

## Response to comments

**Paper #:** essd-2020-182

**Title: GLC\_FCS30:** GLC\_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery

**Journal:** Earth System Science Data

## Reviewer #1

It is good to see the study using time series Landsat to map global land cover and making the product public available. The classification legend is finer (~30 classes) than the currently available global 30 m land cover products. The training data are derived from the existing land cover maps (CCI\_LC) and the Landsat time series temporal metrics were classified using random forest in the GEE platform. The product is validated using reference data collected from different sources for the validation of the existing land cover products and examined by the authors. Validation showed 82.5% overall accuracy in the 9-class level 0 legend and 68.7% accuracy in the ~30 class level-2 legend. Furthermore, the authors also make their global validation dataset public available, which could benefit other map producers. I have a few comments on the clarification of the study. Many sentences are vague including the key information of the methodology.

Great thanks for the positive comments. The manuscript has been improved according to your and another reviewer's comments.

Issue 1: It is unclear to me whether the training data reflectance comes from MODIS or from Landsat. This is the key of the paper. The term 'Global Spatial Temporal Spectral Library' sounds like the training reflectance is from the MODIS data. If the training data reflectance is derived from MODIS NBAR while the trained model is applied on Landsat surface reflectance, there will be some inconsistencies. Both the Landsat across scene viewing geometry variation and the Landsat and MODIS NBAR solar geometry difference will create inconsistency between MODIS NBAR and Landsat reflectance. MODIS NBAR is defined for local noon solar geometry and the Landsat overpass time is 10:30 am local time. Their solar zenith differences can be up to 20 depending on the location and time of the year. Furthermore, there will be spectral band pass difference between the two sensors.

Great thanks for the key comment. The training data reflectance is derived from Landsat imagery in this study. The MCD43A4 NBAR dataset is used for identifying the spectrally homogeneous MODIS-Landsat areas to further guarantee the confidence of the training data. To make the deriving training samples clearer, the corresponding part has been revised as:

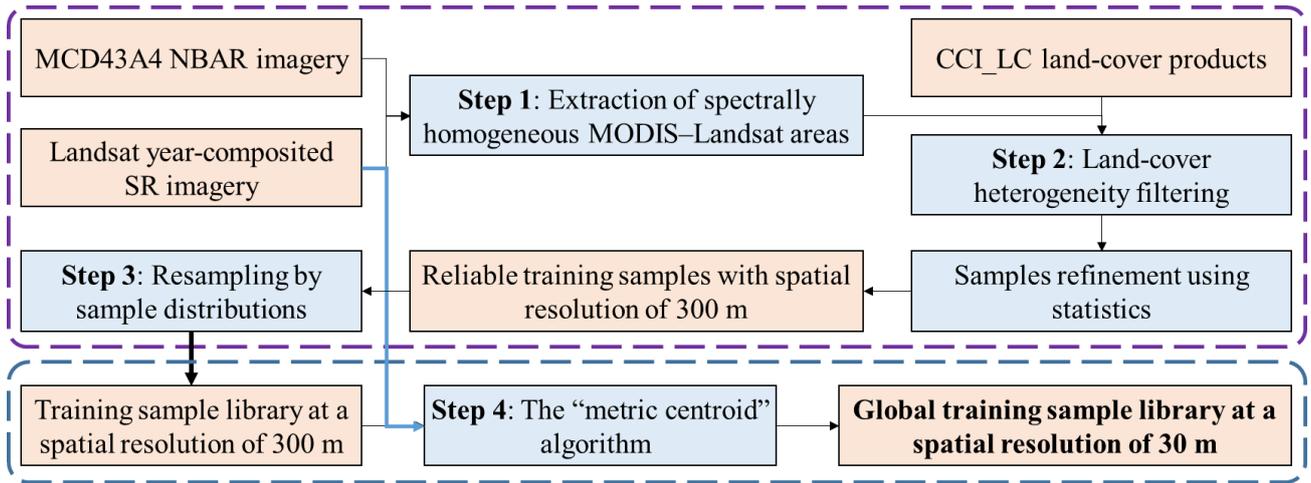


Figure 3. The flowchart of deriving training samples by using multi-source datasets.

