Response to comments
Paper #: essd-2025-73
Title: GLC_FCS10: a global 10-m land-cover dataset with a fine classification system from Sentinel-1 and Sentinel-2 time-series data in Google Earth Engine
Journal: Earth System Science Data

Reviewer #1

This work is of great importance in multiple fields, such as earth science, geography, terrestrial ecosystem. These dataset could be very useful in many places. But some shortcomings in methods, figures, and tables need careful clarification and revisement, here are comments:
Great thanks for the positive comments. The manuscript has been greatly improved based on your and another reviewers' comments and suggestions.

1. Fig 1 is too colorful to demonstrate the core idea. Also, there is no legend to show what are different colors indicating. The relationships between different boxes are confusing too. Suggesting simplifying the figure by summarizing the main and key steps using icon and/or key words, do not put every step and all the datasets in this single figure; be consistent in color, size, and font. More detailed techniques can be in new figure.
Great thanks for the comment. Based on the suggestion, the flowchart of the proposed method has been simplified as:

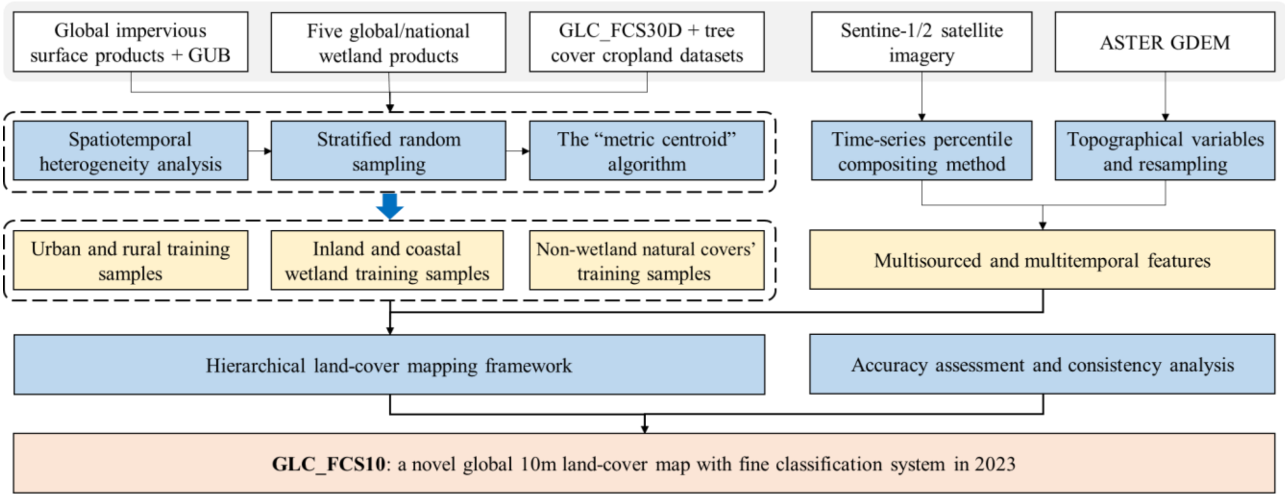

Figure 1. The flowchart of the proposed method for hierarchical land-cover mapping.

2. Table 2. how was the newly added forest- and wetland-related subcategories defined in the new system? What are the quantitative standards for closed vs open forest? Need more justification. These subclasses will of great importance in understanding the diversity of forest and wetland ecosystem, but only if the definition and classification of these subclasses are reasonable and practical.
Great thanks for the comment. Based on your constructive suggestion, the descriptions about the fine classification system have been added in the Table S1 as:

Table S1. The detailed definitions of 30 land-cover types in the fine classification system.

