Remote Sensing of Environment 311 (2024) 114316

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Comparative validation of recent 10 m-resolution global land cover maps

Panpan Xu [a], Nandin-Erdene Tsendbazar [a,*], Martin Herold [a,b], Sytze de Bruin [a], Myke Koopmans [a], Tanya Birch [c], Sarah Carter [d], Steffen Fritz [e], Myroslava Lesiv [e], Elise Mazur [d], Amy Pickens [f], Peter Potapov [f], Fred Stolle [d], Alexandra Tyukavina [f], Ruben Van De Kerchove [g], Daniele Zanaga [g]

[a] Laboratory of Geo-information Science and Remote Sensing, Wageningen University & Research, 6708 PB Wageningen, the Netherlands
[b] Section 1.4 Remote Sensing and Geoinformatics, Helmholtz GFZ German Research Centre of Geosciences, Telegrafenberg, Potsdam, Germany
[c] Google, LLC, 1600 Amphitheatre Pkwy., Mountain View, CA 94043, USA
[d] World Resources Institute, 10G St NE #800, Washington, DC 20002, USA
[e] International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria
[f] University of Maryland, College Park, United States of America
[g] Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium

ARTICLE INFO

Editor: Dr. Marie Weiss

Keywords:
Global land cover
10 m-resolution
Independent validation
Reference data uncertainty
Spatial detail

ABSTRACT

Accurate and high-resolution land cover (LC) information is vital for addressing contemporary environmental challenges. With the advancement of satellite data acquisition, cloud-based processing, and deep learning technology, high-resolution Global Land Cover (GLC) map production has become increasingly feasible. With a growing number of available GLC maps, a comprehensive evaluation and comparison is necessary to assess their accuracy and suitability for diverse uses. This particularly applies to maps lacking statistically robust accuracy assessment or sufficient reported detail on the validation procedures. This study conducts a comparative independent validation of recent 10 m GLC maps, namely ESRI Land Use/Land Cover (LULC), ESA WorldCover, and Google and World Resources Institute (WRI)'s Dynamic World, examining their spatial detail representation and thematic accuracy at global, continental, and national (for 47 larger countries) levels. Since high-resolution map validation is impacted by reference data uncertainty owing to geolocation and labelling errors, five validation approaches dealing with reference data uncertainty were evaluated. Of the considered approaches, validation using the sample label supplemented by majority label within the neighborhood is found to produce more reasonable accuracy estimates compared to the overly optimistic approach of using any label within the neighborhood and the overly pessimistic approach of direct comparison between the map and reference labels. Overall global accuracies of the maps range between 73.4% ± 0.7% (95% confidence interval) to 83.8% ± 0.4% with WorldCover having the highest accuracy followed by Dynamic World and ESRI LULC. The quality of the maps varies across different LC classes, continents, and countries. The maps' spatial detail representation was assessed at various homogeneity levels within a 3 × 3 kernel. Although considered as high-resolution maps, this study reveals that ESRI LULC and Dynamic World have less spatial detail than WorldCover. All maps have lower accuracies in heterogenous landscapes and in some countries such as Mozambique, Tanzania, Nigeria, and Spain. To select the most suitable product, users should consider both the map's accuracy over the area of interest and the spatial detail appropriate for their application. For future high-resolution GLC mapping, producers are encouraged to adopt standardized LC class definitions to ensure comparability across maps. Additionally, the spatial detail and accuracy of GLC maps in heterogeneous landscapes and over some countries are the key features that should be improved in future versions of the maps. Independent validation efforts at regional and national levels, as well as for LC changes, should be strengthened to enhance the utility of GLC maps at these scales and for long-term monitoring.

* Corresponding author at: Wageningen University & Research, P.O. box 47, 6700 AA Wageningen, The Netherlands.
E-mail address: nandin.tsendbazar@wur.nl (N.-E. Tsendbazar).

https://doi.org/10.1016/j.rse.2024.114316
Received 18 December 2023; Received in revised form 25 June 2024; Accepted 9 July 2024
Available online 18 July 2024

1. Introduction

Land cover (LC) maps play a vital role in helping to understand human-induced as well as natural processes in the land system. Their information is invaluable for various applications, such as agriculture, land use planning, nature conservation, climate modeling, and water management, enabling policy- and decision-makers to make informed decisions (Ban et al., 2015). Given the escalating impacts of global climate change and increasing human activity on ecological systems, there is a growing demand for high-resolution, accurate, and up-to-date LC maps to comprehend and effectively address environmental land changes (Szantoi et al., 2020).

Recent advances in satellite data acquisition, cloud-based processing, and the increased use of deep learning approaches in remote sensing have ushered in a new era of Global Land Cover (GLC) mapping, characterized by faster production and higher spatial resolutions. Although the 30 m-resolution GLC maps derived from the extensive Landsat archive continue to serve a substantial user base (Potapov et al., 2022; Yu et al., 2022; Zhang et al., 2021), there is a proliferation of 10 m-resolution GLC maps, leveraging Copernicus Sentinel-1 and 2 data, which includes the FROM-GLC10 map for 2017 by Tsinghua University in China (Gong et al., 2019), the ESRI Land Use/Land Cover maps (ESRI LULC, 2017–2022) (Karra et al., 2021), the European Space Agency (ESA) WorldCover 2020/2021 maps (Zanaga et al., 2022), and the Dynamic World (2015–2023) released by Google and the World Resources Institute (WRI) (Brown et al., 2022). These maps, while sharing the same high spatial resolution, often employ different classification models and methodologies, leading to potentially divergent classification results, which underscores the need for comprehensive evaluation and comparison to ensure their reliability for diverse applications.

With more GLC maps becoming available, users have the possibility to decide which product to use. However, choosing the most suitable product for a given application is not always straightforward. Users may base their selection on reported accuracy from the map producer or on a subjective visual assessment, but this can introduce uncertainties or biases in their further analysis (Kinnebrew et al., 2022; Tsendbazar et al., 2016). Thus, an independent and statistically rigorous comparison of these products against the same validation dataset is crucial to aid users in selecting the most appropriate map for their specific needs. Some regional assessments have been conducted, such as Kang et al.'s (2022) comparison of FROM-GLC10, WorldCover, and ESRI LULC for northwestern China, Chaaban et al.'s (2022) evaluation of WorldCover and ESRI LULC for Syria, and Wang and Mountrakis's (2023) evaluation of eleven 10–30 m global and regional LULC mapping products over the conterminous U.S. However, only limited global level comparison of the high-resolution GLC products is available. Wang et al. (2023b) conducted a comprehensive review of 107 LULC products which included the most recent 10 m products, while an independent validation was not provided by this study. A global validation was done by Venter et al. (2022) by comparing Dynamic World, World Cover, and ESRI LULC at a global level using the Dynamic World validation dataset and within the European Union using the LUCAS database (D'Andrimont et al., 2020). However, the global validation for ESRI LULC was not completely independent, as the authors stated, because the Dynamic World validation dataset was used by ESRI for both training and validation, with unknown distribution of training tiles.

Validating high-resolution LC products is challenging due to potential inherent uncertainties in reference data. Typically, validation datasets are created through visual interpretation of very high-resolution (VHR) imagery, which is susceptible to various sources of error (Pontius, 2000; Tarko et al., 2021). For instance, uncertainty in reference label interpretation can make it difficult to clearly assign a single LC label to a pixel due to multiple classes being present and limited availability of VHR data (Tarko et al., 2021; Tsendbazar et al., 2015). Additionally, geolocation mismatches between map products and validation datasets (Aguilar et al., 2017; Potere, 2008) can significantly affect the estimated map accuracy, particularly at high resolution (Gu and Congalton, 2021; Olofsson et al., 2014). However, most high-resolution GLC maps did not consider reference data uncertainties during validation. For example, Dynamic World was validated through a direct comparison of the map label with expert/non-expert annotations. ESRI LULC did not provided many details of its validation process, leaving users unsure about whether the accuracy assessment of the map is statistically robust or not. Yet, disregarding reference data uncertainties may lead to an unfair comparison of maps (Stehman and Foody, 2019), thus, it is strongly recommended to consider reference data uncertainty when validating the high-resolution GLC maps.

Various approaches have been proposed to address reference data uncertainty during validation. For example, to assess the impact of geolocation errors on map accuracy estimation, Gu and Congalton (2021) discarded sample units in heterogeneous areas, considering that geolocation error has little impact in homogeneous areas. Another approach involves using alternative or secondary labels from the surrounding area of the sample pixel to reduce the effect of reference data uncertainty (Olofsson et al., 2014; Wickham et al., 2021). Understanding the impact of different validation approaches is crucial to achieving consensus on methodologies that improve the comparability of validation estimates for high-resolution LC maps.

With the use of high spatial resolution Sentinel-1 and 2 data at 10 m pixel size, LC is expected to be discerned in greater spatial detail, particularly in heterogeneous landscapes to capture small-scale features and variations within a landscape (Drusch et al., 2012; Torres et al., 2012). However, the added benefit of spatially detailed LC characterization with the high-resolution GLC maps has not been assessed despite the presumed expectation of using high-resolution satellite products as inputs.

The objective of this study is to assess the strengths and weaknesses of recent 10 m-resolution GLC products in terms of accuracy and representation of spatial detail by considering several validation approaches that address reference data uncertainty. To achieve this, we utilized the validation dataset produced by the Copernicus Global Land Service - Land Cover (CGLS-LC) project (Tsendbazar et al., 2021) to compare and assess three recent and openly accessible 10 m GLC products: WorldCover, ESRI LULC, and Dynamic World. We compared five validation methods that account for reference data uncertainty, highlighting their advantages and disadvantages. Additionally, we evaluated the accuracy of the GLC maps across different scales and their capability to represent spatial detail.

2. Methods

2.1. Validation dataset

The multi-purpose Global Land Cover Validation dataset (Tsendbazar et al., 2021) was used for validating the GLC maps. This dataset employs the Sentinel-2 Universal Transverse Mercator (UTM) grid as the geographic base. The dataset is based on a global stratification (combination of Köppen biome and population density), making the stratification independent of LC maps. The dataset has >21,000 primary sampling units (PSUs) at a 100 m-resolution globally with a minimum of 3000 PSUs distributed per continent (Fig. 1).

The CGLS-LC validation dataset offers a robust framework for assessing large-scale LC maps. Each PSU within the validation dataset contains one hundred 10 × 10 m reference pixels (SSUs; secondary sampling units), enabling the assessment of LC maps with resolutions ranging from 10 m to 100 m. The LC information for these locations has been updated annually for the period of 2015–2021, focusing on areas that underwent changes since 2015. For this study, we utilized the data from 2021. Fig. 2 provides an example of the LC labelling in a sample site. The dataset comprises LC validation sample units at a 10 m-resolution, which was contributed by >30 regional experts from around the world and involved extensive review and feedback by international

