Dear Editor and Referees:

We are particularly grateful for your careful reading, and for giving us many constructive comments on this work. According to the comments and suggestions, we have tried our best to improve the previous manuscript ESSD-2022-142 (An improved global land cover mapping in 2015 with 30 m resolution (GLC-2015) based on a multi-source product fusion approach). We believe the revised manuscript accounts for all reviewers' comments, and it was significantly improved as a result. The modified words or sentences are marked as blue color in the revised manuscript. We are providing an item-by-item response to all questions and recommendations.

Thanks very much for your time.

Best regards,

Xiaoping Liu and all co-authors

Reviewer #1:

General comment:

The manuscript has been improved in terms of additional patch-based validation and asset name prefixes. However, due to the following concerns, I am still not convinced that this dataset and manuscript could be candidates for ESSD.

The method is defective when input GLC maps contain errors (see my comments below) and their reliability could not be fully evaluated by the samples. In such a case, the accuracy of some LC maps was overestimated, leading to the wrong contribution to the fusion. This can be confirmed by the misclassification of bare land from water (check my previous comments), where FROM-GLC and GLC_FCS30 identified a volcano as water. This also indicates that input samples are not representative.

Response: Thanks for the comment. These comments are very helpful for revising and improving our paper. The manuscript has been improved according to your and another reviewer's comments. The point-by-point responses are listed below in **blue**. The changes in our manuscript are marked with red.

The DSET method can discount evidence from inaccurate information with a probability mass that reflects the degree of belief (Razi et al., 2019). In this study, we collected over 200,000 point-based samples. We think these samples are representative and enough to evaluate the reliability of each product. Based on the reliability assessment, the accuracy of each product can be appropriately estimated, and the errors from maps can be reduced in the fusion process. The efficiency of the DSET method in discounting wrong information can be demonstrated by the visual comparison in Sundarbans where mangroves are prevalent (see Figure R1). It can be found that GLC-2015 accurately depicted the spatial distribution of mangroves, although the FROM_GLC and GLC_FCS30 performed poorly, with almost no wetlands captured.

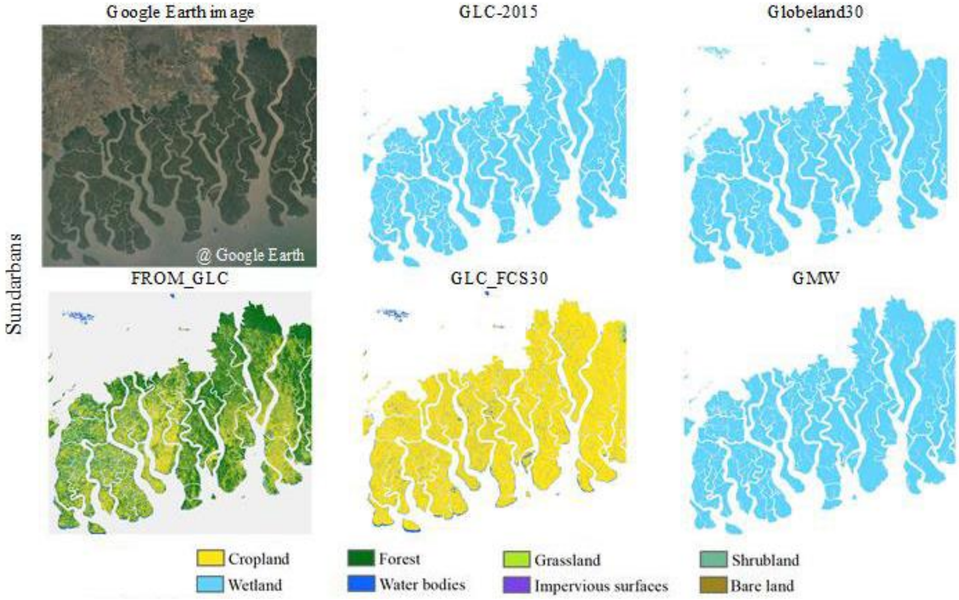

Figure R1. The wetland extent from GLC-2015, Globeland30, FROM_GLC, GLC_FCS30, and the GMW.

GLC mapping at the global scale is challenging, and there are inevitable errors, such as the misclassification of bare land from water. To reduce the error from the input maps, we employed several reliable products to improve our mapping results (the details can be found in our response for Comment #1-5). Accuracy assessment shows the improvement of the GLC-2015 in various LC classes when three national-scale products are used.

References:

Razi, S., Karami Mollaei, M. R., and Ghasemi, J.: A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory, Inf. Sci., 484, 14-26, https://doi.org/10.1016/j.ins.2019.01.053, 2019.

Comment #1-1. The 4° grid is large in the case of global mapping, as many studies have employed smaller grid (such as the 1° grid in GAUD). Generally, smaller grids will help assess each input LC map's accuracy (e.g., geographical variations). Therefore, it's suggested to investigate the relationship between grid size and mapping performance (e.g., global OA as a function of grid size), and find a proper size.

Response: Thanks for the comment. For large-scale or global land cover mapping, previous researchers divided the study area into a lot of sub-regions (Gong et al., 2020; Huang et al.,2021; Jin et al., 2022; Liu et al., 2020; Zhang et al., 2020,2021; Zhao et al.,2021). The shape and size of sub-region vary in previous work, for example, hexagons with a side length of 2°, geographical grids with a size of 1°×1°, 3.5º×3.5°, 5º×5°, or 10º×10°. When applying the DSET method to generate a global hybrid map, the following two factors should be taken into consideration when we decided the size of the sub-region:

(1) Sufficiency of samples for land cover classes. If we generate the samples in a small spatial grid such as a Landsat scene, the size of samples might be insufficient and it was also difficult to obtain samples for the rare land cover classes.

(2) Computation capacity and memory of the GEE platform. The GEE platform provides unprecedented opportunities for global land cover classification tasks due to the access to numerous analysis-ready earth observations datasets and high-performance, intrinsically parallel computation (Gorelick et al., 2017). However, GEE has computation capacity limitations. It is impossible to implement mapping work at a sub-region as large as we want because of the issue of running out of memory.

In the study, we found that when the size was larger than 4º, the execution aborted due to the complex computation exceeding available memory on the GEE platform. To investigate the relationship between grid size and mapping performance of the DSET method, we performed some tests in randomly selected areas. We generated five 4º×4° grids and divided each grid into sub-grids of 2º and 1°. Then, the fusion process was performed in grids of 1 º, 2 º and 4 º. The accuracy comparison shows that the fusion method obtained the highest OA with 4 º gird (Table S1). Therefore, we split the globe into 1507 4º×4° geographical grids and then conducted land cover mapping at the regional scale.

Table S1. Relationship between the overall accuracy of the fusion method and the size of sub-regions.

Grid ID	1 degree	2 degree	4 degree
0119	0.647	0.757	**0.844**
0317	0.641	0.589	**0.872**
0603	0.735	0.765	**0.971**
0817	0.515	0.636	**0.723**
1206	0.929	0.952	**0.976**

Correspondingly, we have added the reason why we divided the world's terrestrial area into 4°×4° grids.

"For large-scale or global land cover mapping, previous researchers divided the study area into a lot of sub-regions and conducted classification in each sub-region on GEE (Gong et al., 2020; Liu et al., 2020; Huang et al., 2021; Jin et al., 2022; Zhang et al., 2021; Zhao et al., 2021). The shape and size of sub-region vary in previous work, such as hexagons with a side length of 2°, geographical grids with a size of 1°×1°, 3.5°×3.5°, 5°×5°, or 10°×10°. When deciding on the size of sub-regions, two important factors should be considered. The size of samples in each sub-region should be sufficient so that the rare land cover classes will not be missed. On the other hand, it is impossible to implement mapping work at a sub-region as large as we want due to memory constraints. To determine the appropriate size, we tested different sizes of the sub-region (see Table S1). Result shows that dividing the study area into 4°×4° grids performed best. Therefore, we split the world's terrestrial area into 1507 4°×4° geographical grids." (Revised manuscript, Line 278-288)

References:

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., and Zhou, Y.: Annual maps of global artificial impervious area (GAIA) between 1985 and 2018, Remote Sens. Environ., 236, 111510, https://doi.org/10.1016/j.rse.2019.111510, 2020.

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

Huang, X., Li, J., Yang, J., Zhang, Z., Li, D., and Liu, X.: 30m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019., Science China Earth Sciences, 64, 1922–1933, https://doi.org/10.1007/s11430-020-9797-9, 2021.

Jin, Q.; Xu, E.; and Zhang, X.: A fusion method for multisource land cover products based on superpixels and statistical extraction for enhancing resolution and improving accuracy, Remote Sens, 14, 1676, https://doi.org/10.3390/rs14071676, 2022.

Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A. D., Chen, A., Gong, P., Chen, J., Hu, G., Chen, Y., Wang, S., Wu, Q., Huang, K., Estes, L., and Zeng, Z.: High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015, Nature Sustainability, 3, 564-570, https://doi.org/10.1038/s41893-020-0521-x, 2020.

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

Zhao, J., Yu, L., Liu, H., Huang, H., Wang, J., and Gong, P.: Towards an open and synergistic framework for mappingglobal land cover, PeerJ, 9, e11877, https://doi.org/10.7717/peerj.11877, 2021.

Comment #1-2. I understand the high quality of GlobeLand30. But I don't think the LC changes caused by the 5-year gap of GlobeLand30 can be largely avoided through the input samples. Due to the simple stratified sampling, these samples could rarely capture the LC changes during the interval. Thus, the accuracy of GlobeLand30 (against the actual LC in 2015) will be overestimated. In such a way, the contribution of Globeland30 (when LC changes) to fusion would not be as small as the authors suggested.

Response: Thanks for the comment. As you are concerned, the LC changes caused by the 5-year interval of the GlobeLand30 might bring uncertainties for the fusion. When we evaluated the reliability of the source maps, we only wanted to know whether these maps provided accurate information about land cover classes for the year 2015 because the data time of our target map (GLC-2015) is 2015. Therefore, for the GlobeLand30, there is no difference between the mismatched information caused by the LC changes and the errors caused by inaccurate classification since neither is consistent with the actual landscape in 2015. In other words, we can assume that the data time of GlobeLand30 is 2015. In this assumption, any mapping result of the GlobeLand30 that was inconsistent with the point-based samples for 2015 was defined as wrong, no matter whether it belongs to inaccurate mapping results due to LC changes or the classification method. In addition, as we mentioned in the previous manuscript, the changed areas of LC caused by the time interval are tiny compared to the global land area. Thus, the GlobeLand30 is still a good choice for the mapping task in 2015 regardless of LC changes.

We performed some experiments in China, a developing country that has undergone significant land cover changes (Yang et al., 2021), to figure out whether the errors from LC changes could be avoided in the fusion process. First, the China's land-use/cover datasets (CLUDs) for 2010 and 2015 were used to derive changed areas in the 5-year temporal interval since the CLUDs had good quality with OA exceeding 90.0% (Liu et al., 2014). The CLUDs datasets show that 2.9% of land cover in China had been changed during the 5 years. Then, a total of 13295 point-based samples in China were filtered from the whole sample set used in the study. Among point-based samples in China, there were 547 samples located in changed areas, accounting for 4.1% of all samples in China. Therefore, the point-based samples we used are enough to capture LC changes between 2010 and 2015 in China.