Similar to our previous works (Zhang et al., 2019; Zhang et al., 2018), four key steps were adopted to guarantee the confidence of each training point, as illustrated in the Figure 3. As in Zhang et al. (2019), the spectrally homogeneous MODIS–Landsat areas were firstly identified based on the variance of a  $3 \times 3$  local window using spectral thresholds of [0.03, 0.03, 0.03, 0.06, 0.03, and 0.03] for the six spectral bands (blue, green, red, NIR, SWIR1, and SWIR2) in the both MCD43A4 NBAR products and Landsat SR imagery (Feng et al., 2012). It should be noted that the year-composited Landsat SR data were downloaded from GEE platform with the sinusoidal projection. As the MCD43A4 NBAR is corrected for view-angle effects and Landsat has a small view angle of  $\pm 7.5^\circ$ , the view-angle difference between MCD43A4 and Landsat SR could be considered negligible.

Before the process of refinement and labeling, the CCI\_LC land-cover products, which had geographical projections, were reprojected to the sinusoidal projection of MCD43A4. The spatial resolution of MCD43A4 was 1.67 times that of the CCI\_LC land-cover product and the spectrally homogeneous MODIS–Landsat areas had been identified in the  $3 \times 3$  local windows. Also, Defourny et al. (2018) and Yang et al. (2017b) found that the CCI\_LC performed better over homogeneous areas; therefore, a larger local  $5 \times 5$  window was applied to the CCI\_LC land-cover product to refine and label each spectrally homogeneous MODIS–Landsat pixel. Specifically, the land-cover heterogeneity in the local  $5 \times 5$  window was calculated as being the percentages of land-cover types occurring within the window (Jokar Arsanjani et al., 2016a). Aware of the possibility of reprojection and classification errors in the CCI\_LC products, the land-cover heterogeneity threshold was empirically selected as approximately 0.95; in other words, if the maximum frequency of dominant land-cover types was less than 22 in the  $5 \times 5$  window, the point was excluded from GSPECLib. After a spatial-spectral filter had been applied to MCD43A4 and a heterogeneity filter to the CCI\_LC product, the points that had homogeneous spectra and land-cover types were retained. In addition, to further remove the abnormal points contaminating by classification error in the CCI\_LC, the homogeneous points were refined based on their spectral statistics distribution, in which the normal samples would form the peak of the distribution whereas the influenced samples were on the long tail (Zhang et al., 2018). It should be noted that the geographical coordinates of each homogeneous point were selected as being the center

of the local window in the CCI\_LC product because this had a higher spatial resolution than that of MCD43A4.

Then, Zhu et al. (2016) and Jin et al. (2014) found that the distribution (proportional to area and equal allocation) and balance of training data had significant impact on classification results, and quantitatively demonstrated that the proportional approach usually achieve higher overall accuracy than the equal allocation distribution. In addition, Zhu et al. (2016) also suggested to extract a minimum of 600 training pixels and a maximum of 8000 training pixels per class for alleviating the problem of unbalancing training data. In this study, the proportional distribution and sample balancing parameters were used to resample these homogeneous points in each GSPECLib 158.85 km×158.85 km geographic grid cell.

Lastly, different from the previous spectrally based classification using MCD43A4 reflectance spectra (Zhang et al., 2019), in this study, we proposed to use the Landsat reflectance spectra, derived by combining the global training samples and time-series Landsat imagery, to produce the global 30 m land-cover mapping. However, as the spatial resolution difference between Landsat SR (30 m) and homogeneous training samples (300 m), therefore, the “metric centroid” algorithm proposed by Zhang and Roy (2017) was used to find the optimal and corresponding training points at a resolution of 30 m. Specifically, as each homogeneous point corresponded to an area equivalent to 10×10 Landsat pixels, the normalized distances (Eq. (2)) between each Landsat pixel and the mean of all 10×10 pixel areas were calculated. The optimal and corresponding training points at 30 m were selected as the ones having the minimum normalized distance,

$$D_i = \left( \rho_i - \frac{1}{n} \sum_{j=1}^n \rho_j \right)^2, i = 1, 2, \dots, n \quad (2)$$

where  $\rho_i$  is a vector representing the annually composited Landsat SR for 2015 and  $n$  is the number of Landsat pixels within a 10×10 local window (defined as 100). If several 30-m pixels had the same minimum  $D_i$  value then one pixel was selected at random.

Issue 2: Does the authors imply that the global land cover uses fine classification system in some region but uses coarse classification system in other regions? If so, please make it more explicit in the paper (abstract and conclusion) and clearer (what region uses fine classification system). This is important for users who consider to use the products. What is the CCI\_LC coverage?