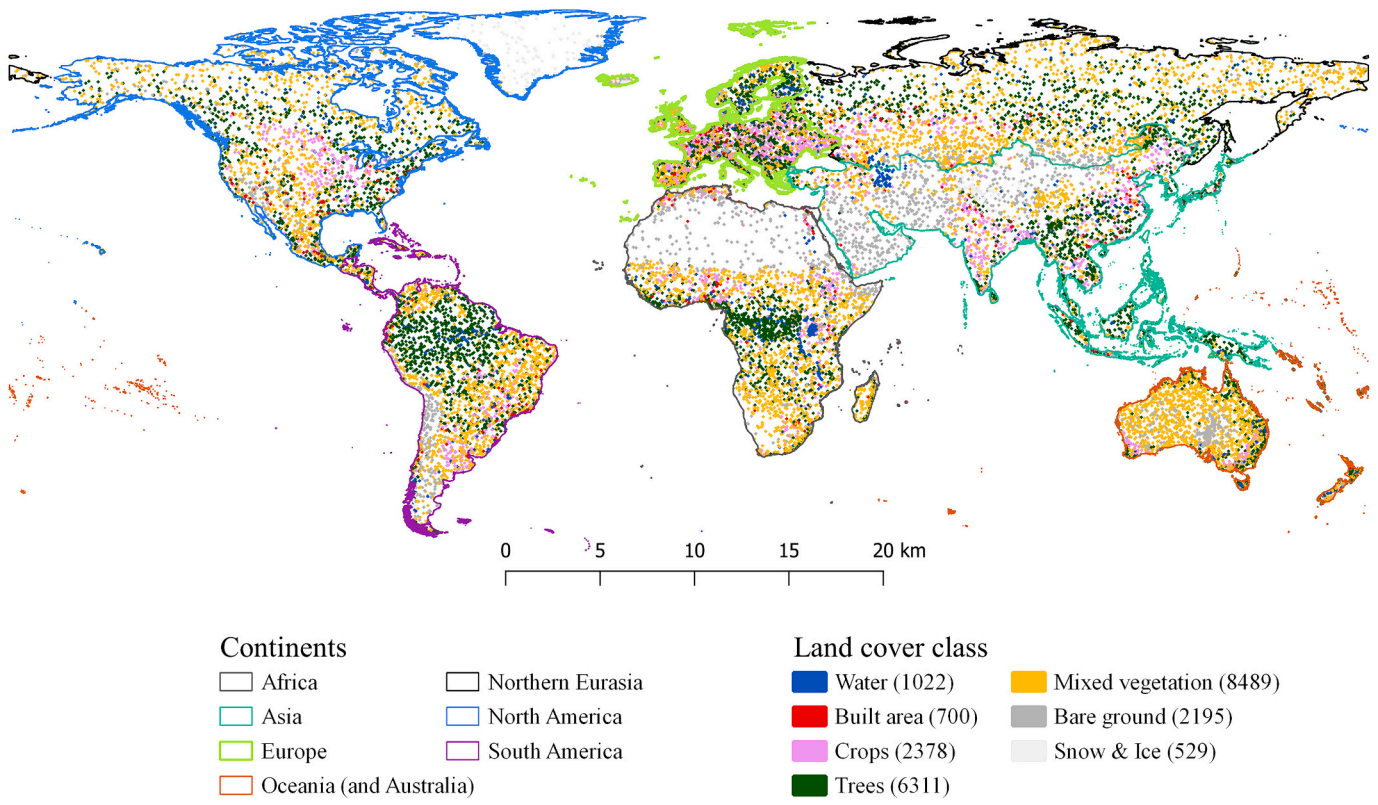

Fig. 1. Distribution of the validation data used for the global and continental accuracy estimation. Classes shown in this figure represent the dominant type in each primary sampling unit (PSU) at 100 m-resolution in 2021. The numbers in the brackets indicate the number of PSUs for the dominant types.

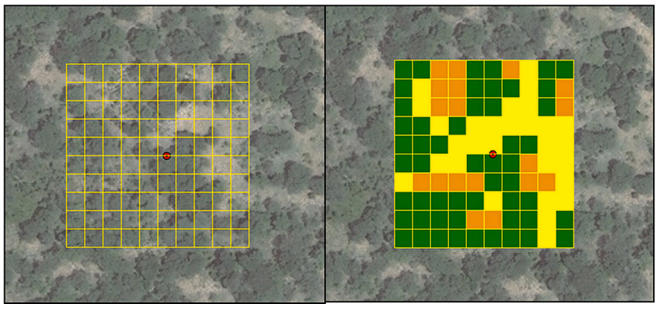

Fig. 2. Screenshot of an example interpretation for a PSU composed of 100 secondary sampling units (SSU). Sub-squares represent SSUs and the colors green, orange and yellow represent trees, shrubs and grassland respectively. Source: Tsendbazar et al., 2018. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

experts (Tarko et al., 2021; Tsendbazar et al., 2021). The LC validation data consist of 10 classes (shown in Table 2), as well as the flooding condition (Tsendbazar et al., 2018). For a detailed description of this validation dataset, such as the sampling design and response design, readers can refer to Tsendbazar et al. (2021).

2.2. Global land cover maps, legend harmonization, and data processing

In this study, we evaluated and compared three GLC maps, namely WorldCover (Zanaga et al., 2022), ESRI LULC (Karra et al., 2021), and Dynamic World (Brown et al., 2022). These maps were selected as they represent the most recent developments among 10 m GLC maps and for their open access to users. The year 2021 was chosen because the WorldCover V200 map was released only for the year 2021, with an improved algorithm and accuracy in comparison to the previous 2020 version. In this study, a one-year composite of Dynamic World was evaluated. More detailed information on the maps is presented in Table 1.

To ensure that the maps are thematically comparable, we reclassified their LC classes into seven types. Table 2 provides an overview of both the harmonized and original classes. The original class definition of each product is shown in Table S1 of the supplementary material. We created a mixed vegetation class as an umbrella category encompassing shrubland, herbaceous vegetation, and wetland herbaceous vegetation from the reference dataset. This reconciliation of classes was undertaken in order to compare the products as the ESRI LULC does not distinguish between grassland and shrublands. Additionally, Dynamic World and ESRI maps include seasonally flooded areas that are a mix of grass/shrub/trees/bare ground as flooded vegetation (Brown et al., 2022; Karra et al., 2021), hence this class of the two maps as well as the mangrove class and herbaceous wetland of WorldCover were attributed to the "mixed vegetation" class (Table 2).

To obtain the map data at the reference locations, the Google Earth Engine (GEE, Gorelick et al., 2017) platform was used. The values of each map at the reference sample sites were extracted using the `reduceRegions()` function in GEE, sampled at the map's original resolution. For ESRI LULC, we used the most recently updated version (July 2023). This data is not available on GEE but can be accessed through Amazon Web Services (AWS), Esri Living Atlas, and Microsoft Azure (see links in Table 1).

To determine a pixel that confidently belongs to a Dynamic World class over time, it is recommended to use the probability bands (Google Earth Engine, 2022), which indicate the estimated likelihood of the original 9 LC classes. We firstly composited the annual probability bands using the median value for each of the LC class, and then the class with the highest probability was chosen as the LC label for that pixel. Nevertheless, the derived annual LC label has some tendency to overestimate snow & ice, crops, and bare ground. Thus, the annual labels

Table 1
Overview of GLC products used in this study.

Dataset name	Data source	Classification model	Period of data	Number of classes	Spatial resolution (m)	Temporal frequency	Reported overall accuracy	Reference	GEE asset ID
WorldCover	Sentinel-1, Sentinel-2	Gradient boosting decision tree (CatBoost)	2020, 2021	11	10	Yearly	76.7%	Zanaga et al., 2022	"ESA/WorldCover/v200"
Dynamic World	Sentinel-2	Fully Convolutional Neural Network (FCNN)	2015–2023	9	10	2–5 days	73.8%	Brown et al., 2022	"GOOGLE/DYNAMICWORLD/V1"
ESRI LULC	Sentinel-2	Convolutional Neural Network - UNet	2017–2022	9	10	Yearly	85.0%	Karra et al., 2021	The most recent version (July 2023) is not available on GEE but can be accessed through AWS[1], Esri Living Atlas[2], and Microsoft Azure[3].

[1] https://registry.opendata.aws/io-lulc/
[2] https://www.arcgis.com/home/item.html?id=cfcb7609de5f478eb7666240902d4d3d
[3] https://planetarycomputer.microsoft.com/dataset/io-lulc-annual-v02

Table 2
Overview of harmonized and original land cover classes of the GLC products. Numbers in the brackets indicate class code. The original class definitions can be found in Table S1 of the supplementary material.

New label	Reference data	WorldCover	Dynamic World	ESRI LULC
Water (0)	Open water (80)	Permanent water bodies (80)	Water (0)	Water (1)
Trees (1)	Closed forest (11); Open forest (12)	Tree cover (10)	Trees (1)	Trees (2)
Mixed vegetation (2)	Shrubs (20); Herbaceous vegetation (30); Wetland herbaceous vegetation (90)	Shrubland (20); Grassland (30); Moss and lichen (100); Herbaceous wetland (90); Mangroves (95)	Grass (2); Shrub & Scrub (5); Flooded vegetation (3)	Rangeland (11); Flooded vegetation (4)
Crops (4)	Cropland (40)	Cropland (40)	Crops (4)	Crops (5)
Built area (6)	Urban/built up (50)	Built-up (50)	Built area (6)	Built Area (7)
Bare ground (7)	Bare/sparse vegetation (60)	Bare / sparse vegetation (60)	Bare ground (7)	Bare ground (8)
Snow & Ice (8)	Snow and ice (70)	Snow and ice (70)	Snow & Ice (8)	Snow/ice (9)

were further adapted using data from the growing season with the following rules:

- The "cropland" in the annual label was changed to a corresponding class if it is "grass", "trees", or "shrubs" during the growing season.
- The "snow & ice" in the annual label was changed to a corresponding class if it is "water", "trees", "grass", "flooded vegetation", "shrubs", "crops", or "bare ground" during the growing season.
- The "bare ground" in the annual label was changed to a corresponding class if it is "trees", "grass", "shrubs", "flooded vegetation", or "crops" during the growing season.

Specifically, the growing season was defined as June–August for the Northern Hemisphere and December–February for the Southern Hemisphere. Here we simplified the definition of growing season to guarantee easier implementation for large-scale map users and to avoid creating harsh/artificial boundaries in the LC representation, while users focused on a smaller scale are recommended to define growing season based on the phenological cycle and climatic conditions of their interested area. The LC type for the growing season was obtained using the same approach applied to obtaining the annual label from probability bands. Through these adaptations, we ensured that the derived LC labels accurately reflected LC on an annual level in the Dynamic World product. Afterwards, the classes were harmonized according to Table 2.

In the definition of built-up areas, the Dynamic World and ESRI LULC products consider urban green (such as lawns, trees, or buildings surrounded by vegetative land covers) as urban, while our reference data do not categorize urban green as built-up. To accommodate our reference data to the two maps with a different definition of built-up areas, we included extra steps in the validation process (see section 2.4).

2.3. Assessing the spatial detail of GLC maps

At the 10 m level, the GLC maps are expected to have greater spatial detail in characterizing various LC types in heterogeneous landscapes compared to maps based on coarser spatial resolution. To assess if the landscape spatial details are adequately reflected in these 10 m maps, we introduced the concept of "homogeneity level".

Based on a moving 3 × 3 kernel, the homogeneity level is defined as the number of pixels in the 3 × 3 kernel that match the LC label of the centre SSU. If none of the surrounding SSU labels matches that of the centre, the homogeneity level is set to "1" and if all labels of the surrounding SSUs match the LC at the centre SSU, the homogeneity level is set to "9" (Fig. 3). The kernel size 3 × 3 was determined in accordance with previous studies (Stehman et al., 2003; Wickham et al., 2021) and accommodates a positional shift of one pixel in the reference data.

The GLC maps were assessed at different homogeneity levels. Firstly, we assessed to what extent the GLC maps and the validation data agree at different homogeneity levels. To do so, we calculated the proportion of sample locations (i.e., SSU) that agree in terms of mapped and reference LC. More specifically, the agreement between the mapped and reference labels was calculated at homogeneity level 1–9, respectively, given both the map and validation data. Secondly, the percentage of sample SSUs in each homogeneity level was calculated for the map and the validation data to identify the proportion of sample units in each homogeneity level.

In this assessment, only complete 3 × 3 kernel grids (nine pixels) were included, excluding the edge SSUs in the 10 × 10 SSUs of each PSU. For all the three maps, the same procedure was applied. Finally, to assess the homogeneity at the PSU-level, we compared the number of LC types present within a 100 × 100 m PSU given both the map and validation data.

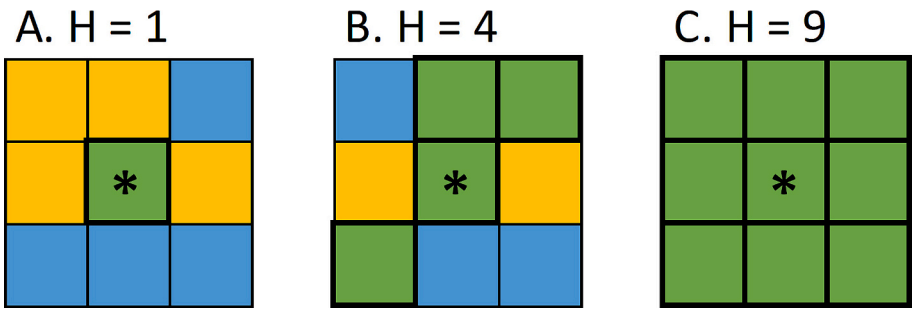

Fig. 3. Example of the homogeneity levels (H) based on the LC labels in a 3 × 3 kernel. A. H = 1: An isolated centre SSU (*). B. H = 4: Three neighboring SSUs match the centre SSU (*) in terms of land cover. C. H = 9: Complete homogeneous for the centre SSU (*).

