satellite mission in 2022. For these AOIs, only scenarios E2, E5, E7 and E9 were calculated.

All models are trained in GEE using the 86 selected spectral and texture indices (window size 17x17) together with their associated statistics. We applied the GEE implementation of a random forest classifier (*ee.Classifier.smileRandomForest*) with 100 trees and a fixed seed to ensure reproducibility. The models are trained on 70% of the sample data and validated on the remaining 30%, which is not seen by the model during training. This process generates two outputs: first, a multitemporal built-up classification for the AOIs (explained in Sect 3.3.3), representing the expansion of built-up areas since 2016; and second, a validation report including overall accuracy, recall, precision, and F1-score. It is important to note that the validation report applies only to the built-up areas classified for 2022, and not to the multi-temporal map. This is because 2022 is the only year for which time-aligned reference data is available. For earlier and later years, validation is not feasible due to the lack of consistent ground truth data. Following the experimental phase, the scenario exhibiting the best performance is implemented in the DIY-BU-mapping tool, as further developed in Sect. 4.4.

Table 4

Description of the nine tested scenarios ordered by data availability (reference and sensor data) from full to low availability (E1-E9). For reference data, AOI means area of interest, AEZ means agroecological zones and A means Africa. For sensor, S1 refers to Sentinel-1, and S2 to Sentinel-2.

	Reference data			
Sensor	**Samples in AOI**	**Samples from the same AEZ in Africa**		**Samples in Africa**
S-1/S-2	E1 (AOI-S1S2) Scenario with the best condition; all sensors and reference data for the AOI are available	E4 (AOI-AEZ-S1S2) All sensors and reference data for the AOI are available; the model is augmented with samples from the same AEZ in Africa	E6 (AEZ-S1S2) All sensors are available; reference data for the AOI are not available; the model is trained with samples from the same AEZ in Africa	E8 (A-S1S2) All sensors are available; reference data for the AOI are not available; the model is trained with samples from Africa
S-2	E2 (AOI-S2) Reference data for the AOI are available; S-1 is unavailable	E5 (AOI-AEZ-S2) Reference data for the AOI are available; S-1 is unavailable; the model is augmented with samples from the same AEZ in Africa	E7 (AEZ-S2) S-1 is unavailable; reference data for the AOI are not available; the model is trained with samples from the same AEZ in Africa	E9 (A-S2) S-1 is unavailable; reference data for the AOI are not available; the model is trained with samples from Africa
S-1	E3 (AOI-S1) Reference data for the AOI are available; S-2 is unavailable; not affected by cloud coverage			

3.3.3. *Multi-temporal classification and pixel change trajectory analysis*

The outputs of the multi-temporal classification are annual binary masks of built-up and non-built-up pixels from 2016 onwards. While validation for 2022 shows high accuracy, some errors persist due to the use of Level-1C S-2 data, model transfer across time, inaccuracies in training data, and spectral confusion related to land cover complexity or building materials. To mitigate temporal inconsistencies, such as abrupt appearance or disappearance of built-up areas, we apply a post-classification pixel change trajectory analysis that assumes a unidirectional transition from non-built-up to built-up. This monotonic urban growth assumption (i.e., a unidirectional growth of built-up areas), widely adopted in the literature (Kühnl et al., 2023; Pesaresi et al., 2024; Taubenböck et al., 2012; Van de Voorde et al., 2016), helps smooth artefacts caused by clouds, data gaps or model instability. We acknowledge that this simplification may not capture processes such as demolition, redevelopment, or the removal of temporary settlements. This could result in overestimation in such cases. However, it provides a more stable and interpretable representation of long-term built-up expansion, which is the tool's primary objective. Users seeking to detect short-term or reversible changes can disable the trajectory correction by modifying a few lines of code in the script (an example is provided in the tool's user guide, Supplementary Material). In doing so, annual classification outputs can be used to explore building removal, though these outputs may be noisier and lack year-specific validation.

Each pixel in our classification takes one of three values: 'non-built-up' (0), 'built-up' (1), or 'no data' (2), the latter indicating persistent clouds throughout the year. With eight annual built-up maps (2016–2023), this yields 6,561 possible temporal trajectories.

To combine these, each annual map is multiplied by $10^{(\text{last year} - \text{current year})}$, and the values are summed, creating a unique trajectory code per pixel. Consistent trajectories (e.g., '000111**1**1′ indicating built-up from 2019 onwards) are considered reliable, while erratic ones (e.g., '110101**1**0′ with alternating built-up and non-built-up years) suggest noise or misclassification. To address such cases, we apply temporal smoothing rules based on a majority filter, prioritizing the 2022 classification, the most robust year due to targeted training (the value for 2022 is highlighted in bold in the example pixel trajectories for reference). In the example above, '110101**1**0′ would be corrected to '111111**1**1′ to enforce temporal consistency. The final output is a categorical map showing the year each pixel was first classified as built-up, with values from 0 to 2023 (e.g., '000111**1**1′ becomes 2019, and '111111**1**1′ becomes 2016).

3.3.4. *Model fine-tuning*

To improve the accuracy of the maps, the model can be fine-tuned by incorporating additional training data in specific locations, following a strategy conceptually similar to active learning methods. In this process, users actively identify areas where the model underperforms, such as regions with systematic misclassifications or challenging urban forms or materials, and provide targeted feedback. Specifically, users draw polygons around these areas to indicate the need for refinement (an example is provided in the step-by-step user guide, Supplementary Material). These polygons serve as spatial input for the generation of new training data, which is sampled using the same procedure described in Sect. 3.2, yielding up to 10,000 new labelled pixels. By focusing on model weaknesses, this guided sampling approach enables more efficient improvement than simply increasing the training set size through random sampling. The model is then retrained for the corresponding AOI using the expanded dataset, resulting in a more context-aware classification. This interactive fine-tuning capability is integrated into the DIY BU mapping tool (Section 4.4), supporting a user-driven, iterative improvement process.

3.4. Accuracy assessment

We assess accuracy in two ways: First, the *model validation* measures the performance of the model on a separate portion of the sample data that was not used for training (i.e., the 30% validation split), providing an estimate of how well the model generalizes to unseen data. This validation focuses solely on the model's performance during training. Second, the *map evaluation* quantifies how well the final classification map matches a reference map, providing a measure of agreement. Unlike the *model validation,* which relies only on part of the data, this evaluation compares the full map against a series of reference maps (see Table 1). We also conduct the *map evaluation* for the SOTA datasets to assess their agreement with the reference maps. All datasets were resampled or aligned to a common 10-m grid using nearest-neighbour resampling in GEE to ensure pixel-wise comparability.

For the evaluation of the maps, we generate rasterized built-up masks from the reference datasets listed in Table 1, at a spatial resolution of 10-m. When confidence scores are available, all buildings with a score ≥ 0.65 are retained. To prevent small buildings from being omitted during the evaluation, we apply a 10% pixel-coverage threshold when rasterizing the footprints (Fig. A2). A sensitivity analysis of alternative rasterization thresholds demonstrated that this parameter has a strong influence on the resulting reference mask and can substantially affect the measured accuracy, highlighting the importance of selecting an appropriate threshold that aligns with the specific definition of built-up areas for the intended application.

We apply a set of common accuracy metrics derived from the confusion matrix, which compares the predicted class with the reference class. In our binary classification, four outcomes are possible: true positive (TP), when the model correctly predicts the 'built-up' class; true negative (TN), when it correctly predicts the 'non-built-up' class; false positive (FP), when it incorrectly predicts the 'built-up' class; and false negative (FN), when it incorrectly predicts the 'non-built-up' class. Based on these, we compute four metrics:

- *Overall accuracy* (OA) measures the overall confidence in the classification, calculated as the ratio of correctly classified pixels to the total number of pixels [(TN + TP) / (TN + TP + FN + FP)].
- *Recall* measures the proportion of correctly classified built-up pixels out of all reference built-up pixels [TP / (TP + FN)].
- *Precision* is the percentage of correctly classified built-up pixels out of all pixels predicted as built-up [TP / (TP + FP)].
- *F1-score* is the harmonic mean of precision and recall, offering a balance measure, especially useful in case of class imbalance [2 · (precision · recall) / (precision + recall)]. The F1-score accounts for both FP and FN, providing an overall measure of the model's ability to correctly identify the built-up class.

Finally, to evaluate the agreement between the GHSL and the building footprint reference maps, we apply a different approach due to the continuous nature of the GHSL data (i.e., built-up area proportion rather than a binary classification). First, we calculate the total building footprint area per hectare using the reference datasets at the same resolution as the GHSL. Then, we compare these values with the GHSL dataset using the Concordance Correlation Coefficient (CCC). The CCC quantifies the agreement between two continuous variables (Lin, 1989) and ranges from −1 to 1, where 1 indicates perfect agreement, 0 no agreement, and −1 perfect inverse agreement.

4. Results

In the following, we present the results of the experimental setup based on the measured accuracies from the model validation for the year 2022 (Sect. 4.1), along with the map evaluation and a comparison with SOTA global products (Sect. 4.2). We then show the improved maps obtained through model fine-tuning (Sect. 4.3) for selected AOIs (Fig. 2). Finally, we introduce the DIY-BU-mapping tool (Sect. 4.4), which implements our best model configuration identified during the calibration and experimental phases.

4.1. Validation of models for test AOIs

Table 5 summarizes the mean accuracy metrics for the test AOIs, overall, per AEZ, and across different urban densities (UD) calculated by aggregating the total built-up area from GHSL and dividing it by the AOI area. Results are reported for all nine model scenarios (E1-E9).

The three models with the highest measured accuracies are those from scenarios E1 (models trained with samples from the AOI, using both S-1 and S-2 data), E2 (samples from the AOI, using S-2 only), and E4 (samples from the AOI, augmented with African samples from the same AEZ, using S-1 and S-2). In contrast, the model with the lowest accuracy is from scenario E3, where only S-1 imagery were used for classification.

These results are consistent with findings from previous studies, as models combining S-1 and S-2 images typically outperform models based on a single sensor (Hafner et al., 2022; Marconcini et al., 2021; Zhu et al., 2022). Additionally, we observed that augmenting the training set with samples from the same AEZ does not lead to accuracy improvement in any scenario, for example, comparing E1 vs. E4, or E2 vs. E5.

As expected, models trained exclusively on samples from entire Africa (E6-E9) perform worse than those trained with site-specific samples, although the high average accuracies still suggest strong generalization capabilities of these global models. Interestingly, the use of AEZ-specific samples in Africa improves model performance slightly compared to using samples from the entire continent (e.g., E6 vs. E8), particularly in arid and humid zones and for low-urban-density AOIs.

Model accuracy also varies by AEZ (Table 5, Fig. A4). In general, models trained with AOI-specific samples (E1-E5) perform better in arid and semiarid zones. Notably, for E3 (the S-1 only scenario), accuracies are substantially higher in arid regions compared to other zones. Similarly, Africa-wide models (E6-E9) tend to perform better in arid and semiarid zones, likely due to a higher frequency of cloud-free images in these areas. Unlike the AOI-trained models (E1-E5), the Africa-wide models benefit from AEZ-specific sampling, as seen when comparing E6 vs. E8 and E7 vs. E9. The comparatively lower accuracies in subhumid regions may be linked to the higher landscape heterogeneity and geographical distance between AOIs in these zones, whereas arid and semiarid areas exhibit more uniform characteristics across Africa.

Accuracy also varies with urban density (Table 5 and Fig. A5). Across all scenarios, models perform best in urban high-density AOIs, while performance decreases slightly in AOIs of medium settlement density and more substantially in low-density settlement AOIs. This pattern is likely due to both the larger number of built-up reference samples available in high-density areas, and the known lower accuracies of building footprint datasets in rural regions (Gevaert et al., 2024).

