Response to Comments of Referee #2

General Comments:

Guan et al. generated a new NDVI dataset, STFLNDVI, by merging the data of MODIS NDVI and AVHRR GIMMS 3g. MODIS NDVI product has good data quality and a high spatial resolution but it is available since the year 2000. AVHRR GIMMS 3g product has been provided since 1982 but it has a relatively coarse spatial resolution (1/12 degree) and relatively poor data quality. The authors then performed the temporal filtering, normalization, and spatial-temporal fusing, making a new NDVI dataset of STFLNDVI with 1-km spatial resolution, covering the period of 1982-2015. Furthermore, the authors checked the temporal consistency, spatial stability, and spatial consistency of the new product during the overlapping periods of MODIS, ANCUS NDVI data. This draft was well-written, but I still have some comments on the algorithms used in this analysis, and I think the novelty is insufficient for a paper in ESD.

Response:

Dear Referee #2,

We are particularly grateful for your careful reading and constructive comments. Although there are still insufficiencies in the STFLNDVI product presented in our study, it is the first attempt to produce the global 1-km long-term NDVI dataset that may be helpful for global change studies. We have taken full consideration of your comments to improve the product and revise the manuscript to make it better. An item-by-item response follows. More experiments are conducted to assess the availability of STFLNDVI on spatial patterns and temporal variations. We are also reproducing the dataset by referring to the mean value calculated from the overlapping MODIS and AVHRR time series, in order to reduce the uncertainties caused by the selected one-year reference data.

Once again, we are particularly grateful for your careful reading and constructive comments. Thanks very much for your time.

Best regards,

Ph.D. Xiaobin Guan

Specific comments:

1. **Doubt on the reliability of the spatial variations at fine resolution. The original AVHRR product at coarse resolution can not provide any spatial variations within 1/12 x 1/12 pixels. The authors claimed that this newly generated NDVI product at 1-km resolution has the information of spatial variations within 1/12 x 1/12 resolution. Such spatial variations for every year are derived from the reference year (2014) of MODIS data. This means that, for STFLNDVI, the spatial variations within 1/12 x 1/12 resolution have no temporal change. This is no realistic, and I think the "high resolution" of STFLNDVI seems like a "pseudo high resolution".**

Response: Thanks for your deep thoughts and comments. It is true that the spatio-temporal fusion process was adapted only referring to one year in this study, such as 2000 or 2001 in producing the STFLNDVI and 2014 in the simulated experiments. Although this is the most commonly used method in the spatio-temporal fusion applications, i.e., improve the resolution of time series data only referring to some unique pairwise data [1-3], there may be insufficient spatial variations in the fusion results. The enhanced spatial patterns rely on the selected reference data at finer resolution, and the temporal changes are all from the time series data at poorer resolution. If there were no land cover changes or other extreme changes in the region, the spatial variations in the fusion results would be reliable because the spatial differences between the MODIS and AVHRR data are consistent in the reference year and the fusion year.

As a result, the reliability of the spatial variations in the fusion results may be doubtable mainly for the regions with obvious land cover changes, because the spatial variations within a coarse pixel from the reference year will not be the same as it is in the fusion year. However, we believe all the methods now can not solve this problem, because there are no actual spatial variations at fine-scale that can be obtained in the fusion year. All the useful spatial information in the fusion year is only from the data at coarse resolution, and the spatial variations from fine resolution data at other time is meaningless due to the land cover change. In this condition, our process can only enhance the spatial variations by combing the spatial patterns at fine resolution in the reference year and the information of land cover changes at coarse resolution in the fusion year. Due to the resample process for the coarse data, the spatial variations of fusion results within a coarse resolution is not only decided by the coarse pixel itself, but also impacted by the surrounded coarse pixels. As a result, there do have temporal changes in the spatial variations within the coarse resolution, which relies on the information provided by the coarse resolution data. If there are apparent differences between the two coarse data at fusion year and reference year, which indicates the obvious land cover changes, the calculation of weights in the applied spatio-temporal fusion method will force the fusion result to be more similar to the coarse data at the fusion year rather than using the spatial patterns from the fine resolution data at the reference year. It is afraid that the interannual variations of spatial patterns are hard to be reproduced at fine resolution, because no data can provide the actual spatial variations at the fusion year during the pre-MODIS period.