2.4. Approaches to deal with reference data uncertainty

In this study, five approaches dealing with reference data uncertainty were compared to assess the effect of validation methods on the map accuracy estimation. These approaches are visually depicted in Fig. 4.

The "Direct" or "Primary" approach (Approach 1, Fig. 4) involved a straightforward pixel-to-pixel comparison between the mapped and reference LC types. In this method, all discrepancies between the reference and map labels were considered map errors. This approach considers the reference class as the absolute truth and does not account for geolocation and reference class ambiguity (Stehman and Foody, 2009).

Next, we explored the incorporation of alternative or secondary labels from the surrounding area (3 × 3 kernel) of each SSU. This approach concerns the step of defining the agreement between the map and the reference label. Here, the use of alternative labels from the surrounding

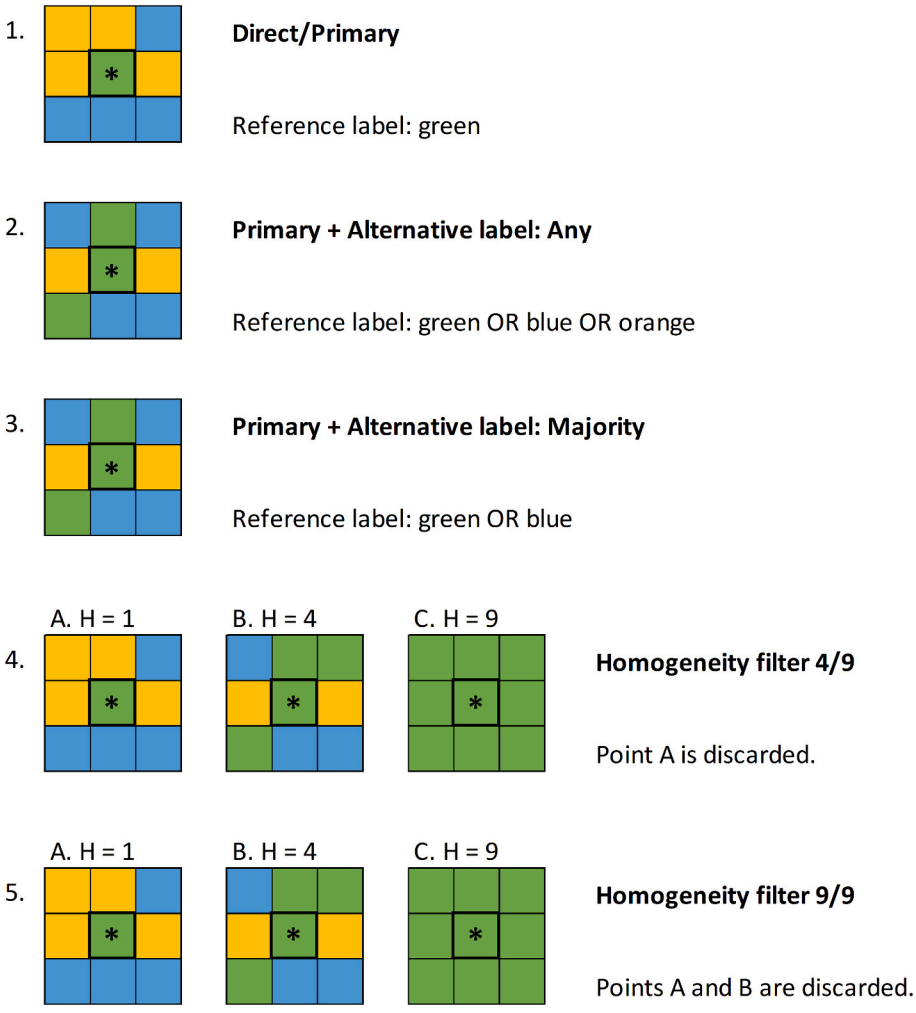

Fig. 4. Five approaches to deal with reference data uncertainties when validating 10 m-resolution GLC maps. Approach 1: "Direct/Primary" uses only the centre SSU (*) as reference label. Approach 2: "Primary + Alternative label: Any" uses any LC class in the 3 × 3 kernel as reference label in addition to the centre pixel label. Approach 3: "Primary + Alternative label: Majority" uses the centre SSU (*) as well as the majority LC class of the 3 × 3 kernel as reference label. Approach 4: "Homogeneity filter 4/9" discards SSU whose homogeneity level (H) < 4 and uses the centre SSU (*) as reference label. Approach 5: "Homogeneity filter 9/9" discards SSU whose homogeneity level (H) < 9.

area aims to reduce the effect of geolocation mismatch or labelling uncertainties such as uncertainties in assigning a dominant label within a small SSU (e.g., 10 × 10 m). Two options, "Primary + Alternative label: Any" and "Primary + Alternative label: Majority" (Approaches 2 and 3 from Fig.4, respectively), were considered. The selection of the kernel size was inspired by previous work such as Wickham et al. (2021), who employed a 3 × 3 kernel of a centre pixel to interpret an alternate label when collecting validation data to assess the National Land Cover Dataset (NLCD) of the USA for 2016. This approach enabled using landscape context to account for the labelling ambiguity and geolocation error, while on the other hand, the use of alternative labels may blur the definition of agreement between the map and reference labels (Stehman et al., 2003).

The "Primary + Alternative label: Any" approach (Approach 2, Fig. 4) takes the LC class of the SSU as the primary label and any of the LC classes in the 3 × 3 kernel grid of the SSU as alternative labels. All primary and alternative labels are then matched with the mapped label at the SSU location (Wickham et al., 2021). In the "Primary + Alternative label: Majority" approach (Approach 3, Fig.4), we considered the majority or modal LC class within the 3 × 3 kernel grid in addition to the reference LC class of each SSU. Stehman et al. (2003) employed a similar concept by utilizing the majority label within a 3 × 3 kernel as alternative label to validate the NLCD products. A mapped label was deemed correct if it matches either the primary LC class of the SSU or the alternative label representing the majority LC class in the kernel (Stehman et al., 2003). Only the clear majority class was used, in case of two or more modal classes, no alternative label was used.

Lastly, we excluded sample units in heterogeneous kernels, assuming higher reference data uncertainty in these areas following Gu and Congalton (2021) who have adapted this approach to assess the effect of reference data uncertainty. We used two homogeneity thresholds. First, the "Homogeneity filter 4/9" retained SSUs with H ≥ 4. Second, the "Homogeneity filter 9/9" retained only SSUs with H = 9. The filtering was done for both the mapped and reference LC labels. Since heterogeneous SSUs were excluded, the resulting map accuracy estimates represent accuracy only for the more homogeneous areas and do not represent the accuracy of the entire mapped area. The numbers of SSUs and PSUs used for each validation method as well as each GLC map are listed in Table S2 of the supplementary material.

To accommodate our reference data for validating built areas of Dynamic World and ESRI LULC, which depict urban green areas as urban (Table S1 of the supplementary material), we included an urban refinement process, which checks the number of built-up pixels in the 3 × 3 kernel. If at least one built-up pixel exists in the 3 × 3 kernel of the reference data, the map label is considered correct.

2.5. Accuracy estimation

Our validation dataset conformed to a stratified, one-stage cluster design (Pengra et al., 2015). Sample inclusion probability, which is the likelihood of a given sample unit being included in the sample, was calculated for the primary sampling units (PSUs) and secondary sampling units (SSUs) following the methods described in Pengra et al. (2015) and Tsendbazar et al. (2018). The PSU is a cluster, and each cluster is assigned to a stratum h. As the selection of PSUs in the validation dataset is based on stratified sampling (see Section 2.1), the inclusion probability (π) per stratum (h) was calculated as $\pi_h = k_h/K_h$, where k_h is the number of PSU sampled in stratum h and K_h is the population size (total possible number of PSUs) for stratum h (Tsendbazar et al., 2021). The inclusion probability for each sampled PSU was available in the validation dataset (Tsendbazar et al., 2021).

Because map accuracy can be expressed as a ratio, the ratio estimator provides a general approach for accuracy estimation for clustered sampling while accounting unequal inclusion probabilities (Pengra et al., 2015). Based on the SSUs within PSUs, following Pengra et al. (2015), the ratio estimator can be calculated as:

$$\widehat{R} = \frac{\sum_{h=1}^{H}\sum_{i=1}^{k_h}\sum_{j=1}^{N_{hi}} \omega_{hij} y_{hij}}{\sum_{h=1}^{H}\sum_{i=1}^{k_h}\sum_{j=1}^{N_{hi}} \omega_{hij} x_{hij}} \quad (1)$$

where j is the index of the SSU ($j = 1,\ldots,N_{hi}$), N_{hi} is the number of SSUs in cluster i (PSU) of stratum h, i is the cluster index in stratum h ($i = 1,2,\ldots,k_h$), h is the stratum index ($h = 1,2,\ldots,H$), and ω_{hij} is the estimation weight (i.e., inverse of the inclusion probability) for SSU j in cluster i (PSU) of stratum h. x_{hij} and y_{hij} are defined to yield the parameter of interest, such as the overall accuracy meaning the total percentage of sample units with correct classification divided by the total number of possible units in the region. The ratio estimates $\widehat{R}$ for the overall accuracy (%) is calculated with y_{hij} defined as 1 if map and reference labels agree and 0 otherwise, and x_{hij} defined as 1, to derive the ratio of correct classification based on all possible sample units.

The variance estimator for $\widehat{R}$ is based on a Taylor series approximation (Pengra et al., 2015)

$$\widehat{V}(\widehat{R}) = \sum_{h=1}^{H} \widehat{V}_h(\widehat{R}) = \sum_{h=1}^{H} \frac{k_h\left(1-\frac{k_h}{K_h}\right)}{k_h - 1} \sum_{i=1}^{k_h} (g_{hi\cdot} - \bar{g}_{h\cdot\cdot})^2 \quad (2)$$

where

$$g_{hi\cdot} = \frac{\sum_{j=1}^{N_{hi}} \omega_{hij}\left(y_{hij} - x_{hij}\widehat{R}\right)}{\sum_{h=1}^{H}\sum_{i=1}^{k_h}\sum_{j=1}^{N_{hi}} \omega_{hij} x_{hij}} \quad (3)$$

and

$$\bar{g}_{h\cdot\cdot} = \frac{\sum_{i=1}^{k_h} g_{hi\cdot}}{k_h} \quad (4)$$

If a stratum has $k_h = 1$, the contribution of that stratum to the estimated variance is 0.

Based on the estimation weight per PSU, each SSU was assigned 1/100th of the weight of the PSU it belongs to, as there are 100 SSUs in each PSU (Fig. 2). Next, using the mapped and reference LC types at each SSU, a confusion matrix was constructed accounting for unequal sample inclusion probabilities (Stehman et al., 2003; Wickham et al., 2021). In the approaches using homogeneity filtering, only the SSUs meeting the filtering thresholds were considered together with their estimation weights. The estimation weights of excluded SSUs were not accounted for in accuracy estimation.

Overall accuracies and their confidence intervals (at a 95% confidence level) were calculated for each of the validation approaches following the eqs. (1–4). Following a similar concept, class specific accuracies were calculated as detailed in Pengra et al. (2015). The accuracy estimation was done globally, per continent (seven sub-continents), and for 47 countries having >100 PSUs based on the initial design of the validation dataset (Tsendbazar et al., 2021).

3. Results

3.1. Approaches dealing with reference data uncertainty

Fig. 5 shows the overall global accuracies for the three GLC maps when different validation approaches are applied. The global overall accuracy varied between 82.8 and 91.4% for WorldCover, 70.8–82.1% for Dynamic World, and 70.3–82.2% for ESRI LULC based on the five validation approaches.