Furthermore, an accuracy comparison between the GLC-2015 and GlobeLand30 was conducted in changed areas of China using 144 validation samples for 2015 (Table R1). Results show that the GLC-2015 performed better in changed areas with OA reaching 74.3%, compared to OA of 43.5% for the GlobeLand30. This indicated to some extent that the GLC-2015 did not inherit inaccurate information about the GlobeLand30 due to LC changes. Especially, the GlobeLand30 showed low PA of 23.8% and 32.1% for impervious surfaces and forest. This is because these two LCs changed significantly from 2010 to 2015 (Liu et al., 2014), so the GlobeLand30 in 2010 failed to describe the changed land cover classes during five years. However, the GLC-2015 improved PA and UA for LCs except for permanent snow and ice compared to the GlobeLand30. This improvement in the changed areas of China is because, with the help of the reliability evaluation of each input map, the fusion depended more on other maps than GlobeLand30. Therefore, it is reasonable to believe that LC changes caused by the 5-year interval of the GlobeLand30 can be largely avoided in the fusion based on the DSET method.

Table R1. Comparison of mapping accuracy for the GLC-2015 and GlobeLand30 in changed areas of China.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Impervious surfaces	Bare land	Permanent snow and ice	OA
GLC-2015	PA	0.750	0.793	0.809	0.50	0.588	1.00	0.826	0.500	1.00	**0.743**
	UA	0.857	0.821	0.531	1.00	0.588	0.714	0.864	0.833	1.00	
Globeland30	PA	0.633	**0.321**	0.583	0	0.250	0.400	**0.238**	0.500	1.00	**0.435**
	UA	0.413	0.750	0.212	0	1.00	0.666	1.00	0.400	1.00	

References:

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., Zhou, W., Zhang, S., Li, R., Yan, C., Wu, S., Shi, X., Jiang, N., Yu, D., Pan, X., and Chi, W.: Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s, Dili Xuebao/Acta Geogr. Sin., 69, 3–14, https://doi.org/10.11821/dlxb201401001, 2014.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data, 13, 3907-3925, https://doi.org/10.5194/essd-13-3907-2021, 2021.

Comment #1-3. It's not sound to determine the location of input samples by FROM_GLC. The results show that FROM_GLC has the lowest UA for shrubs, grasslands, and wetlands, suggesting a large number of omissions. In this way, the area ratio of these LCs will be underestimated, as well as their number within the input samples. Consequently, the mapping performance of the final map will be biased. This may be one of the possible reasons for the poor performance of these LCs in the GLC-2015. Perhaps the GlobeLand30 is a better choice, given its high quality.

Response: Thanks for the comment. We agree that the GlobeLand30 has high quality and has been used to collect samples in many researches (Ma et al., 2017; Zhang et al., 2020; Zhang et al., 2023). However, it was too time-consuming to re-select more than 200,000 point-based samples over the glob using the GlobeLand30 as the reference map. In the study, the FROM_GLC was selected because it has the same date time and similar classification with our target map (GLC-2015). Given that the FROM_GLC shows relatively low accuracy for some special LCs, it was only used to provide approximate information about size and location of samples for each LC class. **Notably, the point-based samples did not inherit the land cover class from the FROM_GLC. Instead, we visually interpreted all the points according to Google Earth high-resolution images and labeled them**.

Moreover, we compared the overall accuracy of the fusion method with samples collected from FROM_GLC and Globeland30 in four grids, as listed in Table R2. It can be found that the accuracy difference between Method 1 and Method 2 was small in most grids. In addition, the fusion method with samples derived from Globeland30 (Method 2) did not perform better in every grid.

Table R2. Comparison the overall accuracy of the two fusion methods. Method 1 and 2 denote the fusion method with samples derived from FROM_GLC and Globleland30, respectively.

Grid ID	Method 1	Method 2
0059	0.785	0.643
0258	0.709	0.730
0608	0.769	0.754
0671	0.968	0.972

References:

Ma, X., Tong, X., Liu, S., Luo, X., Xie, H., and Li, C.: Optimized sample selection in SVM classification by combining with DMSP-OLS, Landsat NDVI and GlobeLand30 products for extracting urban built-up areas, Remote Sens., 9, 236, https://doi.org/10.3390/rs9030236, 2017.

Zhang, X., Liu, L., Wu, C., Chen, X., Gao, Y., Xie, S., and Zhang, B.: Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google

Earth Engine platform, Earth Syst. Sci. Data, 12, 1625–1648, https://doi.org/10.5194/essd-12-1625-2020, 2020.

Zhang, X., Liu, L., Zhao, T., Chen, X., Lin, S., Wang, J., Mi, J., and Liu, W.: GWL_FCS30: a global 30 m wetland map with a fine classification system using multi-sourced and time-series remote sensing imagery in 2020, Earth Syst. Sci. Data, 15, 265–293, https://doi.org/10.5194/essd-15-265-2023, 2023.

Comment #1-4. The incorporation of GLC_FCS30 is problematic, as it adopted a detailed classification system (level-2) in some places. I think this will lead to geographical accuracy biases, which however have not been tested in the current assessment.

Response: Thanks for the comment. In our study, all the level-2 LC classes of GLC_FCS30 were converted into the level-1 classes since we adopted a simple classification system that contains 10 major classes. When we evaluated the accuracy of each GLC product, we focused on the reliability of the level-1 classes. If the GLC_FCS30 has lower quality in level-1 classes, its LC classes will contribute less to the fusion.

Comment #1-5. Finally, I'm not sold by the story that some problems (e.g., some LCs always possess lower accuracies and poor performance in areas with more disagreements) could be solved when more reliable maps are available in the future. It would be more useful for the community to tackle these issues here, wouldn't it?

Response: Thanks for the comment. We have used three national-scale land cover data, including the National Land Cover Database 2016 (NLCD2016) for the year 2016 (Yang et al., 2018), China's land-use/cover datasets (CLUDs) (Liu et al., 2014) for 2015 and the annual China land cover dataset (CLCD) (Yang and Huang, 2021) for 2015, as candidate maps for fusion. To figure out whether the GLC-2015 performed better when three national-scale land cover products were used, we compared mapping results without national-scale data used (previous GLC-2015) and with national-scale data used (updated GLC-2015).

Assessed with point-based samples (Tables R3 – R4), **the OA of the updated GLC-2015 over China and the United States achieved 88.8% and 91.0%, which had an improvement of 8.3% and 11.6% compared to the OA of previous GLC-2015** (80.5% for China and 79.4% for the USA). For each land cover class, the updated GLC-2015 performed better than the previous one in both nations. As for shrubland and wetland, the previous GLC-2015 showed relatively low accuracy. Compared to the previous GLC-2015, the updated one greatly improved the mapping accuracy for these two land cover classes in two nations.

Table R3. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in China.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.795	0.949	0.802	0.263	0.334	0.844	0.818	0.873	0.810	**0.805**
	UA	0.862	0.811	0.738	0.657	0.682	0.730	0.918	0.856	0.870	**(0.763)**
Updated	PA	0.844	0.965	0.968	0.316	0.598	0.896	0.905	0.891	0.793	**0.888**
	UA	0.930	0.928	0.803	0.923	0.870	0.741	0.899	0.962	0.958	**(0.864)**

Table R4. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in USA.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.858	0.972	0.865	0.556	0.685	0.935	0.767	0.875	1.00	**0.794**
	UA	0.921	0.742	0.665	0.975	0.804	0.921	0.891	0.467	0.667	**(0.754)**
Updated	PA	0.890	0.958	0.917	0.869	0.903	0.935	0.867	0.911	1.00	**0.910**
	UA	0.944	0.932	0.815	0.972	0.878	0.977	0.903	0.689	1.00	**(0.893)**

When comparing the mapping performance at global scale (Table R5), it was found that **the OA of the updated GLC-2015 was 1.6% higher than the previous one**. In addition, the PA and OA of the updated GLC-2015 had a slightly improvement for almost all the land cover classes.

Table R5. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 at global scale.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Tundra	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.755	0.925	0.713	0.412	0.395	0.874	0.669	0.857	0.881	0.891	**0.780**
	UA	0.864	0.797	0.504	0.815	0.708	0.852	0.833	0.795	0.776	0.928	**(0.739)**
Updated	PA	0.778	0.910	0.739	0.435	0.622	0.875	0.667	0.869	0.883	0.891	**0.795**
	UA	0.878	0.823	0.535	0.841	0.690	0.863	0.839	0.802	0.789	0.937	**(0.757)**

The quantitative comparison was also conducted in the areas of low inconsistency, moderate inconsistency, and high inconsistency, as listed in Table R6. The updated GLC-2015 performed better than the previous GLC-2015 in three areas, with an accuracy improvement of 0.3% in areas of low inconsistency, 0.7% in areas of moderate inconsistency, and 3.9% in areas of high inconsistency.

Table R6. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in three areas.

	Previous		**Updated**	
	OA	Kappa	OA	Kappa
Areas of low inconsistency	0.948	0.931	0.951	0.938
Areas of moderate inconsistency	0.743	0.701	0.760	0.723
Areas of high inconsistency	0.528	0.450	0.567	0.498

Overall, the updated GLC-2015 obtained higher overall accuracy and had better performance in nearly all the LC classes than the previous one in multiple scales. Therefore, we can conclude that it is helpful to use more national-scale products to improve the performance of the GLC-2015 in some LC classes and regions.

We have employed three national-scale products in the fusion process. The description of three national-scale products had been added in Section 2.1 as follows:

"Land cover products which focus on a national scale are more likely to possess higher accuracy because they were produced by experts who have good knowledge of land cover classes nationally. Thus, the National Land Cover Database 2016 (NLCD 2016) for the year 2016 (Yang et al., 2018), China's land-use/cover datasets (CLUDs) (Liu et al., 2014) for 2015, and the annual China land cover dataset (CLCD)

(Yang and Huang, 2021) for 2015 were also included in the fusion. NLCD 2016 database, which provides continuous and accurate information about land cover and change from 2001 to 2016 at an interval of 2 or 3 years, was produced based on a pixel- and object-based approach and an effective post-classification process (Yang et al., 2018). The level-1 and level-2 overall accuracy of NLCD 2016 database for 2016 was 90.6% and 86.4%, respectively (Wickham et al., 2021). CLUDs, developed by the digital interpretation method using Landsat images, provide land cover information over China from 1980s to 2015. The overall accuracy of CLUDs reached 94.3% and 91.2% for level-1 and level-2 land cover classes, respectively (Liu et al., 2014). CLCD was generated with stable training samples derived from CLUDs and Landsat time series. Assessed with 5463 validation samples, CLCD obtained an overall accuracy of 79.31% (Yang and Huang, 2021).

Table 1. Detailed information of GLC products and national-scale LC products used in this paper.

Product name	Satellite sensors	Year of reference	Access	Literature
Globeland30	Landsat TM/ETM+ HJ-1 A/B	2010	http://www.globallandcover.com/	(Chen et al., 2015)
FROM_GLC	Landsat TM/ETM+/OLI	2015	http://data.ess.tsinghua.edu.cn/	(Gong et al., 2013)
GLC_FCS30	Landsat OLI	2015	https://doi.org/10.5281/zenodo.3986872	(Zhang et al., 2021)
GAUD	Landsat TM/ETM+/OLI	2015	https://doi.org/10.6084/m9.figshare.11513178.v1	(Liu et al., 2020)
GFC	Landsat TM/ETM+	2015	http://earthenginepartners.appspot.com/science-2013-global-forest	(Hansen et al., 2013)
JRC GSW	Landsat TM/ETM+/OLI	2015	http://global-surface-water.appspot.com/	(Pekel et al., 2016)
GMW	ALOS PALSAR Landsat TM/ETM+	2015	https://data.unep-wcmc.org/datasets/45	(Bunting et al., 2018)
NLCD 2016	Landsat TM /OLI	2016	https://www.mrlc.gov/data/nlcd-2016-land-cover-conus	(Yang et al., 2018)
CLUDs	Landsat TM HJ-1 CBERS-1	2015	/	(Liu et al., 2014)
CLCD	Landsat TM/ETM+/OLI	2015	https://doi.org/10.5281/zenodo.4417810	(Yang and Huang, 2021)

” (Revised manuscript, Line 163-177)

The relationship between our classification system and the classification systems of three national-scale land cover products has been added in supplementary material:

“**Table S3. Relationship between our classification system and the classification systems of the three national-scale LC products.**

Id	GLC-2015	CLCD	CLUDs	NLCD 2016
10	Cropland	Cropland	Rice paddy	Pasture
			Bare farmland	Cropland
			Orchard	
20	Forest	Forest	Wooden land	Deciduous forest

				Evergreen forest
				Mixed forest
30	Grassland	Grassland	Grassland, highly-covered	Grassland
			Grassland, medium-covered	
			Grassland, lowly-covered	
40	Shrubland	Shrub	Shrubland	Shrubland
50	Wetland	Wetland	Marshland	Woody wetlands
			Tidal flat	Herbaceous wetlands
			Salt marsh	
			Flooded flat	
60	Water bodies	Water	Rivers	Water
			Lakes	
			Reservoir and ponds	
70	Tundra			
80	Impervious surfaces	Impervious	Urban	Urban, open space
			Rural	Urban, low intensity
			Other construction sites	Urban, med. Intensity
				Urban, high intensity
90	Bare land	Barren	Sandy land	Barren
			Gobi desert	
			Barren	
			Bare rocky land	
100	Permanent snow and ice	Snow/ice	Permanent snow and ice	Ice/snow

” (Supplementary material with change)

Lastly, since we employed three national-scale products in the fusion process, all the related results about the GLC-2015 were updated in the revised manuscript.