Fine classification system	Definition

Herbaceous rainfed cropland	Herbaceous cropland with no irrigation facilities and crops grown by natural precipitation
Tree or shrub covered rainfed cropland (orchard, oil palm…)	Tree or shrub covered rainfed cropland, mainly including orchard, oil palm, etc.
Irrigated cropland	Cropland with guaranteed water sources and irrigation facilities that can be irrigated normally in a typical year
Closed evergreen broadleaved forest	Evergreen broadleaved tree cover, tree height > 3 m, tree-cover percentage > 40%
Open evergreen broadleaved forest	Evergreen broadleaved tree cover, tree height > 3 m, 15%< tree-cover percentage < 40%
Closed deciduous broadleaved forest	Deciduous broadleaved tree cover, tree height > 3 m, tree-cover percentage > 40%
Open deciduous broadleaved forest	Deciduous broadleaved tree cover, tree height > 3 m, 15%< tree-cover percentage < 40%
Closed evergreen needleleaved forest	Evergreen needleleaved tree cover, tree height > 3 m, tree-cover percentage > 40%
Open evergreen needleleaved forest	Evergreen needleleaved tree cover, tree height > 3 m, 15%< tree-cover percentage < 40%
Closed deciduous needleleaved forest	Deciduous needleleaved tree cover, tree height > 3 m, tree-cover percentage > 40%
Open deciduous needleleaved forest	Deciduous needleleaved tree cover, tree height > 3 m, 15%< tree-cover percentage < 40%
Closed mixed-leaf forest	Mixed broadleaved and needleleaved forests, tree height > 3 m, tree-cover percentage > 40%
Open mixed-leaf forest	Mixed broadleaved and needleleaved forests, tree height > 3 m, 15%< tree-cover percentage < 40%
Evergreen shrubland	Vegetation communities dominated by low cover and evergreen dwarf and scrub woodlands
Deciduous shrubland	Vegetation communities dominated by woody shrubs that lose their leaves in winter or the dry season
Grassland	Refers to land where herbaceous plants predominate
Lichens and mosses	Lichens and moss-covered areas
Swamp	The forest or shrubs which grow in the inland freshwater
Marsh	Herbaceous vegetation (grasses, herbs and low shrubs) grows in the freshwater
Lake/river flat	The non-vegetated flooded areas along the rivers and lakes
Saline	Characterized by saline soils and halophytic (salt tolerant) plant species along saline lakes
Mangrove forest	The forest or shrubs which grow in the coastal brackish or saline water
Salt marsh	Herbaceous vegetation (grasses, herbs and low shrubs) in the upper coastal intertidal zone
Tidal flat	The tidal flooded zones between the coastal high and low tide levels including mudflats and sandflats
Urban impervious surfaces	Land covered with buildings and other man-made structures within the urban boundary
Rural impervious surfaces	Land covered with man-made structures outside the urban boundary, mainly including rural residential land, transportation land, etc.
Sparse vegetation	Areas covered by woodland, shrubs and grasses, vegetation-cover percentage < 15%
Bare areas	Refers to land that is largely devoid of vegetation cover
Water	Lakes, rivers and streams that are always flooded
Permanent ice and snow	Areas covered by snow and ice all year round

3. How was the GLC_FCS30D dataset used as the training dataset for the 10m global cover mapping in this study? They are in different spatial resolutions, also, there are uncertainties in the GLC_FCS30D, not to mention the GLC_FCS30D does not cover 2023. How were all the uncertainties in the training dataset evaluated? Without solid evaluation, these training dataset cannot be high-confidence.

Great thanks for the comment. Yes, we completely agree that the high-confidence of training dataset is the key of subsequent land-cover mapping. In this study, three measures were used to identify these spatiotemporal homogeneity and high-quality training areas from GLC_FCS30D: 1) A time-series consistency analysis is applied to the GLC_FCS30D, and only stable areas during 1985–2022 will be retained as *TrainCanArea_NLCs*. 2) The $MaxBound_{imp}$ and $MaxBound_{wet}$ are imported to mask the *TrainCanArea_NLCs*, i.e., the training areas for non-wetland natural land-cover types should be located outside of $MaxBound_{imp}$ and $MaxBound_{wet}$. The aim of this step is to minimize confusion between non-wetland natural land-cover types and these two land-cover types. 3) A morphological erosion filter with a local window of 3 pixels × 3 pixels is used to find the spatially homogeneous areas for non-wetland natural land-cover types.

In terms of the different spatial resolutions between GLC_FCS30D with the need of 10 m training samples, the **"metric centroid"** method is adopted, which had been used to downscale 500-m training samples from MCD12Q1 to 30-m in the work of Zhang and Roy (2017). The detailed descriptions about the **"metric centroid"** method have been explained as:

Second, most high-quality training samples (except for those for impervious surfaces) are derived from the 30-m training areas, so there is also a need to reduce the 30-m training samples to 10-m samples to achieve a global 10-m land-cover map. In this work, the "metric centroid" method is adopted, which had been used to downscale 500-m training samples from MCD12Q1 to 30-m in the work of Zhang and Roy (2017). Specifically, as each 30-m pixel corresponds to 3 × 3 10-m pixels, we first find the centroid from these nine pixels as $P_{centroid}$ through spectral averaging, and then the point with the smallest absolute distance with $P_{centroid}$ was chosen as the optimal downscaled 10-m sample point [**Eq. (4)**].