We added more content on this issue in the section of Discussion to declare the reliability of the spatial variations at fine resolution in the STFLNDVI. Besides, we are also reproducing the STFLNDVI product by referring to the mean data calculated from the overlapping MODIS and AVHRR time series, in order to obtain a more reasonable result and reduce the risk caused by the selected one-year reference data.

2. **Doubt on the reliability of the short-term temporal variations. When doing the normalization (section 3.1.2), the authors just used a linear model to make the multi-year mean value and trend of AVHRR data as same as MODIS data (as shown in Fig 3). The interannual variability or temporal variations within the year of STFLNDVI are from the AVHRR data without any correction. The short-term temporal variations of AVHRR aren't always consistent with those of MODIS data, for example in the regions around the equator. Merging two datasets may lead to some artificial variations.**

Response: Thanks for the comment. Indeed, the aim of producing the STFLNDVI is to using together with the MODIS data, in order to form a 1-km NDVI time series spanning over four decades. Due to the fact that all AVHRR sensors will be retired in the near future, it is imperative to extend NDVI from other sensors to maintain continuity and consistency of this global data set. As there is no other long-term global satellite NDVI dataset available before the 2000s, the temporal variation information from the AVHRR product is the only resource that can be trusted and applied. Besides, MODIS is a superior instrument to AVHRR in vegetation monitoring, so the temporal variation from MODIS is better to represent the vegetation changes after the 2000s [4,5]. As a result, although many studies have declared the inconsistency in the temporal variations between MODIS and AVHRR data, there are still many studies merging the two data for long-term analysis after some corrections [4-6].

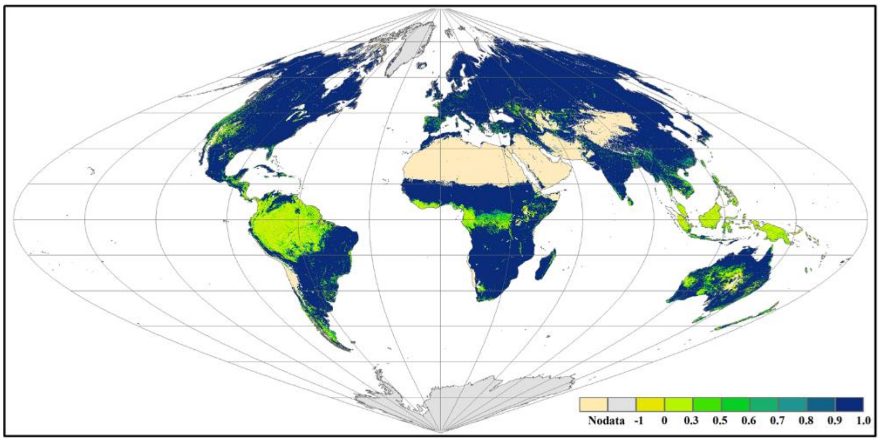

Figure 1. Global distribution of r between MODIS and AVHRR data during the overlapping 2000-2015 time period using a linear model.

