Remote Sensing of Environment 266 (2021) 112686

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Towards operational validation of annual global land cover maps

Tsendbazar N.[a,*], M. Herold[a], L. Li[a], A. Tarko[a], S. de Bruin[a], D. Masiliunas[a], M. Lesiv[b], S. Fritz[b], M. Buchhorn[c], B. Smets[c], R. Van De Kerchove[c], M. Duerauer[b]

[a] *Wageningen University & Research, Laboratory of Geo-Information Science and Remote Sensing, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands*
[b] *International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria*
[c] *Flemish Institute for Technological Research (VITO), Mol, Belgium*

A R T I C L E I N F O

Editor: Marie Weiss

Keywords:
Operational validation
Global land cover monitoring
Validation data update
Stability in product accuracy

A B S T R A C T

Annual global land cover maps (GLC) are being provided by several operational monitoring efforts. However, map validation is lagging, in the sense that the annual land cover maps are often not validated. Concurrently, users such as the climate and land management community require information on the temporal consistency of multi-date GLC maps and stability in their accuracy. In this study, we propose a framework for operational validation of annual global land cover maps using efficient means for updating validation datasets that allow timely map validation according to recommendations in the CEOS Stage-4 validation guidelines. The framework includes a regular update of a validation dataset and continuous map validation. For the regular update of a validation dataset, a partial revision of the validation dataset based on random and targeted rechecking (areas with a high probability of change) is proposed followed by additional validation data collection. For continuous map validation, an accuracy assessment of each map release is proposed including an assessment of stability in map accuracy addressing the user needs on the temporal consistency information of GLC map and map quality. This validation approach was applied to the validation of the Copernicus Global Land Service GLC product (CGLS-LC100). The CGLS-LC100 global validation dataset was updated from 2015 to 2019. The update was done through a partial revision of the validation dataset and an additional collection of sample validation sites. From the global validation dataset, a total of 40% (10% for each update year) was revisited, supplemented by a targeted revision focusing on validation sample locations that were identified as possibly changed using the BFAST time series algorithm. Additionally, 6720 sample sites were collected to represent possible land cover change areas within 2015 and 2019. Through this updating mechanism, we increased the sampling intensity of validation sample sites in possible land cover change areas within the period. Next, the dataset was used to validate the annual GLC maps of the CGLS-LC100 product for 2015–2019. The results showed that the CGLS-LC100 annual GLC maps have about 80% overall accuracy showing high temporal consistency in general. In terms of stability in class accuracy, herbaceous wetland class showed to be the least stable over the period. As more operational land cover monitoring efforts are upcoming, we emphasize the importance of updated map validation and recommend improving the current validation practices and guidelines towards operational map validation so that long-term land cover maps and their uncertainty information are well understood and properly used.

1. Introduction

Land cover represents important biophysical properties of the earth's surface. Changes in land cover can have a significant impact on the earth's ecological and biogeochemical processes. Due to its importance, land cover is regarded as one of the important terrestrial variables for monitoring. It has been selected as one of the Essential Climate Variables (ECV) as part of Climate Change Initiatives (CCI) by the European Space Agency (GCOS, 2010) and it is included in the variables for monitoring by the Copernicus Global Land Service (CGLS) and Copernicus Climate Change Service (C3S) (Buchhorn et al., 2020b; Defourny, 2019). Land cover and its change are directly related to implementing and monitoring Sustainable Development Goal (SDG) indicators such as change in the extent of water-related ecosystems, land consumption rate, and

* Corresponding author.
E-mail address: nandin.tsendbazar@wur.nl (N. Tsendbazar).

https://doi.org/10.1016/j.rse.2021.112686
Received 20 May 2021; Received in revised form 30 August 2021; Accepted 1 September 2021
Available online 8 September 2021

proportion of land that is degraded (Romijn et al., 2016; UN-GGIM, 2019).

Owing to continued interests in land cover monitoring, global land cover (GLC) mapping efforts have seen accelerated progress over the last three decades. Since the first satellite-based GLC map was produced in 1994 (DeFries and Townshend, 1994), several GLC maps have been produced at different resolution and temporal update frequencies. Both the IGBP (International Geosphere–Biosphere Programme) map and GLC2000 map cover a single year at 1 km pixel size (Bartholomé and Belward, 2005; Loveland et al. 2000). In contrast, maps such as the Globcover, CCI-Land Cover, and MODIS (Moderate Resolution Imaging Spectroradiometer) encompass two or more epochs at medium resolution (300–500 m) (CCI Land Cover, 2017; Defourny et al., 2011; Sulla-Menashe et al., 2019). Recent advances in satellite data acquisition and processing capabilities have led to the release of GLC maps at a higher resolution based on Landsat and Sentinel data (Chen et al., 2015; Gong et al., 2013). The first GLC map at 10 m was produced based on Sentinel-2 imagery (Gong et al. 2019). More GLC maps at 10 m are being produced, for example by the ESA's WorldCover project, which aims to release a GLC map using Sentinel-1 and -2 data (ESA-WorldCover, 2020).

One of the important user requirements for GLC mapping is continuity, stemming from the need for continuous monitoring for ongoing global issues, such as climate change, sustainable land management, and SDG implementation monitoring. As such, several mapping efforts provide operational and continuous land cover products. For example, the CCI-Land Cover product offers annual GLC maps between 1992 and 2015 at 300 m pixel size (CCI Land Cover, 2017). As a continuation, 2016–2018 maps were produced as part of the C3S (Defourny 2019). The CGLS also provides operational land cover monitoring at a global scale with the release of annual GLC maps for 2015–2019 containing discrete and fractional land cover layers (CGLS-LC100m) (Buchhorn et al., 2020a). Users interested in operational land cover monitoring emphasize the importance of temporal consistency or stability in land cover characterization over time (Bontemps et al., 2011). Due to classifier uncertainty, noise in the input data, and disturbances such as fire and droughts, multi-temporal maps can be affected by year-to-year variability in the classification results, and this could lead to erroneous detection of land cover changes (Friedl et al., 2010; Sulla-Menashe et al. 2019). The temporal consistency of multi-year land products and stability in their accuracy are therefore highlighted as key requirements for monitoring land cover as an ECV for Global Climate Observing Systems (GCOS) (GCOS, 2011). Particularly, for the climate modelling community, which requires long-term and consistent land cover observations, the stability in product accuracy is an important requirement (Bontemps et al., 2011). Stability assessment of product accuracies has been previously investigated for time series of burnt area products (Padilla et al., 2014). However, stability in product accuracy is not reported for current GLC products, and the methods to assess stability are also lacking.

Evaluating the stability in the product accuracy for continuous GLC monitoring requires an operational validation framework to assess the accuracy of each release or each annual map. However, operational validation lags behind the product generations in GLC monitoring efforts, because GLC products often do not include validation of the annual land cover maps. For instance, the commonly used CCI-Land Cover and MODIS Collection 5 GLC maps do not provide accuracy estimates for each annual GLC map, having the validation limited to a single or limited period (CCI Land Cover, 2017; Friedl et al., 2010).

At the same time, quantitative validation of satellite-based land products is recommended to follow standardized guidelines to provide detailed uncertainty information and to allow product inter-comparison (Strahler et al., 2006). Land product validation guidelines, which also include land cover, were set up by the Land Product Validation (LPV) subgroup of the Committee on Earth Observing Satellites (CEOS) (LPVS-CEOS 2000). Accordingly, most GLC maps are validated following the CEOS-LPV Stage 3 validation guidelines, using statistically rigorous accuracy assessment methods based on good practices (Strahler et al., 2006; Xu et al., 2020). However, current operational GLC mapping does not meet the CEOS-LPV Stage-4 validation guidelines, which recommend a systematic updating of validation results for each new release or time series expansion (LPVS-CEOS 2000). Therefore, to support users' confidence in the continued use of GLC products, operational land monitoring efforts need to expand their product validation into operational validation by updating product uncertainty and consistency information.

Validation datasets used for GLC map assessments are recommended to be based on a probability sampling design which allows unbiased estimation of map accuracy (Mayaux et al., 2006; Olofsson et al., 2014). Probability sampling based validation datasets can also be used for area estimation of land cover classes and the precision of the area estimation can be improved using the mapped classes for stratification (Gallego, 2004). A probability sampling design is therefore followed for the validation of many previous GLC maps (Tsendbazar et al., 2015). The validation datasets are often created through visual interpretation of very high-resolution (VHR) images (Pengra et al., 2020; Tarko et al., 2020). Considering that human interpretation is costly and time-consuming, validation datasets should be designed to be easily adjustable for timely and continuous validation of new releases of land cover products and yet maintaining statistical rigor. Limited research has been done in designing and generating a validation dataset suitable for operational land cover monitoring. At a national scale, Pengra et al. (2020) used a validation design based on simple random sampling to assess a land cover product over the last 30 years by interpreting annual reference land cover between 1984 and 2016. This dataset was used to validate the LCMAP (Land Change Monitoring, Assessment, and Projection) operational land cover monitoring of the United States of America (Brown et al., 2020). Operational and regularly updated validation practices are currently limited for global-scale land cover monitoring.

Multi-temporal or annual land cover products are sometimes used to assess and/or estimate land cover change areas (Li et al., 2018) as an alternative way to direct land cover change monitoring (Woodcock et al., 2020). However, land cover change detection based on post-classification map differencing is not a trivial exercise as any misclassification errors of land cover maps may have a significant impact on land cover change monitoring (Colditz et al., 2012). Therefore the accuracy of land cover change detection needs to be assessed as a separate step. Furthermore, estimation of land cover change areas can be improved by making use of the mapped classes and a statistical sample (Olofsson et al., 2014). These steps require extra attention, in addition to the accuracy assessment of annual or multitemporal maps. The focus of this study is on assessing the accuracy of annual GLC maps.