Fig. 3 shows the distribution of F1-scores across test AOIs using a violin plot, which show both the smoothed distribution and key summary statistics (median, percentiles and outliers). Models trained in

Table 5
Mean accuracy statistics for the models of the tested scenarios ordered by data availability (reference and sensor data) from full to low availability based on the test sample. Mean recall (R), precision (P) and F1-scores (F1) are reported only for the 'built-up' class. Mean F1 are reported for each agroecological zone (AEZ) and three urban density (UD) levels: high 10–1.6% (H), medium 1.6–0.9% (M), low 0.9–0% (L). The AEZ are: arid (A), semiarid (SA), subhumid (SH) and humid (H).

		Mean accuracy metrics				**Mean F1 per AEZ**				**Mean F1 per UD**		
Experiment		**OA**	**R**	**P**	**F1**	**A**	**SA**	**SH**	**H**	**H**	**M**	**L**
E1	AOI-S1S2	0.975	0.905	0.937	0.916	0.910	0.939	0.917	0.892	0.946	0.922	0.875
E2	AOI-S2	0.972	0.897	0.933	0.909	0.909	0.921	0.907	0.893	0.941	0.908	0.865
E3	AOI-S1	0.906	0.635	0.839	0.696	0.769	0.785	0.686	0.517	0.795	0.683	0.590
E4	AOI-AEZ-S1S2	0.973	0.915	0.912	0.912	0.906	0.933	0.913	0.891	0.942	0.916	0.874
E5	AOI-AEZ-S2	0.970	0.909	0.907	0.906	0.903	0.915	0.906	0.892	0.937	0.902	0.865
E6	AEZ-S1S2	0.949	0.884	0.854	0.860	0.833	0.866	0.869	0.866	0.893	0.855	0.825
E7	AEZ-S2	0.938	0.878	0.822	0.837	0.821	0.817	0.851	0.866	0.875	0.819	0.800
E8	A-S1S2	0.948	0.874	0.846	0.849	0.793	0.866	0.868	0.855	0.898	0.850	0.790
E9	A-S2	0.938	0.877	0.818	0.833	0.804	0.819	0.854	0.859	0.885	0.814	0.780

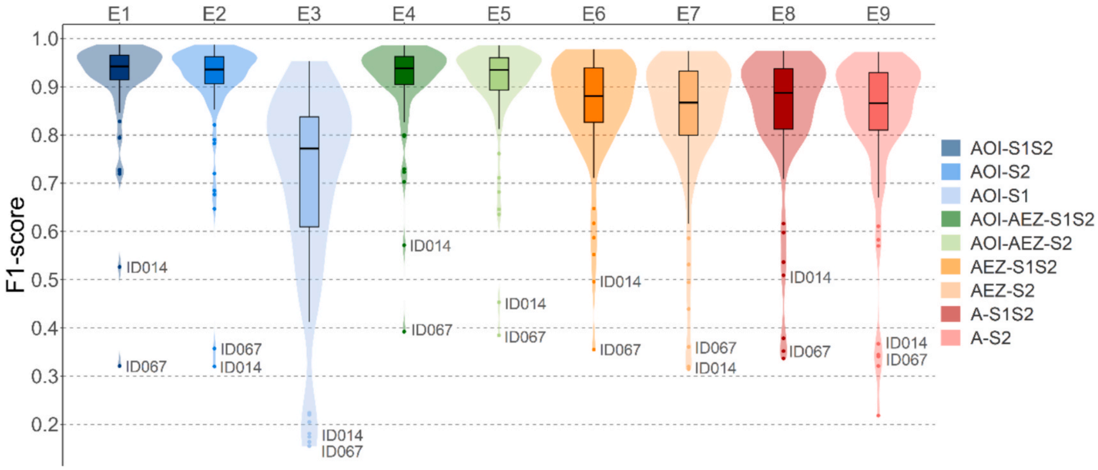


Fig. 3. Violin and boxplots showing the distribution of F1-scores for the test AOIs for the different scenarios using the test sample. Two AOIs are highlighted as they are outliers in most scenarios (ID014 and ID067). Where E1 to E9 are the scenarios with different levels of data availability (reference and sensor data) from full to low availability. For reference data, AOI means area of interest, AEZ means agroecological zones and A means Africa. For sensor, S1 means Sentinel-1, and S2 Sentinel-2.

scenarios E1-E5 consistently outperform those from E6-E9, with most F1-scores exceeding 0.8, except for E3, as previously noted. Outliers are predominantly found in AOIs located in subhumid (e.g., ID014) and humid zones (e.g., ID067), especially in the case of E3, where the outliers are concentrated in humid zones. Fig. A6 illustrates two challenging AOIs: ID014 (Gemmeiza, South Sudan) and ID067 (between Ambo, Democratic Republic of the Congo, and Girili, South Sudan). In both cases, all model scenarios yielded very low classification accuracies. This is likely due to the combination of extremely low urban densities (0.08% and 0.72%, respectively) and their location in subhumid and humid AEZs, where persistent vegetation cover complicates built-up detection. The OB3 dataset in ID014 contains 4,303 building footprints with a mean area of 19 m^2, of which only 95 had a confidence score of at least 0.85. Similarly, ID067 includes 8,886 buildings with a mean area of 20 m^2, but only 329 of these met the high-confidence threshold. These figures highlight the limited availability of high-quality reference data in some AOIs, which restricts the model's capacity to learn and accurately classify built-up areas in such environments.

4.2. Quantitative map evaluation and comparison with state-of-the-art datasets

We quantified the agreement between the built-up area maps generated under scenarios E1-E9 and several SOTA datasets by comparing them against reference building footprint maps (OB3, MBF, OFM, OBT, GBA, and FB).

Among the evaluated datasets in African AOIs, the WSF achieved the highest mean F1-score and precision when evaluated against all reference datasets (Figs. A7-A8, Table 6); however, recall was slightly lower compared to other products. This performance is likely influenced by WSF's binary classification scheme (built-up vs. non-built-up), which differs from the multi-class land cover approaches used by other SOTA datasets. Additionally, WSF applies targeted post-processing steps, including the masking of roads (Marconcini et al., 2021), which improves agreement with building footprint datasets. In contrast, other products frequently misclassify roads as built-up due to spectral similarities and spatial adjacency to buildings. The lower recall observed may be partly due to the fact that the WSF data represent conditions in 2019, whereas the reference data are primarily from 2022. Consequently, WSF does not capture more recent developments, resulting in a higher number of false negatives and reduced recall.

Our built-up area maps, particularly those from scenario E1 (local models trained with S-1 and S-2 samples), achieved the second-highest mean F1-scores in the tested AOIs, slightly outperforming scenario E4 (which also includes AEZ-based African samples) by approximately 1% (Fig. 4, Fig. A7). Among the SOTA datasets, ESA's product ranked next, likely due to its post-processing strategy that refines class labels using auxiliary data, including the WSF (ESA, 2020). Overall, WSF, E1, E4, as well as E2, E5 and ESA provide the highest agreement with the reference datasets for 2022 across our test AOIs in Africa. However, it is important to note that WSF and ESA provide only static snapshots, limiting their utility for monitoring built-up expansion over time. Furthermore, these findings are context-specific and may not generalize to other regions or

Table 6
Mean accuracy metrics in general and per agroecological zone (AEZ) for the full map assessment using FB reference data. Values in bold highlight the two best maps. The Concordance Correlation Coefficient (CCC) only applies to the GHSL product, it is not comparable to the F1-score. These are: arid (A), semiarid (SA), subhumid (SH) and humid (H).

	Mean F1-score / *CCC*					Mean recall					Mean precision				
Model	All	A	SA	SH	H	All	A	SA	SH	H	All	A	SA	SH	H
E1	**0.593**	0.617	**0.594**	**0.590**	**0.570**	0.788	0.759	**0.826**	0.790	0.763	0.505	0.567	0.481	0.491	0.492
E2	0.567	0.579	0.552	0.575	0.569	0.776	0.768	0.774	0.788	0.766	0.481	0.515	0.462	0.474	0.489
E3	0.369	0.474	0.393	0.343	0.269	0.599	0.684	0.689	0.582	0.419	0.326	0.405	0.307	0.303	0.305
E4	0.586	**0.620**	0.584	0.578	0.568	**0.810**	0.768	**0.837**	0.821	0.804	0.478	0.563	0.459	0.454	0.451
E5	0.558	0.578	0.536	0.562	0.567	**0.800**	0.773	0.792	**0.825**	0.803	0.450	0.509	0.424	0.436	0.451
E6	0.496	0.523	0.474	0.482	0.520	0.799	0.758	0.768	0.821	**0.849**	0.387	0.441	0.371	0.366	0.386
E7	0.450	0.460	0.398	0.460	0.523	0.792	0.749	0.760	**0.826**	**0.845**	0.344	0.373	0.298	0.348	0.392
E8	0.491	0.452	0.455	0.508	0.556	0.788	**0.799**	0.798	0.812	0.720	0.383	0.351	0.341	0.387	0.471
E9	0.448	0.410	0.381	0.486	0.555	0.789	**0.795**	0.791	0.816	0.724	0.338	0.300	0.272	0.367	0.466
DW	0.374	0.276	0.372	0.421	0.417	0.650	0.386	0.720	0.749	0.663	0.316	0.359	0.289	0.312	0.323
ESRI	0.410	0.461	0.378	0.416	0.394	0.765	0.734	0.760	0.819	0.714	0.306	0.382	0.280	0.286	0.292
GISA	0.280	0.267	0.223	0.310	0.359	0.234	0.226	0.175	0.265	0.305	**0.578**	0.524	**0.553**	**0.619**	**0.605**
GLC	0.456	0.469	0.417	0.484	0.465	0.518	0.514	0.481	0.549	0.539	0.443	0.428	0.416	0.481	0.445
WSF	**0.660**	**0.668**	**0.649**	**0.672**	**0.647**	0.714	0.709	0.733	0.713	0.687	**0.640**	**0.669**	**0.609**	**0.658**	**0.632**
ESA	0.517	0.517	0.505	0.529	0.515	0.575	0.579	0.562	0.584	0.577	0.525	**0.569**	0.511	0.527	0.490
GISD	0.385	0.396	0.386	0.382	0.372	0.391	0.430	0.383	0.376	0.387	0.509	0.492	0.494	0.556	0.469
GHSL	*0.761*	*0.742*	*0.772*	*0.783*	*0.722*										

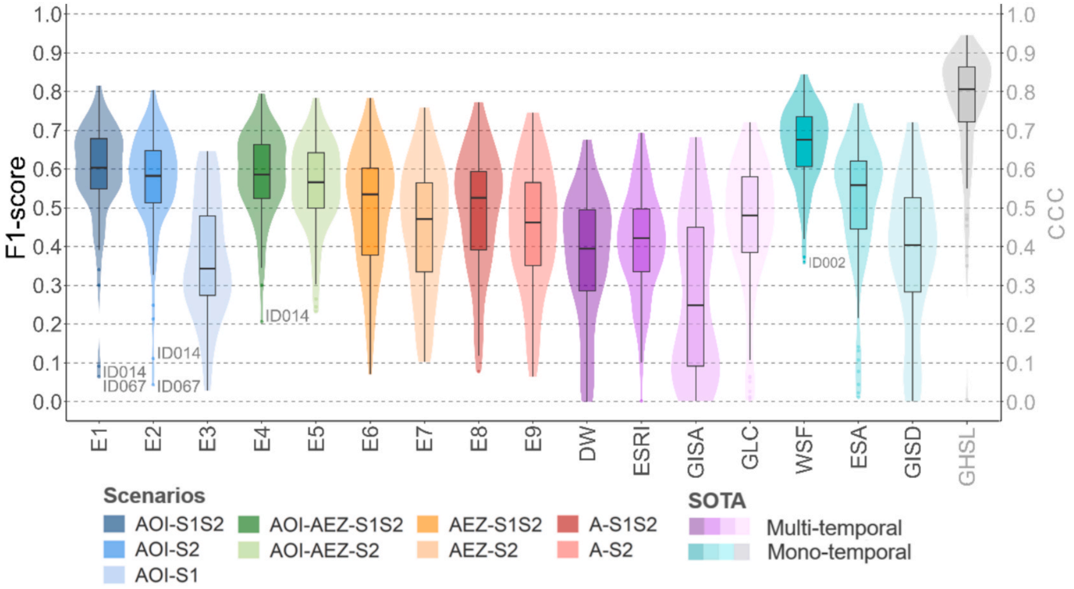


Fig. 4. Violin and boxplots illustrating the distribution of F1-score and Concordance Correlation Coefficient (CCC) for test AOIs across different scenarios and global products using the fused building (FB) reference maps. Scenarios E1 to E9 represent varying levels of data availability, from full to low (reference and sensor data). The global products include Dynamic World (DW), ESRI Land Cover (ESRI), GISA, GLC_FCS30D (GLC), WSF2019 (WSF), WorldCover v200 (ESA), GISD30 (GISD), and GHSL2023 (GHSL), color-coded by the sensors used. Note that CCC applies only to the GHSL product and is not comparable to the F1-score, as it measures agreement in total built-up area per hectare.

datasets not assessed in this study.