References:

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., Zhou, W., Zhang, S., Li, R., Yan, C., Wu, S., Shi, X., Jiang, N., Yu, D., Pan, X., and Chi, W.: Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s, Dili Xuebao/Acta Geogr. Sin., 69, 3-14, https://doi.org/10.11821/dlxb201401001, 2014.

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

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data, 13, 3907-3925, https://doi.org/10.5194/essd-13-3907-2021, 2021.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., and Xian, G.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS J. Photogramm. Remote Sens., 146, 108-123, https://doi.org/10.1016/j.isprsjprs.2018.09.006, 2018.

Reviewer #2:

General comment:

The work put into developing the GLC-2015 product can fulfill the global land cover data pool and may give us more precise information about the Earth system. Before being published, the manuscript still has to go through revisions. Authors should include more product comparisons to further prove the advancement of their product.

Response: Thanks for the comment. These comments are very helpful for revising and improving our paper. The manuscript has been improved according to your and another reviewer's comments. The point-by-point responses are listed below in **blue**. The changes in our manuscript are marked with red.

Comment #2-1. The main concern is the results given in Tables 5 and 6, where it is shown that the accuracy of the GLC-2015 product is not much improved than the other products. It is advised that the authors quantify the area differences for each land cover type at multiple scales, including global, continental, national, and ecoregional scales. Also, the authors should provide more visual comparisons regarding each land cover type with current products, including global-scale data, national-scale data, and other prevalence-used data. The visual comparisons ought to be focused on various vegetation types and climatic zones. For instance, the Amur basin, the Tibetan Plateau, Canada, and coastal mangroves should be taken into account when comparing mapping results for Wetlands. With these comparisons, the authors can state that their product is more robust than other products regarding what land cover types in what regions.

Response: Thanks for the comment. Based on your suggestion, we have quantified the area difference for each land cover class at multiple scales in our revised manuscript.

"**4.3.3 Areal comparison for individual classes**

To assess the similarities and discrepancies between the GLC-2015 and other GLC products, we compared the area of various LC classes at multiple scales, including global, continental, national, and ecoregional scales.

The areal comparison for various classes of different GLC products over the globe is shown in Figure 10. Generally, the areas of water bodies and permanent snow and ice of four GLC products were very similar, which may be related to the similar LC definitions. In contrast, the areas of cropland, forest, grassland, and shrubland in GLC-2015 differed significantly from those in other GLC products. The area of forest in GLC-2015 is much higher than other products. This may be because FROM_GLC and GLC_FCS30 defined forest with tree cover over 15%, while GLC-2015 used a threshold of over 10%. The cropland areas in GLC-2015 and Globeland30 were close, higher than FROM_GLC but lower than GLC_FCS30. Moreover, the FROM_GLC underestimated the cropland area as it had a low producer's accuracy for cropland (see Table 5), which was also demonstrated in previous researches (Liu and Xu, 2021; Zhang et al., 2021). FROM_GLC and Globeland30 shared similar grassland areas since a similar accuracy for grassland was found in these two products. However, the FROM_GLC and Globeland30 significantly overestimated grassland extent, with much bare land misclassified as grassland (Hu et al., 2014). The GLC_FCS30 showed the smallest area for grassland, which might be related to its higher threshold in vegetation cover for grassland. For shrubland, the area difference between GLC-2015 and Globeland30 was minimal, and the areas in FROM_GLC and GLC_FCS30 were similar. Furthermore, the wetland area in FROM_GLC was the lowest among all the products, with a total area of 0.168 million km^2. In contrast, the Globeland30 and GLC_FCS30 exhibited greater wetland extent than GLC-2015 since these two products classified non-wetlands sensitive to water as wetlands (Zhang et al., 2023). In particular, the tundra area in GLC_FCS30 was much smaller than other products. This is mainly because only lichens/mosses in the original classification system of GLC_FCS30 was converted into tundra in the classification system we used, which leads to the omission of tundra. The areas of impervious surfaces in GLC-2015, Globeland30, and GLC_FCS30 were very close and higher than FROM_GLC. For bare land, there was large difference between Globeland30 and other products, while the area in GLC-2015 and GLC_FCS30 was very close.

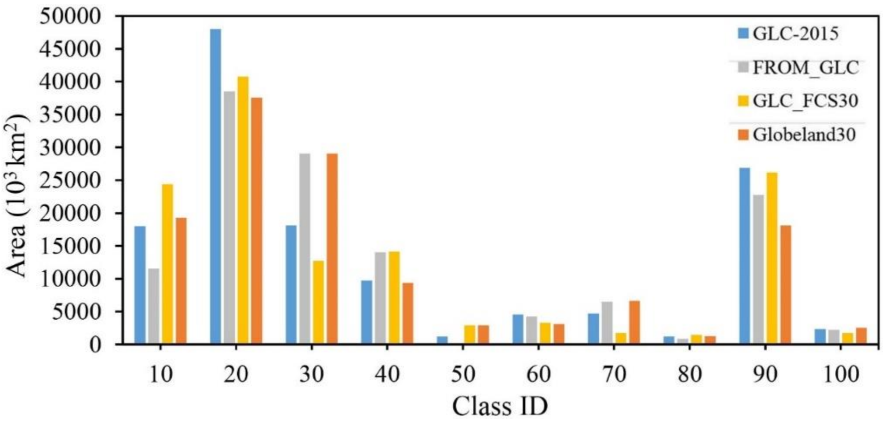

Figure 10. Areal comparison of various land cover classes among GLC products at the global scale. Class IDs 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 denote cropland, forest, grassland, shrubland, wetland, water bodies, tundra, impervious surfaces, bare land, and permanent snow and sea ice, respectively.

The area similarity and difference for various classes of different GLC products were also compared over six continents, the top 40 countries ranked by area, and 21 ecoregions (Figures S5-S7). Overall, the four products showed a similar distribution trend of different classes. For most LC classes, the continental, national, and ecoregional rankings of four products agreed with their ranking at the global scale. Whereas, for grassland and shrubland, the area ranking of four products varied at three different regional scales." (Revised manuscript, Line 576-612)

Table 5. Mapping accuracy of the GLC products with the global point-based samples.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Tundra	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
GLC-2015	PA	0.778	0.910	0.739	0.435	0.622	0.874	0.667	0.870	0.883	0.891	0.795
	UA	0.878	0.823	0.535	0.841	0.690	0.862	0.839	0.802	0.789	0.937	(0.757)
Globeland30	PA	0.752	0.719	0.713	0.245	0.540	0.680	0.769	0.688	0.609	0.821	0.653
	UA	0.786	0.818	0.255	0.428	0.573	0.869	0.577	0.809	0.868	0.905	(0.598)
FROM_GLC	PA	0.389	0.694	0.707	0.411	0.307	0.607	0.712	0.732	0.731	0.881	0.617
	UA	0.671	0.859	0.278	0.422	0.289	0.742	0.686	0.661	0.761	0.773	(0.558)
GLC_FCS30	PA	0.757	0.775	0.452	0.399	0.455	0.604	0.228	0.777	0.809	0.726	0.655
	UA	0.616	0.816	0.384	0.405	0.515	0.808	0.688	0.774	0.645	0.947	(0.591)

"

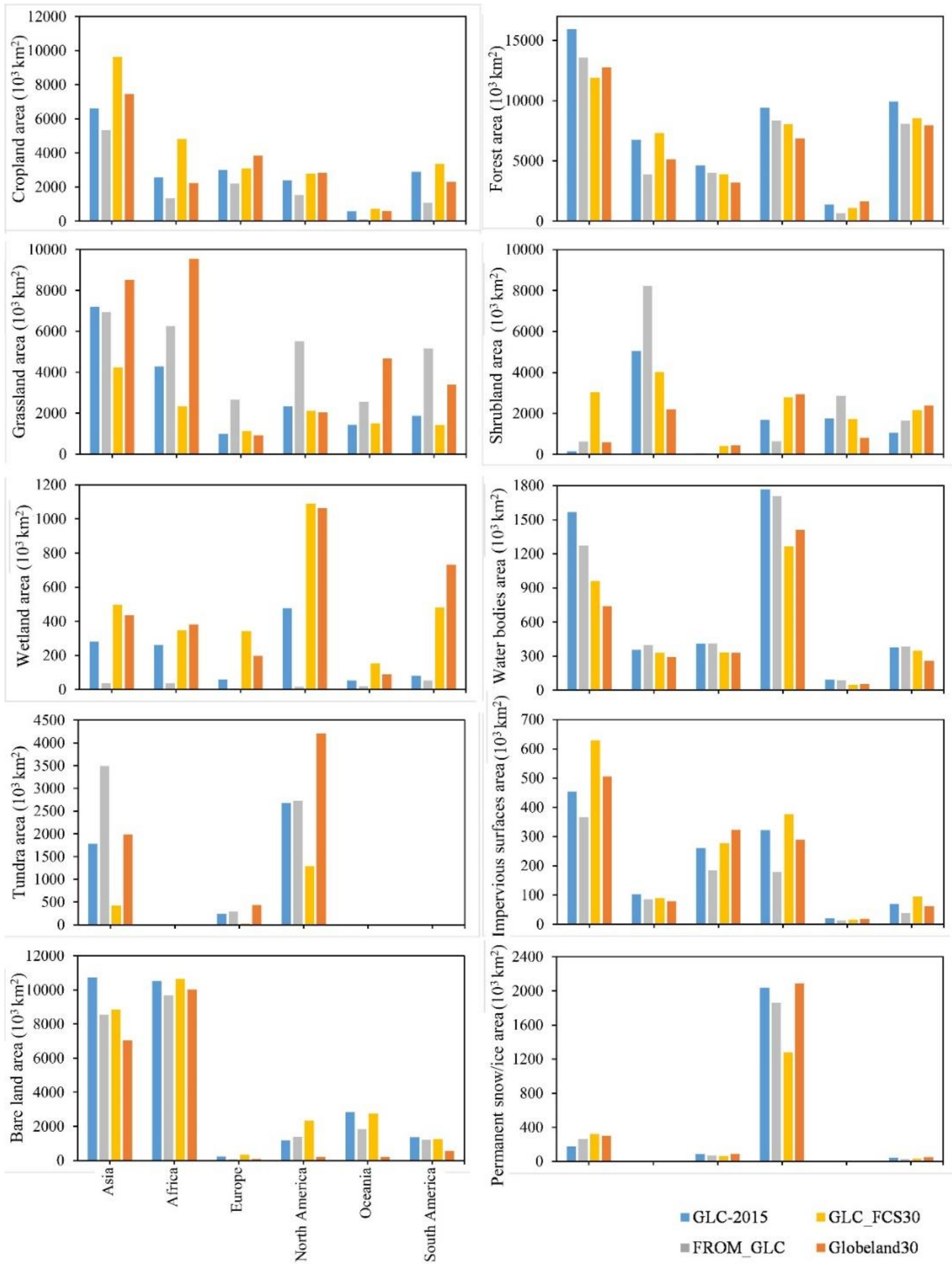

Figure S5. Areal comparison of various land cover classes among GLC products over six continents.