This study presents a framework for operational validation of continuous GLC monitoring and proposes metrics to assess the stability of the accuracy of annual GLC maps. The work is based on the efforts for validating the annual CGLS-LC100 product (Tsendbazar et al., 2020). Tsendbazar et al. (2018) introduced a multi-purpose validation dataset (CGLS-LC100 validation dataset) for Africa based on a design that is suitable for assessing maps with different resolutions (10–100 m). Building upon this dataset, we developed an operational validation protocol suitable for continuous land cover monitoring at a global scale, herein considering cost-effectiveness and timeliness. After implementing it, we assessed the accuracy and the stability of accuracy of the annual CGLS-LC100 product over the period of 2015–2019.

2. Methods

2.1. A framework for operational validation for continuous land cover monitoring

Validating continuous land cover monitoring products requires an up-to-date statistically rigorous validation dataset. This implies validation datasets that need regular updating. To allow map validation without much time lag after the land cover product release, the validation dataset updating needs to be cost-effective without compromising the statistical rigor. Fig. 1 shows the schematic overview of the operational validation framework consisting of two main parts, namely, regular update of the validation dataset and continuous map validation.

A Regular update of a validation dataset

A1 Validation dataset

We base the framework for regular updates of a validation dataset on a stratified random sampling scheme (Fig. 1-A1). This sampling scheme allows unbiased estimation of map accuracy metrics with known associated variance (Olofsson et al., 2014; Stehman, 2009b) and it is commonly used for validation dataset generation for land cover monitoring (Tsendbazar et al., 2015). Furthermore, a stratified random sampling design enables augmenting a validation sample to address specific regions or classes of interest (Stehman et al., 2012).

A2 Revisiting sample sites.

A validation dataset, based on stratified sampling, is collected for a certain reference year (T0). To update the validation dataset to a later period or year (T1), the validity of the reference land cover labels needs to be checked due to possible changes in land cover since the reference year. Since the manual revisiting and interpretation efforts are very costly and time-consuming (Pengra et al., 2020; Tarko et al., 2020), we propose a partial revisit of the validation dataset consisting of a targeted revisit and a random revisit of the validation dataset (Fig. 1-A2). Firstly, limited efforts are targeted to sample sites that have a high possibility of land cover change occurrence since the reference year (T0). Here, a time series change detection algorithm could be conducted to identify sites that are unstable in terms of surface reflectance or vegetation indices, thus having a high possibility of land cover change. Time series change detection algorithms such as Breaks For Additive Season and Trend (BFAST) and Continuous Change Detection and Classification (CCDC) (Verbesselt et al., 2012; Zhu and Woodcock, 2014) are run at validation site locations. Secondly, depending on the available resources, a random subset of the validation dataset is rechecked for possible land cover changes. Such a random revisit is particularly useful to assess whether time series algorithms are omitting any occurrence of the land cover change. Finally, the whole validation dataset is revisited after a certain period (e.g., 5 years) to maintain the quality of the dataset.

A3 Additional sample sites for change areas.

To reduce the time lag between product validation and product

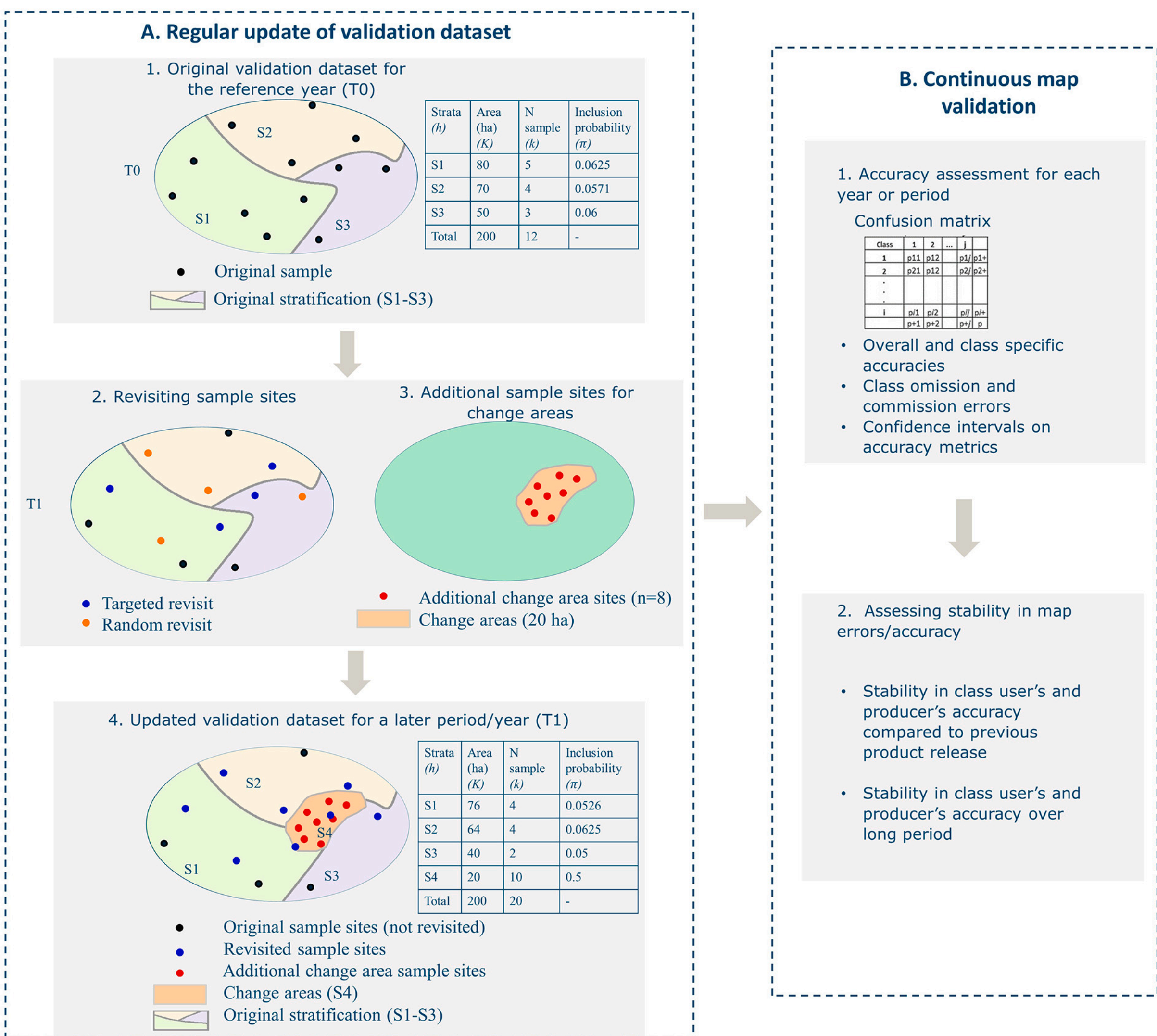

Fig. 1. Framework for operational validation for land cover monitoring consisting of regular update of validation data (A) and continuous map validation (B).

generation, the revisiting step (Fig. 1-A2) is conducted in parallel to the continuous GLC map production as it is independent of the product generation. Once the revisiting is done, the updated validation dataset is used to assess the accuracy of the updated GLC maps. However, if there is an interest in gaining insights on land cover transitions of the operational GLC monitoring product, additional validation sample sites can be collected to increase the sampling intensity of validation sample sites in land cover change areas, identified by the operational GLC monitoring (Fig. 1-A3). As demonstrated by Stehman et al. (2012), a stratified random sampling design is suitable for augmenting the sample size.

A4 Updated validation dataset.

The original validation dataset and the additional validation sample sites are combined to validate updated land cover products for an update year/period (T1) (Fig. 1-A4). To do so, the inclusion probability of a given site being included in the sample needs to be known and must be greater than zero, for both the original sample sites and additional sample sites (Stehman 2000; Stehman et al., 2012). For probability sampling, the inclusion probability (π_h) for sample sites in stratum h is defined as follows:

$$\pi_h = k_h / K_h \tag{1}$$

where k_h is the number of sample sites in stratum h, and K_h is the population size for stratum h (Pengra et al., 2015).

The original stratification is modified by imprinting the change stratum. Here the change stratum is cut out of the original stratification and the areas of some overlapping original strata are reduced. The change stratum overrules the original stratification without producing second level substrata. This was chosen as a practical approach and it allows to give more focus to the change stratum and where changes are likely to have occurred.

The updated validation data (original and additional sample sites) are overlaid on the modified stratification. Subsequently, the area of each stratum is calculated and the inclusion probability is calculated based on the number of sample sites in the modified stratification. For example, as shown in Fig. 1-A1, stratum S1 initially had 5 sample sites for an area of 80 ha. The inclusion probability (π) for this stratum was 0.0625. After adding the change area stratum (Fig. 1-A4), the area of this stratum reduced to 76 ha, and the number of sample sites for this stratum became 4, which resulted in an inclusion probability of 0.0526. Our approach also implies that if an original sample site falls in the added change area stratum, it is considered to be part of the added change stratum regardless of its original stratification(S4). As a result of the revisiting and addition of sample sites, a total of 20 sample sites (including the un-revisited sites) is used for assessing follow-up releases and updates.

For further updates or releases (T2, T3 …), the same procedure is applied (Fig. 1-A4). Here, the additional validation sites collected in Fig. 1-A3 for the previous update (e.g., T1) are not used in validating further releases (e.g., T2). For further updates, the up-to-date original validation dataset (Fig. 1-A4) without the added sites is considered the starting point, and the same procedure is followed to update the dataset further.

B Continuous map validation

B1 Assessing the accuracy of product updates.