In our experimental setup, the E1 configuration (using all available sensors and local reference data) achieved the highest mean accuracy among the tested scenarios. However, F1-scores for E1 varied substantially across AOIs, ranging from as low as 0.07 (in outliers ID014 and ID067, which had very low urban density and limited coverage of high-quality reference data; see Fig. 4 and Fig. A6) to as high as 0.81 (ID055), with a mean F1-score of 0.59 based on the FB reference maps. Results using other reference datasets are presented in Fig. A7. These results are consistent with other studies mapping built-up areas using building footprints as reference data. For instance, Feng et al. (2023) reported F1-scores were between 0.32 and 0.67, Ayala et al. (2021) between 0.23 and 0.82, and Hafner et al. (2022) between 0.57 and 0.69.

As noted during model validation, augmenting local models with AEZ-based African samples introduced additional uncertainty and did not consistently improve classification accuracy. However, we found that this approach was particularly beneficial for outlier AOIs, such as ID014 and ID067, where supplementing local models with AEZ-based training data helped mitigate the lack of local reference data and improved classification performance. This suggests that integrating AEZ-based samples may be a viable strategy in cases where local training data are insufficient or unreliable. Moreover, the use of AEZ-based samples led to slightly better results in models trained with African samples (E6-E9), in scenarios without local reference data.

Table 6 summarizes the mean accuracy values overall and per AEZ using FB as reference. The results reveal substantial difference between recall and precision across all models. Generally, recall values were higher, except in scenario E3, where approximately 52% of built-up areas were omitted (highlighting significant omission errors). The poor performance of E3 can be attributed to the exclusive use of S-1 imagery, which alone is insufficient for reliable built-up detection. Underrepresentation of built-up areas in some SOTA datasets (e.g., GISA, and GISD) may be partially explained by temporal mismatches of up to 2–3 years between the datasets and the reference data. Models and datasets with the highest recall include E2, E5, E7, E9, ESRI and WSF although even these omitted, on average, 25% of the built-up. Overall, there is an unbalanced recall-precision score. Models with high recall often exhibited lower precision (e.g., E6-E9, DW, ESRI), suggesting a tendency to overclassify non-built-up areas as built-up (i.e., high commission errors). Several factors may contribute to these commission errors: First, Sentinel's spatial resolution often results in mixed spectral signatures within a single pixel, particularly in sparsely built-up areas (e.g., ID011). Second, roads are frequently misclassified as built-up due to spectral and spatial similarities. Post-processing (as done in ESA and WSF) could mitigate this issue by removing road segments. Third, isolated trees and shrubs, especially in sparsely vegetated environments (e.g., ID029 and ID072) can mimic the spectral and textural properties of traditional housing architecture, as illustrated in Fig. A6. And fourth, in some AOIs (e.g., ID002, ID007, ID039, ID050), the building footprint reference datasets do not capture all existing structures, often caused by cloud cover, limited coverage in the source imagery, or missed detections of small buildings. In several cases, our model correctly identified built-up areas that were absent from the reference, leading to apparent false positives during validation. This highlights a key challenge: the absence of a true, error-free ground truth dataset. Therefore, validation results must be interpreted in light of these inherent uncertainties.

Despite these limitations, the overall accuracy of our models suggests that the predicted maps in 2022 are reliable representations of built-up areas. Visual evidence of these observations can be explored in the DIY-BU-mapping viewer (https://diybuiltupmapping.users.earthengine.app/view/diy-bu-mapping, accessed on 12/11/2025).

The evaluation also revealed AEZ-specific accuracy patterns (Table 6). Our best-performing model (E1) achieved higher accuracies in arid zones, and lower in humid zones, a trend also observed during model validation. Models trained with African samples (E7-E9), as well as DW and GISA, performed comparatively better in humid and subhumid zones than in arid ones. The lower performance of DW and GISA in arid regions is likely due to the difficulty of distinguishing built-up areas from shrubland, cropland and bare soil (Brown et al., 2022).

WSF, by contrast, performed well across most AEZs, with slightly lower accuracy in semiarid and humid areas, though its static nature limits its utility for temporal high frequency monitoring expansion. For the GHSL, we measured a mean CCC of 0.76, indicating a good agreement at most test sites. The CCC values varied by AEZs. Humid scenarios have a lower agreement with the reference data (0.72), subhumid and semiarid zones have a higher agreement (0.78 and 0.77 respectively), while arid zones have a slightly lower agreement (0.74). This variation underscores the trade-off between spatial accuracy and temporal resolution when using GHSL for built-up area mapping.

Based on the accuracy assessment, we conclude that there is a general tendency, including our best-performing model E1, of higher recall and lower precision, meaning that most built-up areas are successfully identified (low omission). There is also a tendency to classify some non-built-up areas (e.g., bare soil, roads) as built-up (high commission). This results in a slight overestimation of built-up area, particularly in low-density settlements. This may affect fine-scale accuracy; however, it is often preferable in the context of urban expansion monitoring, where missing built-up areas (especially recent development) could be more problematic than slight over-detection.

To qualitatively assess how our model E1 compares with existing SOTA products, we examined results in four representative AOIs across different AEZs. These examples include cities and towns with varying densities, settlement structures, and environmental conditions. Overall, the E1 maps show clear advantages in capturing detailed expansion patterns and small-scale urban structures when compared to global datasets such as DW, ESRI, GISA, and GLC. Notably, differences in performance across datasets often reflect challenges such as dense vegetation, scattered housing, or missing reference data. A detailed visual comparison and discussion of each case is provided in the Supplementary Material (Figs. S1–S4). These figures illustrate how E1 performs relative to other SOTA products in specific contexts, supported by the corresponding evaluation metrics presented in Table S2.

Overall, and as demonstrated in the selected test sites, our results achieved higher accuracy than other global multi-temporal products in the 2022 evaluation (e.g., DW, ESRI, GISA, and GLC; see Fig. A8). While quantitative validation is limited to this year, the multi-temporal outputs were also assessed qualitatively. Visual comparison of urban patterns suggests that our tool provides a more coherent and detailed representation of recent growth across diverse context (Figs. S1–S4). Using a site-specific approach, where training data is drawn from the AOI itself, enables both good accuracy and high temporal resolution when applying the models over time. At the same time, our maps offer the advantage of being easily deployable for any city in Africa using regularly updated satellite data. This is one decisive advantage of the DIY-BU mapping tool, as users can create built-up maps independently, at any time, and under their own conditions, without depending on external product releases. Unlike some alternative global products with higher accuracy (e.g., WSF), our maps provide consistent temporal coverage (Fig. 5). However, the results also reveal significant variations among mapping products, highlighting the complexity and inherent uncertainties in geospatial data. These variations emphasize the need to acknowledge and address uncertainties in any downstream application relying on such data.

4.3. Map improvement through model fine-tuning

While the results of our best-performing model (E1) generally outperform other multi-temporal datasets (cf. Section 4.2), further improvements can be achieved through local fine-tuning. This optional process allows users to adapt the model by incorporating additional training data in specific areas where systematic classification errors are more likely.

To demonstrate this, we applied a fine-tuning procedure to four representative AOIs, as detailed in the Supplementary Material. In each case, we manually added polygons in GEE to delineate areas where the default models underperformed (e.g., roads, mines, quarries, and bare land misclassified as built-up). Within these polygons, new training samples were created, and new models were calculated for each AOI.

Fig. S5 presents a visual comparison between the default scenario (E1) and the fine-tuned result (E1 FT). The fine-tuned results showed marked improvements, including the removal of artefacts such as isolated roads outside urban cores, overclassified quarries, and misclassified bare soils. This demonstrates that adding targeted samples can substantially reduce classification errors and improve output quality. An interactive comparison of these outputs is available in the DIY-BU-mapping viewer: https://diybuiltupmapping.users.earthengine.app/view/diy-bu-mapping (accessed on 12/11/2025).

Although the accuracy metrics for the fine-tuned models are similar to the default E1, the spatial agreement with the reference maps improves significantly (see Table S2). Fine-tuned maps show increases of up to 5% in F1-scores, mainly driven by higher precision. While this improvement may seem modest, it represents a significant reduction in misclassified pixels, particularly given that most errors occurred in small, scattered areas rather than in larger urban conglomerates. Nonetheless, some challenges remain. In AOIs with limited reference data (e.g., ID050), fine-tuning may not substantially increase the sample size, and therefore offers limited benefit. We also observed narrow roads within the urban centers are still often classified as built-up, likely due to the spatial resolution of S-1 and S-2 and the mixed spectral signature, as discussed previously.

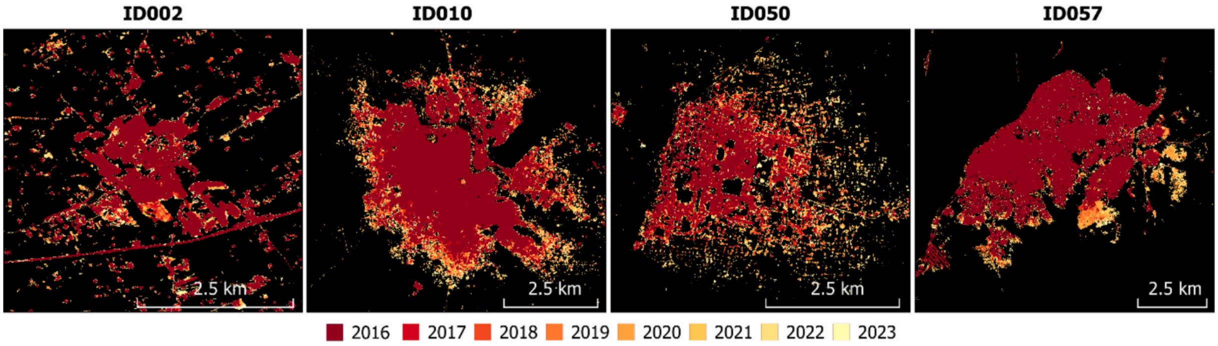


Fig. 5. Examples of built-up area expansion maps generated using scenario E1 for selected AOIs highlighted in Fig. 2. Comparisons with SOTA products are provided in Figs. S1–S4 and Table S2.