$$P_i = \underset{i}{\operatorname{argmin}}\left(\left|\boldsymbol{\rho}_{P_i} - \boldsymbol{\rho}_{P_{centroid}}\right|\right), \boldsymbol{\rho}_{P_{centroid}} = \frac{1}{9}\sum_{j=1}^{9}\frac{\boldsymbol{\rho}_{P_j}}{9} \tag{4}$$

Where $\boldsymbol{\rho}_{P_i}$ is the spectra value of composited Sentinel-2 training features (See Section 3.3) at pixel P_i. If more than one point in the nine pixels has the same minimum absolute distance, then we pick randomly from among them.

Then, as for the uncertainties in these derived training samples, due to the large volume of these globally distributed training samples, we selected approximately 10,000 derived samples from the training sample pool. Upon meticulous inspection, we determined that these chosen samples attained an overall accuracy (O.A.) of 92.18%, with certain uncertainties existing for shrubland and grassland. It has been explained in the Section 4.4.1 as:

A principal difficulty of land-cover mapping is obtaining high-quality training samples (Li et al., 2023; Zhang et al., 2021), in this work, we integrate prior multisource global land-cover products to generate globally distributed training samples. To ensure the confidence of these derived training samples and minimize the classification errors of each prior product, we took the following actions: spatiotemporal consistency checking was used to find homogeneous and stable areas. The intersection of multiple land-cover products minimized the influence of classification errors in each product. A morphological erosion filter was applied to reduce the impact of edge-mixing effects. The accuracy assessment partly demonstrates the reliability of these derived training samples, i.e., GLC_FCS10 achieves satisfactory accuracy metrics and outperforms several other land-cover products. Similarly, Zhang et al. (2021) also used the prior global land-cover products to generate the GLC_FCS30 product with satisfactory performance. Due to the large volume of these globally distributed training samples, we selected approximately 10,000 derived samples from the training sample pool in Section 3.2.4. Upon meticulous inspection, we determined that these chosen samples attained an overall accuracy (O.A.) of 92.18%, with certain uncertainties existing for shrubland and grassland. This result was in accordance with the earlier analysis presented in Table 3.

In addition, as we all know that automatically derived training samples cannot be guaranteed to be completely accurate, so whether these erroneous training samples affect the subsequent mapping accuracy. Therefore, this manuscript further discusses the quantitative relationship between the proportion of erroneous training samples and overall classification accuracy in Section 4.4.1 as:

Moreover, it is still uncertain whether this small amount of erroneous training samples could impact the performance of land-cover mapping, Fig. 9 illustrates the quantitative relationship between the erroneous training samples and the O.A. and kappa coefficients for the basic classification system. Initially, O.A. and the kappa coefficient remain stable as the number of erroneous training samples increases. However, a significant decline occurs when the proportion of erroneous samples exceeds 30%. This indicates that the trained random forest model is robust to the erroneous training samples as long as their proportion remains below 30%. In this work, if the fraction of erroneous samples was kept below 30%, the difference in O.A. is approximately 2% and the decrease in the kappa coefficient is approximately 3%. Gong et al. (2024) also demonstrated that a small number of incorrect samples (approximately 20%) didn't affect the land-cover classification accuracy.

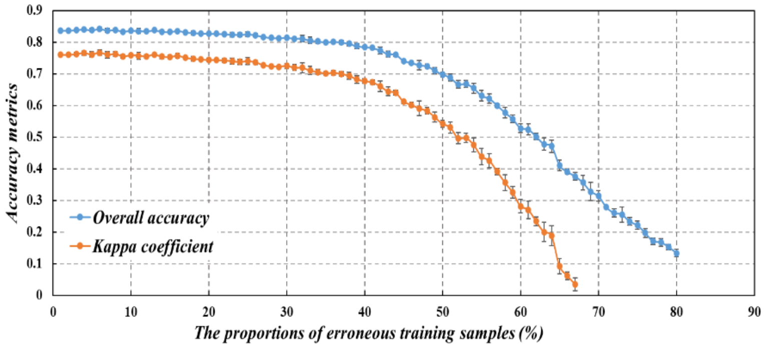

Figure 9. A sensitive analysis of kappa coefficient and O.A. with respect to the proportion of erroneous training samples.