Once the validation dataset is updated, the accuracy of annual product releases and their precision are assessed using suitable estimators. Since we employ a stratified design, a stratified estimator is used (Tsendbazar et al., 2018). The map product accuracy (overall and class-specific accuracies) is then calculated taking un-equal inclusion probability between the sample sites into consideration (Fig. 1-B1). Since the change stratum is added to the stratification, the commonly used stratified estimator by Olofsson et al. (2014) and Card (1982) are not suitable as the strata and the mapped classes are not the same. Instead, a stratified estimator by Stehman (2014) is used which is suitable when the strata are different from the map classes. According to Stehman (2014), the overall accuracy ($\overline{Y}$) and its variance $\widehat{V}\left(\widehat{\overline{Y}}\right)$ are calculated as follows:

$$\widehat{\overline{Y}} = \sum_{h=1}^{H} N_h^* \overline{y_h} \Big/ N \tag{2}$$

Where $\overline{y_h} = \sum_{u \in h} \frac{y_u}{n_h^*}$ is the sample mean of the y_u (y_u is 1 if mapped and reference class matches, otherwise 0 at sample site u) in stratum h, $u \in h$ indicates that sample site u was selected from stratum h, H denotes the number of strata, N is the number of all possible sample sites in the population.

$$\widehat{V}\left(\widehat{\overline{Y}}\right) = \left(\frac{1}{N^2}\right) \sum_{h=1}^{H} N_h^{*2} \left(1 - \frac{n_h^*}{N_h^*}\right) s_{yh}^2 \Big/ n_h^* \tag{3}$$

where n_h* is sample sites selected from the N_h* possible sample sizes in stratum h in the population and the sample variance of the $y_u$values from stratum h is:

$$s_{yh}^2 = \sum_{u \in h} \left(y_u - \overline{y_h}\right)^2 \Big/ \left(n_h^* - 1\right) \tag{4}$$

Class accuracies (user's and producer's) and their variance are calculated as specified in Stehman (2014).

B2 Assessing the stability in product accuracies.

As temporal consistency in multi-year land cover products and stability in their accuracy are important requirements for users such as climate modelers (Bontemps et al., 2011; GCOS, 2011), we propose to include the assessment of stability in product accuracy in the operational validation framework (Fig. 1-B2). According to the climate users (GCOS, 2011), stability in product accuracy is defined as "the extent to which the error of a product remains constant over a longer period". A maximum of 15% of omission and commission error in mapping individual land cover classes together with the stability of 15% is targeted (GCOS, 2011). As stability in the accuracy of land cover products has not been assessed previously, no clear methodologies are available to assess the stability in a descriptive way relating to this GCOS requirement. We propose to estimate the stability in product accuracy by calculating an index from class accuracies at different moments in time.

The stability index for class accuracy $SI_{ca}$is proposed to be expressed as:

$$SI_{ca(t1)} = \frac{|ca_{t1} - ca_{t1-1}|}{ca_{t1-1}} * 100 \tag{5}$$

Where, $SI_{ca(t1)}$ is the stability index for a class accuracy (user's or producer's accuracy) for time $t1$, ca_{t1} is a class accuracy for time $t1$, ca_{t1-1} is a class accuracy for the previous time (t0 or reference year). A low index value indicates that the class accuracy is stable, while higher index values denote the opposite. Eq. (5) is used to calculate the stability index of both the user's and producer's accuracy. The equation could also be adapted to calculate stability index in terms of class errors (omission and commission), by changing the notation to SI_{ce} and ce instead of ca. The omission error is 100 - producer's accuracy, while commission error equals 100 - user's accuracy. With class errors, the denominator in Eq. (5) can be modified to (100-ce_{t1-1}) where ce_{t1-1} indicating the errors of omission or commission for the previous period. This adjustment results in the same stability index values as when class accuracy is used. In this study, we calculated the stability index based on class accuracies.

After assessing the stability in class accuracy between two consecutive years or periods, the stability over a long period (SI_{ol}) is defined as:

$$\mathrm{SI}_{\mathrm{ol}} = max\left(\mathrm{SI}_{\mathrm{ce(t1)}}, \mathrm{SI}_{\mathrm{ce(t2)}}, \mathrm{SI}_{\mathrm{ce(t3)}} \ldots \mathrm{SI}_{\mathrm{ce(tn)}}\right) \tag{6}$$

Taking a maximum of the stability index to estimate stability in map accuracy may be considered a conservative approach and this could be modified to use "mean" or "median", depending on the needs and requirements.

The following subsections describe, how the proposed design was used as part of operational validation of the CGLS-LC100 product.

2.2. Multi-purpose land cover validation data at global scale

The CGLS-LC100 validation dataset builds upon Tsendbazar et al. (2018) which describes the design of the multi-purpose land cover validation dataset for the extent of Africa. Here we describe the dataset expanded to a global scale.

2.2.1. Sampling design

The CGLS-LC100 validation dataset is based on a stratified sample using the same stratification used in Tsendbazar et al. (2018). The stratification is based on the Köppen climate zones and human population density (Olofsson et al., 2012). Here, each climate zone is divided into unpopulated and populated parts (more than 5 persons/km^2). The sample size for each continent was 2700, consistent with the sample size for Africa (Tsendbazar et al., 2018). Asia was divided into 2 sub-continents due to its large size: Northern Eurasia and the rest of Asia. Northern Eurasia includes Russia (both European and Asian parts), Kazakhstan, and Mongolia, while the rest of Asia is referred to as Asia. Including the previously collected sample sites of Africa, a total of 20,019 sample sites were selected across the world. A similar approach as the African validation dataset was used to allocate the sample sites (Tsendbazar et al., 2018).

To increase the sampling intensity in rare land cover types such as wetland, urban, and water, additional sample sites were added for these land cover types. We selected at least 100 sample sites per land cover type for each continent. The extra sample sites were selected based on the CGLS-LC100 V2.0 discrete land cover map (Buchhorn et al., 2020b). The areas of the rare land cover types were added to the original stratification of Olofsson et al. (2012) as additional strata, similar to the adjustments made for African stratification (Tsendbazar et al., 2018). The global stratification consisted of 149 strata in total, divided over seven (sub)continents each having 19–25 strata. The validation dataset contains a total of 21,752 randomly selected sample sites.

2.2.2. Response design

To keep the global validation dataset suitable for multiple purposes, each sample unit (100 m × 100 m) was divided into 100 small blocks/subpixels (10 m × 10 m) (Tsendbazar et al., 2018). The subpixels are aligned to individual Sentinel-2 L1C pixels (Buchhorn et al., 2020b). In each subpixel, the dominant land cover elements were labelled. The land cover elements are trees with different leaf and phenology types, shrubs, grass, crops, built-up areas, water, snow/ice, and lichen/moss (Tsendbazar et al., 2018). Also, regularly flooded areas were marked.

The initial collection of the global validation data was done on a web-interface (Tsendbazar et al., 2018) based on the Geo-Wiki platform. Data were collected between February and August 2018 followed by revision and quality checking processes. Regional experts, who have experience working with satellite-based land cover analysis and image interpretation, visually interpreted the reference land cover for the year 2015 at validation sample sites. Including the regional experts of African data collection, in total 30 regional experts contributed to this process and their efforts were financially compensated. Regional experts and regions they worked on are provided in Fig. S1 in the Supplementary Materials.

The global validation data collection applied several steps to ensure good quality land cover reference data for validation (Fig. 2). A remote training tutorial was conducted to familiarize the experts with the interface and land cover type definitions. As visual interpretation can be subjected to interpreter variability and bias (Strahler et al., 2006), feedback was given in loops of interpretations (Tarko et al., 2020). Here, experts received feedback on their work after they interpreted 10–20, 50, and 100 validation locations and the rest of the validation locations. Experts continued to the next loop when they had resolved the feedback on the previous loop received from validation experts (authors affiliated to Wageningen University). Feedback was given for each sample location. The experts either rebutted the feedback or corrected the interpretations where necessary. Results of the feedback loop and correction rates with the loops are detailed in Tarko et al. (2020).

Next, quality checks and consolidation steps were also conducted to select the revised labels. Here, we compared the reference land cover labels of the validation dataset with land cover products such as Northern American Land Cover product (Latifovic et al., 2004), CORINE (Bossard et al. 2000), Australian Dynamic Land Cover (Geoscience Australia, 2010), and Circumpolar Arctic Vegetation Map (Walker et al., 2005). We rechecked any validation sites which did not match with these datasets for confirmation and consolidation. In total, approximately 32% of the total 21,752 sample sites were updated partially or fully as part of feedback loops and consolidation efforts. The results of the data collection were land cover labels at 10 m resolution subpixels and land cover fraction information of generic land cover elements at 100 m resolution at the sample site locations for the reference year of 2015. At 100 m resolution, the fraction information was then translated into the CGLS-LC100 discrete map legend following the approach used in Tsendbazar et al. (2018).

2.2.3. Accuracy estimation protocol

Since the sample design is the same as that for African extent, a similar accuracy estimation protocol reported in Tsendbazar et al. (2018) could be adopted at a global scale. At 100 m resolution, a stratified estimator by Stehman (2014) is used (Eqs. (2)–(4)). While, at a finer resolution (< 100 m), if multiple sample units are selected within the 100 × 100 m area, a stratified one-stage cluster approach could be applied (Pengra et al., 2015). Furthermore, the land cover fraction maps could also be assessed using this validation data (Tsendbazar et al., 2018).

2.3. Updating the validation dataset for assessing annual land cover maps

We updated the global land cover validation dataset for 2015, described in the previous section, to the subsequent years, namely 2016 to 2019. Since the updates for these years were done at the same time, we considered this as one update. Following the proposed operational validation framework, updates focused on two main parts, i) revisiting and ii) collecting additional validation data in possible change areas (Fig. 1-A).

2.3.1. Revisiting the CGLS-LC100m validation dataset

The revisiting of the CGLS-LC100m validation dataset was done for randomly selected and targeted sample sites following the proposed framework (Fig. 1-A). The revision aimed to confirm the land cover labels for the reference year 2015 and update the labels in case of a change in land cover. A schematic overview of revisiting the CGLS-LC100 validation dataset is provided in Fig. 3.

