Ref.: MS. bg-2022-317 Biogeosciences Reconciling different approaches to quantifying land surface temperature impacts of afforestation using satellite observations

General comments:

This study conducted an interesting research about three influential approaches in evaluating the climatic effects induced by afforestation over China. So far, no such studies have ever compared the three methods simultaneously and investigated the underlying mechanisms that lead to their discrepancies and more importantly, whether the discrepancies can be mitigated or reconciled. I'm happy to see that the authors filled this knowledge gap and gave us a good reference. As far as I know, in previous studies involving both the actual and potential effects (Li Yan, 2016, JGR-A, Shen Wenjuan, 2019, AFM), the two effects, characterized by LST changes (or cooling) were comparable and consistent in magnitude. As a result, their discrepancies attracted less attention. Fortunately, this research emphasized this point by applying the afforestation experiment over China. Coincidentally, I have a pending research (in prepare for subscription) in support of the result (actual effect is largely less than potential effect) in this study.

Overall, I appreciate the authors' efforts to put this question forward and gave a good demonstration.

We appreciate the comments which will help us to improve the manuscript. We are also glad that this researcher reaches a similar conclusion. Please find below the original comments (in black) and our responses (in blue).

Specific Comments:

(1) The distribution of sample grids about the actual and potential effect were not shown. Maybe you can display them in Supplemental Materials, like Peng Shushi et al., 2014, PNAS did.

We will add the distribution of sample grids of the actual and potential effects in the supplemental material in the revised manuscript (MS) (shown below in Fig. R1).

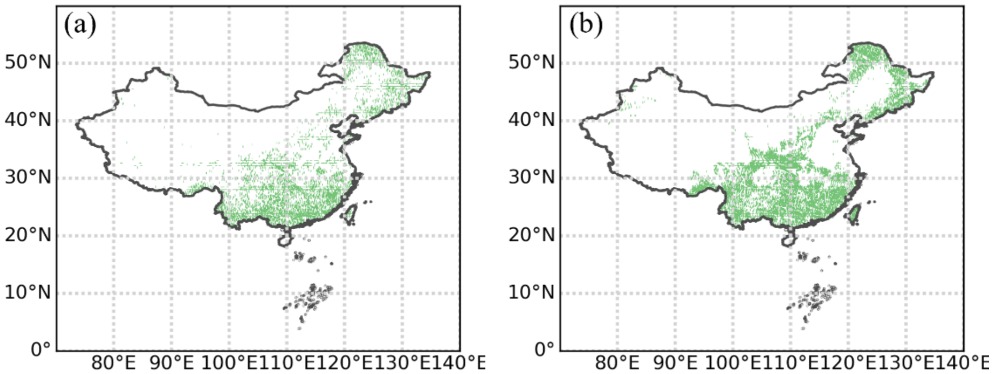

Figure R1. The distributions of the original sample pixels (1km resolution) for (a) actual effect and (b) potential effect.

(2) Line 313: Please explain why GlobeLand30 is not suitable for detecting forest change, instead of just citing Zeng et al., 2021.

To address this comment, we showed in Fig. R2 (below) that GlobeLand30 failed to capture the large-scale forest gain in Southeast China (range of 110°-120°E, 25°-30°N) from 2000 to 2010. Therefore, we did not directly use the GlobeLand30 dataset to detect forest change. Nonetheless, we decide not to include this Fig. R2 in revised MS in order to avoid the redundancy of the Method section.

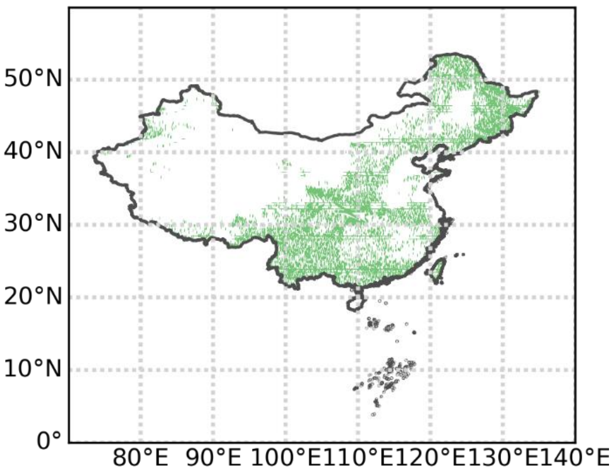

Figure R2. The spatial distribution of forest gain pixels (1km resolution) detected based on GlobeLand30 between 2000-2010.