For all the three GLC maps, the "Direct" or "Primary" approach yielded the lowest accuracy estimates; this is expected as this approach considers all differences between reference and map labels as map errors, regardless of possible reference data errors. Conversely, the

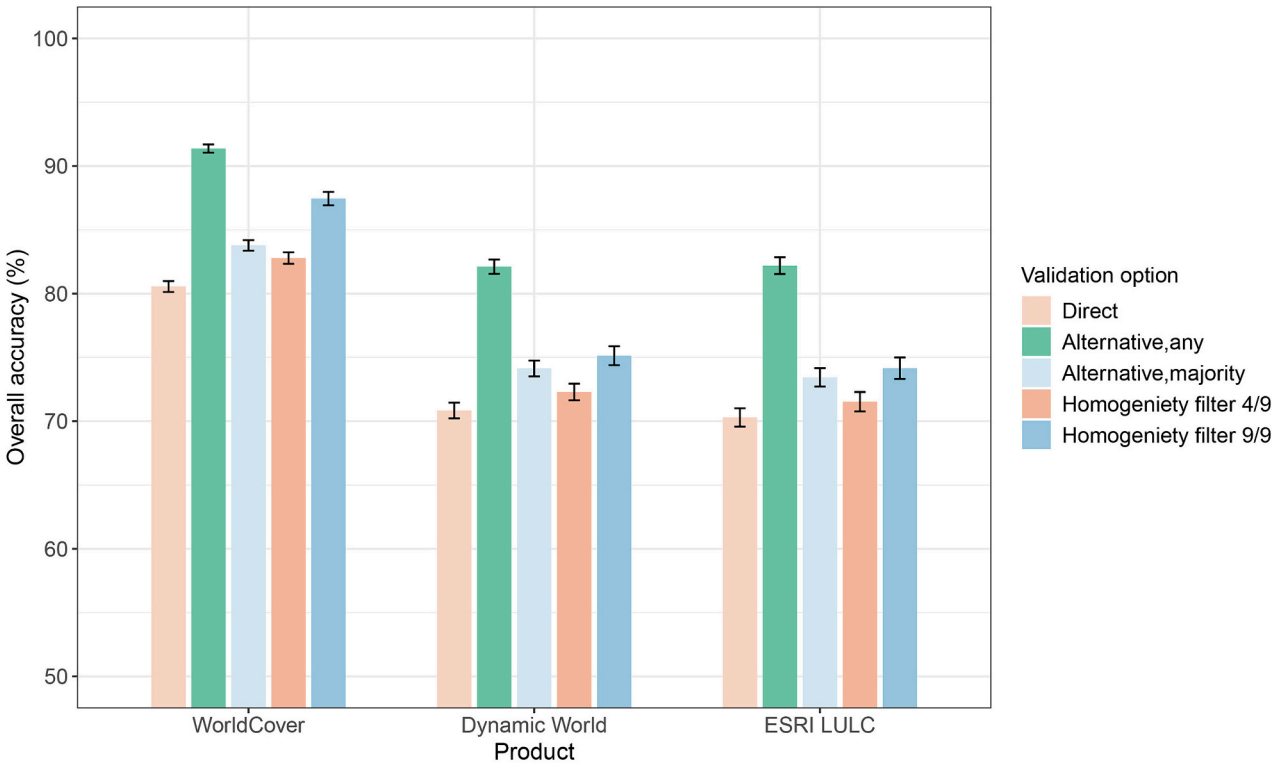

Fig. 5. Global overall accuracy of the GLC maps when applying different validation methods. Error bars represent 95% confidence interval (CI) of the estimated accuracy.

"Primary + Alternative label: Any" approach produced the highest accuracy estimates, followed by those of the "Homogeneity filter 9/9". The "Primary + Alternative label: Majority" and the "Homogeneity filter 4/9" produced comparable accuracy estimates.

The number of SSUs and PSUs used by the two different kinds of approaches, namely, alternative labelling and filtering based on homogeneity, is different (Table S2). More specifically, the "Primary + Alternative label" approaches used all available SSUs and PSUs, while

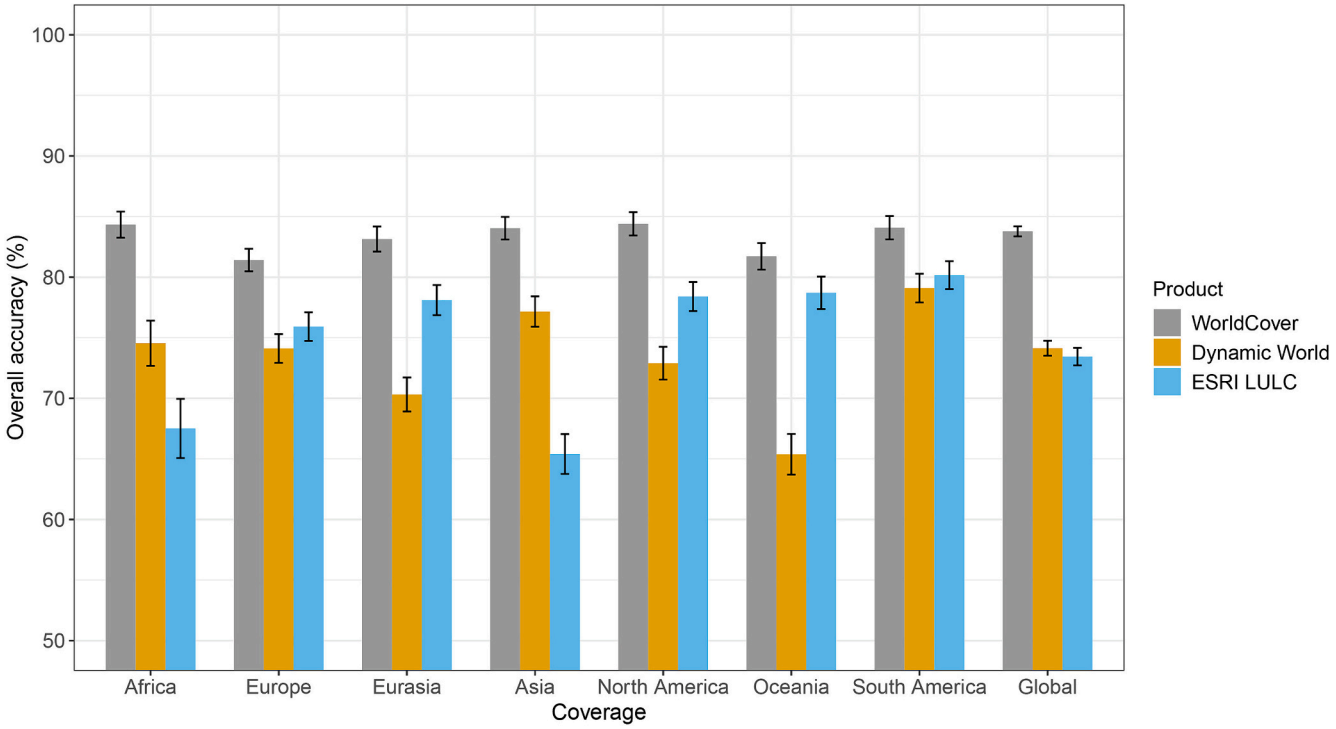

Fig. 6. Global and continental overall accuracies for the GLC maps based on the "Primary + Alternative label: Majority" approach. Error bars represent 95% CI of the estimated accuracy.

the "Homogeneity filter 4/9" discarded 7% - 9% of SSUs and the "Homogeneity filter 9/9" discarded almost 59% - 61% of SSUs (calculated from Table S2; considering all the three GLC maps). Therefore, the confidence interval ranges of the accuracy estimates were larger in the two filtering approaches compared to those of the "Primary + Alternative label" approaches (Fig. 5).

3.2. Accuracy comparison

To further compare the class-specific and overall accuracies of the GLC maps over the globe, per continent, and in the selected countries, we used the "Primary + Alternative label: Majority" approach, because it was expected to produce middle range (neither too optimistic nor too pessimistic) accuracy estimates while utilizing all available SSUs.

Fig. 6 illustrates how the accuracies of the three GLC maps differ between the continents. Globally, the greatest overall accuracy was achieved by WorldCover (83.8% ± 0.4%, 95% *CI*), followed by Dynamic World (74.1% ± 0.6%) and ESRI LULC (73.4% ± 0.7%); thus, the overall accuracy of the different maps varied within about 10%. Among the three maps, WorldCover obtained the highest overall accuracies for all continents. For Dynamic World, South America was found to have the greatest continental accuracy (79.1% ± 1.2%) and Oceania the smallest (65.4% ± 1.7%). The ESRI LULC mapped LC most accurately in South America (80.1% ± 1.2%) but for this map, Asia (65.4% ± 1.6%) was the least accurately mapped continent. The continental accuracies of Africa, Asia, Eurasia, and Oceania showed larger variation compared to the other continents for the three GLC maps (Fig. 6).

Table 3 shows the global class-specific accuracies for the seven LC classes. Among the seven LC types, water and trees were mapped with a relatively high accuracy by all the three GLC maps. For the trees class, the three maps tended to have more error of commission (100% - UA) than omission (100% - PA). Mixed vegetation (including grassland, shrubs, and flooded vegetation) was better mapped by WorldCover, while it was mapped with a low PA (53.5% ± 1.2%) by Dynamic World due to its confusion with trees and bare ground (Table S4 in supplementary material). Built areas were considerably overestimated by ESRI LULC and Dynamic World, for which the misclassification comes at the expense of trees and mixed vegetation mostly (Tables S4-S5), while on the other hand, the PA of built area in the two maps were all higher (> 87%) than that in WorldCover (73.5% ± 2.6%). Crops had considerable confusion with mixed vegetation in Dynamic World and ESRI LULC (Tables S4-S5), while in comparison, WorldCover achieved higher PA and UA for crops (Table 3). ESRI LULC had considerable underestimation (PA: 42.7% ± 3.2%) of bare ground that was misclassified into mixed vegetation (Table S5), while it achieved the highest UA (96.1% ± 0.8%) and Dynamic World achieved the highest PA (87.4% ± 1.1%) for bare ground. Snow & ice was best mapped by WorldCover (both UA and PA > 93%), while the Dynamic World and ESRI LULC maps had considerable overestimation of this class, which resulted from its misclassification between mixed vegetation and bare ground. The confusion matrices of the three GLC maps can be found in Tables S3-S5 of the supplemental material.

Table 3
The global class-specific accuracies for the GLC maps, including 95% CI. The highest PA and UA per class are highlighted in bold.

Class code	LC Type	WorldCover		Dynamic World		ESRI LULC	
		PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
0	Water	**87.1 ± 1.7**	**89.5 ± 1.8**	85.1 ± 2.5	79.8 ± 2.4	80.5 ± 2.8	83.8 ± 2.4
1	Trees	**92.7 ± 0.5**	**80.1 ± 0.7**	92.0 ± 0.5	72.7 ± 0.9	87.0 ± 0.7	78.7 ± 0.9
2	Mixed vegetation	**77.7 ± 0.8**	**83.4 ± 0.7**	53.5 ± 1.2	81.7 ± 0.9	75.5 ± 0.9	67.0 ± 1.2
4	Crops	**79.4 ± 1.5**	**80.7 ± 1.5**	66.5 ± 2.0	64.6 ± 2.3	73.3 ± 2.0	71.9 ± 1.8
6	Built area	73.5 ± 2.6	**65.9 ± 3.3**	87.3 ± 2.1	52.6 ± 2.5	**87.9 ± 2.1**	48.0 ± 2.5
7	Bare ground	83.0 ± 1.2	92.2 ± 0.9	**87.4 ± 1.1**	76.1 ± 1.7	42.7 ± 3.2	**96.1 ± 0.8**
8	Snow & Ice	99.1 ± 0.4	**93.0 ± 2.4**	**99.6 ± 0.4**	57.5 ± 3.3	99.0 ± 0.5	68.3 ± 3.2
Overall accuracy (%)		83.8 ± 0.4		74.1 ± 0.6		73.4 ± 0.7	

The continental class-specific accuracies are listed in Table S6 of the supplemental material. Generally, WorldCover presented better performance than the other two maps in characterizing most of the LC types in most continents. Dynamic World had comparable (and sometimes better) performance in characterizing built area in Africa, Europe, Eurasia, and North America, and bare ground in Africa, South America, and Asia. ESRI LULC accurately mapped mixed vegetation in Oceania (PA and UA > 83%) and crops in Europe (PA > 90%). Bare ground tended to be underestimated in most continents by WorldCover and ESRI LULC.