Great thanks for the comment. Yes, as the land-cover labels came from the CCI\_LC products, the GLC\_FCS30-2015 used the level-1 classification system (containing 16 land-cover types) at global scale, and described by a more detailed legend (14 detailed land-cover types) – where available - to reach a higher level of detail in the legend. The spatial distribution of 14 regional and detailed land-cover types has been added in Section 5.2 as:

The CCI\_LC map used fine classification system in some region but used coarse classification system in other regions (Defourny et al. 2018). Because the training samples were derived from the CCI\_LC land-cover product, our GLC\_FCS30 product inherited these characteristics. Therefore, although the GLC\_FCS30-2015 provided a global 30-m land-cover product with 30 land-cover types (Table 2), the

14 LCCS level-2 detailed land-cover types were applied only for certain regions rather than globe, illustrated in the Figure 12.

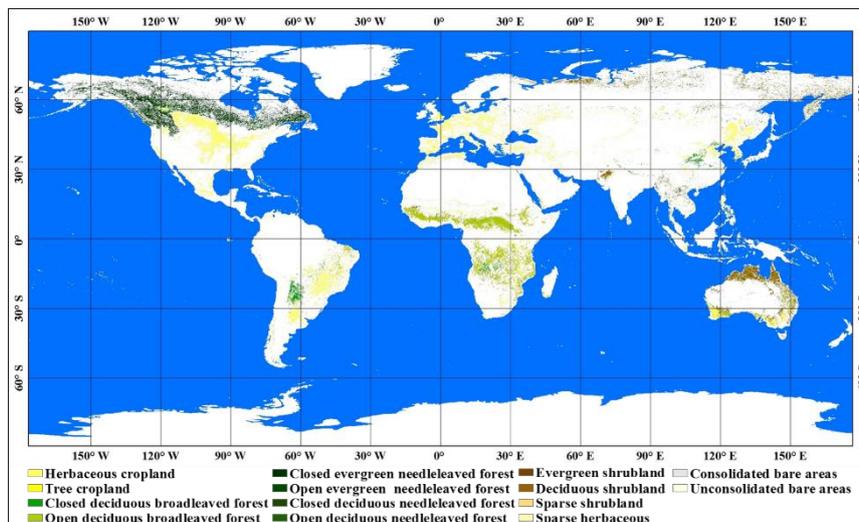


Figure 12. The spatial distributions of 14 detailed regional land-cover types in the GLC\_FCS30-2015 products.

Further, to make it more explicit in the paper, it has been added in the abstract and conclusion section as:

#### Abstract Section

Therefore, it is concluded that the GLC\_FCS30-2015 product is the first global land-cover dataset that provides a fine classification system (**containing 16 global LCCS land-cover types as well as 14 detailed and regional land-cover types**) with high classification accuracy at 30 m. The GLC\_FCS30-2015 global land-cover products produced in this paper is free access at <https://doi.org/10.5281/zenodo.3986871> (Liu et al., 2020).

#### Conclusion Section

“In this study, a global land-cover product for 2015 that had a fine classification system (**containing 16 global LCCS land-cover types as well as 14 detailed and regional land-cover types**) and 30-m spatial resolution (GLC\_FCS30-2015) was developed by combining time-series of Landsat imagery and global training data derived from multi-source datasets”

Lastly, the CCI\_LC and GLC\_FCS30-2015 shares similar spatial distribution for these 14 detailed land-cover types because the training samples are derived from the CCI\_LC and MCD43A4 NBAR products.

Issue 3: Something is wrong about no. of classes: “containing 30 land-cover types” and “(24 fine land cover types).” Later on in Section 3.1, the 34 CCI\_LC classes were “removal of four” and “three wetland land-cover types were further combined into one” so there should be 28 classes?

Great thanks for pointing out this mistake. After carefully checking, the CCI\_LC actually provides the land-cover products containing 36 classes. So our GLC\_FCS30-2015 contained 30 land-cover types. The mistake has been revised as:

“; and 2) the CCI\_LC land-cover product has a detailed classification scheme containing **36** land-cover types, achieves the required classification accuracy over homogeneous areas (75.38% overall), and has a relatively high spatial resolution of 300 m as well as a stable transition between the different annual land-cover products (Defourny et al., 2018; Yang et al., 2017b)...”

Issue 4: For the level-2 classification legend in Table 2, how the level-1 and level-2 classes can be used together for classification. For example, deciduous broadleaved forest 60, closed deciduous broadleaved forest 61, and Open deciduous broadleaved forest 62 cannot be put together for classification. It is either 60 itself OR both 61 and 62. It cannot be all the three together in classification. Great thanks for the comment. As mentioned before, the training samples came from the MCD43A4 NBAR dataset, Landsat year-composited imagery and **CCI\_LC land-cover products which simultaneously contained the LCCS global classification system and detailed regional classification system (containing 14 detailed land-cover types) only for certain regions**, therefore, there will be a phenomenon where global and regional land-cover types coexist at the same time **in these certain regions when training the local random forest models**.