(3) When computing the mixed and full potential effects, what threshold did the authors use to define a 1-km pixel as an afforested pixel using the GFC data? In addition, the method to

process land cover data (GlobeLand30) seems to be ambiguous, since Line 189 described using the majority method to aggregate 30 m to 1km, but Line 309-310 mentioned "vegetation type with area fraction > 50% for every 1km×1km window". In my opinion, majority does not equal > 50%. For instance, one land cover type (i.e., cropland) accounts for 30% can also be designated as the dominated type as long as 30% is the largest area fraction.

For the first question, 1-km pixels with the net forest cover gain > 0 according to the GFC dataset were defined as afforested pixels. We will modify sentences in the Methods section in the revised MS to more clearly define afforestation pixels.

Regarding the second question, we acknowledged that the expression of "majority method" in the MS was misleading. 50% was set as a fraction threshold to define the land cover types based on GlobeLand30. We will revise the related description in MS.

(4) Line 311. What dataset did forest and openland stem from? Based on the early description, forest was only from GLC data and openland only from Globeland30. Please give a clear declaration here. Once more, it's important to clearly elucidate the criterion to define the afforested 1-km pixel when aggregating 30-m pixels. If the authors used 50% as the threshold, then the bars below 50% in Figure 6 seem to be unreasonable because pixels with afforestation fraction below 50% was not afforestation anymore. But if using a lower threshold, would the 1-km pixel stay as an afforestation pixel? Please, give an explicit and consistent explanation.

For the first question, forest and openland stem from the generated land cover map based on GlobeLand30. Forest pixels in this map (Fig. R1b) were selected as samples to obtain potential effects.

For the second question, pixels where forest gained (i.e., afforestation fraction >0%) as detected from GFC (Fig. R1a) were selected to derive the actual effect. Thus, the afforestation fractions in Fig. 6 are reasonable. These two points have been described in the Method section, so we prefer not to make any modifications to MS.

(5) When collecting the sample pixels, did the authors consider the impact of water pixels? As far as I know, the common method is to abandon the grids in which water pixels account for more than a fraction (5% or 10% or 15%...).

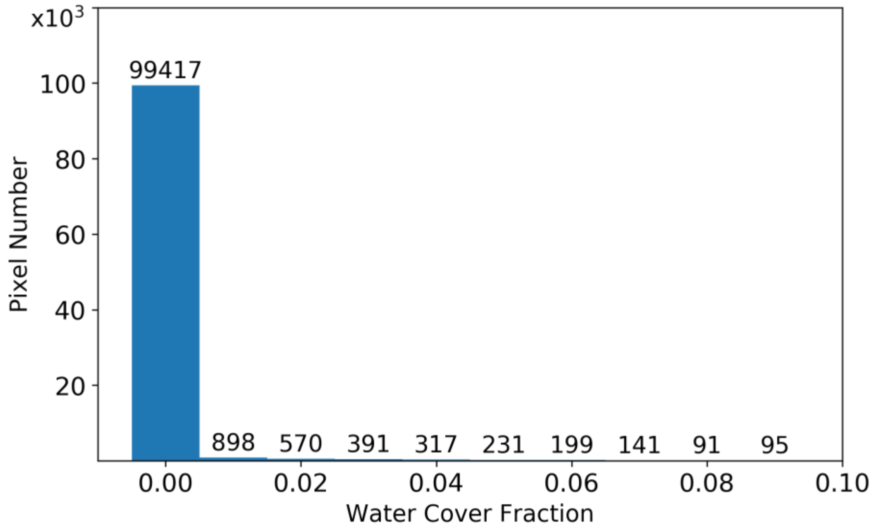

Figure R3. Histogram of water cover fraction of all samples in this study.

In the previous MS, the impact of water cover impact was not considered. As suggested, we took into the "water fraction" and checked the water cover fraction of research pixels based on the land cover fraction map (Fig. 2 in MS) from GlobeLand30. As shown in Fig. R3, the water fraction among all samples in this study is less than 10%, with almost 95% of samples containing no water area (water fraction=0%). In this way, we believe that removing samples where the water fraction is >5% or 10% may have less effect on the findings of this study, therefore we prefer not to change research samples in MS.