The accuracy estimates of the 47 larger countries having >100 PSUs are provided in Table S7 of the supplementary material. WorldCover was found to be more accurate in most of the countries. In contrast, ESRI outperformed the other two maps in Greenland (with an OA of 95.3%), which is dominated by snow and ice, while its performance is comparable to WorldCover in Colombia, Romania, Brazil, and Democratic Republic of the Congo with an OA above 80%. Dynamic World achieved a high OA in Saudi Arabia (95.8%) and Egypt (96.6%), which are dominated by desert, and it also performed good in Greenland (90.4%) and Peru (84.4%). Countries such as Mozambique, Tanzania, Nigeria, and Spain were mapped with an OA of <70% by all the three maps. Moreover, Nigeria, Iran, and Pakistan were mapped with an OA of <50% by ESRI LULC.

3.3. Map assessment at different homogeneity levels

Fig. 7 shows, at different homogeneity levels, the agreement between the map and validation data (left column) and the proportion of sample locations (i.e., SSUs) of the map and reference data (right column). As can be seen from the green matrices (Fig. 7), the agreement between the map and the reference data was higher at higher homogeneity levels (i.e., levels 7–9), which applies to all maps. In general, the agreement was greater when the homogeneity levels of the map and reference data matched (diagonal values in Fig. 7a, c, e) and smaller when the map and reference data did not agree on the level of homogeneity (off-diagonal values in Fig. 7a, c, e). This pattern was more visible in WorldCover (Fig. 7e) which showed closer agreement with validation data in terms of LC characterization, also at lower homogeneity levels (homogeneity levels 1–3) that represent more heterogeneous areas. Some relatively high agreement values were noticeable in the lower levels of homogeneity (such as the highlighted green cells in map homogeneity level 1 of Dynamic World in Fig. 7a), however, these could be related to the negligible number of SSUs in these homogeneity levels (Fig. 7b).

A difference in the total number of SSUs corresponding to different homogeneity levels can be observed among the GLC maps. As can be seen in Fig. 7b, d, f, WorldCover has the SSUs most spread out over different homogeneity levels, resulting in more SSUs at lower homogeneity levels compared to the other two products. The ESRI LULC shows the least number of SSUs in the lower homogeneity levels. Completely homogeneous SSUs (map homogeneity 9) account for 88.54%, 85.40%, and 78.79% of the sample locations for ESRI LULC, Dynamic World, and WorldCover, respectively, indicating that ESRI LULC tends to map less spatial detail than the other maps.

To assess the effect of land cover heterogeneity on the performance of the maps, we evaluated the accuracy of the maps at three different homogeneity levels (Low: H = 1–3, Medium: H = 4–6, and High: H = 7–9) according to the respective map homogeneity as shown in Fig. 7. The results of the validation are shown in Fig. 8. Compared with the global OA obtained using the entire validation dataset (Fig. 6), all the

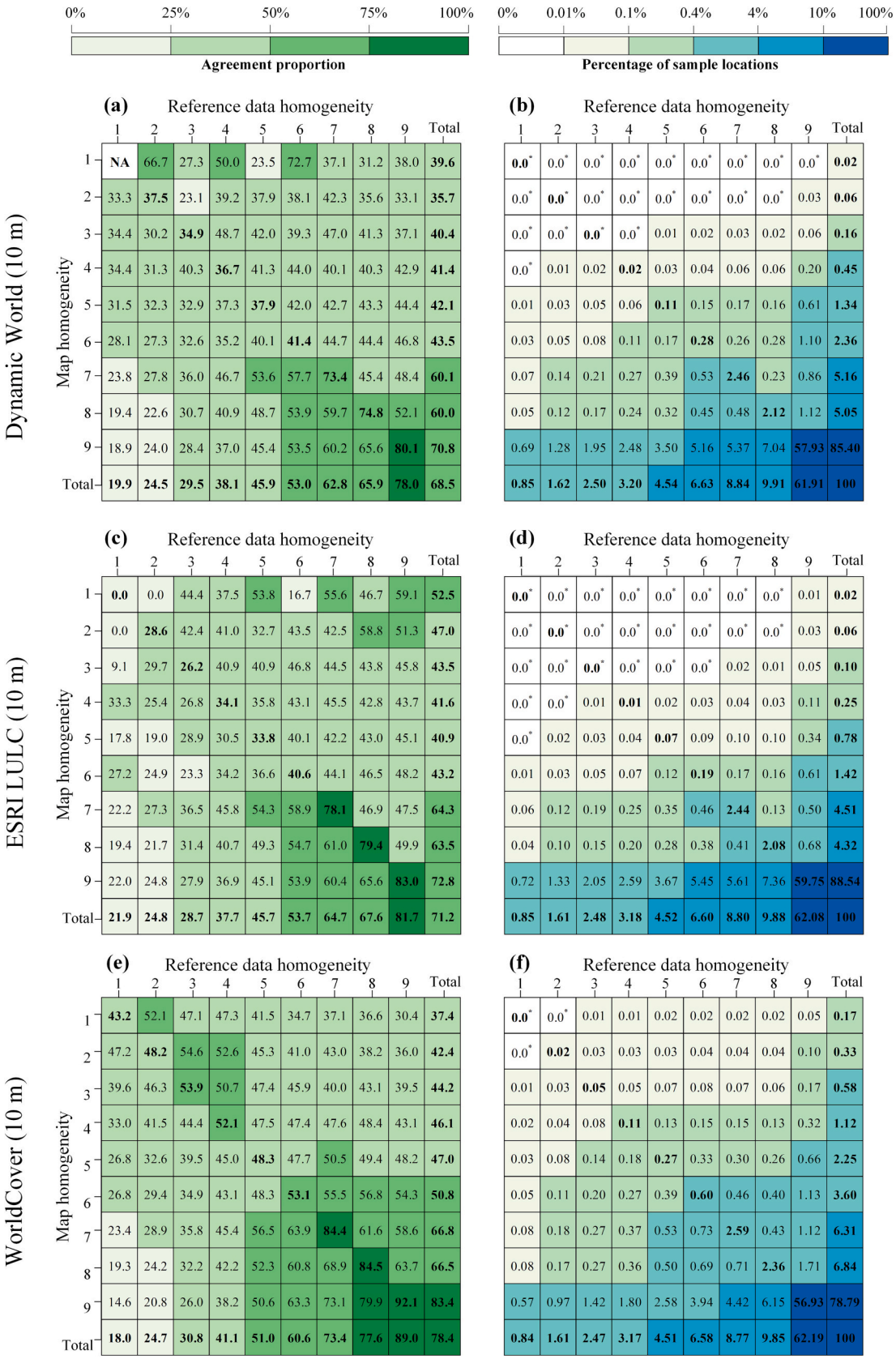

Fig. 7. Agreement proportion of sample locations (SSUs) for homogeneity levels of the map and reference data (a, c, e) and the percentage of sample locations at different homogeneity levels of the map and reference data (b, d, f). 0.0* indicates that the percentage of sample locations is <0.01%.

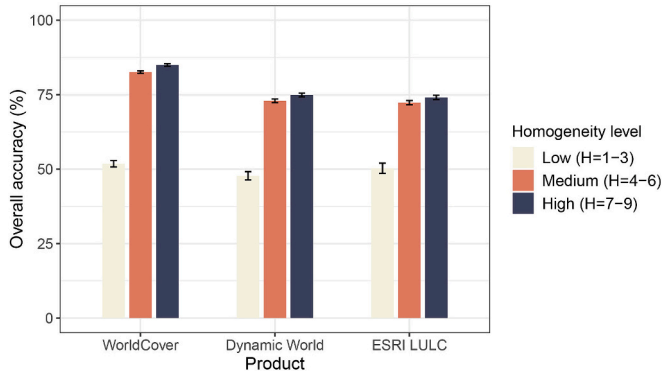

Fig. 8. Overall accuracies of the GLC maps at different homogeneity levels (H). The homogeneity levels Low (H = 1–3), Medium (H = 4–6), and High (H = 7–9) are based on the map homogeneity calculated in Fig. 7.

maps had reduced accuracy in heterogenous areas (H = 1–3) with an OA ranging between 47.8% and 51.8% (Fig. 8). On the contrary, the OA of the maps in strongly homogeneous areas (H = 7–9) was greater than the global OA calculated using the entire validation dataset. This indicates that the homogeneity level influences the validation accuracy and the evaluated GLC maps are less accurate in heterogeneous landscapes.

For both the GLC maps and the validation data, we also assessed their homogeneity at PSU-level and Table 4 shows the frequency of PSUs with different number of LC classes being present in a 100 × 100 m PSU. Note that Fig. 7 is based on 3 × 3 kernels at the resolution of the maps, while Table 4 reflects how many LC types are mapped within a PSU regardless of the map's spatial resolution. Results showed that all maps more often predict a single LC type in a PSU than observed in the reference data. PSUs with a single LC type accounted for 41.1% of the validation data, 59.3% for WorldCover, 71.3% for Dynamic World, and 82.2% for ESRI LULC. ESRI LULC had a greater tendency to predict only one class in a PSU, while its multi-class (i.e., two or more LC types) proportion was much lower compared to that of Dynamic World and WorldCover.

4. Discussion

4.1. Effect of the applied validation approach

The validation methods produced widely different accuracy estimates, showing that the way of dealing with reference data uncertainty has an impact on the final validation results. The "Primary + Alternative label: Any" option reported the greatest accuracies for the three GLC maps. By allowing any LC types within the 3 × 3 kernel as an alternative label, this approach is the most lenient towards the map to produce a "match" with the reference data. It is noted that the relatively small number of LC classes taken into consideration in this study made the validation results to be rather optimistic. With only seven classes, there is a great chance that the map class matches one of the LC classes in a 3 × 3 kernel of the validation data, particularly in heterogeneous areas. For products with a larger number of LC classes such as the Corine land cover map (Büttner et al., 2004)–which has 44 LC classes–the leniency towards the map will be much less.

Table 4
Frequency of PSUs with different number of LC classes present in a 100 × 100 m area for the validation data and the GLC product.

Data	PSUs with a single LC class (%)	PSUs with 2 LC classes (%)	PSUs with 3 LC classes (%)	PSUs with 4 LC classes (%)	PSUs with 5 LC classes (%)	PSUs with 6 LC classes (%)
Validation data	41.11	39.23	13.56	4.82	1.25	0.03
WorldCover	59.31	30.52	8.11	1.80	0.24	0.00
Dynamic World	71.33	24.46	3.76	0.42	0.01	0.02
ESRI LULC	82.15	15.69	2.04	0.12	0.01	0.00

The accuracies based on the "Primary + Alternative label: Majority" were lower than those of the "Primary + Alternative label: Any" approach (Fig. 5). Likely, validation based on any alternative label within the 3 × 3 kernel leads to overly optimistic accuracy estimates. Both approaches use an alternative labelling principle, introducing a problem of blurring the agreement determination between the map and the validation labels. This impedes estimating the area of LC classes (Stehman, 2013).

Although the "Homogeneity filter 4/9" approach generated similar accuracies as the "Primary + Alternative label: Majority", the former only uses a subset of the original reference data. Removal of sampling units in heterogeneous areas introduces zero inclusion probability in heterogeneous areas and therefore invalidates the probability sampling requirement of "known and non-zero inclusion probability" (Olofsson et al., 2014). As such, the estimated accuracy does not represent the entire mapped area, but rather only more homogeneous areas, which is a major limitation of this approach. Another problem with this approach is that the determination of the homogeneity benchmark is arbitrary because the level of homogeneity as well as the kernel size can be varied in different assessment cases.