Therefore, our ongoing works are combining quantitative retrieval models and multi-source datasets to improve the diversity of global land-cover types in GLC\_FCS30-2015, by using the Fractional Vegetation Cover (FVC) estimation models to retrieve the annual maximum FVC and then distinguish between open and closed broadleaved or needleleaved forests, combining the time-series NDVI to split the evergreen and deciduous shrublands. It has been revised in the Section 5.2 as:

“In future work, quantitative retrieval models and multi-source datasets should be combined to improve the diversity of global land-cover types in GLC\_FCS30-2015 and further avoid the existence of global LCCS classification system and detailed regional land-cover classification system. This could be done, for example, by using the Fractional Vegetation Cover (FVC) estimation models (Yang et al., 2017a) to retrieve the annual maximum FVC and then distinguish between open and closed broadleaved or needleleaved forests, combining the time-series NDVI to split the evergreen and deciduous shrublands, as well as integrating the GLCNMO training dataset to further distinguish consolidated from unconsolidated bare areas (Tateishi et al., 2014; Tateishi et al., 2011).”

## **Specific comments**

Introduction “stamping effect was noticeable” it is unclear what is stamping effect? Use the term which has been used in the literature.

Great thanks for the comment. According to your suggestion, the sentence has been revised by referencing the original literature:

“...as the overall accuracy for the detailed land-cover types was only 52.76% and the **patch effects was noticeable caused by the temporal differences among the Landsat scenes...**”

Figure 1, the Landsat end overlap (row overlaps) cannot be considered as two observations. Great thanks for the comment. Yes, the areas where there was overlap cannot be considered as two observations. This figure was used to explain the spatial distributions of total clear observations. To avoid the confusion, the corresponding paragraph has been revised as:

“Fig. 1 illustrates the clear-sky Landsat-8 SR temporal frequency after the cloud, cloud shadow and saturated pixels have been masked out. The statistical results indicated that: 1) most land areas, except for tropical areas, had a high availability of clear-sky Landsat imagery; and 2) areas with a low frequency of clear-sky Landsat SR were mainly located in rainforest areas including the Amazon rainforest, African rainforests and Indian–Malay rainforests, which are areas mainly covered by evergreen broadleaved forests.”

Section 2.2 Define what is the GImpS-2015 product.

Great thanks for pointing out this mistake. The ‘GImpS-2015’ has been revised as ‘MSMT\_IS30-2015’, so the revised sentence was:

“The validation results indicated that the MSMT\_IS30-2015 product achieved an overall accuracy of 95.1% and a kappa coefficient of 0.898 using 11,942 validation samples from fifteen representative regions.”

Section 3.1. This step is not conducted in GEE? The authors stated the Landsat data “were reprojected to the sinusoidal projection of MCD43A4.” The “metric centroid” algorithm is proposed in Zhang and Roy 2017 NOT by Roy and Kumar (2016). I don’t quite follow what is the purpose of the “metric centroid” algorithm since the training reflectance is derived from MODIS rather than Landsat. The “metric centroid” algorithm is used if the training reflectance is from Landsat and the training class label from MODIS.

Great thanks for the comment. Yes, this step was conducted in the localhost computation environment instead of the GEE platform, and the Landsat were reprojected to the sinusoidal projection of MCD43A4 to extract the spectrally homogeneous MODIS-Landsat areas (Step 1) and derive the training sample library at 30 m using the “metric centroid” method (Step 4). To clarify the process of deriving the training sample, the part has been revised as:

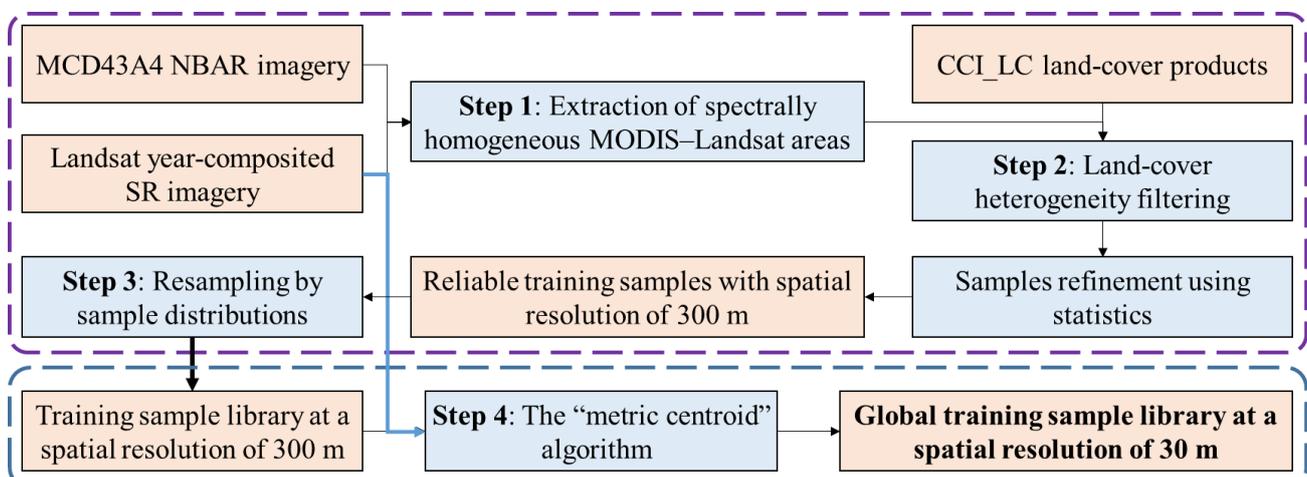


Figure 3. The flowchart of deriving training samples by using multi-source datasets.

Similar to our previous works (Zhang et al., 2019; Zhang et al., 2018), four key steps were adopted to guarantee the confidence of each training point, as illustrated in the Figure 3. As in Zhang et al. (2019), the spectrally homogeneous MODIS–Landsat areas were firstly identified based on the variance of a  $3\times 3$  local window using spectral thresholds of [0.03, 0.03, 0.03, 0.06, 0.03, and 0.03] for the six spectral bands (blue, green, red, NIR, SWIR1, and SWIR2) in the both MCD43A4 NBAR products and Landsat SR imagery (Feng et al., 2012). It should be noted that the year-composited Landsat SR data were downloaded from GEE platform with the sinusoidal projection. As the MCD43A4 NBAR is corrected for view-angle effects and Landsat has a small view angle of  $\pm 7.5^\circ$ , the view-angle difference between MCD43A4 and Landsat SR could be considered negligible.