Firstly, randomly selected 40% of the total sample sites were revisited for each continent over the update period (2016–2019). Since, we wanted to update the datasets through four years (2016, 2017, 2018, and 2019), a 10% random revisit for each year was deemed appropriate given the available resources. Secondly, sample sites that had a high possibility of land cover change occurrence were targeted. Here, we identified sample sites that were tagged as "unstable" or "with breaks" in

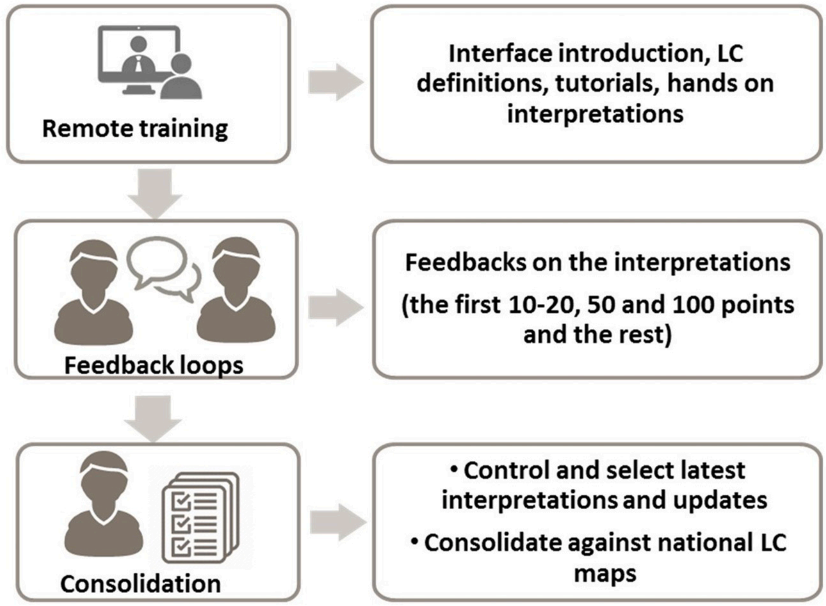

Fig. 2. Validation data collection and feefdback process.

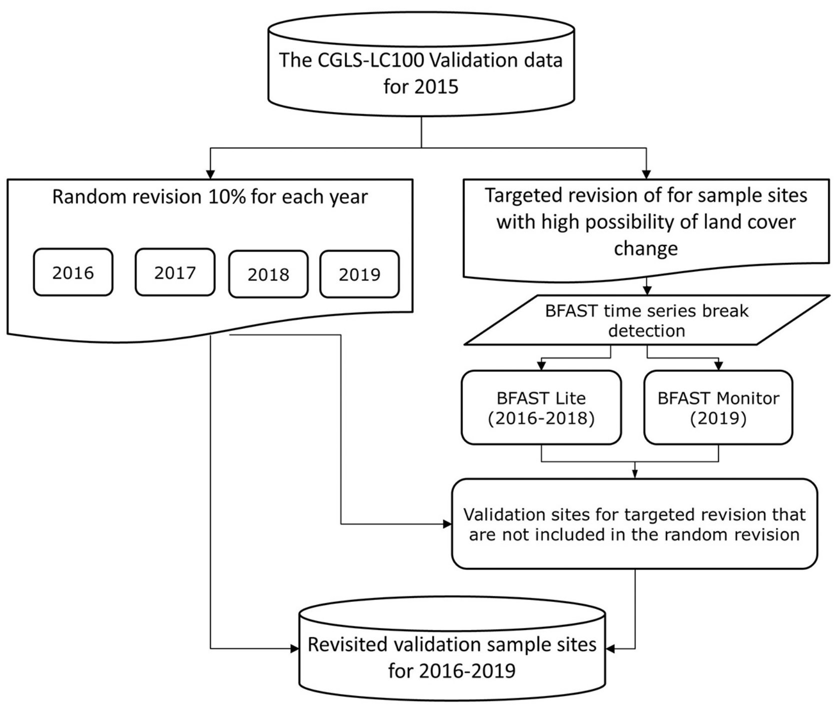

Fig. 3. A schematic overview of the revisiting done for the CGLS-LC100 validation dataset for 2016–2019.

terms of long-term time-series satellite data. Breaks detected during the update period may indicate possible land cover change. BFAST-Lite and BFAST Monitor change detection algorithms were run on MODIS NIRv (Near Infrared Reflectance of Vegetation) time series from 2009 until 2020 at 300 m pixel size (Buchhorn et al., 2020a; Masiliunas et al., 2021; Verbesselt et al., 2010). The BFAST-Lite algorithm was used for the first three years (2016–2018), and the BFAST Monitor algorithm for the last year (2019). The latter was specifically designed for detecting breaks at the end of the time series (Verbesselt et al., 2010). We created yearly break maps between 2016 and 2019 for the globe and identified sample sites that were within the break areas. The algorithm could also be run at sample locations. As it could be possible to have breaks for multiple years for some locations (Masiliunas et al., 2021), we selected sample locations with at least one break during the period. In total, around 1300 sample locations (6% of the total sample sites) were identified to be unstable or “with breaks”. From these points, some sample sites were already included in the 40% random revisits. Therefore additional 3.5% of sample sites were revisited, making the total revisited sites 9465 (43.5% of the total sites). The revision was done by experts who worked on the generation of the CGLS-LC100 validation dataset for 2015. Besides previously available images on the web interface, experts could also consult VHR images that were purchased from the Digital Globe repository for visual interpretation purposes. Each image covered a 500 × 500 m area centred on the sample sites. Where available, at least one image was displayed for each year (2016–2019).

As a result of the revision, for the reference year 2015, 607 sample sites had at least one subpixel updated, and at 204 out of those, the updates were substantial enough to modify the discrete map legend at 100 m level. The modifications of the reference land cover labels for the reference year 2015 were caused by the availability of the new VHR images and extra images made available in the background layers such as Bing and ESRI images. The modifications were made mostly in regions where satellite image coverage is scarce, namely the Siberia region.

2.3.2. Collecting additional validation sample sites in possible change areas

We collected additional validation sites to increase the sample size in possible change areas since 2015 (Fig. 1-A3).

2.3.2.1. Additional sample site selection. We identified possible land cover change areas to use as a stratification for the additional sample site selection. The possible land cover change stratification was created for each pair of years for the update period (2015–2016, 2016–2017, 2017–2018, and 2018–2019). First, based on the post-classification map differencing of the annual CGLS-LC100 V3.0 land cover maps (Buchhorn et al., 2020a), possible areas of change were identified. Second, to reduce spurious changes due to possible classification errors, the change area was further refined using the global break maps for each year (2016–2019). Any possible change areas outside the break masks were removed. Third, the possible change areas were limited to land cover change transitions that are deemed probable within the period (See Table S1 in the Supplementary Material). Lastly, to remove any spurious changes caused by salt and pepper pixel effects, and to be consistent with MODIS data on which the break maps were generated, all possible change areas that were less than 3 ha (3 adjacent pixels) were also removed.

Next, for each pair of years (e.g. 2015–2016), and for each continent, 240 sample sites were randomly selected based on the possible change stratification. A total of 1680 for each pair of years (e.g., 2015–2016, 2016–2017) and 6720 sample sites for the update period (2015–2019) were selected. Sample allocation was based on the land cover type after a change, so each set of 240 sample sites was evenly distributed between the land cover type of the latter year (e.g., land cover class of 2016 in case of pair 2015–2016) within the possible change stratification. For each continent, there were 7–9 land cover types after a change for each pair year, which resulted in 31–35 strata for four pair years. By combining the seven (sub)continents, a total of 230 strata was included in the possible change stratification.

The original CGLS-LC100 validation dataset was then combined with the additional sample sites. Similarly, the original stratification was modified by imprinting the change stratification. The modified stratification included 379 (149 + 230) strata. As this affects the original stratification, the areas of all strata were recalculated. Sample inclusion probability and design weights of the validation sites were calculated using the area of the strata (the modified stratification) and the number of sample sites per stratum (Eq. (1)).

Fig. 4 shows the spatial distribution of the additional validation sample sites, including the original validation sample sites.

2.3.2.2. Reference data collection. The additional validation sample sites were visually interpreted by experienced experts involved in the global validation data collection (Table S3, in the Supplementary Material). There were 960 sample sites for each (sub)continent. The experts used the same validation web-interface, based on the Geo-Wiki platform. Experts received feedback on their first 10–20 collected sample sites. Online assistance and revision of at least 10% of collected sample sites were provided by validation experts from Wageningen University. The collection and revision lasted from May to July 2020.

At each sample unit, the land cover elements were labelled at 10 × 10 m subpixel level for each reference year 2015–2019. In the case of sample sites where all subpixels had the same land cover for this period, the reference land cover was labelled once. In the case of sample sites where land cover for at least one subpixel was changed, the land cover was labelled for each reference year in the period 2015–2019 accordingly (five times). Experts collected information on land cover interpretation confidence for each sample site. Fig. 5 shows screenshots of an exemplary sample site with interpreted land cover for each year.

Digital Globe VHR images, different background layers, Sentinel-2 time series thumbnails, along with time series NDVI profiles displayed for visual interpretation in the validation data collection interface (Section 2.2), played an important role in visually interpreting land cover change. A screenshot of false-colour combinations of Sentinel-2 time series data on the validation data collection interface is shown in Fig. 6.

Finally, the additional validation sample sites targeting potential change areas and the revisited version of the validation dataset were combined (Fig. 1-A4) to come up with a validation dataset (total of 28,321 sites) for each reference year in the update period, 2016 to 2019.