The "Homogeneity filter 9/9" deals with reference data uncertainty by only accounting for locations where all the SSUs in a 3 × 3 kernel have the same LC class. Similar to the "Homogeneity filter 4/9", this approach considers that the reference data error is smaller in homogeneous areas. However, it eliminates a considerable number of sample units in heterogeneous areas. In this study, the "Homogeneity filter 9/9" discarded 59% to 61% of the SSUs (calculated from Table S2 for the three GLC maps). Especially for the maps depicting higher heterogeneity, this leads to the removal of a great number of sample units and the validation would therefore include a relatively small subset, which consequently leads to biased accuracy estimates representing only a small subset of the mapped area (Stehman et al., 2003).

The direct approach uses a single-pixel (SSU) comparison and does not account for possible reference data uncertainties owing to positional errors and labelling ambiguity (Gu and Congalton, 2020). This produces an underestimation of accuracy, especially in heterogeneous areas where the positional shift of a pixel and ambiguity in labelling a dominant LC class in a small sized sample unit can have a large impact (Stehman and Foody, 2009). Our current study tested the impact of approaches dealing with reference data uncertainty when validating 10 m-resolution LC maps. However, the tested approaches did not address interpretation variability or interpretation errors (e.g., labelling the entire PSU as shrubs instead of trees) (McRoberts et al., 2018; Tarko et al., 2021). Ideally these reference data error variability should be quantified, and its individual and combined impacts need to be analysed and accounted for in accuracy estimation in future studies (Stehman et al., 2022).

In Table 5, we summarize the most noticeable strengths and limitations of each validation method. Considering the advantages and disadvantages of each validation approach to address reference data uncertainty (Table 5), a careful assessment is required when applying a method to validate a high-resolution LC product. The "direct" approach is likely too pessimistic whereas the "Primary + Alternative any" is likely too optimistic about achieved accuracy. The major downside of the homogeneity filter approaches was the exclusion of sample units; hence the estimated accuracies do not represent the entire mapped area. The "Primary + Alternative majority" is somewhere in between the two extremes and from that perspective a reasonable choice.

The large differences between the validation results indicate that future studies should consider the reference data uncertainty when estimating the accuracy of high-resolution LC maps (Stehman and

Table 5
The major advantages and disadvantages of the five applied validation approaches.

Validation approach	Advantages (+) and disadvantages (−)
Direct/Primary	+ No alterations of the data and all sample points are included. - Does not take reference data uncertainties into account, all errors are deemed map errors. - Tends to be too pessimistic.
Primary + Alternative label: Any	+ Uses all sample units in the validation. + Reduces the potential effect of reference data uncertainty. - Less suitable for products with a relatively small number of land cover classes. - Uses alternative reference labels, hence class area estimation based on reference data may become problematic. - Tends to be too optimistic.
Primary+ Alternative label: Majority	+ Uses all sample points in the validation. + Limits the potential effect of reference data uncertainty. - Uses alternative reference labels, hence class area estimation based on reference data may become problematic.
Homogeneity filter 4/9	+ Limits the potential effect of reference data uncertainty. - Eliminates sample units that do not match the homogeneity criteria, which invalidates the probability sampling requirement: “non-zero” probability of being selected in the sample for all units in the population. Thus, the estimated map accuracy does not represent the entire mapped area. - Selection of the homogeneity level is relatively arbitrarily; only areas meeting the homogeneity criterion are validated.
Homogeneity filter 9/9	+ Minimizes the potential effect of reference data uncertainty. - Eliminates sample units that do not match the homogeneity criteria, which invalidates the probability sampling requirement: “nonzero” probability of being selected in the sample for all units in the population. Thus, the estimated map accuracy does not represent the entire mapped area. - Validates only completely homogeneous areas.

Foody, 2019). The variation in the estimated accuracies also highlights the importance of transparency in reporting. The selected validation method greatly influenced the quantitative results, even though the relative ranking of the compared maps was mostly consistent across methods. Guidance on validating high-resolution LC maps would ensure easier comparison of these products for both users and producers. Although some guidelines on LC validation are available, such as the good practices proposed by Olofsson et al. (2014) and the approaches for estimating map accuracy dealing with ground reference data error suggested by Foody (2010), they should be updated to account for the challenges of high-resolution LC products.

4.2. Evaluation of recent 10 m-resolution GLC maps

This study presents a comprehensive, independent, and statistically rigorous accuracy assessment for the recent 10 m-resolution GLC maps following the internationally recommended validation guidelines (Strahler et al., 2006). According to the validation results, WorldCover achieved the highest global overall accuracy, followed by the Dynamic World and ESRI LULC maps (Fig. 6). The accuracy estimates reported in this study are different from those reported by map producers (Table 1) and some previous studies, which stems from the used validation methods, validation datasets, year of the map, data processing (e.g., resampling), and the reclassification to seven LC classes performed in the current study.

Our validation results conflict with a previous comparison of the same three GLC maps (WorldCover, Dynamic World, and ESRI LULC) for the year 2020 done by Venter et al. (2022), where ESRI LULC was reported to have the highest overall accuracy among the 10 m-resolution maps. However, since they used the validation dataset provided by the Dynamic World team, the global assessment cannot be considered independent for validating ESRI LULC because some of the Dynamic World’s validation data may have been used in the training process of the ESRI LULC map production (Venter et al., 2022). Nevertheless, their regional assessment across Europe based on an independent validation dataset (ground truth data from the European Union’s Land Use/Cover Area frame Survey (LUCAS)), where WorldCover showed the highest accuracy, aligns with our results. Moreover, our validation results are consistent with several findings of previous national comparisons. For example, Wang and Mountrakis, 2023 found that WorldCover and ESRI LULC performed better than Dynamic World in the conterminous U.S., which alignes with our results for North America (Table S6) when examining class-specific accuracies. Kang et al. (2022) found that WorldCover performed better than ESRI LULC over northwestern China, which is consistent with our findings for China (Table S7) when considering the overall accuracy.

The mapping accuracy of the three GLC maps varies across the considered LC classes and continents. Generally, the three maps have relatively high accuracy for water and trees, and lower accuracies for mixed vegetation, crops, bare ground, and built-up areas. WorldCover outperforms Dynamic World and ESRI LULC at a global scale, but for some LC classes and over some continents and countries, their performances are comparable. For instance, Dynamic World has relatively high accuracies in characterizing the built-up areas in North America and ESRI LULC is good at mapping the mixed vegetation in Oceania. Despite that, users should be aware of the weaknesses of the maps for overall mapping at the global scale. Notably, the three maps tend to overestimate trees; Dynamic World overestimates bare ground, while ESRI LULC and WorldCover underestimate this class globally. All the three maps have lower accuracies in areas with greater heterogeneity (Fig. 8), revealing that current GLC products are still limited in accurately identifying LC types in heterogeneous areas. The maps’ quality varies across regions and countries. For instance, the OA of Iran decreases from 85.3% ± 3.5% (by WorldCover) to 42.9% ± 7.5% (by ESRI LULC). All maps have low accuracies (<70%) in countries such as Mozambique, Tanzania, Nigeria, and Spain.

The map assessment at different homogeneity levels reveals that the ESRI LULC exhibits the least spatial detail with 88.54% of SSUs at completely homogeneous areas (Fig. 7d) compared to the other GLC maps. Maps are expected to capture the level of spatial/landscape detail discernible with the spatial resolution of the data used for their creation. However, ESRI’s map characterization does not match its spatial resolution of 10 m. Compared to ESRI LULC and Dynamic World, WorldCover has a much smaller percentage (78.8%) of SSUs in completely homogeneous 3 × 3 kernels (Fig. 7f), indicating that it depicts a greater landscape spatial detail than the other two 10 m maps. This can also be seen from Fig. 9, where WorldCover depicts more landscape detail than ESRI LULC and Dynamic World, preserving small features and complex patterns, such as roads and individual agricultural fields.

The spatial detail representation among the three 10 m-resolution maps—Dynamic World, ESRI LULC, and WorldCover—varies substantially owing to differences in their data sources, classification methods, and the training data. In examining classification methodologies and input data sources, WorldCover employs gradient boosting decision tree algorithms (i.e., CatBoost), integrating Sentinel-1 and Sentinel-2 data as well as their temporal dynamics. These dynamics are derived from 10-day median composites of the data, enabling the capturing of seasonal variations. Radar data penetrates cloud cover and captures surface structure, which can enhance LC classification, particularly in cloudy or

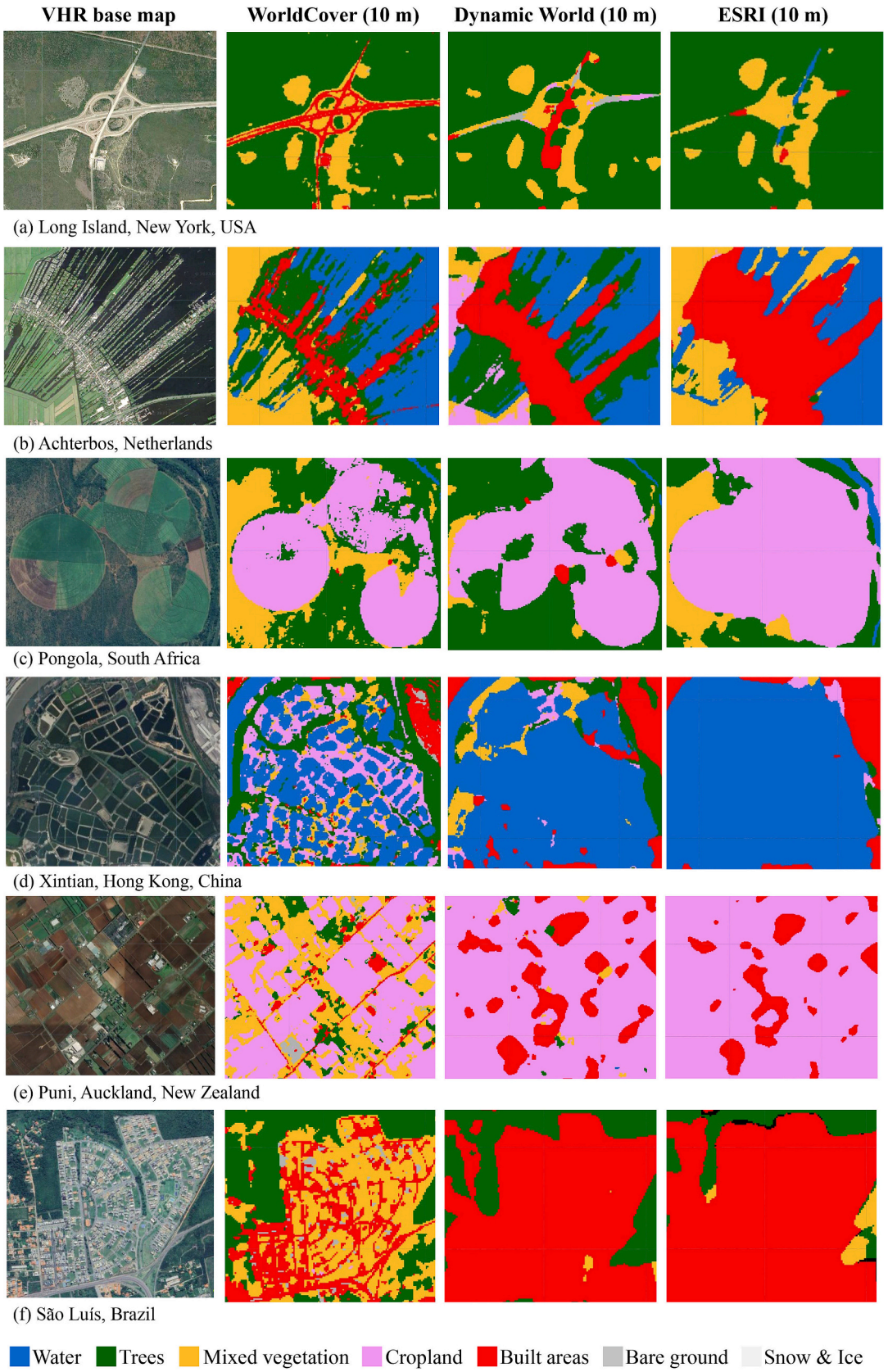

Fig. 9. Comparison of the spatial representation of GLC maps in a) Long Island, New York, USA (72.6°W, 40.9°N), b) Achterbos, Netherlands (4.9°E, 52.2°N), c) Pongola, South Africa (31.8°E, 27.4°S), d) Xintian, Hong Kong, China (114.1°E, 22.5°N), e) Puni, Auckland, New Zealand (174.9°E, 37.2°S), and f) São Luís, Brazil (44.2°W, 2.5°S).

mixed vegetated areas where optical data alone are insufficient (Slagter et al., 2020; Xu et al., 2022). Using Sentinel-2 data as inputs, Dynamic World uses a Fully Convolutional Neural Network (FCNN) model that classifies each Sentine-2 imagery individually (Brown et al., 2022) and ESRI LULC utilizes a UNet model that incorporates multiple observations primarily in cloudier regions. Both models did not utilize the temporal characteristics of the LC types. The minimum mapping unit (MMU) of training data directly impact the model's ability to classify LC types accurately. Dynamic World and ESRI LULC share a training dataset annotated in 50 × 50 m units, while the training data for WorldCover was annotated at a 10 m-resolution, which is more capable of delineating LC transitions and smaller landscape features, contributing to the superior spatial detail observed in WorldCover's classifications.