(6) Section 2.4, I wonder about the significance and necessity of using Bonferroni correction in this study. Many audiences including me seem not to be familiar with this operation. The authors may give a more detailed explanation.

We performed paired samples t-test to examine the differences in the afforestation effects on land surface temperature (LST) based on three approaches, which involved three hypotheses, i.e., $\Delta T_a = \Delta T_{p1}$; $\Delta T_a = \Delta T_{p2}$; $\Delta T_{p1} = \Delta T_{p2}$ in this study. Here, we employed three paired comparisons to test these hypotheses.

For each comparison, if we use significance level (p-value) =0.05 to determine that the means of a pair of conditions (e.g., ΔT_a and ΔT_{p1}) are statistically different from each other, we will have a 5% chance of committing a Type I error when we reject the null hypothesis (H_0: $\Delta T_a = \Delta T_{p1}$). When conducting three comparisons, the possibility of committing a Type I error for comparisons can be estimated as 0.05×3. Bonferroni correction was applied in this study to adjust the p-value to mitigate the increasing type I error when making multiple paired-samples t-tests (Lee and Lee, 2018; UC Berkely, 2008). We will cite related literature on Bonferroni correction in the Method section in the revised MS, but the explanation here will not be added to the revised MS.

(7) Figure 6. When the fraction of afforestation reached (50, 60], why the mixed potential effect exceeded the full potential effect. It seems strange and no explanation about this phenomenon was seen. In addition, the significant linear trend can be found for actual effect (as displayed in Figure 5), but it seems that this significant trend was not found in mixed potential especially the full potential effect. May the authors give an explanation about this?

As for the first question, we checked the data processing script and found Fig. 6 in the original manuscript was wrong and will be corrected as Fig. R4. In Fig. R4, when the fraction of afforestation reached (50, 60], the mixed potential effect is still smaller than the full potential effect. The relevant description in the Results section of our original manuscript still applies and will not be changed.

For the second question, ΔT_{p1} and ΔT_{p2} are grouped into intensity bins not because they have any relation to the afforestation intensity bins. Conceptually, they represent a complete shift from openland to forest, so it's expected they do not show any trend here.

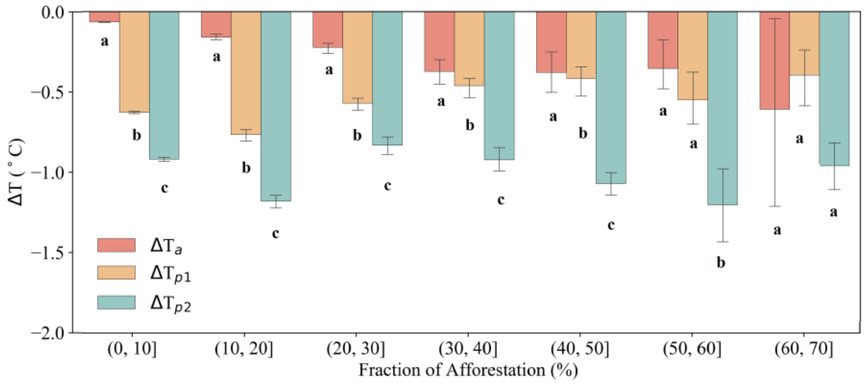

Figure R4. Comparison of ΔT for the three approaches for bins of afforestation intensity.

(8) The reconciliation was reached when increasing the fraction to 100% for the actual effect. But why the fraction increase (through linear extrapolation) was only implemented for actual effect rather than both actual and mixed potential effect. It seems unfair because the author compared the 100% fraction-based actual effect with not 100% based (mixed) potential effect.

We believe this comment is related to Comment #7. From the concept of ΔT_{p1} and ΔT_{p2}, it does not make sense to make regression between ΔT_{p1} and ΔT_{p2} and F_{aff}.

(9) What is the difference between Figure 8 and Figure A6? Mean values of all grids for Figure 8 and gross values of all grids for Figure A6? Do the cumulative biophysical changes only refer to delta_LE? Because the numbers in Line 586-587 corresponded to delta_LE in Figure. A6.

For Figure 8, afforestation-induced changes in surface energy fluxes referred to the flux change per unit area (W m^2), whilst cumulative surface energy effect (f_{cum}) in Figure A6 referred the sum of the flux change (J) from all the samples after considering the forest change area (m^2). More specifically, the cumulative surface energy change (f_{cum}) can be calculated from the equation R1:

$$f_{cum} = \sum_{i=1}^{i=n}(area_i \times F_i) \qquad \text{(R1)}$$

where F_i is the flux change in per unit area (W m^2) for pixel i, n is the total number of samples, and $area_i$ is the forest change area in pixel i. This part will be added in Supplementary Material.