2.4. Assessing the accuracy of annual GLC maps

The third version (V3.0) of the Dynamic Land Cover product at 100 m (CGLS-LC100m) was released recently by the Copernicus Global Land

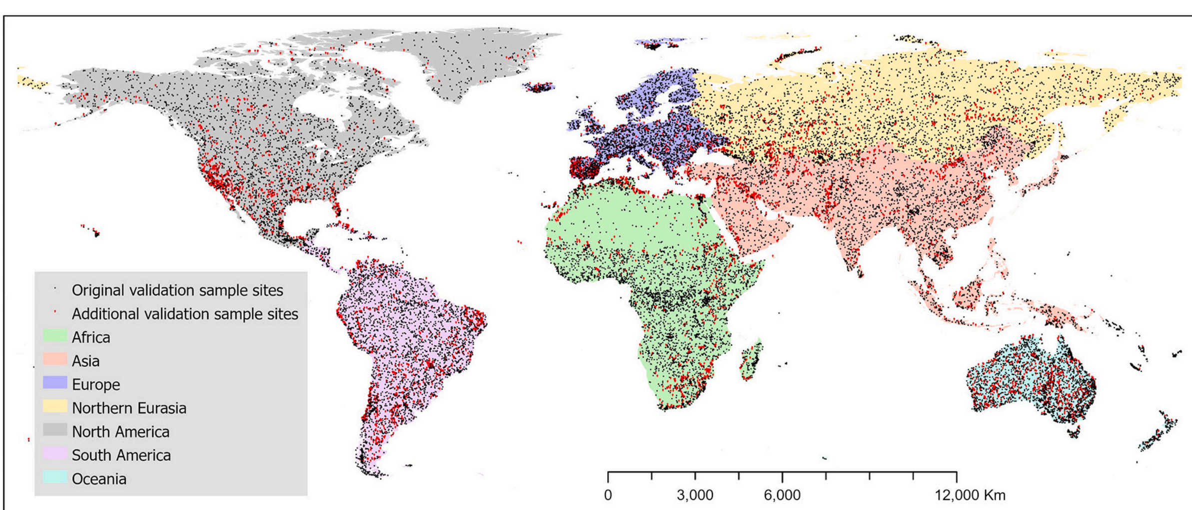

Fig. 4. (Sub) continental distribution of the validation sample sites (original and additional).

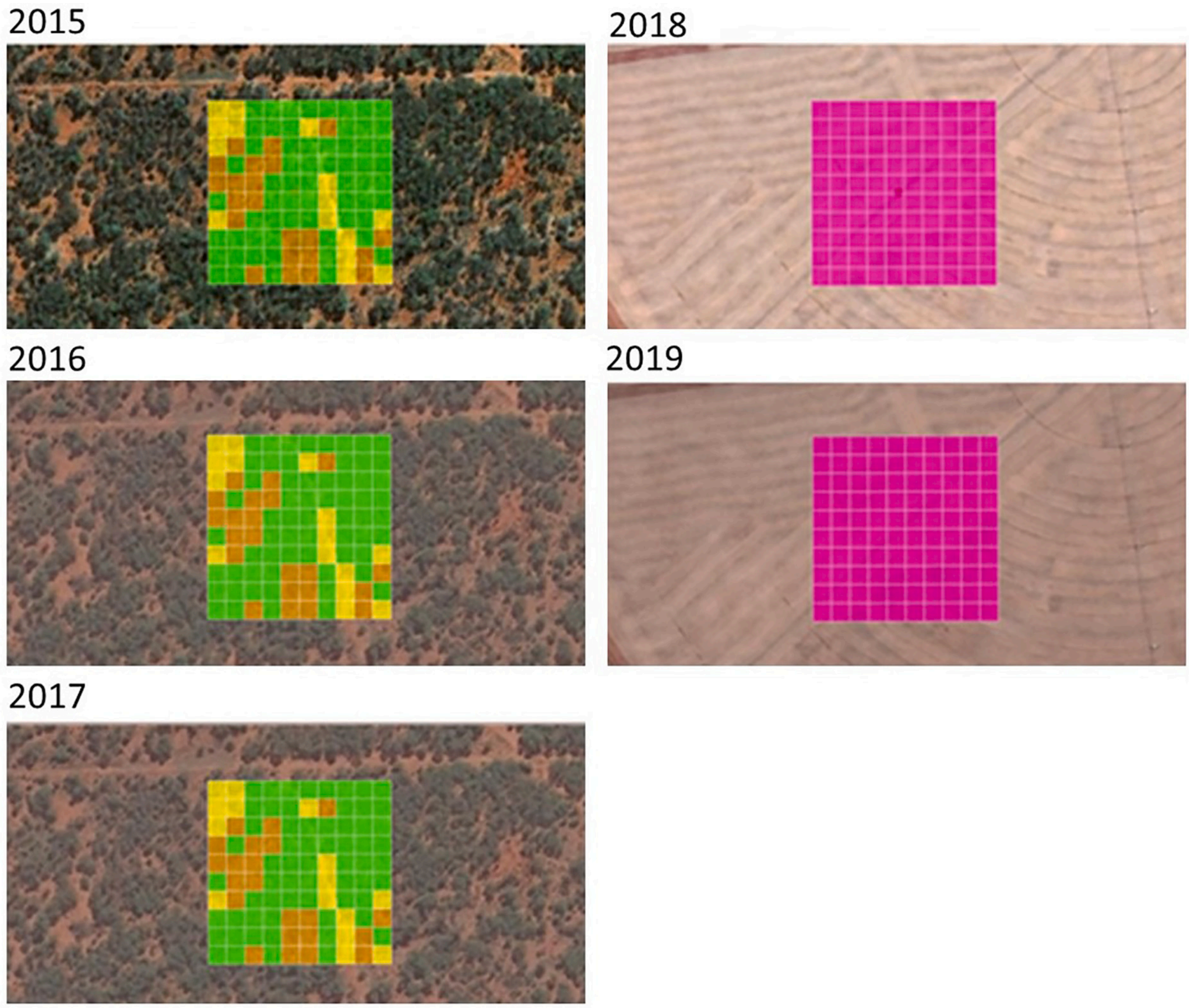

Fig. 5. Screenshots of an example sample interpretation in the Geo-Wiki interface for each year (2015–2019). 2015–2017: the land cover was a mix of trees, shrubs, and grass (green, orange and yellow colour respectively). A land cover change occurred around the fourth quarter of the year 2017 (see Fig. 6. 2018 and 2019: the land cover was cropland (magenta colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

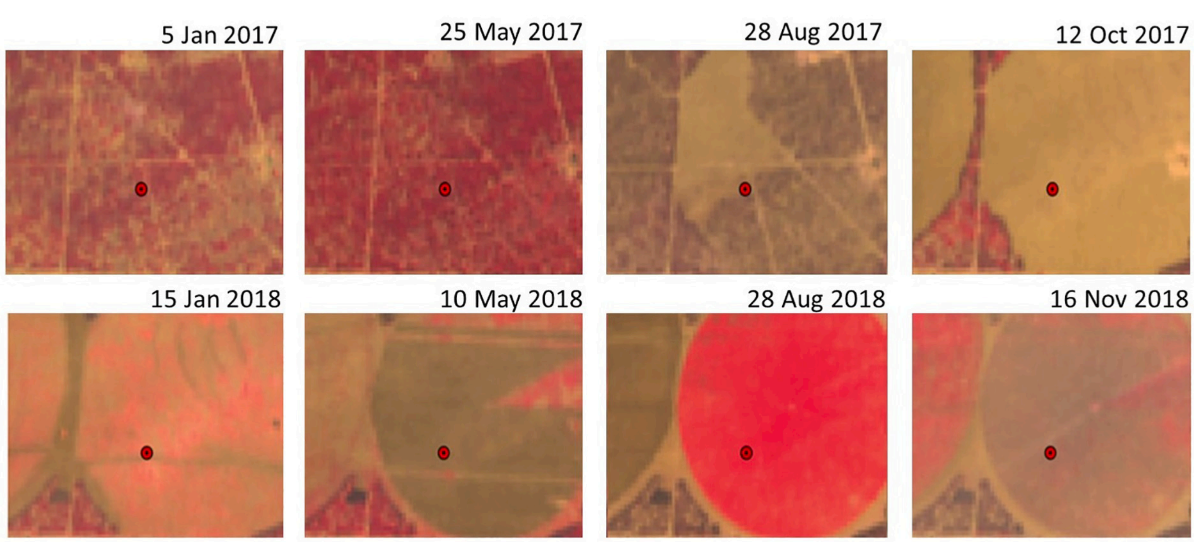

Fig. 6. Screenshot of Sentinel-2 false colour composite (NIR, red, green) time series for 2017 and 2018 depicting land cover changes from a mix of trees, shrubs and grass into cropland. Sentinel-2 images are retrieved from Google Earth Engine through the Geo-Wiki interface. The images depict the same area as shown in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Service (Copernicus Global Land Service., 2020). The product includes GLC discrete and fraction layers for the reference years, 2015–2019. This product was generated using the PROBA-V 100 m time-series, a database of high-quality land cover training data, and several ancillary datasets (Buchhorn et al., 2020a). The global discrete map includes 23 land cover classes at 3 different levels. The "level 1" legend contains generic land cover classes, such as forest, shrubs, herbaceous vegetation, cropland, urban/built-up, herbaceous wetland, lichen/moss, snow/ice, and water. The other two levels provide more detailed classes related to forests. The fraction layers show percentages of the generic land cover types (except herbaceous wetland) within each map pixel area (100 × 100 m). The maps for 2015 (discrete and fraction) had a "Base" mode, a year for which the map processing chain was trained for. The maps for 2016, 2017, and 2018 were produced using longer-term satellite data (3 years). The map for 2019 was in NRT (near-real-time) mode, as longer-term satellite data was not yet available after the end of that year.

