Response to Anonymous Referee #1 Comments

Dear Editors and Referee:

Thanks again for your kind comments concerning our manuscript. Those comments are all valuable and very helpful for improving our paper, and we have made correction which we hope meet with approval. Revised portion are marked in the manuscript. The main corrections in the paper and the responds to the reviewer's comments are as flowing:

Anonymous Referee #1:

Q 1: The article has improved a lot. But I still have concerns about the method for the classification of forested land areas. The authors used a threshold-based method (or decision tree) to extract the forested areas. They used constant thresholds of NDVI (0.61), RVI (6), and NIR (0.38) to extract the forest areas for the different years. This is questionable. Firstly, the Landsat TM and Landsat OLI data are inherently slightly different (wavelengths in each band); secondly, the acquisition time of the images is not consistent from year to year, some in the summer and some in October; and finally the Qin, et. al. (2015) may be a wrong reference and I did not find that they used the same method. I would expect slightly varying thresholds over years. A better approach is to use the RF model directly to classify coniferous forests, broadleaf forests, and non-forestry forests for each year.

Reply: Thanks for your careful suggestion. Qin et al. (2015) showed that the annual maximum NDVI ($NDVI_{max}$) values of built-up areas, barren lands, and sparsely vegetated lands are usually lower than 0.30, whereas forest $NDVI_{max}$ values are usually higher than 0.50. However, empirical NDVI values for different times and regions are not universal (Ma et al., 2019). For this study area, we determined the decision tree classification rules based on sample training: NDVI values greater than 0.62 and RVI values greater than 6.0 were selected as vegetation land, otherwise land was regarded as non-vegetation land. According to your suggestion, we have added the methodological limitations in the discussion section.

Line 142-151

In this classification, two vegetation indices, NDVI and ratio vegetation index (RVI), were used for the discrimination between forest and nonforest land. NDVI can effectively weaken the effects of complex terrain in image information extraction, and enhance the distinction between vegetation and other land types, which is helpful for improving the accuracy and credibility of forest information extraction. The RVI can better reflect the difference of vegetation growth and coverage, and is suitable for vegetation monitoring in areas with vigorous vegetation growth and high coverage. The annual maximum NDVI ($NDVI_{max}$) values of built-up areas, barren lands, and sparsely vegetated lands are usually lower than 0.30, whereas forest $NDVI_{max}$ values are usually higher than 0.50 (Qin et al., 2015). Subsequently, we determined the decision tree classification rules based on sample training: NDVI values greater than 0.62 and RVI values greater than 6.0 were selected as vegetation land, otherwise land was regarded as non-vegetation land.

Line 185-188

However, empirical NDVI values for different times and regions are not universal (Ma et al., 2019). Therefore, the results of the threshold-based method (decision tree) used in this study for the classification of forested and non-forested land may have a certain uncertainty. Future research could use the machine learning or deep learning methods to classify forests to improve the accuracy of classification.

Q 2: For the training sample, I would not agree that it is very difficult to obtain for each year. First of all the forest area has stability, so the samples acquired in 2015, after checking, can be used for other years as well (some samples may need to be added and removed appropriately). This process can be done with the help of Google Earth history high-resolution images, or Landsat images. It is not a challenge for classifying only three categories (Conifer, broadleaf, and non-forest).

Reply: Thanks. Yes. As you said training samples can be obtained with the help of Google Earth history high-resolution images, or Landsat images. In this study, we selected the sample points used for the classification for different years based on Landsat images refer to GF-2 images and Google Earth images (Gong et al., 2013). (Line 158-160)

Q 3: Furthermore, the description in the methods section needs further refinement. For example, "it was found that different plants have different spectral reflectance peaks in the near-infrared band" -- I would say the peak reflectance of different plants may be also different in some other spectral regions. This sentence has no meaning.

"This band is highly sensitive to the differences in reflectance that result from different types of leaves having different internal structures" --- not only the structure but also the color, etc. Also missing a reference here.

Reply: Thanks for your careful suggestion. According to your comments, we revised the corresponding sentences and added references.

Line 151-154

In this work, it was found that different plants have different spectral reflectance peaks in the near-infrared band; this band is highly sensitive to the differences in reflectance that results from different types of leaves having different internal structures and colors (Lewis 2002).