Lastly, as for the GLC_FCS30D does not cover 2023, yes, the released GLC_FCS30D only covered the period of 1985-2022, actually, the land cover change mask between 2022-2023 has also been produced using the continuous change detection method, but not yet shared with the public, and that change mask dataset was also used to generate the training samples in this manuscript to ensure the temporal consistency between GLC_FCS30D and training samples. It has been added in the Section 3.2.3 as:

In addition, it should be noted that the TrainCanArea_NLCs represents the stable areas during 1985-2022, and there is still one-year interval with the land-cover mapping year in 2023. Fortunately, the ongoing updating of GLC_FCS30D is still in progress, the land-cover change masks during 2022-2023 have been finished, and which are also used to remove these changed and low-confidence areas.

4. Line 277, how as the percentiles be quantified, by date? by quality or what? So was the VV and VH percentiles in Line 280.

Great thanks for the comment. The quantile-based compositing method rearranges intra-annual time-series reflectance according to **mathematical magnitude and take the corresponding quartiles to reflect the phenological variation of the time-series.** Figure S1 illustrates the schematic of how to extract quantile features from time-series satellite observations, and we can find that these percentiles can efficiently capture the variations of phenology.

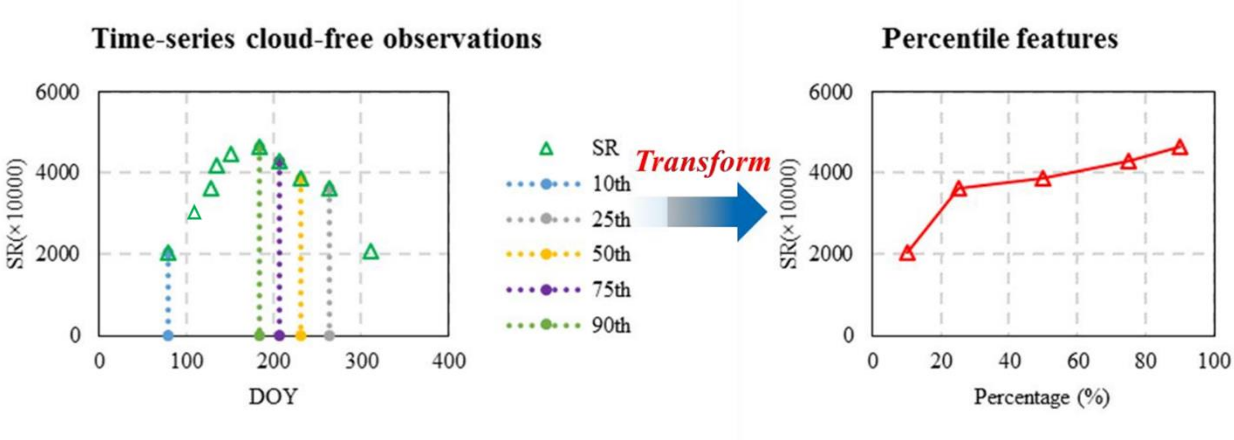

Figure S1. The schematic of how to extract quantile features from time-series satellite observations.

The basic principle of the quantile-based compositing method has been added as:

The basic principle of this method is to rearrange intra-annual time-series reflectance according to mathematical magnitude and take the corresponding quartiles to reflect the phenological variation of the time-series and suppress the noise interference such as clouds and shadows (Hansen et al., 2014). Some previous studies have also demonstrated its ability to flexibly balance noise removal and signal retention, reflecting the surface normality and capturing peak features (e.g., high percentile), adapting to different monitoring needs (Hansen et al., 2014; Zhang and Roy, 2017). Thus, in this study, time-series Sentinel-2 images are composited into the 10th, 30th, 50th, 70th, and 90th percentiles for their 10 optical bands from visible to shortwave infrared and three typical indexes [NDVI, NDWI, and LSWI in **Eq. (5)**] using the percentile-based compositing method.