Based on the accuracy estimation protocol of the validation dataset, we assessed the accuracy of the CGLS-LC100 V3.0 maps for 2015–2019. Since validation sample sites were not allocated proportionally to the strata areas, sample unequal inclusion probabilities between different strata were taken into consideration (Section 2.1). At validation sample locations, the reference land cover and the mapped land cover types were compared for each year. To estimate map accuracy and its confidence intervals, Eqs. (2)–(4) were used following the approach by Stehman (2014). Accuracies and their confidence intervals of the discrete map were calculated at the global and (sub) continent levels, focusing on the Level 1 generic land cover classes (at 95% confidence level).

We used the revised version (Section 2.3) of the original global validation data for 2015 (21,752 sample sites) for assessing the 2015 GLC discrete map. To assess the updated GLC maps (2016–2019), we used the updated global validation data which combines the original validation dataset with the additional validation dataset (28,472 sample sites) (Section 2.3, Fig. 1-A4).

Stability in map accuracy was also estimated following the proposed framework for operational validation (Fig. 1-B2). Stability in class accuracies was calculated between the adjacent pairs of years using Eqs. (5) and (6). Over the whole period of 5 years, the stability was calculated by estimating the maximum and mean stability index between the pairs of years (2015–2019).

3. Results

3.1. Validation dataset updates

The number of validation sites updated through 2015–2019 is listed in Table 1 for the different updating steps. As a result of the random and targeted revisits of the 9465 sample sites (Section 2.3), 277 sites were identified as changed for at least one subpixel in the sample site requiring update for the follow-up years 2016–2019. Among these, for 151 (1.6%) sample sites, the changes were significant enough to have a change in the CGLS-LC100 map legend (Level1) between the years 2015 and 2019. The change rate was 3% of sample sites that were targeted as unstable based on the BFAST time series algorithms.

In contrast, among the additional sample sites, for the same period, 34.8% of sites were reported to have changes in land cover at the CGLS-LC100 map legend level. A very high level of land cover change rate within the additional sample sites indicates that the added change stratification was successful in targeting land cover changes that occurred between 2015 and 2019 as opposed to the change rate in the original validation sites as part of revisiting.

Table 1
Number of sample sites reviewed or collected as having a change in land cover between 2015 and 2019

	Targeted revision	Total revision including the targeted revision	Additional sample collection
Total reviewed	1305	9465	6720
Change in any subpixel	79	277	2515
Change in the CGLS-LC100 legend	39	151	2344
Rate of change in the CGLS-LC100 legend	3%	1.6%	34.8%

3.2. Annual map validation

Annual map validation results showed that the discrete CGLS-LC100m V3.0 Level 1 map had an overall accuracy of 80.6 ± 0.7% at the global level for 2015, at 95% confidence level(Table 2). The annual updates (2016 - 2019) were assessed with around 80.3–80.4 ± 0.7% accuracy, with the 2019 map in NRT mode having 80.3 ± 0.7% accuracy at global level.

For 2015, the overall accuracy at the continental level was around 80%, with the highest accuracy of 83.5% for Asia and the lowest accuracy of 77.6% for North America (Table 2). The same trend could be observed for the maps for 2016 - 2019 with a maximum fluctuation in the overall accuracy of 0.8% for Northern Eurasia. Across five years, the 2019 map had lower overall accuracies compared with the previous years. The lower accuracy could be explained by the 2019 NRT mode, which uses less data after the year has ended. Nevertheless, at the continental level, the land cover was mapped with high consistency. Despite the additional validation sites for change areas, the confidence intervals of the accuracy estimates (Table 2) are similar over the period. This is because the additional sample sites were selected only in change areas which occupy a relatively small portion of the mapped area, as opposed to other non-change strata.

Fig. 7 shows the user's and producer's accuracies of different land cover types of the discrete CGLS-LC100m V3.0 2015–2019 maps. Generally, forest, bare/sparse vegetation, snow/ice, and permanent water were mapped with high accuracies (> 85%). Herbaceous vegetation, croplands, and urban were mapped with moderate accuracy (65% - 85%). In contrast, shrubs and herbaceous wetland classes were mapped with the lowest accuracies (< 65%). Class accuracies also show consistency between the annual products, as class-specific accuracies varied less than 2% for most classes within the period, with the exception of the herbaceous wetland class.

Table 2
Overall accuracies of the discrete CGLS-LC100m V3.0 2015–2019 maps at 95% confidence level, for global and continental levels.

	2015	2016	2017	2018	2019
Africa	80.3 ± 1.9	80.2 ± 2	80.4 ± 1.9	80.1 ± 2	80.1 ± 2
Asia	83.5 ± 1.4	83.3 ± 1.4	83.4 ± 1.4	83.5 ± 1.4	83.5 ± 1.4
Northern Eurasia	80.9 ± 1.5	80.5 ± 1.5	80.5 ± 1.6	80.2 ± 1.6	80.1 ± 1.6
Europe	80.2 ± 1.6	80.2 ± 1.6	80 ± 1.6	79.9 ± 1.6	79.9 ± 1.6
North America	77.6 ± 1.7	77.7 ± 1.7	77.6 ± 1.7	77.5 ± 1.7	77.5 ± 1.7
Oceania & Australia	81.5 ± 1.9	80 ± 1.9	79.9 ± 1.9	80.2 ± 1.8	79.9 ± 1.9
South America	80.1 ± 1.5	80 ± 1.5	80.1 ± 1.5	79.9 ± 1.5	80 ± 1.5
Global	**80.6 ± 0.7**	**80.4 ± 0.7**	**80.4 ± 0.7**	**80.3 ± 0.7**	**80.3 ± 0.7**

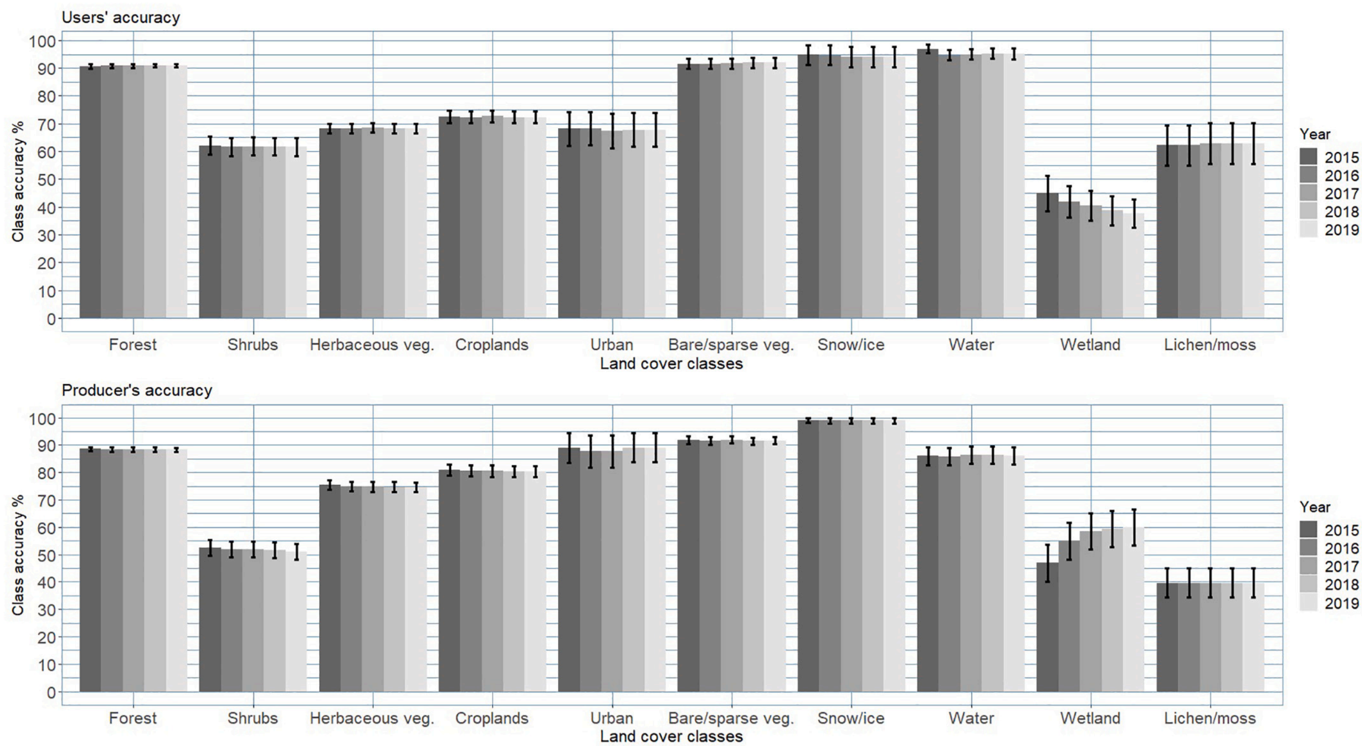

Fig. 7. User's and producer's accuracies of different land cover types for the discrete CGLS-LC100m V3.0 2015–2019 maps.

3.3. Stability in map accuracy

The stability in class accuracies was estimated for the 5 years for the CGLS-LC100 product (Fig. 8).

The results show that the CGLS-LC100m product was stable for the 5 years, for most of the classes (stability index <15%). In terms of both the user's and producer's accuracy, the herbaceous wetland class was shown to be the least stable, although the user's accuracy was slightly more stable. The herbaceous wetland class also had the lowest class accuracies (Fig. 7).

Next, water, urban and shrubs classes showed higher relative instability while forest, bare/sparse vegetation, and snow/ice classes showed the least relative instability.