(10) Uncertainty about the Global Forest Cover dataset should be discussed. References can be found in recent papers published by Dr. Zeng Zhenzhong.

We assumed the reviewer was most likely referring to the GFC accuracy discussion. Comparing forest gained area from GFC to forest area statistics reported in Forest Resource Assessment (FRA), LiDAR detection (Geoscience Laser Altimeter System), and MODIS NDVI time series, the GFC product demonstrated a global accuracy of greater than 99% (Hansen et al., 2013). Chen et al. (2020) applied the global land cover validation data from the United States Geological Survey (USGS) to evaluate the accuracies of the selected land cover datasets while the correlations between the GFC dataset and the validation data were the highest (0.77). Zeng et al. (2021) also demonstrated that GFC can achieve an overall accuracy of 98.4% in Southeast

Asia. We will add some description of GFC uncertainty in the Methods section.

(11) The reasons leading to the discrepancies between actual and potential effects were not considered and discussed thoroughly.

1) Actual effect was calculated using the LST data from two years (target and reference year), but the potential effect used the LST from the same year (2012 in this study).
2) When computing the actual effect, the control pixels were constant or stable unchanged forests, however, as for potential effect, the reference pixels were cropland or grassland pixels.
3) Even though the author adopted the same sample pixels (same locations) for the three approaches, the inherent afforestation fraction was not consistent because different criteria were adopted.

Please give a detailed explanation and discussion about the above aspects.

We believe the point (1) and (2) refer to the method difference of "space-and-time" between "space-for-time". We will add a detailed description in the Supplementary Material to further explain the methodological differences to clarify the results of our research are not limited to fraction differences. While for point (3), we have never claimed that the inherent afforestation fraction for three methods is consistent, and the specific reasons can be found in our responses to comment #7. Based on previous research (Li et al., 2016), it could be reasonably suspected that the differences in estimated land surface cooling by afforestation by different approaches could be potentially due to the afforestation fraction, but this suspect was only proved until our current work.

References used in the responses:

Chen, H., Zeng, Z., Wu, J., Peng, L., Lakshmi, V., Yang, H., and Liu, J.: Large Uncertainty on Forest Area Change in the Early 21st Century among Widely Used Global Land Cover Datasets, Remote Sensing, 12, 3502, https://doi.org/10.3390/rs12213502, 2020.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., and Loveland, T. R.: High-resolution global maps of 21stcentury forest cover change, science, 342, 850–853, 2013.

Lee, S. and Lee, D. K.: What is the proper way to apply the multiple comparison test?, Korean J Anesthesiol, 71, 353–360, https://doi.org/10.4097/kja.d.18.00242, 2018.

Zeng, Z., Wang, D., Yang, L., Wu, J., Ziegler, A. D., Liu, M., Ciais, P., Searchinger, T. D., Yang, Z.-L., Chen, D., Chen, A., Li, L. Z. X., Piao, S., Taylor, D., Cai, X., Pan, M., Peng, L., Lin, P., Gower, D., Feng, Y., Zheng, C., Guan, K., Lian, X., Wang, T., Wang, L., Jeong, S.-J., Wei, Z., Sheffield, J., Caylor, K., and Wood, E. F.: Deforestation-induced warming over tropical mountain regions regulated by elevation, Nature Geoscience, 14, 23–29, https://doi.org/10.1038/s41561-020-00666-0, 2021.

Duveiller, G., Hooker, J., and Cescatti, A.: The mark of vegetation change on Earth's surface energy balance, Nat Commun, 9, 679, https://doi.org/10.1038/s41467-017-02810-8, 2018.

Li, Y., Zhao, M., Mildrexler, D. J., Motesharrei, S., Mu, Q., Kalnay, E., Zhao, F., Li, S., and Wang, K.: Potential and Actual impacts of deforestation and afforestation on land surface temperature: IMPACTS OF FOREST CHANGE ON TEMPERATURE, J. Geophys. Res. Atmos., 121, 14,372-14,386, https://doi.org/10.1002/2016JD024969, 2016.

UC Berkely. Spring 2008 - Stat C141/ Bioeng C141 - Statistics for Bioinformatics