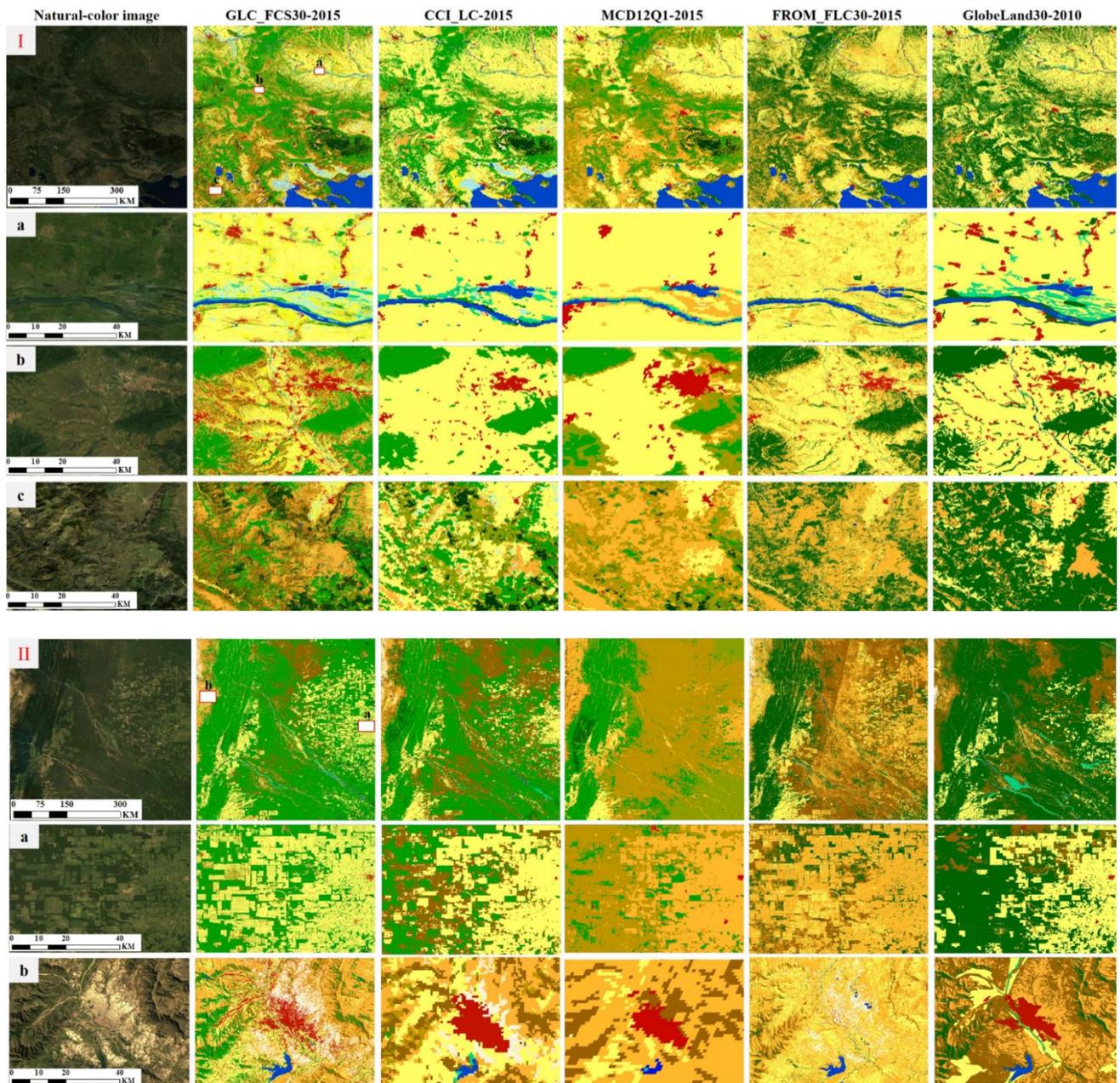
Before the process of refinement and labeling, the CCI\_LC land-cover products, which had geographical projections, were reprojected to the sinusoidal projection of MCD43A4. The spatial resolution of MCD43A4 was 1.67 times that of the CCI\_LC land-cover product and the spectrally homogeneous MODIS–Landsat areas had been identified in the  $3\times 3$  local windows. Also, Defourny et al. (2018) and Yang et al. (2017b) found that the CCI\_LC performed better over homogeneous areas; therefore, a larger local  $5\times 5$  window was applied to the CCI\_LC land-cover product to refine and label each spectrally homogeneous MODIS–Landsat pixel. Specifically, the land-cover heterogeneity in the local  $5\times 5$  window was calculated as being the percentages of land-cover types occurring within the window (Jokar Arsanjani et al., 2016a). Aware of the possibility of reprojection and classification errors in the CCI\_LC products, the land-cover heterogeneity threshold was empirically selected as approximately 0.95; in other words, if the maximum frequency of dominant land-cover types was less than 22 in the  $5\times 5$  window, the point was excluded from GSPECLib. After a spatial–spectral filter had been applied to MCD43A4 and a heterogeneity filter to the CCI\_LC product, the points that had homogeneous spectra and land-cover types were retained. In addition, to further remove the abnormal points contaminating by classification error in the CCI\_LC, the homogeneous points were refined based on their spectral statistics distribution, in which the normal samples would form the peak of the distribution whereas the influenced samples were on the long tail (Zhang et al., 2018). It should be noted that the geographical coordinates of each homogeneous point were selected as being the center of the local window in the CCI\_LC product because this had a higher spatial resolution than that of MCD43A4.

Then, Zhu et al. (2016) and Jin et al. (2014) found that the distribution (proportional to area and equal allocation) and balance of training data had significant impact on classification results, and quantitatively demonstrated that the proportional approach usually achieve higher overall accuracy than the equal allocation distribution. In addition, Zhu et al. (2016) also suggested to extract a minimum of 600 training pixels and a maximum of 8000 training pixels per class for alleviating the problem of unbalancing training data. In this study, the proportional distribution and sample balancing parameters were used to resample these homogeneous points in each GSPECLib  $158.85\text{ km}\times 158.85\text{ km}$  geographic grid cell.

The “metric centroid” algorithm has been revised as:



The size of the enlargement figures is 40km×60 km, the information has been added in the title of Figure 8 and the scale bars also have been added in the corresponding figures in the following:



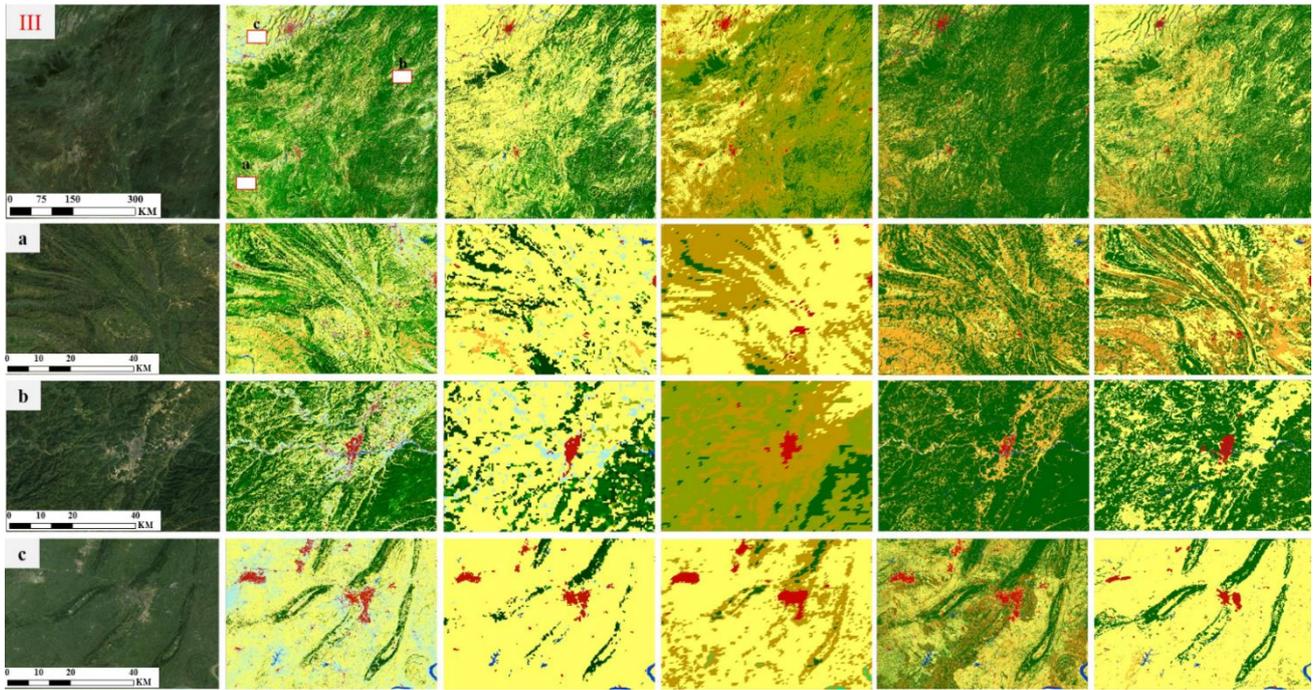


Figure 8. Comparison between GLC\_FCS30-2015 and other land-cover products (CCI\_LC-2015 products developed by (Defourny et al., 2018), the MCD12Q1-2015 developed by (Friedl et al., 2010), the FROM\_GLC-2015 developed by (Gong et al., 2013) and the GlobeLand30 developed by (Chen et al., 2015)) in three  $5^{\circ} \times 5^{\circ}$  regions. In each case, 2–3 local enlargements **with the size of 40km $\times$ 60 km** (a-c) were used to reveal further details of each land-cover product.

Lines 470-475, I would suggest deleting this paragraph. This is a little aggressive.

Great thanks for the suggestion. The aggressive paragraph has been removed in the revised manuscript.

Discussion. It is good to see Figure 8. However, it is a little misleading. If Figure 8 only shows the number of training samples, why “where there are relatively uniform land-cover types, there are fewer training samples”. I would think the other way around.

Great thanks for the comment. The reason why “where there are relatively uniform land-cover types, there are fewer training samples” is because we use the resample to balance the training samples in Section 3.1, therefore, the homogeneous areas have relatively fewer training samples comparing these transition areas. The part has been added as:

“Then, Zhu et al. (2016) and Jin et al. (2014) found that the distribution (proportional to area and equal allocation) and balance of training data had significant impact on classification results, and quantitatively demonstrated that the proportional approach usually achieve higher overall accuracy than the equal allocation distribution. In addition, Zhu et al. (2016) also suggested to extract a minimum of 600 training pixels and a maximum of 8000 training pixels per class for alleviating the problem of unbalancing training data. In this study, the proportional distribution and sample balancing parameters were used to resample these homogeneous points in each GSPECLib 158.85 km $\times$ 158.85 km geographic grid cell.”

In order to avoid the misleading, the sentence of “where there are relatively uniform land-cover types, there are fewer training samples” has been deleted in the revised manuscript as:

“Figure 9 illustrates the number of global training samples in each  $1^{\circ} \times 1^{\circ}$  geographical grid cell. The statistics are generally consistent with the land-cover patterns shown in Fig. 5. In addition, in contrast to other studies that used manual interpretation of samples for global land-cover mapping (Friedl et al., 2010; Gong et al., 2013; Tateishi et al., 2014), the total number of training samples in this study reaching 27,858,258 points and so was tens to hundreds of times higher than that used in these global land-cover classifications.”

For each 5 by 5 degree local training, does the authors also use some training samples outside the 3 by 3 tiles if there is insufficient samples in the 3 by 3 tiles? If so make it clearer in the paper.