4.4. DIY-BU-mapping tool

Based on the results of our experimental comparison and evaluation across 100 test sites, we selected scenario E1 (using local reference data and S-1 and S-2 imagery) as the most suitable configuration for preparing the DIY-BU-mapping tool. This scenario demonstrated consistent performance relative to other tested approaches and products, and has now been implemented in GEE. This free, open, and cloud-based tool leverages multimodal datasets to generate annual, high-resolution (10-m), temporally consistent maps of built-up expansion for African cities using random forest classification and Sentinel imagery. These maps are quantitatively validated against 2022 reference data and extend from 2016 to the present using a monotonic growth trajectory.

The DIY-BU-mapping tool supports the full classification workflow, from data collection and preparation to the generation of final maps, using a sampling-based approach. To improve computational efficiency and avoid memory issues, the tool is divided into two components: (1) design of the sampling dataset for model training and validation, and (2) classification of built-up areas over time.

Access to the tool is available via the GEE repository at https://code.earthengine.google.com/?accept_repo=users/diybuiltupmapping/diy_bu_tool (accessed on 12/11/2025). The two components are organized as follows:

(1) The first component focuses on generating the reference sampling data, following the workflow described in Sect. 3.2. This is implemented in the script "DIY_step1_sampling_data". Users specify an AOI, and the sampling data is generated automatically and stored in the user's GEE account for later use.

(2) The second component is used to train the model and classify built-up areas. This is implemented in the script "DIY_step2_mapping_builtup". It carries out data collection, preparation, and pre-processing (Sect. 3.1), followed by the supervised classification and multi-temporal mapping described in Sects. 3.3.2. and 3.3.3. For the temporal analysis, pixel trajectory rules are applied. These rules currently span the period 2016–2025 and will be updated in GEE as new Sentinel data becomes available.

In terms of computing time, the first component typically completes within 5–20 min. The second component takes approximately 30 min to 3 h, depending on the size of the AOI and the current load on GEE, as it is a shared cloud resource and system performance may vary accordingly.

The tool produces a validation report that includes the overall accuracy, precision, recall, and F1-score, reflecting the model's performance on unseen data. This validation applies only to the 2022 classification, as building footprints for other years are not available.

Several parameters, such as the OB3 confidence threshold, sample size, cloud probability, and rasterization threshold, are user-configurable and not hard-coded. The default values reflect those tested during our experiments, but users may adjust them based on local conditions. We strongly recommend that users validate their results if changes are made. As a primary quality control measure, users modifying parameters should consult the accuracy report generated by the tool's built-in automated validation module to ensure that performance is not degraded. We also recommend that advanced users verify their results by conducting a spatial evaluation against reference data for significant methodological adjustments, similar to the analysis performed in this study. Optional fine-tuning, by adding polygons in underperforming areas, is also supported and explained in the user guide provided in the Supplementary Material.

5. Discussion

In this study, we aimed to identify the most effective model configuration for mapping built-up areas for integration into the DIY-BU-mapping tool. To reflect the diverse data availability scenarios across Africa, where access to data is often limited, we tested multiple combinations of sensor and reference data. This experimental approach allowed us to evaluate the trade-offs of each setup and better understand their limitations. Ultimately, we selected the configuration that combined local reference data with S-1 and S-2 imagery, as this showed the best overall performance. While this setup does not surpass the static WSF in terms of accuracy, it outperforms all current multi-temporal global datasets (e.g., DW, ESRI, GISA, GLC and GISD) when evaluated against 2022 reference data.

A key innovation of the DIY-BU-mapping tool lies in its ability to produce temporally continuous, annual maps of built-up areas from 2016 onwards, something not currently possible with most global datasets. In addition, the tool is designed to be site-specific and adaptable, enabling users to produce tailored, high-resolution outputs for any location in Africa. This flexibility enhances the tool's relevance and usability in diverse contexts and supports a wide range of policy applications, including the monitoring of informal settlements, infrastructure development, disaster risk reduction, and climate adaptation. By contributing to urban resilience planning and SDG 11 (Sustainable Cities and Communities), the tool offers a practical solution to bridge persistent data and capacity gaps in fast-growing urban regions across the continent.

In addition to its strengths, the DIY-BU-mapping tool has certain limitations, particularly its reliance on third-party datasets, such as Google's Open buildings and Microsoft's Building footprints. This dependency constrains our analysis, as it inherits the accuracy and availability limitations of these sources. While misclassifications in the reference data cannot be directly quantified, we mitigate potential bias by evaluating model performance against multiple reference datasets. Still, reference data inherently involve uncertainty, and even expert manual digitization is prone to considerable variability, especially in morphological complex areas (Kraff et al., 2020). The lack of high-quality reference data remains a key limitation across Africa. To address this challenge, future updates of the tool could incorporate a unified reference layer that combines OB3, MBF, and GBA datasets. This integration could enhance both the availability and representativeness of training samples, improving model robustness across a wider range of urban forms and geographic settings.

Temporal misalignments also pose challenges. Although data providers aim to use the most recent imagery, acquisition dates can vary significantly across regions. This was especially evident in humid zones, where persistent cloud cover limits image availability. For this study, we assumed all building footprints correspond to 2022. However, in practice, some newly developed areas from 2020 to 2022 were not captured. In a few AOIs, no buildings were detected in either reference datasets, making local model training impossible. In these cases, we used models trained with Africa-wide samples, which, as shown in our experiments, tend to yield lower accuracies, with F1-score drops averaging 9–11%, and up to 16% in some AEZs. The performance decline was most notable in arid and semi-arid regions (~10%). These spatial and temporal uncertainties affect both training and evaluation, highlighting the need of critically assess reference data.

We also observed that map accuracy varied significantly with AEZ and urban density, which highlights the model's limitations across different geographic contexts. Specifically, the model showed the best overall performance in the arid zones and high urban density areas. In humid and subhumid zones, lower accuracy may be due to persistent vegetation cover, which creates spectral confusion, especially where buildings are made from low-reflectance materials like wood or thatch. Frequent cloud cover further limits the availability of clear Sentinel-2 imagery. In low-density areas, limited built-up coverage reduces training sample availability, making it more difficult for the model to capture representative patterns.

Another set of limitations relates to data availability and platform constraints. The tool is designed for annual updates as new Sentinel imagery becomes available. However, we rely on Level-1C S-2 data, which, unlike Level-2A, is consistently available in Africa since late 2015. Level-2A, which includes atmospheric correction, was only released for northern regions in 2017 and for the rest of the continent in late 2018. To ensure consistent temporal coverage, we opted for Level-1C data, acknowledging that this may reduce temporal transferability. To mitigate this, we applied a pixel trajectory smoothing step, prioritizing the 2022 classification as the based year aligned with available reference data. This limitation could be addressed in future updates as more Level-2A data becomes available for earlier years. Similarly, the reduced availability of Sentinel-1 following the end of the S-1B mission in 2022, has resulted in gaps in radar coverage across certain regions of Africa (e.g., Algeria, Libya, and Angola). For these regions, users can switch to a S-2-only configuration (scenario E2), which is supported in the DIY-BU-mapping tool with minor code modifications. In this case, an average accuracy drop of about 2.6% in the F1-score can be expected. Fortunately, the launch of Sentinel-1C satellite in December 2024 restored the S-1 coverage.

Lastly, the performance of the tool is also shaped by GEE computing limitations. When extracting a large number of features (indices and statistics across years) for tens of thousands of samples, memory and execution time can become a bottleneck. We addressed this by limiting sample size and reducing features using data reduction techniques. However, this can affect classification quality, especially for large or high heterogeneous AOIs. In such cases, we recommend users reduce the AOI size to maintain accuracy.

Despite the limitations outlined above, the DIY-BU-mapping tool presents several compelling strengths for a wide range of users, from researchers and educators to students and technicians with basic remote sensing skills. The tool was developed with a user-centred approach in mind, i.e., providing gridded map outputs, high temporal frequency to capture urban dynamics, and detailed maps at local scale, all within a framework that ensures fair and open access to analysis-ready data, bridging digital divides for decision-makers and stakeholders (Kuffer et al., 2021). On the one hand, users with general geospatial knowledge but limited programming skills can follow the step-by-step tutorial (provided in the supplementary material) to generate built-up maps with minimal manual input. On the other hand, more advanced users can modify the code to adapt it for different applications, use individual components, or enhance the methodology further. This ensures the community benefits not only from the tool and its outputs but also from the overall workflow and methodology. For instance, the sampling strategy can be re-used for other classification tasks, the annual cloud- and shadow-free mosaics can support visualization or additional mapping efforts, and the calculated spectral and textural indices can be used for other applications. The workflow is also designed to be transferable to other target variables aligned with the SDGs, for example, mapping urban green spaces (Tohoun et al., 2023), analysing land use and land cover changes, or identifying economically deprived areas using night time-light imagery (Sapena et al., 2023), assuming sufficient reference data is available.

A further strength is the continuity of the underlying data sources. The Copernicus program ensures long-term data availability through upcoming missions such as Sentinel-2C and Sentinel-2D, along with the planned new-generation Sentinel-2 satellites (ESA, 2023b). This allows the DIY-BU-mapping tool to be updated annually, enabling users to generate updated maps with the latest available data. For long-term analyses, results can be complemented with historical data from WSF Evolution (1985–2015), although differences in spatial resolution must be considered (Sapena et al., 2023). Regarding the tool's implementation, certain concerns apply. While GEE is currently powerful and free for research purposes, it is a proprietary service and its long-term continuity cannot be guaranteed. Changes to the access policy, pricing, or terms of use in the future could affect the sustainability of the tool. This risk is partly mitigated by our transparent approach: all scripts, parameters, and methodological steps are openly available. While the current tool is bound to GEE, this ensures that the mapping framework is reproducible and could be migrated to alternative cloud ecosystems (e.g., Digital Earth Africa, SEPAL) albeit technical adaptation would be necessary.

While the tool was primarily developed for Africa, it can be applied to other regions where suitable reference data is available. Google's open building dataset covers also parts of South Asia, South-East Asia, Latin America and the Caribbean, representing 1.8 billion buildings across 58 million km^2 (Google, 2023). Other countries also provide official or open-access building footprint datasets (Milojevic-Dupont et al., 2023; Uhl et al., 2023), opening opportunities for use in the Global North. Furthermore, recent advances in VHR image segmentation offer promising alternatives in data-scarce context, although these typically represent a single time step and may fall short in highly dynamic urban environments. Recently, Google introduced a new 2.5D building dataset that includes a temporal component, an important step toward multi-temporal urban mapping. However, this dataset is still very recent and has not yet been thoroughly validated or tested for local-scale applications and it remains unclear if and when up-dates will be produced and published. While it holds strong potential, its performance and usability remain to be assessed in diverse urban contexts. As building footprint datasets continue to improve rapidly users increasingly benefit from greater data availability. However, most of these datasets remain static or do not provide access to the underlying algorithms, limiting reproducibility and preventing users from generating tailored outputs for specific contexts. In this light, the DIY-BU-mapping tool remains highly relevant, offering a transparent, customizable workflow that empowers users to produce multi-temporal, analysis-ready built-up maps using open satellite imagery at any time. This flexibility ensures that the tool complements rather than competes with emerging datasets, providing a flexible solution that can be fine-tuned to different user needs and geographic contexts. Its proven usability at the local scale makes it particularly valuable for supporting decision-making and research in underserved and overlooked regions.