Since the short-term temporal variations before the 2000s are from the STFLNDVI (i.e., from the AVHRR data only with a linear correction) and the short-term temporal variations after the 2000s are all from the MODIS data, the only problem in the short-term variations may exist around 2000. The applied normalization process is used to correct the short-term variations of AVHRR to be consistent with MODIS. Although the linear model seems quite simple, the results showed in Figure 3 in the manuscript are satisfactory. The results in Figure 1 further supported the satisfactory results of the linear model, which is the linear r between MODIS and AVHRR data over the overlapping 2000-2015 period. A value higher than 0.9 can be found in most of the world except for the region around the equator. Previous studies also showed that although the interpretation of GIMMS NDVI trends in humid areas should be done with certain reservations, it is well-suited with the MODIS for long-term vegetation studies in the non-humid areas [7]. Our result further supports these conclusions that the linear model can well build the relationship between AVHRR and MODIS data in most areas in the world, except for the regions with humid climates, such as the regions around the equator. In these areas, both the temporal variations in the MODIS and AVHRR data cannot be believed due to the cloudy climate-induced frequent noise in the time series. We have also tried some more methods in the normalization process, such as the moving window correction and many other non-linear functions. However, the results indicated that the linear model has the best performance in building the relationship between the two datasets worldwide.