5. Line 297-300 add a figure to show how was the hierarchical land cover constructed.

Great thanks for the comment. Based on your suggestion, the flowchart of how to build the hierarchical land-cover mapping models has been added as:

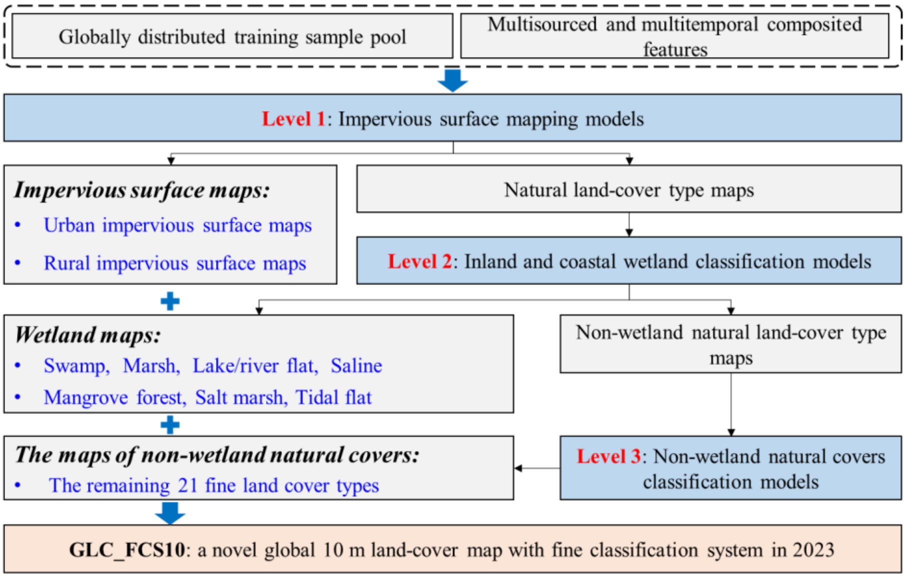

Figure 2. The detailed flowchart of hierarchical land-cover mapping algorithm by integrating globally distributed training samples and multisourced composited features.

6. Line 307 what are the 5 × 5 geographical tiles indicating? how large is the tile, and why choosing 5×5?

Great thanks for pointing out the issue. The 5° ×5° geographical tile is a regional unit for local adaptive modeling, and the size of each 5° ×5° geographical tile is equivalent to **556 km × 556 km on the equator**, and the Figure S2 gives the overview of these 983 5° ×5° geographical tiles used for local adaptive modeling.

In terms of why we choose these 5° ×5° geographical tiles, the reasons are concluded as: 1) the local adaptive modeling strategy has been demonstrated to achieve the superior performance than the single land-cover global modeling; 2) the previous works of Zhang et al., (2017) and Zhang et al., (2019) have explained that the training samples of sparse land-cover types in a small geographical grid were usually missed or greatly sparse, and the training samples from neighboring 3-by-3 tiles were also imported; 3) the GEE platform also had some limitations for computation capacity and memory; 4) the previous works in global 30 m land-cover mapping and change monitoring also demonstrated the efficiency and accuracy of these 5° ×5° geographical tiles. In summary, after balancing the accuracy performance, computation efficiency, and training sample volume, the 5° × 5° geographical tile is used to train the local adaptive classification models.