In summary, while the proposed approach has its limitations, the

DIY-BU-mapping tool stands out as a robust, transparent, and flexible solution for monitoring built-up area expansion in Africa and beyond. By leveraging third-party data and addressing computational challenges within the GEE environment, the tool provides a valuable resource for a wide range of users. Its adaptability, sustainability, and commitment to open access make it a unique contribution to the mapping and research community, encouraging collaboration and innovation. Looking ahead, continued updates with new satellite data will ensure the tool remains relevant, supporting SDGs monitoring by offering scalable, analysis-ready data and workflows that address themes such as poverty, land use, and sustainable urban development.

6. Conclusions

In this paper, we introduced the DIY-BU-mapping tool, a cloud-based solution for generating analysis-ready data on built-up area expansion in African cities. The tool is designed to meet the growing demand for objective, accurate, frequent, and timely data in rapidly developing urban environments, while also encouraging the widespread use of Sentinel satellite imagery. It is intended for a wide range of users, including researchers, educators, students, and technicians who support decision-makers and urban analysts. By automating complex processing steps, the tool reduces the technical barrier for users with basic geospatial knowledge, while offering experts the flexibility to refine the analysis.

The DIY-BU-mapping tool offers several key advantages, and we hope it becomes a valuable asset in the field of built-up area mapping: (1) it generates site-specific maps trained with samples from an AOI; (2) it produces up-to-date maps that users can refresh as soon as new data are available; (3) its outputs yielded higher F1-scores than those of multi-temporal global SOTA products, such as Dynamic World and ESRI land cover, in the 2022 comparison for the 100 test sites in Africa; (4) the maps can be further improved through local fine-tuning; (5) the methodology is open-source and currently implemented via GEE, offering flexibility for methodological adaptation; (6) it integrates S-1 and S-2 imagery, leveraging the complementary strengths of radar and optical data; and (7) although originally designed for Africa, the tool is transferable to other continents and urban contexts provided that the necessary reference data is available.

While the tool has several strengths, it also has limitations. Its performance depends on the availability and quality of reference building footprint data, which can vary in space and time. Incomplete or outdated footprints may reduce the accuracy of classifications or the quality of training. Additionally, the tool relies on uncorrected Level-1C imagery and is susceptible to cloud and shadow contamination in optical data, particularly in humid regions. The spatial resolution of Sentinel imagery also limits the detection of small or sparsely distributed buildings. Finally, the multi-temporal map is derived using a monotonic trajectory based on the 2022 validation. While this approach ensures temporal consistency, it restricts the tool's capacity to identify de-urbanisation or demolition events. Despite these challenges and the reliance on a third-party platform, the tool offers a transparent, scalable and adaptable approach to generating timely built-up area maps, thereby contributing to the bridging of key data gaps in support of SDG 11 and broader urban resilience initiatives.

CRediT authorship contribution statement

Marta Sapena: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Johannes Mast:** Writing – review & editing. **Elisabeth Schoepfer:** Writing – review & editing, Supervision, Funding acquisition. **Hannes Taubenböck:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Funding

This study has been conducted as part of the European Union's Caroline Herschel Framework Partnership Agreement on Copernicus User Uptake (Grant No. FPA 275/G/GRO/COPE/17/10042), project FPCUP, Action 2019–2-11, and the project MIGRAWARE (Grant No. 01LG2082C), funded by the German Federal Ministry of Education and Research (BMBF) as part of the programme WASCAL WRAP 2.0.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We want to thank all participants from various African institutions for taking part in our organized workshops and for their users' recommendations and feedback on the tool.

Appendix

Table A1
Summary table showing the total number of test sites with the latest available VHR imagery in Google Earth Pro, per month and per year, for the Open Building dataset version 3 (v3).

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Sum
Year	**2018**							1						1
	2019						1		1			1	1	4
	2020			1								1	2	4
	2021	2	1			2	3	1		2	4	1	2	18
	2022	2	3	2	1	5	3	3	1	2	5	5	5	37
	2023	3	1	4	16	12								36
	Sum	7	5	7	17	19	7	5	2	4	9	8	10	100

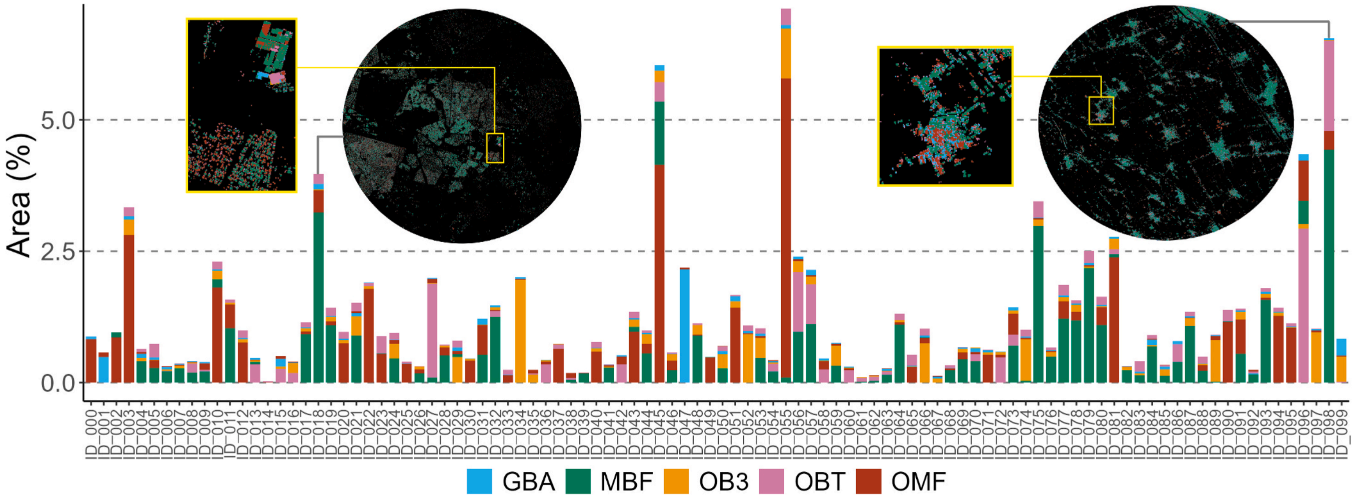


Fig. A1. Percentage of the areas of interest covered by built-up areas based on five reference building footprint datasets for each of the 100 test sites presented in Sect. 3.3. Where GlobalBuildingAtlas (GBA), Microsoft building footprints (MBF), Open buildings v3 (OB3), Open buildings 2.5D temporal (OBT), and Overture maps foundation (OMF). An example is shown for sites ID018 and ID098.

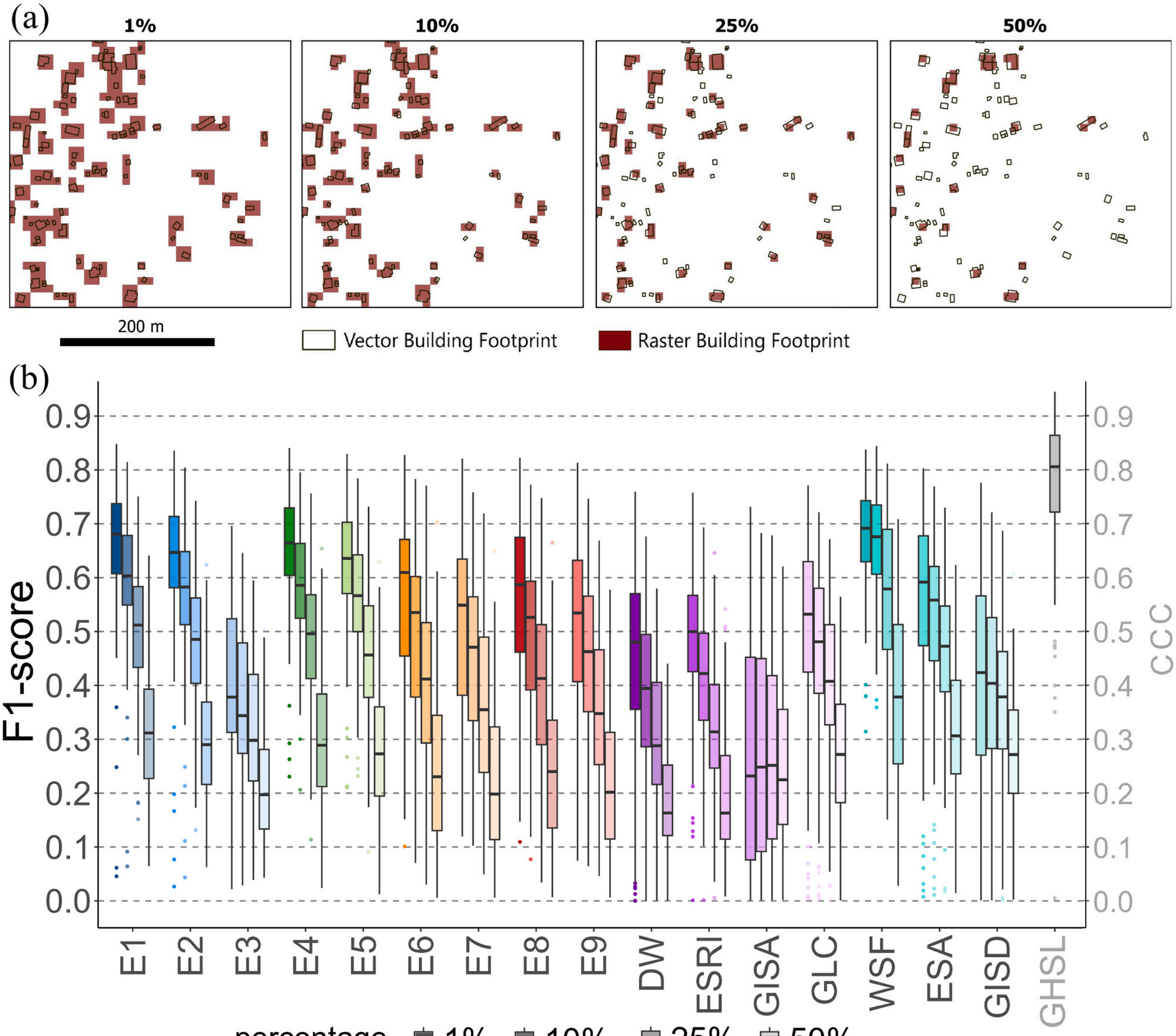


Fig. A2. (a) Illustration of the effect of different rasterization thresholds (1%, 10%, 25%, and 50%) used to convert vector building footprints into 10-m raster masks. Lower thresholds include more pixels per building footprint, potentially over-representing marginal overlaps (e.g., 1% or 10%). At higher thresholds (e.g., 25% or 50%), smaller buildings are increasingly excluded from the mask. (b) Sensitivity of the evaluation process to different rasterization thresholds, shown through the F1-score distribution across our experimental maps and SOTA products. This analysis demonstrates how the choice of threshold used to generate the reference built-up mask influences the measured accuracy. Reference data is derived from the fusion of multiple building footprint datasets (FB).