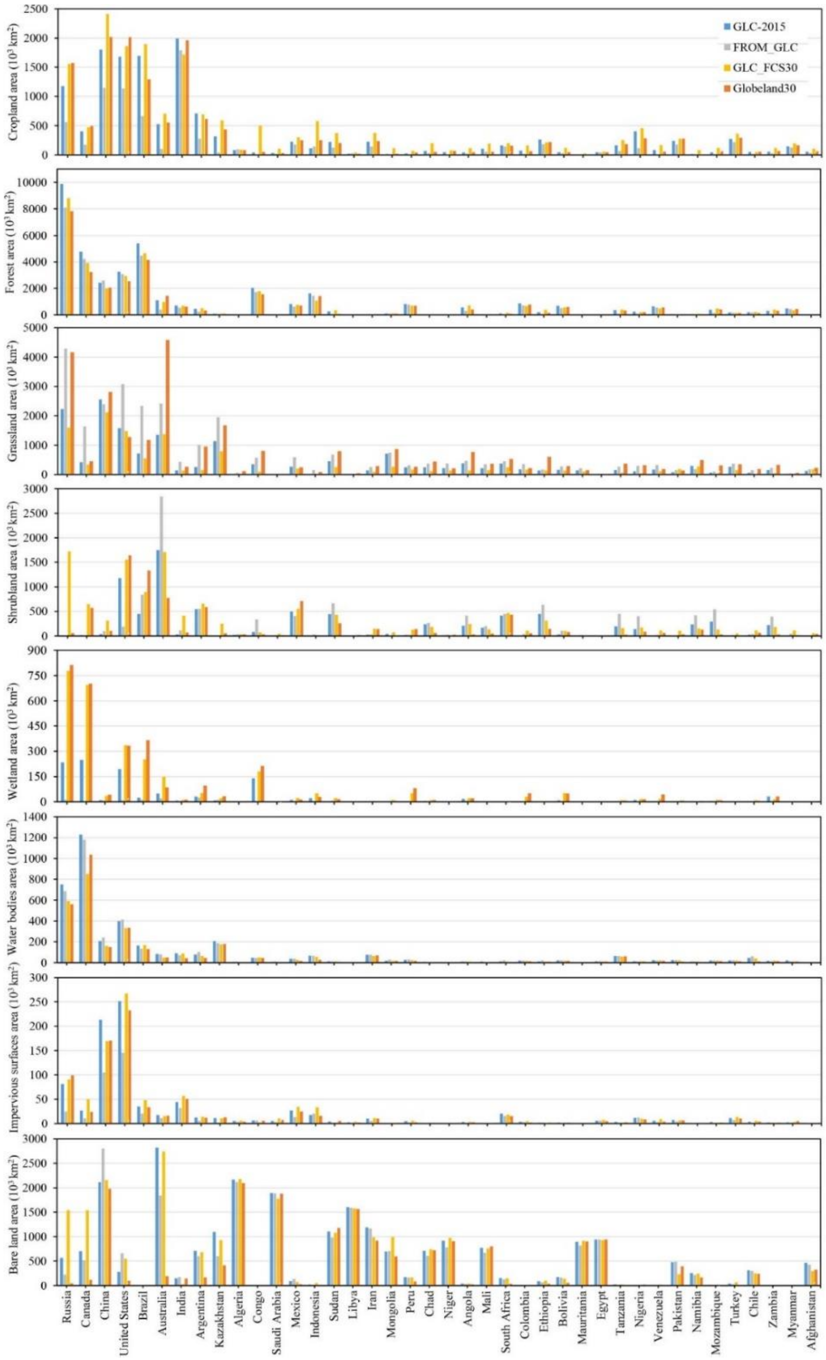

Figure S6. Areal comparison of various land cover classes among GLC products over the top40 countries.

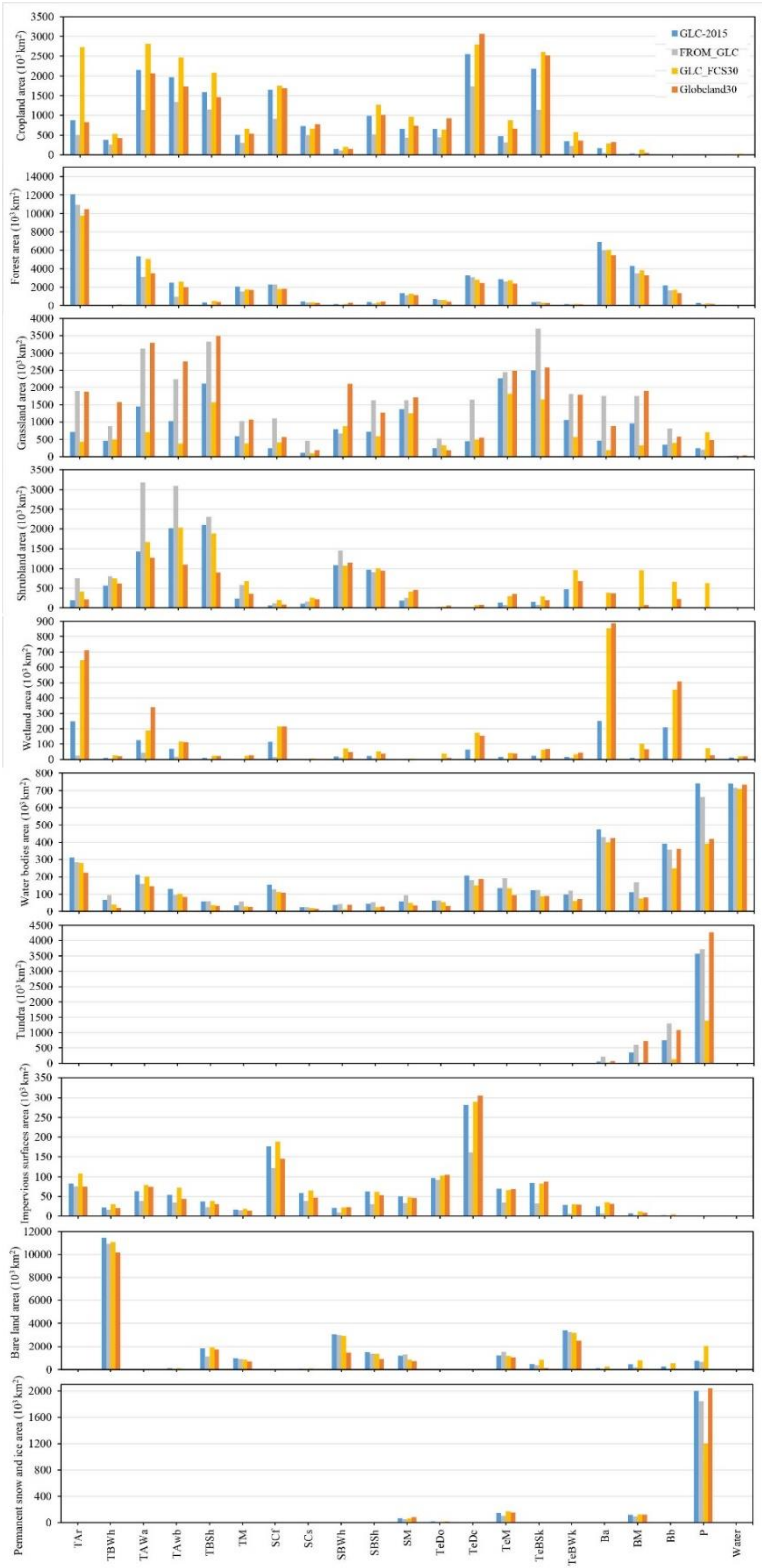

Figure S7. Areal comparison of various land cover classes among GLC products over different ecoregions."
(Supplementary material with change)

Also, we have added the visual comparison for individual classes between GLC-2015 and other widely-used products in our revised manuscript.

“**4.3.4 Visual inter-comparison for individual classes**

The visual comparison of cropland in GLC-2015, Globeland30, FROM_GLC, GLC_FCS30, GSFAD30 (Xiong et al., 2017; Teluguntla et al., 2018), and other national-scale maps was conducted in three local regions (Figure S8). In the Egyptian agricultural area, GLC-2015, FROM_GLC, and GLC-FCS30 shared similar delineation of the cropland and had a good representation of cropland with fine spatial details. Since the date time of the Google Earth image is 2015, Globeland30 missed the newly cultivated cropland. GFASD30 had the largest cropland area among five products but misclassified bare land as cropland. In the agricultural area of Southeastern China, GLC-2015 had an agreement with GFSAD30 and CLCD. Globeland30 and GLC_FCS30 overestimated the area of cropland. As for FROM_GLC, it failed to depict the spatial distribution of cropland and had many omissions. In cropland-dominated areas of the United States, FROM_GLC significantly underestimated the extent of cropland. The other five products exhibited a similar delineation of cropland, but there were little differences in some small areas. For example, Globeland30 misclassified some grassland into cropland, and NLCD 2016 had a good ability to distinguish the farm rack.” (Revised manuscript, Line 613-626)

“

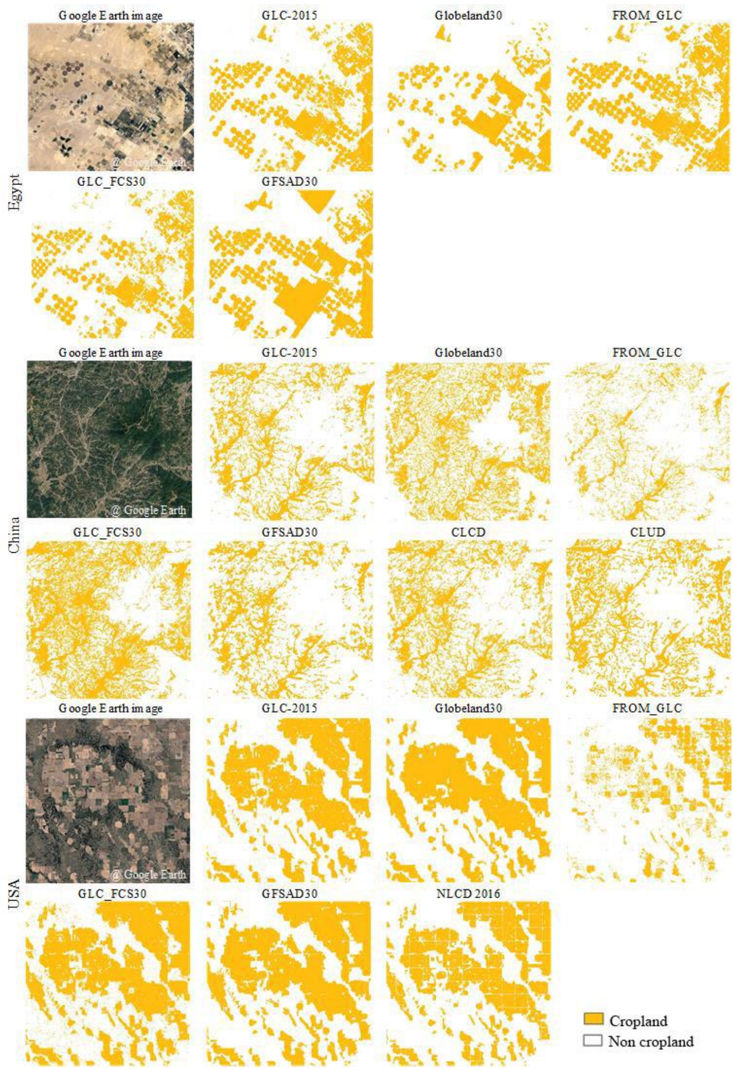

Figure S8. Comparing the crop extent from GLC-2015 and other widely used products in three agricultural regions.” (Supplementary material with change)