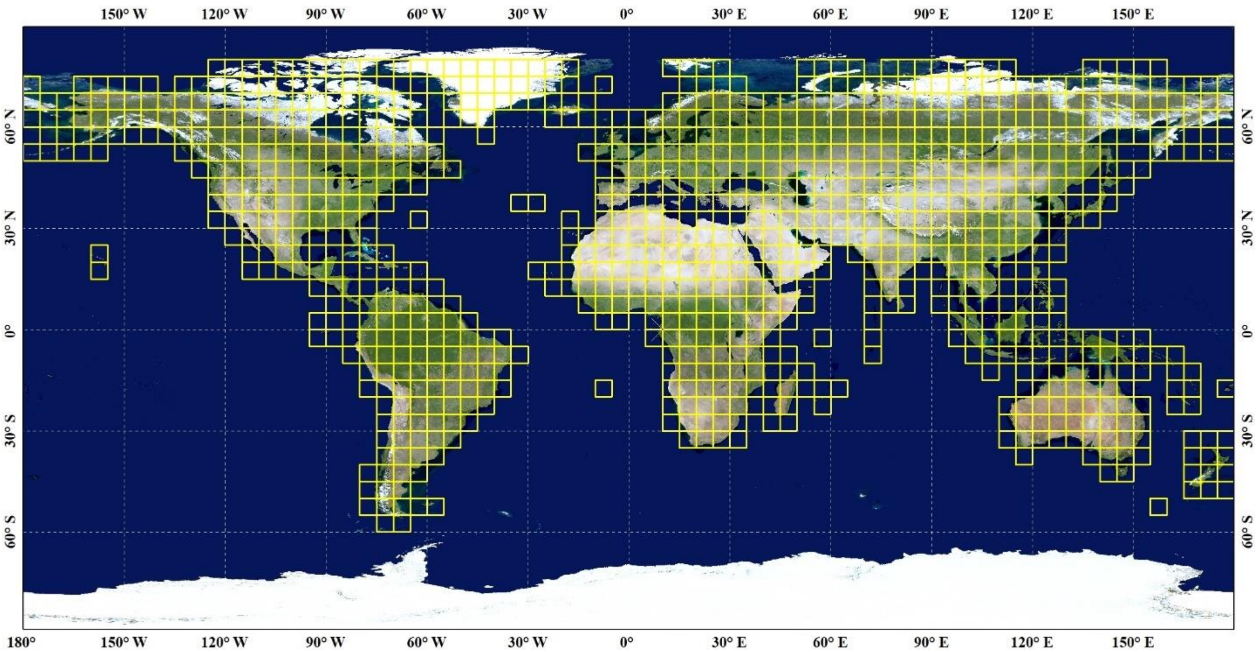

Figure S2. Overview of the 5° × 5° geographical tiles used for local adaptive modeling; the globe land area was split into 983 5° ×5° geographical tiles.

In the revised manuscript, the explanation of why we choose these 5° ×5° geographical tiles have been added as:

Then, **we split the globe into 984 5 × 5 geographical tiles (approximately 556 km × 556 km on the equator, illustrating on the Figure S1), because some studies emphasized that the local adaptive modeling usually achieves better mapping accuracy than single land-cover global modeling (Zhang et al., 2021), and previous works of Zhang and Roy (2017) and Zhang et al. (2019) have explained that the training samples of sparse land-cover types in a small geographical grid were usually missed or greatly sparse. Thus, after balancing the training sample volume, mapping accuracy, and the limitation of GEE platform, the local modeling tile size of 5 ×5 , similar to the works of (Zhang et al., 2021; Zhang et al., 2024), were used.** Then, when building the training model for each 5 × 5 geographical tile, we also import training samples within their spatial neighborhood of 3 × 3 tiles to ensure spatial consistency over the adjacent tiles.

7. Fig 3 is a map not maps.

Great thanks for the comment. It has been revised over the manuscript.

8. Table 3. why only 10 types evaluated, what about the other 20 types?

Great thanks for the comment. To comprehensively evaluate the accuracy metrics of GLC_FCS10, we designed two level of validation system (containing various land-cover details).

First, the **10 major land-cover types** correspond to the basic classification system, we can find a mapping relationship between the 30 fine types and the 10 base types in the Table 2, i.e., **the 30 fine land-cover types are merged into 10 major land-cover types in the Table 3**. The aim of Table 3 is to give the overall accuracy metrics under basic classification system.

At present, based only on current visual interpretation means and information, we can't decipher the global validation sample points for all 30 fine land-cover categories, and therefore, the confusion matrix under the 30 types cannot be given at this time. Thus, this limitation has been discussed on our Section 4.5 as:

We collected a globally distributed validation dataset and one third-party validation dataset (LCMAP_Val) for the purpose of quantifying the performance of the GLC_FCS10. However, the accuracy metrics of GLC_FCS10 for the fine classification system (containing 30 land-cover types) is still unknown. Actually, some previous studies have emphasized that collecting a large-area validation dataset is quite challenging (Tsendbazar et al., 2021; Xu et al., 2020), especially as this study also needed to focus on 30 fine land-cover types. Fortunately, over the past decades, many previous works have collected high-quality validation points at global or regional scales (d'Andrimont et al., 2020; Li et al., 2017; Stanimirova et al., 2023; Stehman et al., 2012; Zhao et al., 2023). Making full use of these prior knowledge bases to refine the globally distributed validation points into 30 fine land-cover types will be another focus for ongoing work.

what are the bottom line of OA. Kappa mean, why they are different from others?

To the best of our ability, we comprehensively evaluated the performance of GLC_FCS10 at different levels (i.e., merged into different class systems), utilizing both global and third-party regional validation sample data. The O.A. and Kappa in Table 3 give the overall accuracy metrics under basic classification system, i.e., the GLC_FCS10 has been transformed into basic classification system (10 basic land-cover types).

The reasons why they are different from the metrics in the Table 4~6 are: 1) the Table 3 and Table 4 correspond to different validation systems, i.e., Table 4 further refines forest and cropland; 2) Table 3 applies a different validation data source than the subsequent Tables 5&6, with the latter applying the third-party LCMAP_Val dataset.