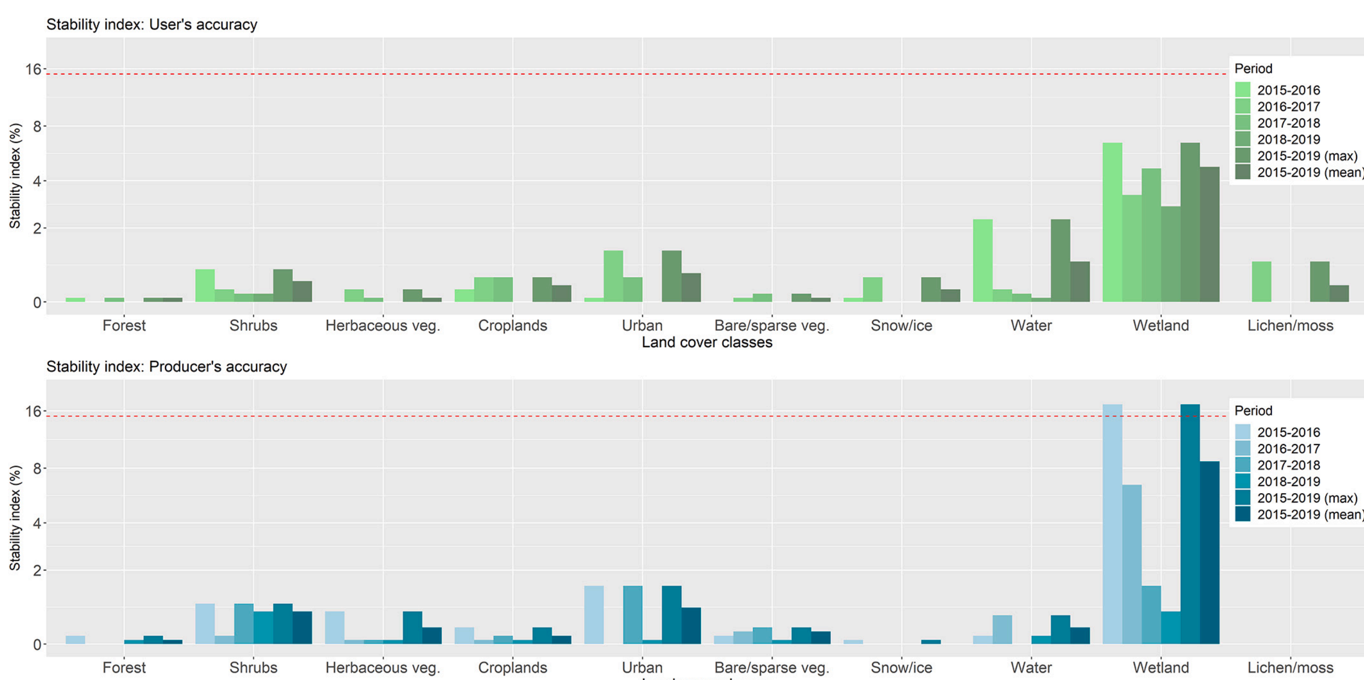

Fig. 8. Stability index for class accuracies (user's and producer's) in logarithmic increments for the CGLS-LC100 product 2015–2019; red dashed lines show targeted stability by the users according to GCOS (GCOS, 2011). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

This study proposes a framework for operational validation of GLC monitoring products that aimed to reduce the lag between map production updates and their validation. The framework has been implemented as part of the CGLS-LC100 GLC monitoring efforts (Copernicus Global Land Service, 2017). As a result, we were able to timely assess the accuracy of the new collection of the CGLS-LC100 product including its annual land cover maps (Buchhorn et al., 2020a). In doing so, our validation addressed the Stage-4 validation requirements by the CEOS-LPV (LPVS-CEOS, 2000) and also provided estimations on the stability of map accuracy, deemed important by the GLC map users (GCOS, 2011). The work presented in this study brings possibilities of unbiased area estimation of land cover types and land cover change estimation, which can be targeted in further studies.

An essential part of operational validation is the regular update of the validation data to recent dates. This was addressed by a partial revisit of the original validation data and augmentation of sample sites in selected areas (Fig. 1-A). A partial revisit instead of a full revisit is proposed considering the efforts and time required for manual revisit and interpretation of validation sites (Pengra et al., 2020; Tarko et al., 2020) and reducing delays between map accuracy estimation and map production release. Efficient use of available time and resources could be achieved by targeting the revisions to validation sites that have a high possibility of undergoing land cover change. We used the BFAST family algorithms to identify validation sample sites with breaks in time series based on MODIS data (Masiliunas et al., 2021; Verbesselt et al., 2010). Other change detection algorithms (Asokan and Anitha, 2019) could also be used to improve the detection rate of land cover changes for validation sites. Furthermore, higher resolution data such as Landsat and Sentinel-2 could also be applied to detect land cover changes, provided that there is sufficient history of time series data to accurately detect land cover change. Complementary to the targeted revisit, we further revisited a random fraction of the original validation dataset. We revisited 10% of the sample sites for each reference year (2016–2019). The random subset can be different depending on the available resources and confidence in the change detection algorithms for the targeted revisit.

The practical approach of partially revisiting instead of a full revisit does not come without caveats as it is plausible that the change detection algorithm has missed validation sites with land cover change and an un-revisited site has seen a change in land cover. This can have an impact on the quality of the validation data. Such issues could be alleviated by improving the quality of the land cover change detection algorithm. At the same time, possible errors of the validation data could be taken into account when assessing land cover map accuracies (Stehman and Foody, 2019). Depending on whether the accuracy of the reference land cover type is known or not, a few methods were proposed by Foody (2010) to account for errors in reference data in assessing the accuracy of land cover change mapping. This study did not account for the possible uncertainty of the un-revisited sites. Therefore, future studies could investigate this issue for example based on the methods by Foody (2010). Regardless, we recommend a full revisit of the validation dataset after a certain period (e.g., 5 years) to maintain the quality of the validation dataset.

We collected additional validation sites to increase the sampling intensity of validation sites in possible change areas since 2015. Our results revealed that the rate of land cover change in the additional validation sites is 10-fold larger than the rate of change in the revisited sites (Table 1). This result conforms our intention to increase the sampling intensity in change areas. Such sample augmentation (addition) was possible thanks to the flexibility of the stratified random sampling design (Stehman et al., 2012). However, other sampling schemes such as simple random sampling could also be modified to allow additional sites without compromising the statistical rigor.

General stability in terms of accuracy of the annual CGLS-LC100 maps was observed (Fig. 8). The stability in product accuracy was estimated for the first time descriptively in this study to address the user requirement of stability in product accuracy of land cover maps (GCOS, 2011). For assessing the stability over a longer period, we used the maximum and mean of stability index over the assessing period (Eq. (6)). The choice could depend on the strictness of the stability requirement by different users. Furthermore, in this study, the stability of product accuracy was assessed relative to the product accuracy of the previous year (or period). For users such as the climate modelling community that has a strong emphasis on the consistency of product and stability of product accuracy over time (Bontemps et al., 2011), assessments based on relative stability index is deemed to be informative. However, other users may be more interested in stability requirements in absolute terms, for example, accuracy fluctuations not exceeding 5% over time. With the increased availability of operational land cover monitoring providing land cover data over time, further guidelines and best practices need to be developed to assess the long-term stability in product accuracy, considering different user needs.

We proposed an adjacent year approach (Eq. (5)) to assess the stability in product accuracy and applied this for the annual CGLS-LC100 maps from 2015 to 2019. For longer time series of annual maps, e.g., when different sensors are introduced, other methods of assessing stability in product accuracy can be investigated. For example, the trend in product accuracy (Padilla et al., 2014) could be assessed, particularly when higher accuracy is expected due to algorithm development and the availability of higher spatial and temporal resolution satellite data. Note that higher accuracy due to improved mapping might lead to decreased stability in product accuracy over the long period. In addition, next to the stability assessment on the general class accuracy level, the stability can also be assessed on more detailed spatial levels, e.g., continental, national and regional. Similarly, in addition to stability assessment, the temporal consistency assessment focusing on year-to-year variability in the mapped products is useful for product users (Sulla-Menashe et al. 2019).

This study focused on assessing the accuracy of annual land cover maps in the context of operational land cover monitoring. As many users are interested in land cover change monitoring (Lesiv et al., 2016; Szantoi et al., 2020), the accuracy of land cover change detection based on annual land cover maps can be analyzed in future studies. A significant amount of sites with land cover change in our validation data (Table 1) raises the possibility of using this dataset for accuracy and area estimation of land cover change. Although not the focus of this study, the sampling design of the CGLS-LC100 validation dataset is suitable for further adjustments to achieve statistically rigorous land cover change estimations, for example by making use of the mapped classes to improve the precision of the area estimates. Similarly, future efforts should also include a statistical estimation for areas of each land cover class (Stehman, 2009a). The fraction layers provided by the CGLS-LC100 could be a useful option for improving area estimation of land cover classes (Buchhorn et al., 2020a).

For the sake of keeping the independence of this operational land cover validation dataset, the CGLS-LC100 validation dataset is currently not foreseen to be publicly released. However, we envision making the dataset available via platforms such as LACO-Wiki (The Land Cover Validation Platform) (See et al., 2015) to be used to validate other land cover products. The LACO-Wiki platform allows users to upload their land cover products and assess their accuracy by using self-generated or existing validation datasets. This ensures that the validation dataset is used for validation purposes rather than training or calibration of image classification (Tsendbazar et al., 2018).

5. Conclusion

The continuous monitoring of land cover at a global scale is recognized to be useful for policies focusing on adaptations to challenges facing today's society such as climate change and sustainable development. This study focused on advancing the operational and continuous validation of GLC products, which is an intrinsic part of land cover monitoring. We presented a framework for operational validation of GLC products aiming to find an efficient and timely way to update a validation dataset and to update the validation for the subsequent release of GLC products towards meeting the CEOS Stage-4 validation guidelines.