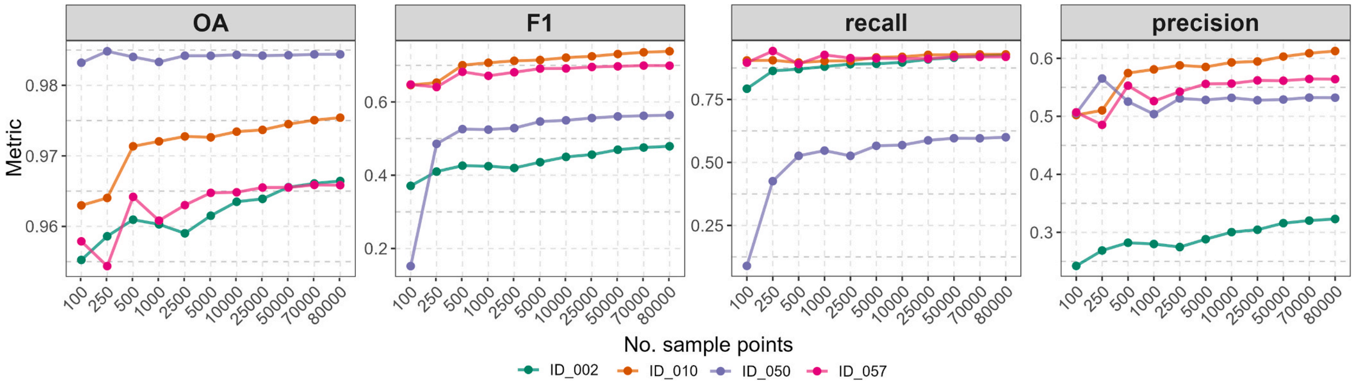


Fig. A3. Effect of sample size on the evaluation metrics of the built-up maps for four selected AOIs using experiment E1 and the fused building footprints as reference data. Each panel shows how the accuracy metrics vary as the number of evaluation sample points increases (from 100 to 80,000). While larger sample sizes lead to moderate accuracy improvements, these become less substantial beyond 50,000 points.

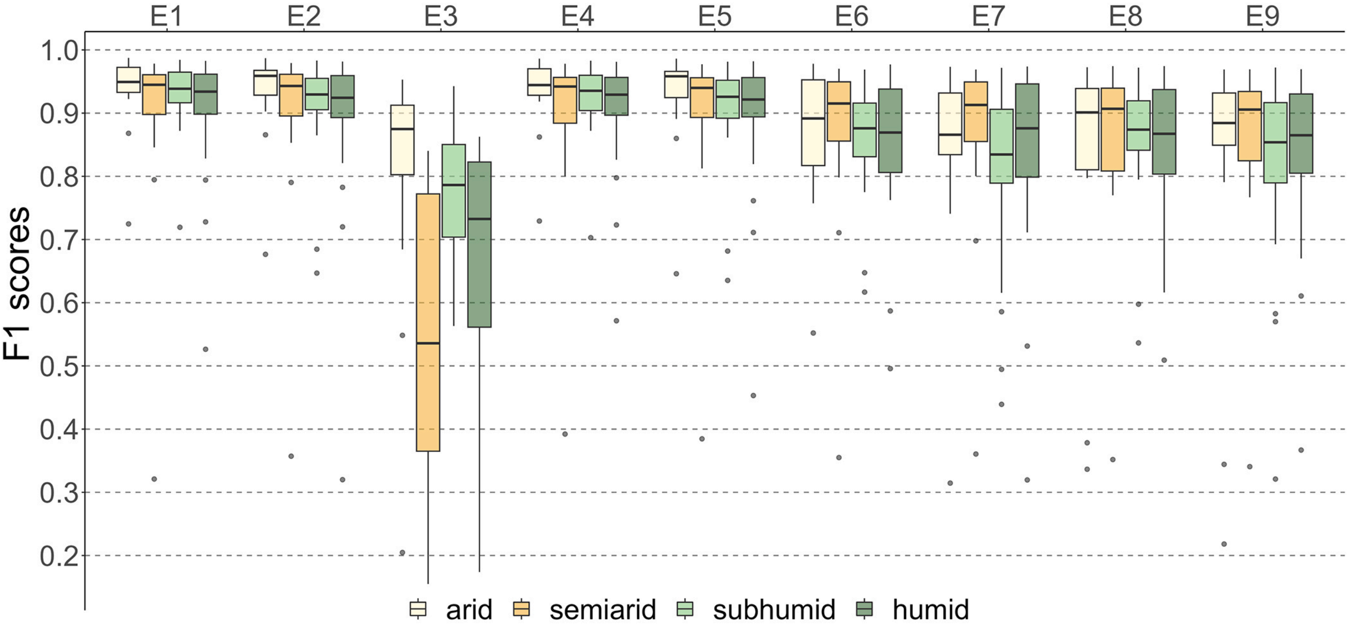


Fig. A4. Model F1-scores per agroecological zone for each scenario.

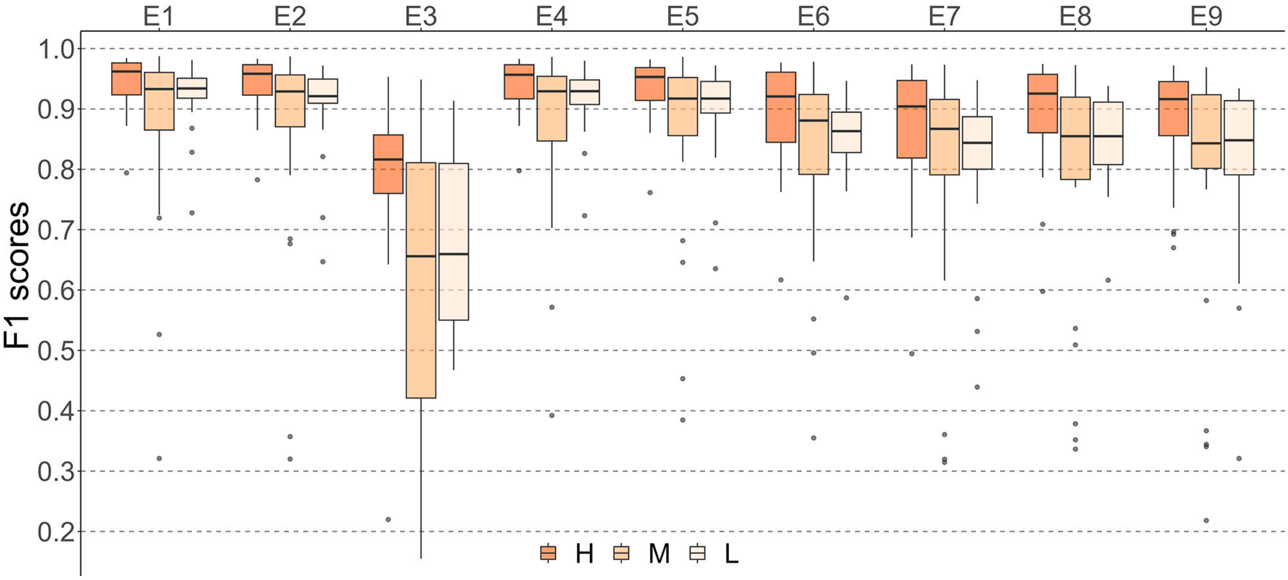


Fig. A5. Model F1-scores per urban density level (high, medium and low) for each scenario.

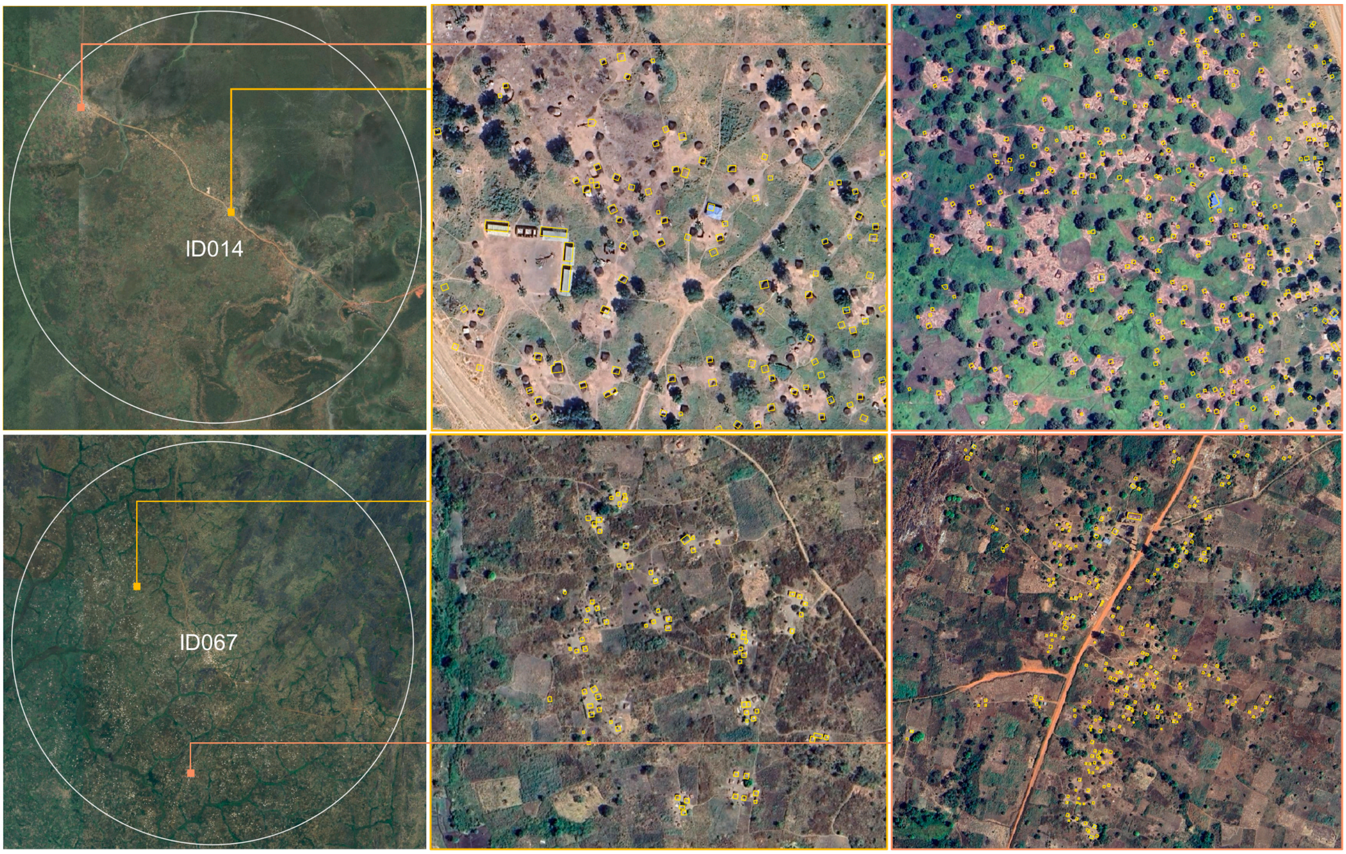


Fig. A6. Visual representation of sparse and very low urban density test sites: (1) Gemmeiza, South Sudan (ID014) and (2) the area between Ambo, Democratic Republic of the Congo, and Girili, South Sudan (ID067). Two representative areas with buildings are highlighted, with building footprints overlaid in yellow. **Source:** Map data ©2015 Google; Open Buildings V3.