"We also compared the performance in the forest of different products in three forest-prevalent regions of Congo, China and the United States (Figure S 9). Overall, GLC-2015 and Globeland30 showed accurate delineation in three regions. FROM_GLC also had good performance for the forest in Congo and USA but overestimated the forest in China, mislabeling shrubland and grassland as forest. Furthermore, GFC tended to miss sparse trees in China, and GLC_FCS30 underestimated the extent of forest in both three regions. As for national-scale products, CLCD and NLCD 2016 had a good ability to identify the details of forest, while CLUD dramatically missed both dense and sparse woodlands." (Revised manuscript, Line 627-633)

"

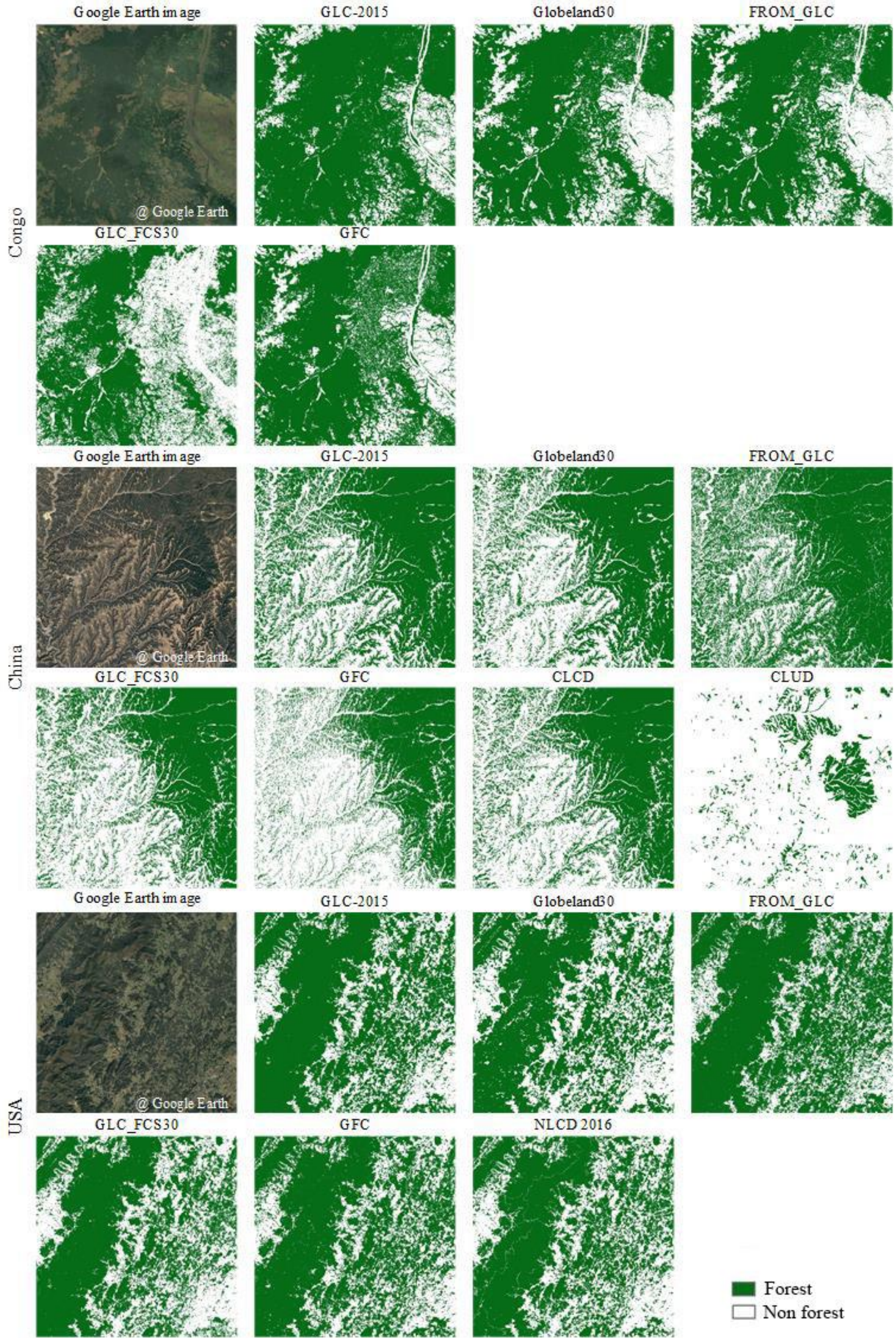

Figure S9. Comparing the forest extent from GLC-2015 and other widely used products in three forest-

dominated regions." (Supplementary material with change)

"Furthermore, to compare the performance in the wetland of GLC-2015 with other global and national-scale products, three wetland regions in South-central Canada, coastal America, and Sundarbans were selected. It can be found that GLC-2015 and Globeland30 had similar representation and performed well in identifying wetlands over three regions (Figure S10). Unexpectedly, FROM_GLC performed poorly in each region, with almost no wetlands captured. GLC_FCS30 also showed unstable quality in three regions. For example, it highly underestimated the wetland area in coastal America and completely mislabeled the mangrove as cropland in Sundarbans. NLCD 2016 and GMW accurately demonstrated the spatial pattern of wetlands, while the CA_wetlands map underestimated the wetland extent because it defined wetlands by wetland frequency of no less than 80% from 2000 to 2016 (Wulder et al., 2018). "
(Revised manuscript, Line 634-642)

"

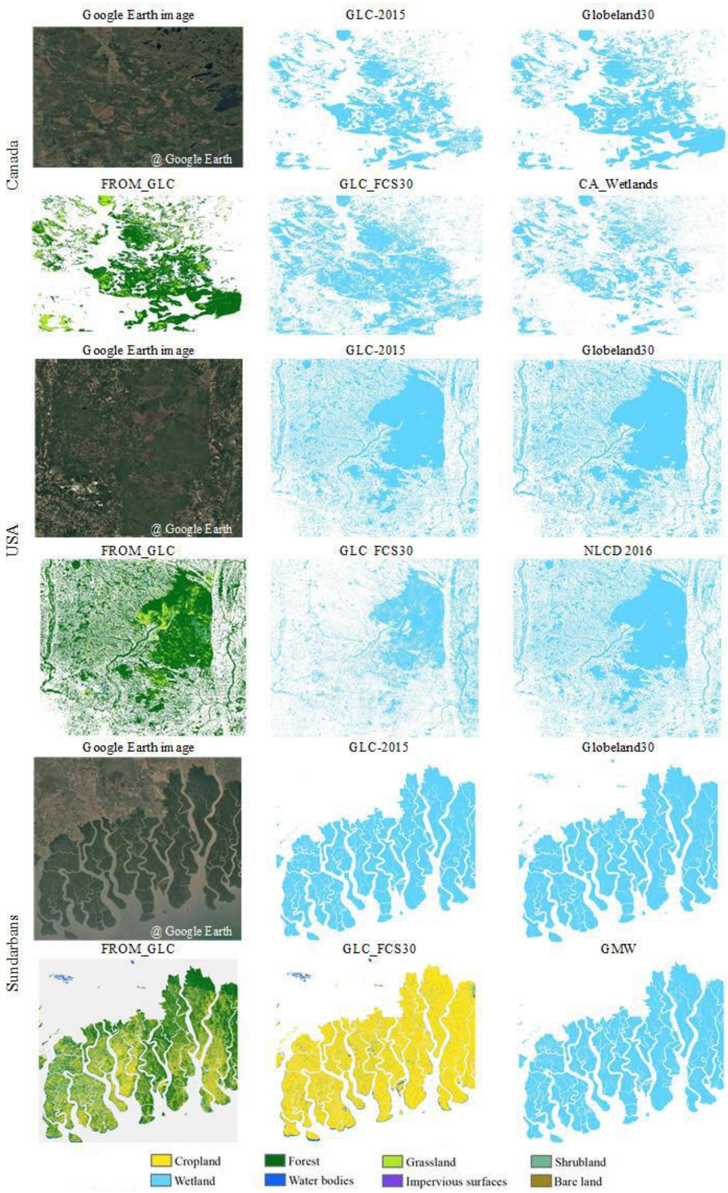

Figure S10. Comparing the wetland extent from GLC-2015 and other widely used products in three wetland-dominated regions." (Supplementary material with change)

"To understand the spatial distribution of impervious surfaces in different products, a comparison of mapping results for three megacities, including Tokyo, Shanghai, and New York, was shown in Figure S11. In Tokyo, a high consistency was found between GLC-2015, FROM_GLC, and GAUD, and both successfully captured the impervious surfaces in peri-urban areas. GLC_FCS30 showed the largest area for impervious surfaces because it misclassified many croplands into impervious surfaces. In Shanghai, GLC_FCS30 underestimated the central city, and CLUD lost the details of impervious surfaces because it was developed using the visual interpretation method. Other products generally had the similar representation and accurately demonstrated the spatial distribution of the city. For New York, the FROM_GLC, GLC_FCS30, and GAUD agreed well with GLC-2015, while Globeland30 and NLCD 2016 had high impervious areas than others." (Revised manuscript, Line 643-652)

"

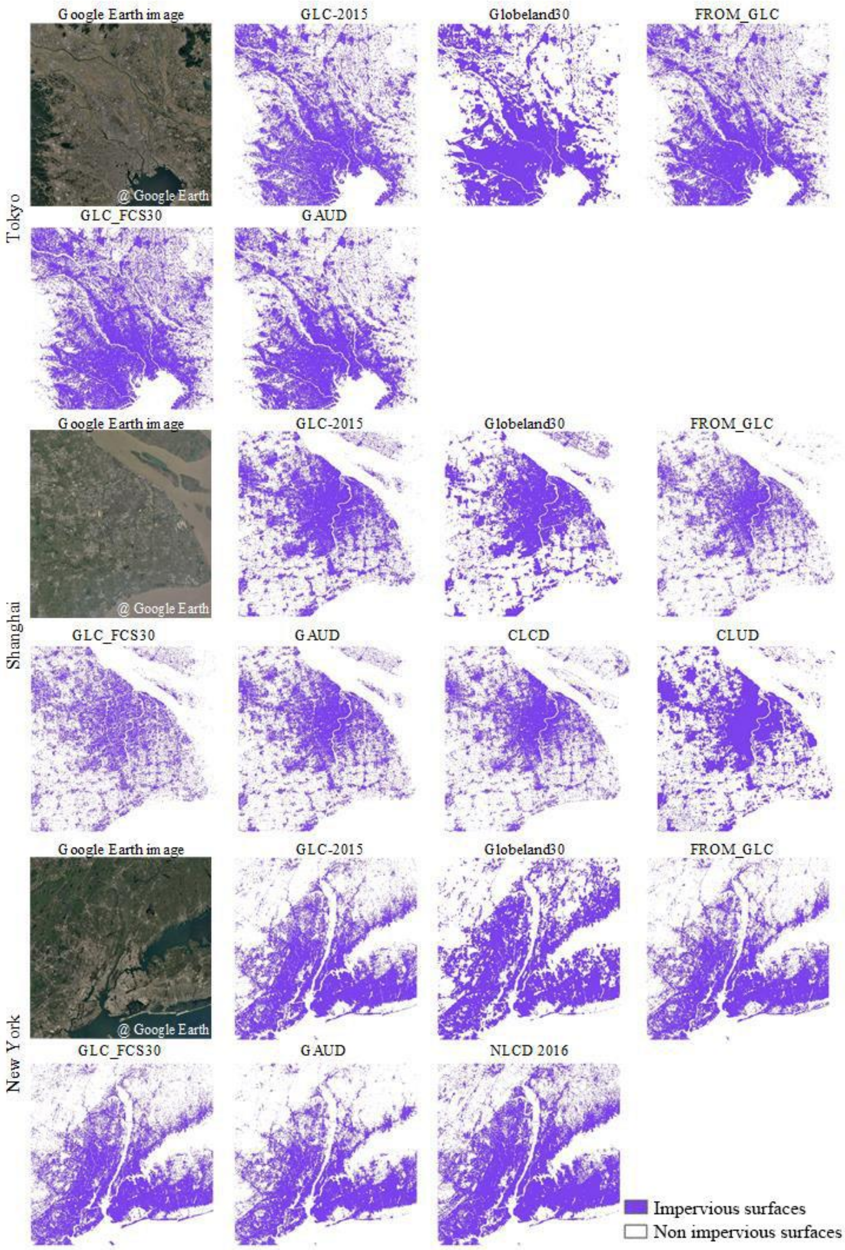

Figure S11. Comparing the impervious extent from GLC-2015 and other widely used products in three megacities." (Supplementary material with change)

References:

Hu, L., Chen, Y., Xu, Y., Zhao, Y., Yu, L., Wang, J., and Gong, P.: A 30 meter land cover mapping of China with an efficient clustering algorithm CBEST, Science China Earth Sciences, 57, 2293-2304, https://doi.org/10.1007/s11430-014-4917-1, 2014.