When choosing the most suitable LC map for specific application goals, users should consider both the quantitative accuracy and the representation of spatial detail. It's crucial not to base the decision solely on the map's resolution. Instead, assess how accurately the map captures the landscape's detail. For an overview of the accuracies, users are referred to the global, continental (Table S6), and national (Table S7) validation metrics in the supplementary materials. When combining different GLC maps in an analysis, users should bear in mind the disagreement in class definitions between the maps (see Table S1). For the use of Dynamic World, it is advised to select an appropriate way to derive the LC labels from its near-real-time maps and their probability layers for specific use cases. For snow & ice, crops, and bare ground, data from the growing season are useful to determine the LC type on a yearly basis.

4.3. Limitation of the current study and implications for future GLC mapping and validation

This study assessed three 10 m-resolution GLC maps in terms of accuracy and their ability to represent spatial detail of landscape. Accordingly, comparative accuracy estimates at global, continental, and national (for 47 countries) levels are provided to aid users in selecting maps with the best accuracy for their class or area of interest. The findings of this study offer insights into the future of GLC mapping and the associated validation processes. Several key implications emerge from the results and observations, which can guide future endeavors in this field.

4.3.1. Standardization of LC class definitions

One of the limitations of the current study is that the grouping of some classes, especially those that are challenging to separate using remote sensing data such as shrubs, herbaceous vegetation, and flooded vegetation, will lead to inflation in the accuracy for the evaluated maps. This is caused by inconsistencies in the legend definition among the maps and it underscores the need for future GLC mapping efforts to standardize LC class definitions. Disagreements in class definitions across different maps, such as "built area" or "crops" can also lead to substantial inconsistencies for comparative assessments. For instance, urban green areas (e.g., parks and lawns) are classified as built areas in the evaluated GLC maps except WorldCover and the validation data. Despite the adjustments made in the validation data to assess urban classes for Dynamic World and ESRI LULC, there may still be inconsistencies in the reference data for characterizing "urban" instead of "built-up areas" (Liu et al., 2014), leading to the low user's accuracy of built areas for the two maps. In the case of cropland, fallow plots are included as crops in Dynamic World and ESRI LULC while they are considered "herbaceous vegetation" in the validation data, which might explain higher confusion rates between crops and mixed vegetation in the validation results. Thus, addressing the discrepancies and aligning class definitions will be pivotal for enhancing the consistency and comparability of LC maps over large scales.

4.3.2. Improving LC characterization and spatial detail representation

Firstly, efforts should be directed towards improving classes that are difficult to characterize, particularly in heterogeneous landscapes. Similar LC types, such as crops and mixed vegetation (i.e., grassland, flooded vegetation, and shrubs), are spectrally and physically alike, leading to misclassification. Multitemporal information and land surface phenology that are intrinsic to the definition of these classes can contribute to differentiating such LC categories (Ienco et al., 2019). As discussed in section 4.2, Convolutional Neural Network (CNN)-based approaches (e.g., ESRI LULC, Dynamic World) should consider including temporal dynamics in their models.

Secondly, the capability of high-resolution maps for representing LC detail should be improved to match its respective spatial resolution. Although recent studies have demonstrated the superiority of deep learners over traditional machine learning classifiers (Mountrakis and Heydari, 2023), in the current study, it is observed that the pixel-based machine learning classification achieved better accuracy and spatial detail representation than the kernel-based CNN approaches, indicating room for further improvements for the latter approaches to map LC at high resolutions. The MMU of training data is crucial for model performance. Annotating training data at a finer MMU and ensuring a diverse representation of LC types especially in heterogeneous areas, will enable more detailed LC classification. Integrating multiple data sources could complement single data source. These approaches, as demonstrated by WorldCover and other recent studies (Venter and Sydenham, 2021; Xu et al., 2022), can enhance the accuracy of LC classifications. Regarding the CNN models, future efforts could explore adapting existing networks (Wang et al., 2023a), training new networks (Li et al., 2022), or using hybrid models that combine the strengths of different architectures. For instance, integrating the global contextual understanding of visual transformer (ViT) with the local detail captured by CNN could offer a balanced approach to capturing both broad LC patterns and fine spatial details (Yue et al., 2024), and combining CNN with Long Short-Term Memory (LSTM) network can capture both spatial and temporal information (Masolele et al., 2021; Mountrakis and Heydari, 2023). Enhancement in the above aspects will improve the spatial detail representation of the maps and benefit the accurate characterization of LC types in heterogeneous areas (Mountrakis and Heydari, 2023).

4.3.3. Reducing reference data uncertainty

As shown in this study, varied accuracy estimates were obtained when choosing different ways of dealing with reference data uncertainty. To address reference data uncertainty effectively in high-resolution LC validation, it is crucial to quantify and account for reference data errors in map validation (Stehman and Foody, 2019) and select validation methods that consider neighboring pixels and spatial context. This requires moving from reference data at a single pixel (as sample unit) to collecting neighborhood or spatial heterogeneity information for validation data targeting high-resolution maps. Such approaches enhance the reliability of validation efforts, particularly in regions known for their heterogeneity. In addition, land dynamics should be considered in the reference data collection process. The interpretation of cropland/fallow dynamics, snow & ice, or bare ground should use both intra- and inter-annual time series data to determine the maximum/minimum coverage of the LC type (Potapov et al., 2022). Furthermore, transparency in reporting and validation methods is paramount to ensure the delivery of accurate and meaningful results for diverse applications.

4.3.4. Next steps: addressing validation at regional/national levels and for LC changes

The practical utility of GLC maps often concerns regional and national levels. Countries with limited capacities for national monitoring have a high tendency to rely on these global maps. However, national-level validation in the current study was conducted exclusively for

larger countries with >100 PSUs and our validation dataset does not support conclusions regarding smaller countries. Currently, the acquisition of validation data that is representative at regional or national scales for global products is still challenging. This difficulty arises from the extensive time and financial resources required for data collection at these scales, coupled with the need for an in-depth understanding of regional LC characteristics that necessitates local expertise. To facilitate validation at regional/national level in the future, collaborative initiatives should be undertaken to gather high-quality validation data that accurately represent LC conditions at these finer scales. One potential strategy involves promoting open access to existing datasets through collaboration with local and national agencies and research institutions. Furthermore, leveraging public engagement and data collection platforms, such as Collect Earth Online (Saah et al., 2019), could significantly contribute to this endeavor. Such efforts will improve the applicability and reliability of LC maps for specific geographic regions and facilitate informed decision-making at various levels.

Advancements in satellite data availability enable near-real-time mapping and time-series analysis of LC changes. These capabilities are vital for monitoring LC types with significant temporal dynamics and tracking land changes. Except WorldCover, the evaluated maps are updated annually (i.e., ESRI LULC) or weekly (i.e., Dynamic World). However, LC changes suggested by these datasets are not validated. The challenge in validating LC changes, particularly on a sub-annual basis, largely stems from the limited availability of high-frequency validation datasets (Lamarche et al., 2021). The existing validation dataset is primarily updated on an annual level and with limited change information, which reduces its effectiveness in assessing the accuracy of maps that are updated more frequently. Hence, there is a pressing need to develop comprehensive validation datasets that capture the intra-annual dynamics of LC changes as well as changes over longer periods. Such datasets will be instrumental in ensuring the accuracy and reliability of time-series GLC maps.

As we journey forward, high-resolution GLC mapping holds the potential to significantly enhance environmental monitoring and informed decision-making. Incorporating these implications into future GLC mapping and validation endeavors will contribute to producing more accurate, consistent, and reliable LC maps.

5. Conclusion

This paper presents a comprehensive and independent evaluation of the accuracy and spatial detail representation of three recent 10 m-resolution GLC maps, i.e., ESA WorldCover, Google-WRI Dynamic World, and ESRI LULC. We compared five approaches to deal with reference data uncertainty, an essential factor in assessing high-resolution maps. The "Direct/Primary" approach, "Primary + Alternative label" approaches, and "Homogeneity filter" approaches produced widely different accuracy estimates, underscoring the significance of accounting for reference data uncertainty in high-resolution map validation. We propose to use the direct label supplemented by alternative labels from the majority of neighboring pixels ("Primary + Alternative label: Majority") to account for reference data uncertainties.

Our analysis reveals differences in the performance among the GLC maps for different LC classes and geographic regions. Overall, WorldCover has achieved the highest global accuracy, followed by Dynamic World and ESRI LULC. For some LC classes over some continents, the latter two maps have comparable or better performance than WorldCover. The spatial homogeneity analysis shows that ESRI LULC and Dynamic World have less spatial detail compared to WorldCover, tending to generate more homogeneous LC characterizations. All the evaluated maps have low accuracy in heterogeneous areas, with an OA of around 50%. The maps' performance varies in countries, and some countries are mapped with low accuracies, which needs to be considered with care before applying the data in a national context.

Based on these findings, we recommend the following to the global community of GLC map users and producers:

For Users: When selecting a GLC map, consider not only the accuracy for the extent of the area of interest (e.g., global, continental, national) but also the spatial detail appropriate for the application, noting that the level of spatial detail varies across the maps despite all being at 10 m resolution.

For Map Producers:

- Use standardized LC class definitions to ensure consistency and comparability across different GLC maps. This will help address discrepancies and facilitate the integration of multiple maps.
- Improve the spatial detail and accuracy of the GLC maps for difficult classes (such as mixed vegetation), heterogeneous landscapes, and at the national level by enhancing deep learning models. Annotating training data at finer MMUs and incorporating temporal dynamics and multi-source data can aid in this effort.
- Account for reference data uncertainty in high-resolution map validation. Include assessments of regional and national-level validation and for LC changes in future validation efforts. Collaborative initiatives and leveraging public engagement platforms can help acquire high-quality validation data at these scales.

With advances in satellite data acquisition, improved computation, and deep learning approaches in remote sensing, the production of high-resolution GLC products is accelerating. Anticipating the availability of more high-resolution GLC products in the future, independent validation will be crucial to ensure the credibility and reliability of these maps and support diverse applications at global, regional, and national levels.

CRediT authorship contribution statement

Panpan Xu: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Nandin-Erdene Tsendbazar:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Martin Herold:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Sytze de Bruin:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Myke Koopmans:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Tanya Birch:** Writing – review & editing, Methodology, Investigation. **Sarah Carter:** Writing – review & editing, Visualization, Investigation. **Steffen Fritz:** Writing – review & editing, Visualization, Investigation. **Myroslava Lesiv:** Writing – review & editing, Visualization, Methodology, Investigation. **Elise Mazur:** Writing – review & editing, Visualization, Software, Methodology, Investigation. **Amy Pickens:** Writing – review & editing, Methodology, Investigation. **Peter Potapov:** Writing – review & editing, Methodology, Investigation. **Fred Stolle:** Writing – review & editing, Visualization, Methodology, Investigation, Funding acquisition. **Alexandra Tyukavina:** Writing – review & editing, Visualization, Methodology, Investigation. **Ruben Van De Kerchove:** Writing – review & editing, Visualization, Methodology, Investigation. **Daniele Zanaga:** Writing – review & editing, Visualization, Methodology, Investigation.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the World Resources Institute (No. G2914), the ESA SEN4LDN project, the CGIAR/CIAT MITIGATE+ project, and the Open-Earth-Monitor Cyberinfrastructure project funded by the European Union's Horizon Europe Research and Innovation Program (No. 101059548). We thank Alex Kovac for providing assistance with aggregating Dynamic World annual composites. Additionally, we thank Martina Duerauer for providing validation data collection platform service. We also thank the anonymous reviewers for their constructive comments that helped us improve the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2024.114316.