Q 4: "Six bands, Landsat TM bands 1–5 and 7, and Landsat OLI bands 2–7 were selected as characteristic spectral variables, and meanwhile NDVI, the normalized difference index (NDI) (Rodríguez-Moreno and Bullock, 2014) " --- wrong reference.

Reply: Thanks for pointing out this. We revised this error and added a reference.

Line 160-162

Six bands, Landsat TM bands 1–5 and 7, and Landsat OLI bands 2–7, were selected as characteristic spectral variables, and meanwhile NDVI, the normalized difference built-up index (NDBI) (Cha et al., 2003) and the RVI were also selected as index characteristic variables for classification in RF.

Q 5: "The overall accuracy was found to be 90.37%, and the F1-scores for the broad-leaved, coniferous forest and non-forest land were 0.85, 0.93, and 0.91, respectively (Pontius and Millones, 2011)." --- why the result has a reference???

Reply: Thanks for your careful suggestion. Previous anonymous referee pointed out that the Kappa may is not a reliable accuracy metric. He recommended to us a very good paper, i.e., Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment (Pontius and Millones, 2011), and suggested using F1 scores (the harmonic means of the user's and the producer's accuracies) for assessing classification accuracy. This reference is the F1 score citation. we revised the corresponding sentences and added references.

Line 165-166

The overall accuracy was found to be 90.37%, and the F1-scores (Chen et al., 2021; Pontius and Millones, 2011) for the broad-leaved, coniferous forest and non-forest land were 0.85, 0.93, and 0.91, respectively.

Q 6: "Following this, a haze optimized transformation (HOT) algorithm was used to identify and remove noise due to thin clouds" ---- Have you filled the gaps caused by cloud removal? If the cloud pixels that you removed happen to be forested areas, will that affect the results of your change detection?

Reply: Thanks for your good suggestion. The haze optimized transformation (HOT) algorithm was used to identify and remove noise due to thin clouds. We used the adjacent date images to fill the gaps caused by thick cloud cover. This has a minor effect on the results of change detection in this study.

Q 7: Please revise the method section carefully by referring to remote sensing-related papers!!!

In summary, I understand that it would be difficult (or might be not necessary) to use new methods. But since the classification products serve as the basis for the subsequent analysis, the methodological limitations must be discussed in the discussion section at least, otherwise, I do not consider it suitable for publication.

Reply: Thanks for your constructive suggestion. According to your suggestion, we have added the methodological limitations in the discussion section. (Line 185-188)

All above revisions are highlighted in the manuscript information. We hope you will be satisfied with our changes. Thanks again for your good suggestions.

References

Cha, Y., Ni, S.X., and Yang, S.: An Effective Approach to Automatically Extract Urban Land-use from TM Imagery. Natl. Remote Sens. Bull., 1, 37–40, https://doi.org/10.3321/j.issn:1007-4619.2003.01.007, 2003.

Chen, J., Yuan, Z.Y., Peng, J., Chen, L., Huang, H.Z., Zhu, J.W., Liu, Y., and Li, H.F.: DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. IEEE J. Stars., 14, 1194–1206, https://doi.org/10.1109/JSTARS.2020.3037893, 2021.

Gong, P., Wang, J., Yu, L., Zhao, Y.C., Zhao, Y.Y., Liang, L., Niu, Z.G., Huang, X.M., Fu, H.H., and Liu, S.: Fine resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens., 34, 2607–2654, https://doi.org/10.1080/01431161.2012.748992, 2013.

Lewis, M.: Spectral characterization of Australian arid zone plants. Can. J. Remote Sens., 28(2), 219–230, https://doi.org/10.5589/m02-023, 2002.

Ma, X.P., Bai, H.Y., Deng, C.H., and Wu, T.: Sensitivity of Vegetation on Alpine and Subalpine Timberline in Qinling Mountains to Temperature Change. Forests, 10, 1105, https://doi.org/10.3390/f10121105, 2019.

Pontius, R.G., and Millones, M.: Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens., 32, 4407–4429. https://doi.org/10.1080/01431161.2011.552923, 2011.

Qin, Y., Xiao, X., Dong, J., Zhang, G., Shimada, M., and Liu, J.: Forest cover maps of china in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI. ISPRS J. Photogramm. Remote Sens., 109, 1–16, https://doi.org/10.1016/j.isprsjprs.2015.08.010, 2015.