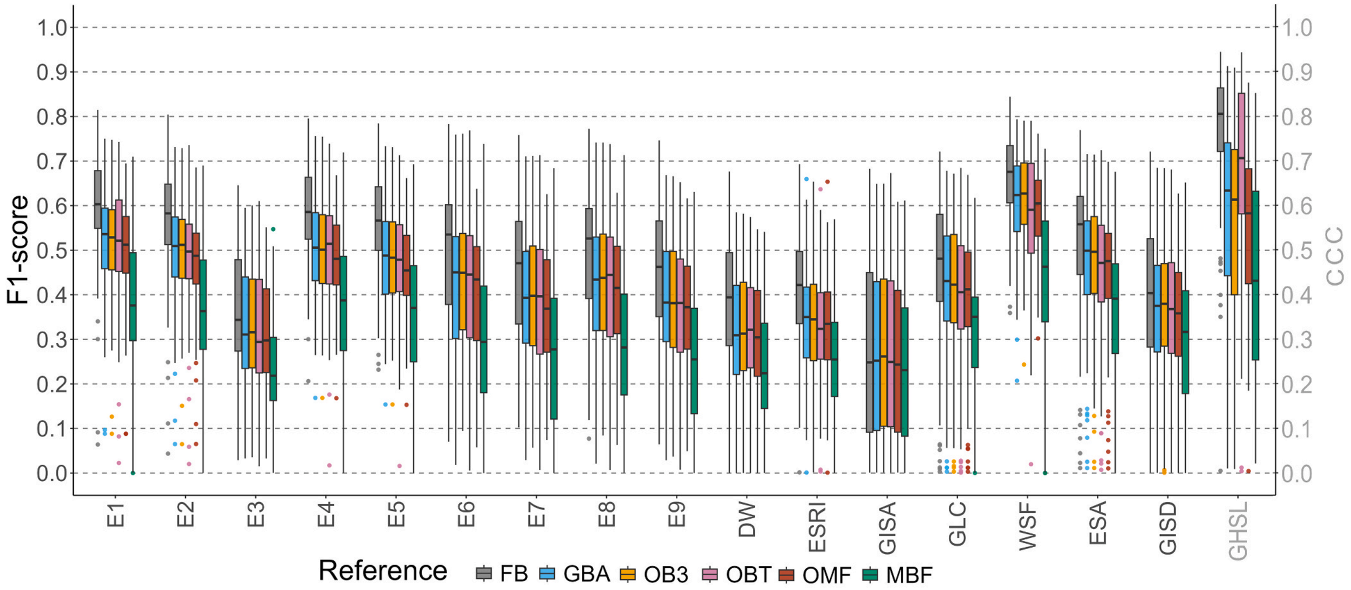


Fig. A7. Boxplot with the F1-scores for the map evaluation of each scenario (E1-E9) and state-of-the-art global maps, Dynamic World (DW), ESRI land cover (ESRI), GISA, GLC_FCS30D (GLC), WSF2019 (WSF), WorldCover v200 (ESA), GISD30 (GISD) and GHSL2023 (GHSL), using six reference building footprints: Open buildings v3 (OB3), Open buildings 2.5D temporal (OBT), Overture maps foundation (OMF), Microsoft building footprints (MBF), GlobalBuildingAtlas (GBA), and Fusion Buildings (FB).

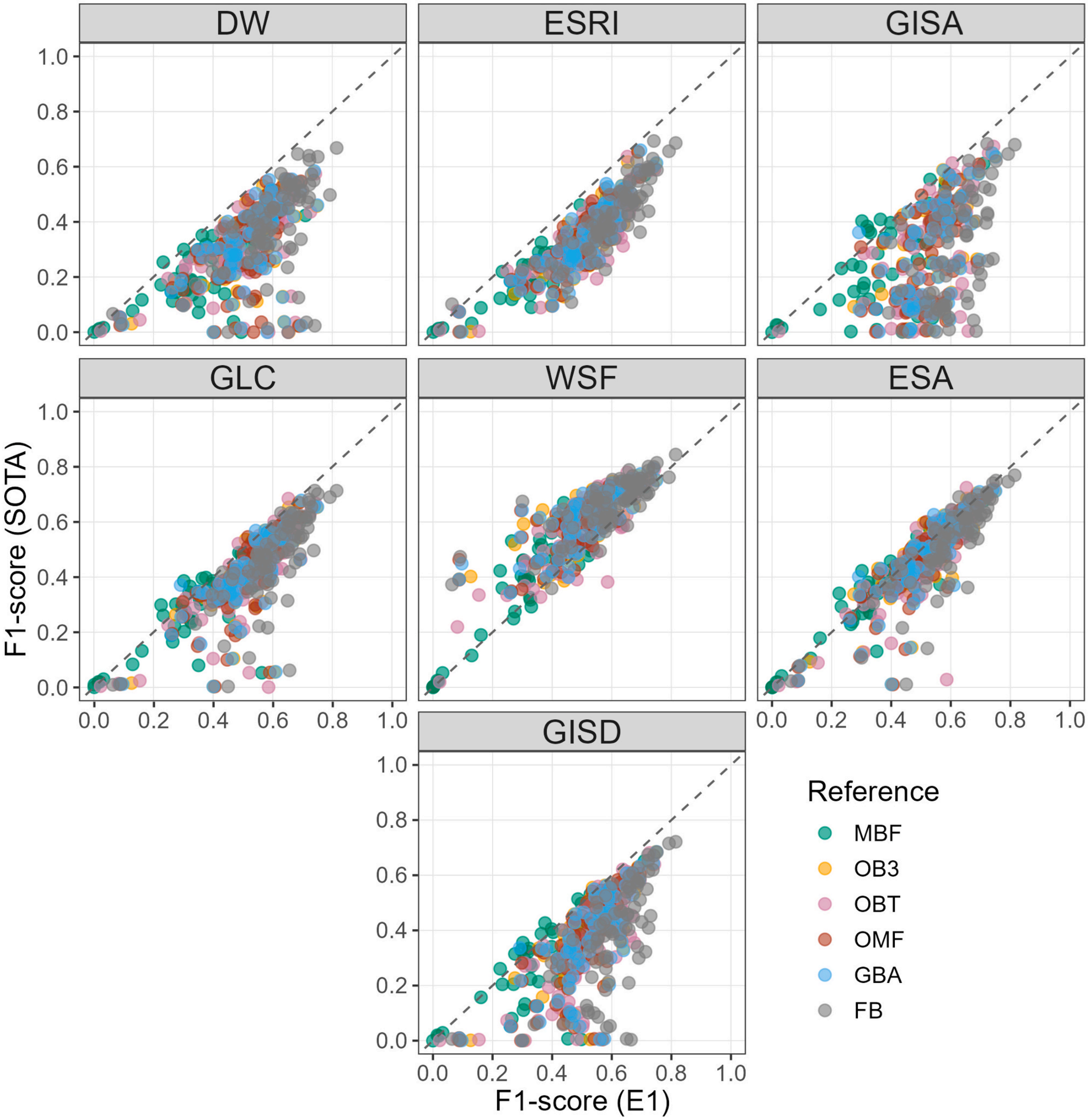


Fig. A8. Comparison of F1-scores for the 100 test sites for each combination of our scenario E1 and the state-of-the-art global maps (SOTA): Dynamic World (DW), ESRI land cover (ESRI), GISA, GLC_FCS30D (GLC), WSF2019 (WSF), WorldCover v200 (ESA), GISD30 (GISD). The x-axis corresponds to the F1-score of our model E1, the y-axis to the F1-score of each global map, the colour indicates the reference building footprint dataset used to calculate the F1-scores: Open buildings v32 (OB3), Open buildings 2.5D temporal (OBT), Overture maps foundation (OMF), Microsoft building footprints (MBF), GlobalBuildingAtlas (GBA), and Fusion Buildings (FB). Points above the line of equality indicate higher accuracy in the global maps, while points below the line indicate higher accuracy in our maps.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2026.105153.

Data availability

The research outputs generated during this study can be visualised using the DIY-BU-mapping viewer, which is available at https://diybuiltupmapping.users.earthengine.app/view/diy-bu-mapping. The DIY-BU-mapping tool and underlying code can be accessed via the Google Earth Engine repository: https://code.earthengine.google.com/?accept_repo=users/diybuiltupmapping/diy_bu_tool.

References

Acuto, M., Seto, K., Parnell, S., Contestabile, M., Allen, A., Attia, S., Xuemei Bai, Batty, M., Bettencourt, L., Birch, E.L., Bulkeley, H., Maruxa Cardama, Ebikeme, C., Elmqvist, T., Elsheshtawy, Y., Kickbusch, I., Shuaib Lwasa, McCann, J., McCarney, P., McPhearson, T., Patel, S., Pelling, M., Pieterse, E., Ratti, C., Aromar Revi, Sampson, R., Satterthwaite, D., Sennett, R., Tyler, N., Yongguan Zhu, 2018. Science and the Future of Cities. DOI: 10.13140/RG.2.2.27706.64969.

Amarsaikhan, E., Enkhjargal, D., Jargaldalai, E., Amarsaikhan, D., 2024. Comparison of machine learning and parametric methods for the discrimination of urban land cover types. Geocarto Int. 39, 2380372. https://doi.org/10.1080/10106049.2024.2380372.

Ayala, C., Aranda, C., Galar, M., 2021. Multi-class strategies for joint building footprint and road detection in remote sensing. Appl. Sci. 11, 8340. https://doi.org/10.3390/app11188340.

Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., Espinoza, W., Prudencio, F., Llactayo, V., Montero, D., Sudmanns, M., Tiede, D., Mateo-García, G., Gómez-Chova, L., 2022. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. Sci. Data 9, 782. https://doi.org/10.1038/s41597-022-01878-2.

Braaten, J., 2023. Sentinel-2 Cloud Masking with s2cloudless. Url: https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless (accessed 05.12.2023).

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4.

Chakraborty, T., Venter, Z.S., Demuzere, M., Zhan, W., Gao, J., Zhao, L., Qian, Y., 2024. Large disagreements in estimates of urban land across scales and their implications. Nat. Commun. 15, 9165. https://doi.org/10.1038/s41467-024-52241-5.

Chamberlain, H.R., Darin, E., Adewole, W.A., Jochem, W.C., Lazar, A.N., Tatem, A.J., 2024. Building footprint data for countries in Africa: to what extent are existing data products comparable? Comput. Environ. Urban Syst. 110, 102104. https://doi.org/10.1016/j.compenvurbsys.2024.102104.

Chen, T.-H.-K., Pandey, B., Seto, K.C., 2023. Detecting subpixel human settlements in mountains using deep learning: a case of the Hindu Kush Himalaya 1990–2020. Remote Sens. Environ. 294, 113625. https://doi.org/10.1016/j.rse.2023.113625.

Copernicus, 2024. Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © AirbusDefence and Space GmbH 2014-2018 provided under COPERNICUS by the European Union and ESA; all rights reserved. doi:10.5270/ESA-c5d3d65.

Denis, E., Moriconi-Ebrard, F., Harre-Roger, D., Thiam, O., 2008. Africapolis. Url: https://hal.archives-ouvertes.fr/hal-00357271 (accessed 20/11/2025).

Digital Earth Africa, 2024. Data and Services Roadmap 2024. url: https://digitalearthafrica.org/wp-content/uploads/DE-Africa-Data-and-Services-Roadmap-version-3.0-Final.pdf (accessed 20/11/2025).

Digital Earth Africa, 2022. Analysis ready data: A smart way to use Earth observation for Africa's rising nations. Insight report.

Dodman, D.B., Hayward, M., Pelling, V., Castan Broto, W., et al., 2022. Cities, Settlements and Key Infrastructure., in: Clim. Change 2022 – Impacts Adapt. Vulnerability Work. Group II Contrib. Sixth Assess. Rep. Intergov. Panel Clim. Change H-O Pörtner DC Roberts M Tignor ES Poloczanska K Mintenbeck Alegría M Craig Langsdorf Löschke V Möller Okem B Rama Eds. Cambridge University Press, pp. 907–1040.

Ecopia, 2023. Ecopia, global feature extraction. Url: https://www.ecopiatech.com/global-feature-extraction.

ESA, 2023a. ESA Copernicus program. User Guides. URL https://sentinels.copernicus.eu/web/sentinel/user-guides (accessed 10/07/2023).

ESA, 2023b. Sentinel-2 mission guide. Url: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2 (accessed 23/11/2023).

ESA, 2020. WorldCover Product User Manual. Url: https://esa-worldcover.org/en/data-access (accessed 05/10/2023).