Great thanks for the comment. We didn't import the training samples outside the 3 by 3 tiles. Actually, we have built a backup training sample library to avoid missing training samples of sparse land-cover types, however, after using the training samples from neighboring 3 by 3  $5^{\circ} \times 5^{\circ}$  geographical tiles, the missing training samples in the central tile almost were supplemented by neighboring  $3 \times 3$  tiles, which caused the backup library to lose its function.

“Therefore, it can be assumed that the training data derived from the updated GSPECLib were accurate and suitable for large-area land-cover mapping at 30 m.” If the GSPECLib's contribution is only to identify homogenous locations, do not overemphasize in discussion or conclusion. Use something like derivation of training data from existing land cover products.

Great thanks for the comment. Based on the suggestion, the statement has been revised as:

“Therefore, it can be assumed that the training data, derived **by combining the MCD43A4 NBAR and CCI\_LC land-cover products**, were accurate and suitable for large-area land-cover mapping at 30 m.”

Line 525, “applied only for certain regions”, which region? Users deserve to know before using the data.

Great thanks for the comment. Yes, it is necessary to provide the spatial distribution for the 14 LCCS level-2 detailed land-cover types. According to the suggestion, the spatial distributions of 14 detailed land-cover types has been added as:

“The CCI\_LC map used fine classification system in some region but used coarse classification system in other regions (Defourny et al. 2018). Because the training samples were derived from the CCI\_LC land-cover product, our GLC\_FCS30 product inherited these characteristics. Therefore, although the GLC\_FCS30-2015 provided a global 30-m land-cover product with 30 land-cover types (Table 2), the 14 LCCS level-2 detailed land-cover types were applied only for certain regions rather than globe, illustrated in the Figure 12.”

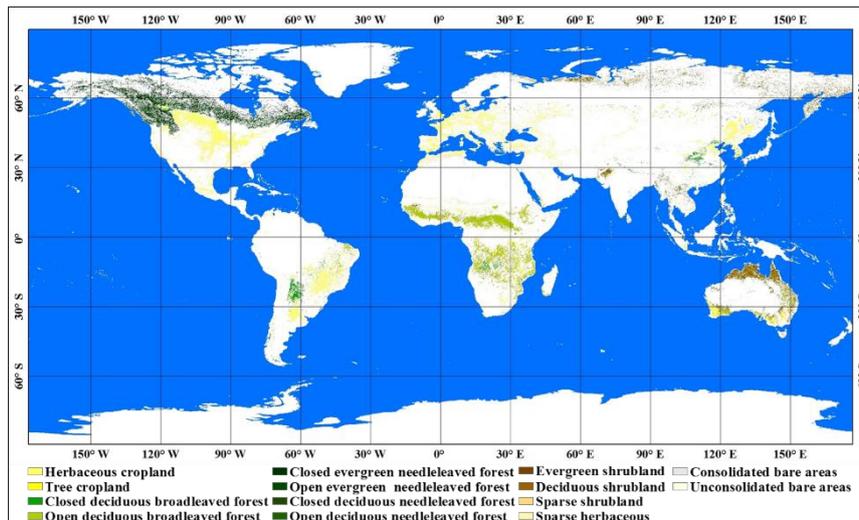


Figure 12. The spatial distributions of 14 detailed land-cover types in the GLC\_FCS30-2015 products.

Data availability Make it explicit that the validation dataset is also public available. I believe it is an important contribution to the community.

Thanks for the suggestion. According to your suggestion, the free access of validation dataset has been added in the Data availability section as:

“The corresponding validation dataset, producing by integrating existing prior datasets, high-resolution Google Earth imagery, time-series of NDVI values for each vegetated point and visual checking by several interpreters, is available at <http://doi.org/10.5281/zenodo.3551994> (Liu et al., 2019).”

Conclusion “global training data derived from GSPECLib”. It is a little misleading if the GSPECLib is only to identify homogeneous locations. Use something like derivation of training data from existing land cover products.

Great thanks for the comment. Based on the suggestion, the sentence has been revised as:

“In this study, a global land-cover product for 2015 that had a fine classification system (containing 16 global LCCS land-cover types as well as 14 detailed and regional land-cover types) and 30-m spatial resolution (GLC\_FCS30-2015) was developed by **combining time-series of Landsat imagery and global training data derived from multi-source datasets**. Specifically, by combining MCD43A4 NBAR, CCI\_LC land-cover products and Landsat imagery, the difficulties of collecting sufficient reliable training data were easily solved and the fine classification system was also made use of.”