As part of the framework, a partial revision of the validation dataset supplemented by an additional validation data collection was proposed to update validation datasets. This was implemented to update the CGLS-LC100 validation dataset for 2016–2019. Driven by the costs involved in visually reviewing validation sites for possible changes in land cover and the aim to update map validation without much time lag, a partial revision of the validation dataset supported by change detection algorithms at validation site locations was shown to be a useful practical solution. With this partial revision, more efforts could be spent to collect extra validation sample sites to increase the sampling intensity in possible land cover change areas. The same strategy can be used to update the validation dataset further (2020 and beyond).

Further, we proposed metrics to assess the stability of product accuracy, which is required by product users and which has not been provided to date for land cover products. Information on the stability of the product and its accuracy is important for users who require long-term land cover monitoring. The lack of methods to provide this information had been noted in this study, and we proposed a descriptive approach to assess stability in product accuracy. As more operational land cover monitoring efforts are upcoming, we emphasize the importance of updated map validation and recommend improving the current validation practices and guidelines towards operational map validation so that long-term land cover maps and their uncertainty information are well understood and properly used.

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.

Acknowledgments

This work was supported by the European Commission - Copernicus programme. The authors thank the regional experts for their contributions in collecting the validation dataset. We 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.2021.112686.

References

Asokan, A., Anitha, J., 2019. Change detection techniques for remote sensing applications: a survey. Earth Sci. Inf. 12, 143–160.

Bartholomé, E., Belward, A., 2005. GLC2000: a new approach to global land cover mapping from earth observation data. Int. J. Remote Sens. 26, 1959–1977.

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., Defourny, P., 2011. Revisiting land cover observations to address the needs of the climate modelling community. Biogeosci. Discuss. 8, 7713–7740.

Bossard, M., Feranec, J., Otahel, J., 2000. CORINE Land Cover Technical Guide - Addendum 2000. European Environment Agency, Copenhagen. https://www.eea.europa.eu/publications/tech40add.

Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R., Rover, J., 2020. Lessons learned implementing an operational continuous United States national land change monitoring capability: the Land change monitoring, assessment, and projection (LCMAP) approach. Remote Sens. Environ. 238, 111356.

Buchhorn, M., Bertels, L., Smets, B., De Roo, B., Lesiv, M., Tsendbazar, N.-E., Li, L., 2020a. Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015–2019: Algorithm Theoretical Basis Document. https://doi.org/10.5281/zenodo.3938968.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020b. Copernicus global Land Cover layers—collection 2. Remote Sens. 12, 1044.

Card, D.H., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogramm. Eng. Remote. Sens. 48, 431–439.

CCI Land Cover, 2017. Release of a 1992–2015 Time Series of Annual Global Land Cover Maps at 300 m. https://www.esa-landcover-cci.org/index.php?q=webfm_send/88.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., 2015. Global land cover mapping at 30m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27.

Colditz, R.R., Acosta-Velázquez, J., Díaz Gallegos, J.R., Vázquez Lule, A.D., Rodríguez-Zúñiga, M.T., Maeda, P., Cruz López, M.I., Ressl, R., 2012. Potential effects in multi-resolution post-classification change detection. Int. J. Remote Sens. 33, 6426–6445.

Copernicus Global Land Service, 2017. Copernicus Global Land Service- *Land Cover*: VITO.2018,March 12. https://land.copernicus.eu/global/products/lc.

Copernicus Global Land Service, 2020. *Land Cover*.2020,December 7. https://land.copernicus.eu/global/products/lc.

Defourny, P., Bontemps, S., Schouten, L., Bartalev, S., Cacetta, P., de Wit, A., di Bella, C., Gérard, B., Giri, C., Gond, V., Hazeu, G., Heinimann, A., Herold, M., Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mücher, S., Nonguierma, A., Stibig, H., Shimabakuro, Y., Van Bogaert, E., Vancutsem, C., Bicheron, P., Leroy, M., Arino, O., 2011. GLOBCOVER 2005 and GLOBCOVER 2009 validation : learnt lessons. In: *GOFC-GOLD Global Land Cover & Change Validation Workshop*. Laxenburg, Austria.

Defourny, P., Lamarche, Celine, Marrissiaux, Quentin, Brockman, Carsten, Boettcher, Martin, Kirches, Grit, 2019. *Product User Guide and Specification: ICDR Land Cover 2016–2019*: UCLouvain. https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D3.3.12-v1.3_PUGS_ICDR_LC_v2.1.x_PRODUCTS_v1.3.pdf.

DeFries, R., Townshend, J., 1994. NDVI-derived land cover classifications at a global scale. Int. J. Remote Sens. 15, 3567–3586.

ESA-WorldCover, 2020. *Worldwide Land Cover Mapping*: VITO NV.2021. https://esa-worldcover.org/en.

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

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X.M., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.

Gallego, F.J., 2004. Remote sensing and land cover area estimation. Int. J. Remote Sens. 25, 3019–3047.

GCOS, 2010. Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update). http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf.

GCOS, 2011. Systematic Observation Requirements for Satellite-Based Data Products for Climate: Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. https://library.wmo.int/doc_num.php?explnum_id=3710.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., Chen, J., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34, 2607–2654.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., 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. 64, 370–373.

Latifovic, R., Zhu, Z.-L., Cihlar, J., Giri, C., Olthof, I., 2004. Land cover mapping of north and Central America—global Land Cover 2000. Remote Sens. Environ. 89, 116–127.

Lesiv, M., Moorthy, I., Fritz, S., Herold, M., Tsendbazar, N.E., Smets, B., Van De Kerchove, R., 2016. GCLOPS: User Requirement Document - Dynamic Land Cover.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R.A., Peng, S., 2018. Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). Earth Syst. Sci. Data 10, 219–234.

Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L., Merchant, J., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int. J. Remote Sens. 21, 1303–1330.

LPVS-CEOS, 2000. Land product validation subgroup- CEOS validation hierarchy. 2021 (January 15). https://lpvs.gsfc.nasa.gov/index.html.

Masiliunas, D., Tsendbazar, N., Herold, M., Verbesselt, J., 2021. BFAST lite: a lightweight break detection method for time series analysis. Remote Sens. 13.

Mayaux, P., Eva, H., Gallego, J., Strahler, A.H., Herold, M., Agrawal, S., Naumov, S., De Miranda, E.E., Di Bella, C.M., Ordoyne, C., Kopin, Y., Roy, P.S., 2006. Validation of the global land cover 2000 map. In: Geoscience and Remote Sensing, IEEE Transactions on, 44, pp. 1728–1739.

Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J. D., Friedl, M.A., Herold, M., 2012. A global land-cover validation data set, part I: fundamental design principles. Int. J. Remote Sens. 33, 5768–5788.

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.
Padilla, M., Stehman, S.V., Litago, J., Chuvieco, E., 2014. Assessing the temporal stability of the accuracy of a time series of burned area products. Remote Sens. 6, 2050–2068.
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.
Pengra, B.W., Stehman, S.V., Horton, J.A., Dockter, D.J., Schroeder, T.A., Yang, Z., Cohen, W.B., Healey, S.P., Loveland, T.R., 2020. Quality control and assessment of interpreter consistency of annual land cover reference data in an operational national monitoring program. Remote Sens. Environ. 238, 111261.
Romijn, E., Herold, M., Mora, B., Briggs, S., Seifert, F.M., Paganini, M., 2016. Monitoring Progress towards: Sustainable Development Goals the Role of Land Monitoring. Wageningen, Netherlands. http://www.gofcgold.wur.nl/documents/newsletter/Sustainable_Development_Goals-infobrief.pdf.
See, L., Perger, C., Hofer, M., Weichselbaum, J., Dresel, C., Fritz, S., 2015. LACO-WIKI: an open access online portal for land cover validation. In: *ISPRS Annals of Photogrammetry,* Remote Sensing and Spatial Information Sciences*,* II-3/W5, pp. 167–171.
Stehman, S.V., 2000. Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sens. Environ. 72, 35–45.
Stehman, S.V., 2009a. Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing. Remote Sens. Environ. 113, 2455–2462.
Stehman, S.V., 2009b. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30, 5243–5272.
Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. Int. J. Remote Sens. 35, 4923–4939.
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., Olofsson, P., Woodcock, C.E., Herold, M., Friedl, M.A., 2012. A global land-cover validation data set, II: augmenting a stratified sampling design to estimate accuracy by region and land-cover class. Int. J. Remote Sens. 33, 6975–6993.
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. Eur. Communities Luxemb. 51.
Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P., Friedl, M.A., 2019. Hierarchical mapping of annual global land cover 2001 to present: the MODIS collection 6 Land Cover product. Remote Sens. Environ. 222, 183–194.
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.
Tarko, A., Tsendbazar, N.-E., de Bruin, S., Bregt, A.K., 2020. Producing consistent visually interpreted land cover reference data: learning from feedback. Int. J. Digital Earth 1–19.
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., 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.
Tsendbazar, N., Tarko, A., Li, L., Herold, M., Lesiv, M., Fritz, S., V M, 2020. Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015–2019: Validation Report.
UN-GGIM, 2019. Global and Complementary (Non-authoritative) Geospatial Data for SDGs: Role and Utilisation. https://ggim.un.org/documents/Report_Global_and_Complementary_Geospatial_Data_for_SDGs.pdf.
Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115.
Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. Remote Sens. Environ. 123, 98–108.
Walker, D.A., Raynolds, M.K., Daniëls, F.J., Einarsson, E., Elvebakk, A., Gould, W.A., Katenin, A.E., Kholod, S.S., Markon, C.J., Melnikov, E.S., 2005. The circumpolar Arctic vegetation map. J. Veg. Sci. 16, 267–282.
Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E., 2020. Transitioning from change detection to monitoring with remote sensing: a paradigm shift. Remote Sens. Environ. 238, 111558.
Xu, P., Herold, M., Tsendbazar, N., Clevers, J., 2020. Towards a comprehensive and consistent global aquatic land cover characterization framework addressing multiple user needs. Remote Sens. Environ. 250, 112034.
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.