Ettehadi Osgouei, P., Kaya, S., Sertel, E., Alganci, U., 2019. Separating built-up areas from bare land in mediterranean cities using Sentinel-2A imagery. Remote Sens. 11, 345. https://doi.org/10.3390/rs11030345.

Feng, L., Xu, P., Tang, H., Liu, Z., Hou, P., 2023. National-scale mapping of building footprints using feature super-resolution semantic segmentation of Sentinel-2 images. Giscience Remote Sens. 60, 2196154. https://doi.org/10.1080/15481603.2023.2196154.

Gevaert, C.M., Buunk, T., van den Homberg, M.J.C., 2024. Auditing geospatial datasets for biases: using global building datasets for disaster risk management. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 12579–12590. https://doi.org/10.1109/JSTARS.2024.3422503.

Google, 2023. Open Buildings: A dataset of building footprints to support social good applications. Url: ttps://sites.research.google/open-buildings/#dataformat (accessed: 23/11/2023).

Google for developers, 2023. Machine Learning in Earth Engine. Url: https://developers.google.com/earth-engine/guides/machine-learning (accessed 21/08/2023). URL https://developers.google.com/earth-engine/guides/machine-learning.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Hafner, S., Ban, Y., Nascetti, A., 2022. Unsupervised domain adaptation for global urban extraction using Sentinel-1 SAR and Sentinel-2 MSI data. Remote Sens. Environ. 280, 113192. https://doi.org/10.1016/j.rse.2022.113192.

He, C., Shi, P., Xie, D., Zhao, Y., 2010. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. Remote Sens. Lett. 1, 213–221. https://doi.org/10.1080/01431161.2010.481681.

Huang, X., Song, Y., Yang, J., Wang, W., Ren, H., Dong, M., Feng, Y., Yin, H., Li, J., 2022. Toward accurate mapping of 30-m time-series global impervious surface area (GISA). Int. J. Appl. Earth Obs. Geoinformation 109, 102787. https://doi.org/10.1016/j.jag.2022.102787.

IPCC, 2023. Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st ed. Cambridge University Press. DOI: 10.1017/97810093.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use / land cover with Sentinel 2 and deep learning, in: 2021 IEEE Int. Geosci. Remote Sens. Symp. IGARSS. Presented at the IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Brussels, Belgium, pp. 4704–4707. DOI: 10.1109/IGARSS47720.2021.9553499.

Kraff, N.J., Wurm, M., Taubenbock, H., 2020. Uncertainties of human perception in visual image interpretation in complex urban environments. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4229–4241. https://doi.org/10.1109/JSTARS.2020.3011543.

Kuffer, M., Wang, J., Thomson, D.R., Georganos, S., Abascal, A., Owusu, M., Vanhuysse, S., 2021. Spatial Information gaps on deprived urban areas (Slums) in low-and-middle-income-countries: a user-centered approach. Urban Sci. 5, 72. https://doi.org/10.3390/urbansci5040072.

Kühnl, M., Sapena, M., Wurm, M., Geiß, C., Taubenböck, H., 2023. Multitemporal landslide exposure and vulnerability assessment in Medellín. Colombia. Nat. Hazards 119, 883–906. https://doi.org/10.1007/s11069-022-05679-z.

Lin, L.I.K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255. https://doi.org/10.2307/2532051.

Marconcini, M., Metz- Marconcini, A., Esch, T., Gorelick, N., 2021. Understanding Current Trends in Global Urbanisation - The World Settlement Footprint Suite. GI_Forum 9, 33–38. DOI: 10.1553/giscience2021_01_s33.

Microsoft, 2023. Worldwide building footprints derived from satellite imagery. Url: https://github.com/microsoft/GlobalMLBuildingFootprints.

Milojevic-Dupont, N., Wagner, F., Nachtigall, F., Hu, J., Brüser, G.B., Zumwald, M., Biljecki, F., Heeren, N., Kaack, L.H., Pichler, P.-P., Creutzig, F., 2023. EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings. Sci. Data 10, 147. https://doi.org/10.1038/s41597-023-02040-2.

Mu, H., Li, X., Zhou, Y., Gong, P., Huang, J., Du, X., Guo, J., Cao, W., Sun, Z., Xu, C., Liu, D., 2022. Identifying discrepant regions in urban mapping from historical and projected global urban extents. Earth 34, 167–178. https://doi.org/10.1080/27669645.2022.2104990.

Mück, M., Klotz, M., Taubenbock, H., 2017. Validation of the DLR Global Urban Footprint in rural areas: A case study for Burkina Faso, in: 2017 Jt. Urban Remote Sens. Event JURSE. Presented at the 2017 Joint Urban Remote Sensing Event (JURSE), IEEE, Dubai, United Arab Emirates, pp. 1–4. DOI: 10.1109/JURSE.2017.7924618.

OECD, 2024. Built-up area (indicator), DOI: 10.1787/7c06b772-en (accessed: 24/04/2024).

OECD/SWAC, 2020. Africa's Urbanisation Dynamics 2020: Africapolis, Mapping a New Urban Geography. OECD Publishing, Paris. DOI: 10.1787/b6bccb81-en.

Ouma, Y.O., Keitsile, A., Nkwae, B., Odirile, P., Moalafhi, D., Qi, J., 2023. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. Eur. J. Remote Sens. 56, 2173659. https://doi.org/10.1080/22797254.2023.2173659.

Overture Maps Foundation, 2024. Overture Maps Foundation. Url: https://overturemaps.org/ (accessed: 14/11/2024).

Patino, J.E., Duque, J.C., 2013. A review of regional science applications of satellite remote sensing in urban settings. Comput. Environ. Urban Syst. 37, 1–17. https://doi.org/10.1016/j.compenvurbsys.2012.06.003.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. https://doi.org/10.1038/nature20584.

Pesaresi, M., Politis, P., 2023. GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC).

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J.H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., Friedrich, H.K., Gao, J., Leyk, S., Lu, L., Maffenini, L., Mari-Rivero, I., Melchiorri, M., Syrris, V., Van Den Hoek, J., Kemper, T., 2024. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. Int. J. Digit. Earth 17, 2390454. https://doi.org/10.1080/17538947.2024.2390454.

Rasul, A., Balzter, H., Ibrahim, G., Hameed, H., Wheeler, J., Adamu, B., Ibrahim, S., Najmaddin, P., 2018. Applying built-up and bare-soil indices from landsat 8 to cities in dry climates. Land 7, 81. https://doi.org/10.3390/land7030081.

RCMRD, 2015. Agroecological Zones for Africa. Url: http://geoportal.rcmrd.org/layers/servir:africa_agroecological_zoning (accessed 11/01/2024).

Reba, M., Seto, K.C., 2020. A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change. Remote Sens. Environ. 242, 111739. https://doi.org/10.1016/j.rse.2020.111739.

Roy, S., Swetnam, T., Saah, A., 2024. samapriya/awesome-gee-community-datasets. Community Catalog. https://doi.org/10.5281/ZENODO.14042069.

Sapena, M., 2020. Development and analysis of land-use/land-cover spatio-temporal metrics in urban environments: Exploring urban growth patterns and linkages to socio-economic factors. Universitat Politècnica de València, Valencia (Spain). DOI: 10.4995/Thesis/10251/158626.

Sapena, M., Inkoom, J.N., Umaru, E.T., Mast, J., Okhimamhe, A.A., Taubenböck, H., 2023. Integrating multi-source remote sensing data for monitoring urban development in West Africa, in: 2023 Jt. Urban Remote Sens. Event JURSE. Presented at the 2023 Joint Urban Remote Sensing Event (JURSE), IEEE, Heraklion, Greece, pp. 1–4. DOI: 10.1109/JURSE57346.2023.10144165.

SEPAL, 2025. System for Earth Observation Data Access, Processing and Monitoring Analysis for Land Monitoring (SEPAL). Url: https://sepal.io/ (accessed 15/11/2025).

Sirko, W., Brempong, E.A., Marcos, J.T.C., Annkah, A., Korme, A., Hassen, M.A., Sapkota, K., Shekel, T., Diack, A., Nevo, S., Hickey, J., Quinn, J., 2023. High-Resolution Building and Road Detection from Sentinel-2. DOI: 10.48550/ARXIV.2310.11622.

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., Quinn, J., 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud mask intercomparison eXercise (CMIX): an evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens. Environ. 274, 112990. https://doi.org/10.1016/j.rse.2022.112990.

Taubenböck, H., 2019. Remote Sensing for the Analysis of Global Urbanization. DLR-Research Report 2019. Habilitation Thesis, University Würzburg.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. Remote Sens. Environ. 117, 162–176. https://doi.org/10.1016/j.rse.2011.09.015.

Taubenböck, H., Mast, J., Geiß, C., Wurm, M., Esch, T., Seto, K.C., 2024. Global differences in urbanization dynamics from 1985 to 2015 and outlook considering IPCC climate scenarios. Cities 151, 105117. https://doi.org/10.1016/j.cities.2024.105117.

Tohoun, B.A., Sapena, M., Mast, J., Taubenböck, H., Haruna, I., Orekan, V., Okhimamhe, A.A., 2023. Are citizens' perceptions on urban green spaces influenced by their immediate environment? The case of Grand Nokoue, Benin Republic, in: 2023 Jt. Urban Remote Sens. Event JURSE. Presented at the 2023 Joint Urban Remote Sensing Event (JURSE), IEEE, Heraklion, Greece, pp. 1–4. DOI: 10.1109/JURSE57346.2023.10144198.

Uhl, J.H., Royé, D., Burghardt, K., Aldrey Vázquez, J.A., Borobio Sanchiz, M., Leyk, S., 2023. HISDAC-ES: historical settlement data compilation for Spain (1900–2020). Earth Syst. Sci. Data 15, 4713–4747. https://doi.org/10.5194/essd-15-4713-2023.

UN, 2023. The Sustainable Development Goals Report 2023: special edition. United Nations, S.l.

Van de Voorde, T., van der Kwast, J., Poelmans, L., Canters, F., Binard, M., Cornet, Y., Engelen, G., Uljee, I., Shahumyan, H., Williams, B., Convery, S., Lavalle, C., 2016. Projecting alternative urban growth patterns: the development and application of a remote sensing assisted calibration framework for the Greater Dublin Area. Ecol. Ind. 60, 1056–1069. https://doi.org/10.1016/j.ecolind.2015.08.035.

Van Den Hoek, J., Friedrich, H.K., 2021. Satellite-based human settlement datasets inadequately detect refugee settlements: a critical assessment at thirty refugee settlements in Uganda. Remote Sens. 13, 3574. https://doi.org/10.3390/rs13183574.

Waqar, M.M., Mirza, J.F., Mumtaz, R., Hussain, E., 2012. Development of new indices for extraction of built-up area & bare soil from landsat data. Sci. Rep.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., 2022. ESA WorldCover 10 m 2021 v200. Url: https://zenodo.org/record/7254221.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. Earth Syst. Sci. Data 13, 2753–2776. https://doi.org/10.5194/essd-13-2753-2021.

Zhang, X., Liu, L., Zhao, T., Gao, Y., Chen, X., Mi, J., 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 Syst. Sci. Data 14, 1831–1856. https://doi.org/10.5194/essd-14-1831-2022.

Zhu, X.X., Chen, S., Zhang, F., Shi, Y., Wang, Y., 2025. GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models. DOI: 10.5194/essd-2025-327.

Zhu, X.X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., Taubenböck, H., 2022. The urban morphology on our planet – Global perspectives from space. Remote Sens. Environ. 269, 112794. https://doi.org/10.1016/j.rse.2021.112794.