References

Aguilar, M.A., Nemmaoui, A., Aguilar, F.J., Novelli, A., García Lorca, A., 2017. Improving georeferencing accuracy of very high resolution satellite imagery using freely available ancillary data at global coverage. Int J Digit Earth 10, 1055–1069. https://doi.org/10.1080/17538947.2017.1280549.

Ban, Y., Gong, P., Giri, C., 2015. Global land cover mapping using earth observation satellite data: recent progresses and challenges. ISPRS J. Photogramm. Remote Sens. 103, 1–6. https://doi.org/10.1016/j.isprsjprs.2015.01.001.

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., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci Data 9, 251.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., Soukup, T., 2004. The CORINE land cover 2000 project. EARSeL eProceedings 3, 331–346.

Chaaban, F., El Khattabi, J., Darwishe, H., 2022. Accuracy assessment of ESA WorldCover 2020 and ESRI 2020 land cover maps for a region in Syria. J. Geovis. Spat. Anal. 6, 31. https://doi.org/10.1007/s41651-022-00126-w.

D'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges, C., Lemoine, G., van der Velde, M., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Sci Data 7, 352. https://doi.org/10.1038/s41597-020-00675-z.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution Mission for GMES operational services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026.

Foody, G.M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. Remote Sens. Environ. 114, 2271–2285.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Liu, Q., Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci Bull (Beijing) 64, 370–373. https://doi.org/10.1016/j.scib.2019.03.002.

Google Earth Engine, 2022. Dynamic World V1. URL. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1 (accessed 6.9.22).

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.

Gu, J., Congalton, R.G., 2020. Analysis of the impact of positional accuracy when using a single pixel for thematic accuracy assessment. Remote Sens. 12, 4093.

Gu, J., Congalton, R.G., 2021. Analysis of the impact of positional accuracy when using a block of pixels for thematic accuracy assessment. Geographies 1, 143–165.

Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22. https://doi.org/10.1016/J.ISPRSJPRS.2019.09.016.

Kang, J., Yang, X., Wang, Z., Cheng, H., Wang, J., Tang, H., Li, Y., Bian, Z., Bai, Z., 2022. Comparison of three ten meter land cover products in a drought region: a case study in northwestern China. Land (Basel) 11. https://doi.org/10.3390/land11030427.

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 International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 4704–4707.

Kinnebrew, E., Ochoa-Brito, J.I., French, M., Mills-Novoa, M., Shoffner, E., Siegel, K., 2022. Biases and limitations of global Forest change and author-generated land cover maps in detecting deforestation in the Amazon. PLoS One 17. https://doi.org/10.1371/journal.pone.0268970.

Lamarche, C., Bontemps, S., Marissiaux, Q., Defourny, P., Arino, O., 2021. Towards a Multi-Level Sampling Scheme for Land Cover and Land Cover Change Validation. Lessons Learned from the Land Cover Climate Change Initiative. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1986–1989. https://doi.org/10.1109/IGARSS47720.2021.9553898.

Li, Z., Zhang, H., Lu, F., Xue, R., Yang, G., Zhang, L., 2022. Breaking the resolution barrier: a low-to-high network for large-scale high-resolution land-cover mapping using low-resolution labels. ISPRS J. Photogramm. Remote Sens. 192, 244–267. https://doi.org/10.1016/j.isprsjprs.2022.08.008.

Liu, Z., He, C., Zhou, Y., Wu, J., 2014. How much of the world's land has been urbanized, really? A hierarchical framework for avoiding confusion. Landsc. Ecol. 29, 763–771. https://doi.org/10.1007/s10980-014-0034-y.

Masolele, R.N., De Sy, V., Herold, M., Marcos Gonzalez, D., Verbesselt, J., Gieseke, F., Mullissa, A.G., Martius, C., 2021. Spatial and temporal deep learning methods for deriving land-use following deforestation: a pan-tropical case study using Landsat time series. Remote Sens. Environ. 264, 112600 https://doi.org/10.1016/J.RSE.2021.112600.

McRoberts, R.E., Stehman, S.V., Liknes, G.C., Næsset, E., Sannier, C., Walters, B.F., 2018. The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. ISPRS J. Photogramm. Remote Sens. 142, 292–300. https://doi.org/10.1016/J.ISPRSJPRS.2018.06.002.

Mountrakis, G., Heydari, S.S., 2023. Harvesting the Landsat archive for land cover land use classification using deep neural networks: comparison with traditional classifiers and multi-sensor benefits. ISPRS J. Photogramm. Remote Sens. 200, 106–119. https://doi.org/10.1016/j.isprsjprs.2023.05.005.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57.

Pengra, B., Long, J., Dahal, D., Stehman, S.V., Loveland, T.R., 2015. A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30m continuous field tree cover data. Remote Sens. Environ. 165, 234–248. https://doi.org/10.1016/j.rse.2015.01.018.

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. Photogramm. Eng. Remote. Sens. 66, 1011.

Potapov, P., Hansen, M.C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, X.-P., Baggett, A., Kommareddy, I., Kommareddy, A., 2022. The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. Frontiers in Remote Sensing 3.

Potere, D., 2008. Horizontal positional accuracy of Google Earth's high-resolution imagery archive. Sensors 8, 7973–7981.

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., Poortinga, A., Markert, K., Quyen, N.H., San Aung, K., Schlichting, L., Matin, M., Uddin, K., Aryal, R.R., Dilger, J., Lee Ellenburg, W., Flores-Anderson, A.I., Wiell, D., Lindquist, E., Goldstein, J., Clinton, N., Chishtie, F., 2019. Collect earth: an online tool for systematic reference data collection in land cover and use applications. Environ. Model Softw. 118, 166–171. https://doi.org/10.1016/J.ENVSOFT.2019.05.004.

Slagter, B., Tsendbazar, N.E., Vollrath, A., Reiche, J., 2020. Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: a case study in the St. Lucia wetlands, South Africa. Int. J. Appl. Earth Obs. Geoinf. 86 https://doi.org/10.1016/j.jag.2019.102009.

Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. Remote Sens. Environ. 132, 202–211. https://doi.org/10.1016/j.rse.2013.01.016.

Stehman, S.V., Foody, G.M., 2009. Accuracy Assessment (The SAGE handbook of remote sensing).

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199.

Stehman, S.V., Wickham, J.D., Smith, J.H., Yang, L., 2003. Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: statistical methodology and regional results. Remote Sens. Environ. 86, 500–516. https://doi.org/10.1016/S0034-4257(03)00128-7.

Stehman, S.V., Mousoupetros, J., McRoberts, R.E., Næsset, E., Pengra, B.W., Xing, D., Horton, J.A., 2022. Incorporating interpreter variability into estimation of the total variance of land cover area estimates under simple random sampling. Remote Sens. Environ. 269, 112806 https://doi.org/10.1016/j.rse.2021.112806.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., Woodcock, C.E., 2006. Global land cover validation: recommendations for evaluation and accuracy assessment of global land cover maps. European Communities, Luxembourg 51, 1–60.

Szantoi, Z., Geller, G.N., Tsendbazar, N.-E., See, L., Griffiths, P., Fritz, S., Gong, P., Herold, M., Mora, B., Obregón, A., 2020. Addressing the need for improved land cover map products for policy support. Environ. Sci. Pol. 112, 28–35. https://doi.org/10.1016/j.envsci.2020.04.005.

Tarko, A., Tsendbazar, N.-E., de Bruin, S., Bregt, A.K., 2021. Producing consistent visually interpreted land cover reference data: learning from feedback. Int J Digit Earth 14, 52–70. https://doi.org/10.1080/17538947.2020.1729878.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., 2012. GMES Sentinel-1 mission. Remote Sens. Environ. 120, 9–24.

Tsendbazar, N.E., De Bruin, S., Herold, M., 2015. Assessing global land cover reference datasets for different user communities. ISPRS J. Photogramm. Remote Sens. 103, 93–114.

Tsendbazar, N.E., de Bruin, S., Mora, B., Schouten, L., Herold, M., 2016. Comparative assessment of thematic accuracy of GLC maps for specific applications using existing reference data. Int. J. Appl. Earth Obs. Geoinf. 44, 124–135. https://doi.org/10.1016/j.jag.2015.08.009.

Tsendbazar, N.-E., Herold, M., de Bruin, S., Lesiv, M., Fritz, S., Van De Kerchove, R., Buchhorn, M., Duerauer, M., Szantoi, Z., Pekel, J.-F., 2018. Developing and applying a multi-purpose land cover validation dataset for Africa. Remote Sens. Environ. 219, 298–309. https://doi.org/10.1016/j.rse.2018.10.025.

Tsendbazar, N., Herold, M., Li, L., Tarko, A., de Bruin, S., Masiliunas, D., Lesiv, M., Fritz, S., Buchhorn, M., Smets, B., Van De Kerchove, R., Duerauer, M., 2021. Towards operational validation of annual global land cover maps. Remote Sens. Environ. 266, 112686 https://doi.org/10.1016/j.rse.2021.112686.

Venter, Z.S., Sydenham, M.A.K., 2021. Continental-scale land cover mapping at 10 m resolution over Europe (ELC10). Remote Sens. 13, 2301. https://doi.org/10.3390/rs13122301.

Venter, Z.S., Barton, D.N., Chakraborty, T., Simensen, T., Singh, G., 2022. Global 10 m land use land cover datasets: a comparison of dynamic world, world cover and Esri land cover. Remote Sens. 14 https://doi.org/10.3390/rs14164101.

Wang, X., Hu, Z., Shi, S., Hou, M., Xu, L., Zhang, X., 2023a. A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet. Sci. Rep. 13, 7600. https://doi.org/10.1038/s41598-023-34379-2.

Wang, Z., Mountrakis, G., 2023. Accuracy assessment of eleven medium resolution global and regional land cover land use products: a case study over the conterminous United States. Remote Sens. 15, 3186.

Wang, Y., Sun, Y., Cao, X., Wang, Yihan, Zhang, W., Cheng, X., 2023b. A review of regional and global scale land use/land cover (LULC) mapping products generated from satellite remote sensing. ISPRS J. Photogramm. Remote Sens. 206, 311–334. https://doi.org/10.1016/j.isprsjprs.2023.11.014.

Wickham, J., Stehman, S.V., Sorenson, D.G., Gass, L., Dewitz, J.A., 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sens. Environ. 257, 112357 https://doi.org/10.1016/j.rse.2021.112357.

Xu, P., Tsendbazar, N.E., Herold, M., Clevers, J.G.P.W., Li, L., 2022. Improving the characterization of global aquatic land cover types using multi-source earth observation data. Remote Sens. Environ. 278, 113103 https://doi.org/10.1016/J.RSE.2022.113103.

Yu, L., Du, Z., Dong, R., Zheng, J., Tu, Y., Chen, X., Hao, P., Zhong, B., Peng, D., Zhao, J., Li, X., Yang, J., Fu, H., Yang, G., Gong, P., 2022. FROM-GLC plus: toward near real-time and multi-resolution land cover mapping. GIsci Remote Sens 59, 1026–1047. https://doi.org/10.1080/15481603.2022.2096184.

Yue, H., Qing, L., Zhang, Z., Wang, Z., Guo, L., Peng, Y., 2024. MSE-net: a novel master–slave encoding network for remote sensing scene classification. Eng. Appl. Artif. Intell. 132, 107909 https://doi.org/10.1016/j.engappai.2024.107909.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., 2022. ESA WorldCover 10 m 2021 v200.

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.