Liu, K. and Xu, E.: Fusion and correction of multi-source land cover products based on spatial detection and uncertainty reasoning methods in Central Asia, Remote Sens., 13, 244, https://doi.org/10.3390/rs13020244, 2021.

Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., Yadav, K., and Huete, A.: A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform, ISPRS J. Photogramm. Remote Sens., 144, 325-340, https://doi.org/10.1016/j.isprsjprs.2018.07.017, 2018.

Wulder, M. A., Li, Z., Campbell, E. M., White, J. C., Hobart, G., Hermosilla, T., and Coops, N. C.: A National Assessment of Wetland Status and Trends for Canada's Forested Ecosystems Using 33 Years of Earth Observation Satellite Data, 10.3390/rs10101623, 2018.

Xiong, J., Thenkabail, P. S., Tilton, J. C., Gumma, M. K., Teluguntla, P., Oliphant, A., Congalton, R. G., Yadav, K., and Gorelick, N.: Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine, 10.3390/rs9101065, 2017.

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

Zhang, X., Liu, L., Zhao, T., Chen, X., Lin, S., Wang, J., Mi, J., and Liu, W.: GWL_FCS30: a global 30 m wetland map with a fine classification system using multi-sourced and time-series remote sensing imagery in 2020, Earth Syst. Sci. Data, 15, 265-293, https://doi.org/10.5194/essd-15-265-2023, 2023.

Comment #2-2. Why not use national-scale land cover data (e.g., CLUD, CLCD, NLCD) as candidates if the authors have acknowledged their value in lines 756 to 762? National-scale land cover products with the participation of local experts are more accurate than global products and should therefore be considered for inclusion in the development of GLC-2015.

Response: Thanks for the comment. We think it is an excellent suggestion. We have used three national-scale land cover data, including CLUD2015, CLCD2015 for China, and NLCD2016 for America, as candidate maps for fusion to improve the mapping performance of GLC-2015. Since the classification system of CLUD2015 and NLCD2016 is different with our classification system, the land cover classes of CLUD2015 and NLCD2016 were reclassified according to the classification system we adopted. Then, we re-performed land cover mapping with national-scale products included.

To figure out whether the GLC-2015 performed better when three national-scale land cover products were used, we compared mapping results without national-scale data (previous GLC-2015) and with national-scale data (updated GLC-2015) at the national scale and global scale.

Assessed with point-based samples at national scale (Tables R3-4), the OA of the updated GLC-2015 over China and the United States achieved 88.8% and 91.0%, which had an improvement of 8.3% and 11.6% compared to the OA of previous GLC-2015 (80.5% for China and 79.4% for the USA). For each land cover class, the updated GLC-2015 performed better than the previous one in both nations. As for shrubland and wetland, the previous GLC-2015 showed relatively low accuracy. Compared to the previous GLC-2015, the updated one greatly improved the mapping accuracy for these two land cover classes in two nations.

Table R3. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in China.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.795	0.949	0.802	0.263	0.334	0.844	0.818	0.873	0.810	**0.805**
	UA	0.862	0.811	0.738	0.657	0.682	0.730	0.918	0.856	0.870	**(0.763)**
Updated	PA	0.844	0.965	0.968	0.316	0.598	0.896	0.905	0.891	0.793	**0.888**
	UA	0.930	0.928	0.803	0.923	0.870	0.741	0.899	0.962	0.958	**(0.864)**

Table R4. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in USA.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.858	0.972	0.865	0.556	0.685	0.935	0.767	0.875	1.00	**0.794**
	UA	0.921	0.742	0.665	0.975	0.804	0.921	0.891	0.467	0.667	**(0.754)**
Updated	PA	0.890	0.958	0.917	0.869	0.903	0.935	0.867	0.911	1.00	**0.910**
	UA	0.944	0.932	0.815	0.972	0.878	0.977	0.903	0.689	1.00	**(0.893)**

When comparing the mapping performance at global scale (Table R5), it was found that **the OA of the updated GLC-2015 was 1.6% higher than the pervious one**. In addition, the PA and OA of the updated GLC-2015 had a slightly improvement for almost all the land cover classes.

Table R5. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 at global scale.

		Cropland	Forest	Grassland	Shrubland	Wetland	Water bodies	Tundra	Impervious surfaces	Bare land	Permanent snow and ice	OA (Kappa coefficient)
Previous	PA	0.755	0.925	0.713	0.412	0.395	0.874	0.669	0.857	0.881	0.891	**0.780**
	UA	0.864	0.797	0.504	0.815	0.708	0.852	0.833	0.795	0.776	0.928	**(0.739)**
Updated	PA	0.778	0.910	0.739	0.435	0.622	0.875	0.667	0.869	0.883	0.891	**0.795**
	UA	0.878	0.823	0.535	0.841	0.690	0.863	0.839	0.802	0.789	0.937	**(0.757)**

The quantitative comparison was also conducted in the areas of low inconsistency, moderate inconsistency, and high inconsistency, as listed in Table R6. The updated GLC-2015 performed better than the previous GLC-2015 in three areas, with an accuracy improvement of 0.3% in areas of low inconsistency, 0.7% in areas of moderate inconsistency, and 3.9% in areas of high inconsistency.

Table R6. Comparison of mapping accuracy for the previous GLC-2015 and updated GLC-2015 in three areas.

	Previous		**Updated**	
	OA	Kappa	OA	Kappa
Areas of low inconsistency	0.948	0.931	0.951	0.938
Areas of moderate inconsistency	0.743	0.701	0.760	0.723

Areas of high inconsistency	0.528	0.450	0.567	0.498

Correspondingly, we have added the description of three national-scale products in Section 2.1 as follows:

"Land cover products which focus on a national scale are more likely to possess higher accuracy because they were produced by experts who have good knowledge of land cover classes nationally. Thus, the National Land Cover Database 2016 (NLCD 2016) for the year 2016 (Yang et al., 2018), China's land-use/cover datasets (CLUDs) (Liu et al., 2014) for 2015, and the annual China land cover dataset (CLCD) (Yang and Huang, 2021) for 2015 were also included in the fusion. NLCD 2016 database, which provides continuous and accurate information about land cover and change from 2001 to 2016 at an interval of 2 or 3 years, was produced based on a pixel- and object-based approach and an effective post-classification process (Yang et al., 2018). The level-1 and level-2 overall accuracy of NLCD 2016 database for 2016 was 90.6% and 86.4%, respectively (Wickham et al., 2021). CLUDs, developed by the digital interpretation method using Landsat images, provide land cover information over China from 1980s to 2015. The overall accuracy of CLUDs reached 94.3% and 91.2% for level-1 and level-2 land cover classes, respectively (Liu et al., 2014). CLCD was generated with stable training samples derived from CLUDs and Landsat time series. Assessed with 5463 validation samples, CLCD obtained an overall accuracy of 79.31% (Yang and Huang, 2021).

Table 1. Detailed information of GLC products and national-scale LC products used in this paper.

Product name	Satellite sensors	Year of reference	Access	Literature
Globeland30	Landsat TM/ETM+ HJ-1 A/B	2010	http://www.globallandcover.com/	(Chen et al., 2015)
FROM_GLC	Landsat TM/ETM+/OLI	2015	http://data.ess.tsinghua.edu.cn/	(Gong et al., 2013)
GLC_FCS30	Landsat OLI	2015	https://doi.org/10.5281/zenodo.3986872	(Zhang et al., 2021)
GAUD	Landsat TM/ETM+/OLI	2015	https://doi.org/10.6084/m9.figshare.11513178.v1	(Liu et al., 2020)
GFC	Landsat TM/ETM+	2015	http://earthenginepartners.appspot.com/science-2013-global-forest	(Hansen et al., 2013)
JRC GSW	Landsat TM/ETM+/OLI	2015	http://global-surface-water.appspot.com/	(Pekel et al., 2016)
GMW	ALOS PALSAR Landsat TM/ETM+	2015	https://data.unep-wcmc.org/datasets/45	(Bunting et al., 2018)
NLCD 2016	Landsat TM /OLI	2016	https://www.mrlc.gov/data/nlcd-2016-land-cover-conus	(Yang et al., 2018)
CLUDs	Landsat TM HJ-1 CBERS-1	2015	/	(Liu et al., 2014)
CLCD	Landsat TM/ETM+/OLI	2015	https://doi.org/10.5281/zenodo.4417810	(Yang and Huang, 2021)

" (Revised manuscript, Line 163-177)

The relationship between our classification system and the classification systems of three national-scale land cover products has been added in supplementary material:

"**Table S3. Relationship between our classification system and the classification systems of the three national-scale LC products.**

Id	GLC-2015	CLCD	CLUDs	NLCD 2016
10	Cropland	Cropland	Rice paddy	Pasture
			Bare farmland	Cropland
			Orchard	
20	Forest	Forest	Wooden land	Deciduous forest
				Evergreen forest
				Mixed forest
30	Grassland	Grassland	Grassland, highly-covered	Grassland
			Grassland, medium-covered	
			Grassland, lowly-covered	
40	Shrubland	Shrub	Shrubland	Shrubland
50	Wetland	Wetland	Marshland	Woody wetlands
			Tidal flat	Herbaceous wetlands
			Salt marsh	
			Flooded flat	
60	Water bodies	Water	Rivers	Water
			Lakes	
			Reservoir and ponds	
70	Tundra			
80	Impervious surfaces	Impervious	Urban	Urban, open space
			Rural	Urban, low intensity
			Other construction sites	Urban, med. Intensity
				Urban, high intensity
90	Bare land	Barren	Sandy land	Barren
			Gobi desert	
			Barren	
			Bare rocky land	
100	Permanent snow and ice	Snow/ice	Permanent snow and ice	Ice/snow

” (Supplementary material with change)

Lastly, since we employed three national-scale products in the fusion process, all the related results about the GLC-2015 were updated in the revised manuscript. Meanwhile, we have re-uploaded the mapping results and changed the access to the GLC-2015 as follows:

“The improved global land cover map in 2015 with 30 m resolution is available at https://doi.org/10.6084/m9.figshare.22358143.v2 (Li et al., 2022).” (Revised manuscript, Line 847-848)

References:

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., Zhou, W., Zhang, S., Li, R., Yan, C., Wu, S., Shi, X., Jiang, N., Yu, D., Pan, X., and Chi, W.: Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s, Dili Xuebao/Acta Geogr. Sin., 69, 3-14, https://doi.org/10.11821/dlxb201401001, 2014.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data, 13, 3907-3925, https://doi.org/10.5194/essd-13-3907-2021, 2021.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., and Xian, G.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS J. Photogramm. Remote Sens., 146, 108-123, https://doi.org/10.1016/j.isprsjprs.2018.09.006, 2018.

Comment #2-3. One of the motivations behind the development of global land use land cover data is the expectation of a more detailed classification scheme (e.g., GLC_FCS30’s scheme). The authors may provide insights on how to meet this expectation with the proposed fusion method.

Response: Thanks for suggestion. The classification system used in our study was determined with the classification system of all the input maps taken into consideration. All single-class land cover products provided information for only one class with no subclass. For example, the GFC for 2015 provided the extent of forest but could not tell users where were broadleaf and needleleaf trees. For multi-class land cover products, most had a simple classification system. For example, the GlobeLand30 used a classification system that contained only 10 first-level classes. These input maps with a simple classification scheme had no contribution to the level-2 detailed land cover classes. So, the existing products used may be not enough to generate a new GLC product with a detailed classification system using the proposed fusion method. Therefore, we adopted a classification system that contains 10 major land cover classes. In future work, efforts will be made to improve the diversity of land cover classes in our GLC-2015 product. Since the data fusion method we proposed can be easily used to integrate a wide variety of land cover maps for different regions, an improved global land cover product with a fine classification system rather than a simple one-level classification system will be developed when land cover products with more diverse land cover classes are available. In addition, when there are enough source maps with detailed land cover classes, the large discrepancies in the definition and criteria to distinct level-2 land cover classes might still hinder the transformation into a uniform system. So, a feasible framework for the conversion of different level-2 classification systems into a uniform system should be created in the future. For example, the semantic similarity between each input map’s scheme and the target classification system may facilitate the harmonization (Gao et al., 2020).

We have added the discussion about the classification system and the further work to improve the diversity of land cover classes in our map. The detailed revision can be seen below.

"Lastly, most candidate LC products used a simple classification system without a level-2 classification system, so they made no contributions to a more detailed classification system when they served as source data for data fusion. Although some maps provided detailed LC classification results, such as the GLC_FCS30 and FROM_GLC for 2015, there might be several challenges in the standardization and uniformity of level-2 classification systems due to the large discrepancies in the definition and criteria. Therefore, the GLC-2015 adopted a simple classification system containing 10 major LC classes. In future work, measures will be taken to meet the expectation of a more detailed classification system for GLC mapping. An improved GLC product with a detailed classification system rather than a simple one-level classification system can be further developed based on the highly applicable and general DSET method whenever more products with diverse LC classes are available. Additionally, a feasible framework for the conversion of different level-2 classification systems into a uniform system should be developed." (Revised manuscript, Line 834-845)

References:

Gao, Y., Liu, L., Zhang, X., Chen, X., Mi, J., and Xie, S.: Consistency Analysis and Accuracy Assessment of Three Global 30-m Land-Cover Products over the European Union using the LUCAS Dataset, Remote Sen., 12, 3479, https://doi.org/10.3390/rs12213479, 2020.

Comment #2-4. The format of the reference needs to be standardized.

Response: Thanks for suggestion. We have standardized the format of the reference.

Reviewer #3:

General comment:

The other two reviewers have provided very professional suggestions for data and manuscript improvements. I fully agree with their opinions. Generally, I think this data has some value but I have two major concerns.

Response: Thanks for the comment. These comments are very helpful for revising and improving our paper. The manuscript has been improved according to your and another reviewer's comments. The point-by-point responses are listed below in **blue**. The changes in our manuscript are marked with red.

Comment #3-1. The first is about the classification system used. This study assimilated various information from other land use data products. However, the definitions of different land types greatly differ in each classification system. For example, in your data, 'cropland' is defined as 'Land areas used for food production and animal feed.', which means pasture was classified as cropland. However, pasture is classified as grassland in GLC_FCS30. When all these signals were combined, the approach used could cause problems depending on the LC definitions adopted. Moreover, during the validation process, how this definition differences were treated? I think, for example for the cropland, it does not indicate that GLC_FCS30 was wrong (as claimed in Line565), but because of the differed definitions. Besides, it is unclear how you deal with fallow land, which is often been mixed with grassland in classifications (need to be separated by examine the temporal information). Another example, in FROM_GLC, the forest was defined as tree cover >=10%, and GlobeLand30 defined forest as land with tree cover above 30% and also include sparse woodland with tree cover between 10%-30%. These are different from the fixed threshold of 30% adopted in this study. In other words, I do not agree that other datasets are 'wrong' (as claimed in Line565 and other places).

Response: Thank you very much for the insightful comment on the classification system adopted in the study. We are sorry for the unclear explanation in the previous manuscript. We agree that LC definitions in each GLC product have great difference, which might cause uncertainties when integrating various products. We have made every effort to reduce the uncertainties from the discrepancy between various classification systems.

First, we employed a simple classification system containing 10 major LC classes for GLC-2015 (Table 2). This classification system is the same as that adopted by Globeland30. We sincerely apologize for not clearly describing the definitions of forest and shrubland in the classification system used in the study, which confused the reviewers and our readers. Correspondingly, we have supplemented the definitions of these two categories in Table 2. **In our classification system, the forest includes trees with a tree canopy cover over 30% and sparse trees with a tree canopy cover between 10% - 30%, which is the same as Globeland30.**

"**Table 2. Classification system adopted in this paper.**

Id	LC class	Definition
10	Cropland	Land areas used for food production and animal feed.
20	Forest	Land areas dominated by trees with tree canopy cover over 30%, and sparse trees with tree canopy cover between 10%-30%.
30	Grassland	Land areas dominated by natural grass with a cover over 10%.
40	Shrubland	Land areas dominated by shrubs with a cover over 30%, including

		mountain shrubs, deciduous shrubs, evergreen shrubs and desert shrubs with a cover over 10%.
50	Wetland	Land areas dominated by wetland plants and water bodies.
60	Water bodies	Land areas covered with accumulated liquid water.
70	Tundra	Land areas dominated by lichen, moss, hardly perennial herb and shrubs in the polar regions.
80	Impervious surfaces	Land areas covered with artificial structures.
90	Bare land	Land areas with scarce vegetation with a cover lower than 10%.
100	Permanent snow and ice	Land areas dominated by permanent snow, glacier and icecap.

” (Revised manuscript, Line)

Second, according to the classification system adopted in the study, the original LC classes of FROM_GLC and GLC_FCS30 were converted into the 10 target land cover classes based on the similarity of LC definition (Table S2). It can be found that 10 level-1 classes of the FROM_GLC and 9 level-0 classes of the GLC_FCS30 are the same as the LC classes used in the target classification system despite that the definitions of some classes differ. For the FROM_GLC, all the level-2 classes, excluding pasture, were aggregated into their corresponding level-1 classes. Note that the cropland in our classification system was defined as land areas for food production and animal feed. **Therefore, “pasture” in level-2 classes of the FROM_GLC was converted into cropland rather than grassland.** For the GLC_FCS30, all fine LC classes excluding lichens/mosses were aggregated into their corresponding 9 level-0 classes. Although the level-0 classification system of the GLC_FCS30 lacks tundra, lichens/mosses in the level-2 detailed LC classes has little distinction with tundra. **Separately, we transformed Lichens/mosses into the tundra, one of the major classes in our classification system**.

Table S2. Relationship between our classification system and the classification systems of the three GLC products.

Id	GLC-2015	Globeland30	FROM_GLC	GLC_FCS30
10	Cropland	Cultivated land	Rice paddy	Rain-fed cropland
			Greenhouse	Herbaceous cover
			Other/orchard	Tree or shrub cover (orchard)
			Bare farmland	Irrigated cropland
			Pasture	
20	Forest	Forest	Broadleaf, leaf-on	Evergreen broadleaved forest
			Broadleaf, leaf-off	Deciduous broadleaved forest
				Open/closed deciduous broadleaved forest
			Needleleaf, leaf-on	Evergreen needleleaved forest
				Open/closed evergreen needleleaved forest
			Needleleaf, leaf-off	Deciduous needleleaved forest

				Open/closed deciduous needleleaved forest
			Mixed leaf, leaf-on	Mixed leaf forest
			Mixed leaf, leaf-off	
30	Grassland	Grassland	Natural grassland	Grassland
			Grassland, leaf-off	
40	Shrubland	Shrubland		Shrubland
			Shrubland, leaf-on	Evergreen shrubland
			Shrubland, leaf-off	Deciduous shrubland
50	Wetland	Wetland	Marshland	Wetlands
			Mudflat	
			Marshland, leaf-off	
60	Water bodies	Water bodies	Water	Water body
70	Tundra	Tundra	Shrub and brush tundra	
			Herbaceous tundra	Lichens/ mosses
80	Impervious surfaces	Artificial surfaces	Impervious surfaces	Impervious surfaces
90	Bare land	Bare land	Bare land	Sparse vegetation
				Sparse shrubland
				Sparse herbaceous cover
				Bare areas
				Consolidated/unconsolidated bare areas
100	Permanent snow and ice	Permanent snow and ice	Snow	Permanent ice and snow
			Ice	

By carefully considering the original LC definitions in each product and the similarity between various classification systems, we managed to transform these various classification systems into a uniform one with the principle of minimizing potential errors and inconsistencies caused by different classification systems. Even though, there are still uncertainties caused by the harmonization of classification systems. For example, in GLC_FCS30, the sparse herbaceous cover with a vegetation cover below 15% was directly transformed into bare land regardless that our classification system distinguished grassland using vegetation cover threshold of 10%. In this case, herbaceous cover between 10%-15% in GLC_FCS30 was inappropriately transformed into bare land rather than grassland in the study. Due to the different LC definitions, these uncertainties in classification system conversion are inevitable (Zhang et al., 2017). However, we conducted a reliability evaluation of the candidate maps to reduce the effects of uncertainties in classification system conversion on the fusion using the DSET. **Note that all the point-based samples used for reliability evaluation were labeled referring to the LC definitions in our classification system.** When evaluating the reliability of candidate maps for BPA construction in the fusion, **all the maps were assessed under the criterion of the classification system we used**. For instance, herbaceous cover between 10%-15% in GLC_FCS30 was transformed into bare land, while point-based samples in areas with herbaceous cover between 10%-15% were labeled as grassland. In this case, bare land with the threshold between 10%-15% from GLC_FCS30 was confirmed to mismatch our classification system, and the GLC_FCS30 was assessed to have lower accuracy for areas where the mismatched information existed. When we integrated all the maps grid by grid, the mismatched information would contribute less to the output map. **Similarly, during the validation process, the mapping accuracy of Globeland30, FROM_GLC, and GLC_FCS30 was assessed under the criterion of the classification system we used.** In this case, the accuracy assessment results represented the consistency of each product with validation samples labeled referring to our classification system.

As for the fallow land you are concerned with, we treated it as cropland according to the definition in our classification system. So, the fallow land from any other products was converted into cropland. If the candidate maps showed confusion with fallow land and grassland, this misclassification might bias our mapping results since the GLC-2015 was developed based on the integration of the candidate maps.

Corresponding, we have added how we harmonized the different classification systems in the revised manuscript.

"According to the classification system adopted in the study, the original LC classes of FROM_GLC and GLC_FCS30, CLUD for 2015, and NLCD 2016 for 2016 were converted into the 10 target land cover classes based on the similarity of LC definition. Note that cropland in our classification system was defined as land areas for food production and animal feed. Therefore, pasture in level-2 classes of the FROM_GLC was converted into cropland rather than grassland. In addition, lichens/mosses in the level-2 detailed LC classes of GLC_FCS30 was converted into tundra." (Revised manuscript, Line 307-312)

In addition, we have added the discussion about the uncertainties brought by the LC definition differences:

" Third, due to the different LC definitions, uncertainties in classification system conversion are inevitable (Zhang et al., 2017), which might cause problems for the fusion based on the DSET method. However, we conducted a reliability evaluation of the candidate maps to reduce the influence of uncertainties in classification system conversion on the fusion. The point-based samples used for reliability evaluation were labeled referring to the LC definitions in our classification system so that all the maps were evaluated under the criterion of the classification system we used. By the reliability evaluation, the candidate maps were assessed to have lower accuracy for areas with mismatched information. When integrating all the maps grid by grid, the mismatched information would contribute less to the fusion." (Revised manuscript, Line 826-833)

References:

Zhang, M., Ma, M., De Maeyer, P., and Kurban, A.: Uncertainties in classification system conversion and an analysis of inconsistencies in global land cover products, ISPRS Int. J. Geo Inf., 6, 112, https://doi.org/10.3390/ijgi6040112, 2017.

Comment #3-2. My second major concern is the current manuscript reads really arrogant. The authors' attitudes against other data products are not polite. I strongly suggest the authors tone down their remarks against other data products. As I said, the differences are partially from LC various definitions in each product. Please be moderate and modest. You actually built your data relying on these products. The words such as "wrongly", "worst performance" should be avoided, they are too harsh (see a few examples below).

Response: Thank you very much for pointing out this issue. We are really sorry for the inappropriate words against other GLC products. This was not our intention. We have tried our best to revise our manuscript to address this concern. **We carefully checked our words and replaced those impolite remarks with more modest ones in the revised manuscript**. For example, “worst performance”, “wrongly”, “overwhelming superiority”, “great superiority” were replace with “lowest accuracy”, “misclassified”, “better mapping performance”.

Meanwhile, we agree with you that the accuracy differences are partially from the discrepancy of LC definition in each classification system. In the LC products comparison, we chose a classification system containing 10 major LC classes as the basic system and reclassified the detailed LC classes of FROM_GLC and GLC_FCS30. During the classification system conversion, uncertainties are inevitable. However, the mapping accuracy of different GLC products was assessed with the criterion of the classification system we used. Thus, some LC classes in FROM_GLC and GLC_FCS30 regarded as accurate classification based on their original classification system disagreed with the validation samples and obtained relatively low accuracy.

In addition, the Globeland30, FROM_GLC, and GLC_FCS30 are excellent and indispensable GLC products that provide comprehensive and reliable information about the Earth’s surface. These products have been widely used by researchers, policymakers, and other stakeholders worldwide. As GLC products at 30m resolution, the Globeland30, FROM_GLC, and GLC_FCS30 have provided the fundamental information for various applications, such as biodiversity conservation (Wu et al., 2020; Meng et al., 2023), climate change (Kim et al., 2016; Xue et al., 2021; Zheng et al., 2022), and land management (Shafizadeh-Moghadam et al., 2019), despite that they show unstable performance in certain LC classes and some specific areas (Sun et al., 2016; Kang et al., 2020). Thanks to these products, we developed the GLC-2015. Although data inter-comparison has shown that the GLC-2015 had some improvements, there are still some limitation, such as inaccurate mapping results for grassland, shrubland, and wetland. These issues should be the focus of future work. In any case, the GLC-2015 can complete the 30m-resolution GLC product pool and provide better data support for global change research and sustainable development in conjunction with the existing products.

Correspondingly, we have added the recognition of the Globeland30, FROM_GLC, and GLC_FCS30 in Introduction Section:

“The Globeland30, FROM_GLC, and GLC_FCS30 are excellent and indispensable GLC products which have contributed much to various researches, such as biodiversity conservation (Wu et al., 2020; Meng et al., 2023), climate change (Kim et al., 2016; Xue et al., 2021; Zheng et al., 2022), and land management (Shafizadeh-Moghadam et al., 2019).” (Revised manuscript, Line 70-74)

References:

Kang, J., Wang, Z., Sui, L., Yang, X., Ma, Y., and Wang, J.: Consistency analysis of remote sensing land cover products in the tropical rainforest climate region: A case study of Indonesia, Remote Sens., 12, 1410, https://doi.org/10.3390/rs12091410, 2020.

Kim, D., Lim, C.-H., Song, C., Lee, W.-K., Piao, D., Heo, S., and Jeon, S.: Estimation of future carbon budget with climate change and reforestation scenario in North Korea, Adv. Space Res., 58, 1002-1016, https://doi.org/10.1016/j.asr.2016.05.049, 2016.

Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J.: Finer-resolution mapping of global land cover: Recent developments, consistency analysis, and prospects, Journal of Remote Sensing, 2021, 5289697, https://doi.org/10.34133/2021/5289697, 2021.

Meng, Z., Dong, J., Ellis, E. C., Metternicht, G., Qin, Y., Song, X.-P., Löfqvist, S., Garrett, R. D., Jia, X., and Xiao, X.: Post-2020 biodiversity framework challenged by cropland expansion in protected areas, Nature Sustainability, https://doi.org/10.1038/s41893-023-01093-w, 2023.

Shafizadeh-Moghadam, H., Minaei, M., Feng, Y., and Pontius, R. G.: GlobeLand30 maps show four times larger gross than net land change from 2000 to 2010 in Asia, Int. J. Appl. Earth Observ., 78, 240-248, https://doi.org/10.1016/j.jag.2019.01.003, 2019.

Sun, B., Chen, X., and Zhou, Q.: Uncertainty assessment of GlobeLand30 land cover data set over central Asia, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., 41, 1313, https://doi.org/10.5194/isprs-archives-XLI-B8-1313-2016, 2016.

Wu, J., Wang, X., Zhong, B., Yang, A., Jue, K., Wu, J., Zhang, L., Xu, W., Wu, S., Zhang, N., and Liu, Q.: Ecological environment assessment for Greater Mekong Subregion based on Pressure-State-Response framework by remote sensing, Ecol. Indic., 117, 106521, https://doi.org/10.1016/j.ecolind.2020.106521, 2020.

Xue, J., Wang, Y., Teng, H., Wang, N., Li, D., Peng, J., Biswas, A., and Shi, Z.: Dynamics of vegetation greenness and Iits response to climate change in Xinjiang over the past two decades, Remote Sens., 13, 4063, https://doi.org/10.3390/rs13204063, 2021.

Yang, Y., Xiao, P., Feng, X., and Li, H.: Accuracy assessment of seven global land cover datasets over China, ISPRS J. Photogramm. Remote Sens., 125, 156-173, https://doi.org/10.1016/j.isprsjprs.2017.01.016, 2017.

Zheng, W., Liu, Y., Yang, X., and Fan, W.: Spatiotemporal variations of forest vegetation phenology and its response to climate change in northeast China, Remote Sens., 14, 2909, https://doi.org/10.3390/rs14122909, 2022.

Other suggestions:

Comment #3-3. Are the validation samples checked and used separately in 2015 and 2010? For example, the GlobeLand30 in 2010 was used, in which the validation samples of 2015 might not be appropriate to be used (as claimed to have been visually checked using google earth).

Response: Thanks for the comment. In our study, the global point-based and patch-based samples were only visually interpreted using Google Earth images in 2015. Since the data time of our target map is 2015, we are concerned about the mapping performance of various GLC products in 2015. In the accuracy comparison between our map and other products, we regarded Globeland30 as a GLC map developed for 2015 but not for 2010. Under this assumption, we evaluated the consistency between

Globeland30 and the actual landscape in 2015. Any mapping result of GlobeLand30 that was inconsistent with the validation samples for 2015 was defined as misclassification.

Comment #3-4. Line282, why 75% and 25% been chosen? what if these weights changed?

Response: Thanks for the comment. We used the local adaptive fusion model to combine the existing products for each grid. To avoid the inequacy in the size of local samples for rare land cover classes, we also used the global samples to evaluate products' reliability. Since the local samples play a more critical role in the local accuracy assessment, a higher weight should be assigned to the local samples in the construction of the BPA for each grid. To define the weights for the local and global samples, we randomly selected 8 geographical grids of 4º×4° to conduct a pre-test. The weight of the local samples was set from 60% to 90%, with 5% as an interval. The accuracies of the mapping results for each grid with different weights were calculated (FigS1). It was found that in some grinds, the performance of the fusion method was influenced by the weights. When the local samples counted for 75% of the whole sample set and the global samples counted for 25%, the fusion method exhibited robust performance and achieved relatively high accuracy.

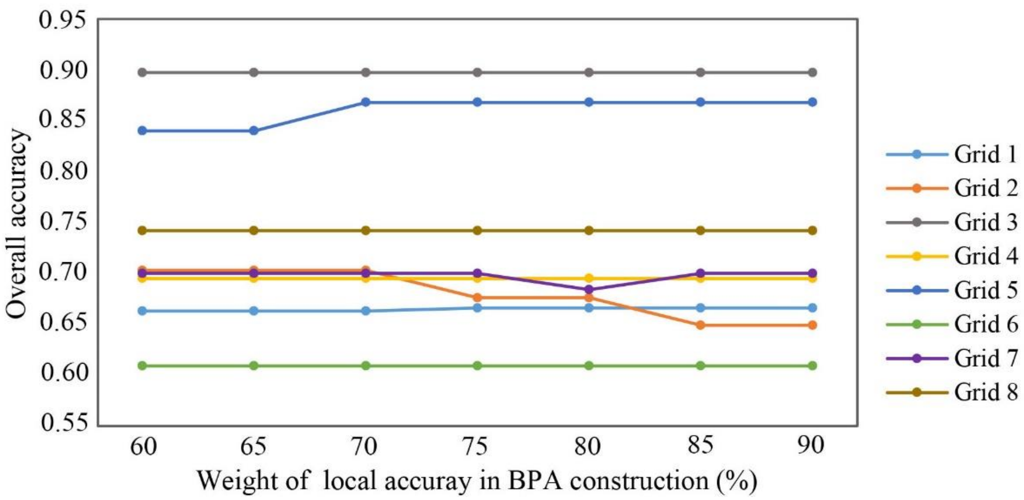

Figure S1. The relationship between the overall accuracy and the weight of local accuracy in BPA construction

Correspondingly, we have explained why we used 75% and 25% as two thresholds in our manuscript.

"Given that the local accuracy for a 4º×4° grid was not able to adequately reflect the actual land cover landscape, especially for the rare LC classes, the global accuracy was incorporated into the construction of the BPA to avoid uncertainties from a local point of view. Since the assessment based on local samples plays a more critical role in BPA construction for a local grid, a higher weight should be assigned to local accuracy. To identify the best weight, we tested different weights of the local accuracy (see Figure S1). The result shows that using 75% performed robustly and obtained relatively higher overall accuracy. Therefore, we chose 75% as the weight for local accuracy and 25% for global accuracy." (Revised manuscript, Line 352-359)

Comment #3-5. Line315, error matrix?

Response: Thanks for comment. We are sorry for this slip of the pen. Instead, we have revised the wrong phrase as “confusion matrix” throughout the manuscript.

Comment #3-6. Lines453 and 462, 'worst' should be replaced by, for example, lowest.

Response: Thanks for suggestion. We have revised as suggested.

Comment #3-7. Line537, 'overwhelming superiority' is too far. As I mentioned, it might because of the definition differences.

Response: Thanks for comment. We agree that the definition difference might cause uncertainties in accuracy comparison between GLC products. We have replaced the original phrase with “outperformance”.

Comment #3-8. Line601, 'show great superiority over others', same as above.

Response: Thanks for comment. The original words have been revised as “show better performance than others”.

Comment #3-9. Line585, 'capture most human activity'. I don't agree, 'capture the footprint of human activities' might be better.

Response: Thanks for suggestion. The original expression has been replaced by “capture the footprint of human activities”.