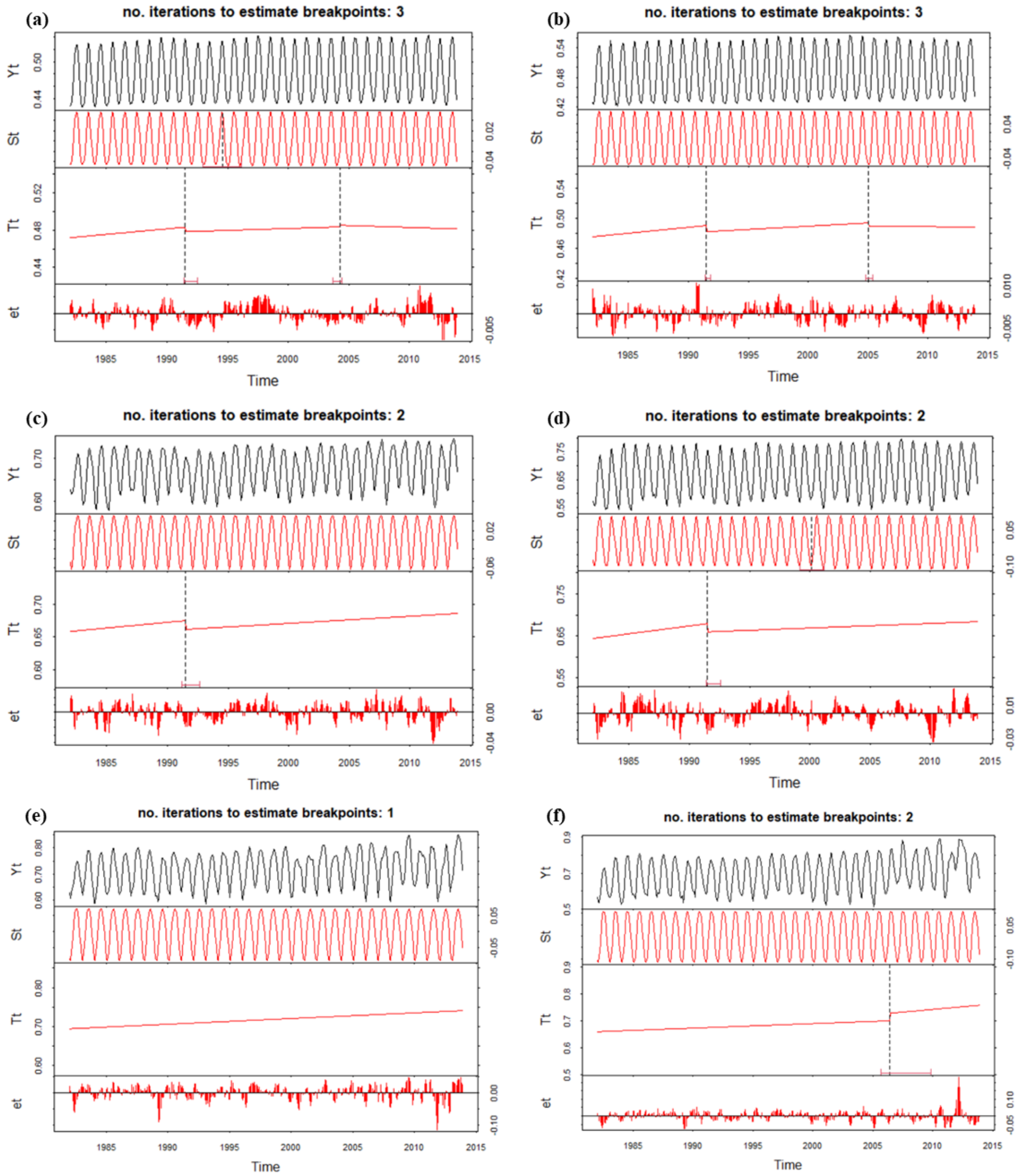

Figure 2. Comparison of the break points detected by the BFAST method between original AVHRR data and the combined time series using STFLNDVI and MODIS data. The left column are the results for AVHRR time series, and the right column are the results for the combined time series using STFLNDVI

and MODIS data. From up to bottom are the results for NDVI at different scales: (a) and (b) are the results for the mean NDVI over the world; (c) and (d) are the mean NDVI for the tile h27v06; and (e) and (f) are the NDVI for one pixel in tile h27v06 with a location of (100,200).

In order to further evaluate the short-term temporal variations of NDVI time series when combining the STFLNDVI and MODIS data, we compared the break points detected in the original AVHRR data and the combined time series. The BFAST breakpoint detection algorithm is applied, with the same parameters for both the two time series (minimum segment size h=0.3, season model = 'harmonic' and maximum iteration max.iter=5). In order to compare the short-term temporal variations at different scales, we compared the results of mean NDVI over the world, mean NDVI over a unique tile, and NDVI of an individual pixel. The results are shown in Figure 2. It can be observed that the break points detected in the two long-term NDVI time series are almost the same for the mean NDVI at the world scale and the single tile scale. No additional break points around the year 2000 can be observed in all the three scales after using the combination of STFLNDVI and MODIS time series, which indicates a reliable short-term temporal variation. Although the combined time series using STFLNDVI and MODIS data shows one additional break point in the time series of the pixel NDVI, it is occurred in 2006 due to the different variations of MODIS NDVI compared to AVHRR data after the 2000s. As a result, the break point of the combined time series using STFLNDVI and MODIS data is not caused by short-term temporal variations issues. Considering the better data quality of MODIS NDVI, the break point here may be a more reasonable result than using the AVHRR data only.

Some more discussion and explanation on the reliability of the short-term temporal variations between STFLNDVI and MODIS data have been added in the section of Discussion in our revised manuscript.

3. **Ln 300: I don't understand why the authors show the mean NDVI of the year 1990 here? How about the other years?**

Response: The reason why we only show the mean NDVI in the year 1990 is that it is impossible to show the results in all the years from 1982 to 1999 in the manuscript. As a result, we only show the result in one year that in the middle of 1982 and 1999. We also compared the mean NDVI from STFLNDVI and MODIS in many other years, a similar spatial distribution can also be observed. We will provide the comparison of mean NDVI every five years during the period 1982 and 1999 (i.e., 1982, 1985, 1990, 1995, 1999) in the supplementary materials.

4. **Ln 336: Why r in Europe (forest) and South of China (forest) is relatively low (Fig 4a)?**

Response: It is a good question. The main reason for the relatively low *r* between fusion results and actual MODIS data in Europe and the South of China is the relatively poor relationship between the original MODIS and AVHRR data, as shown in Figure 14 in the manuscript. Our analysis has indicated that the accuracy of fusion results highly relies on the original relationship between AVHRR and MODIS data. Thanks for your comments. We further analyzed the distribution of the regions with a poor relationship between the original MODIS and AVHRR data, and found it may mainly be caused by the cloudy climate in these regions. As shown in Figure 3, which is the mean precipitation calculated based on the MERRA-2 reanalysis data from 1980 to 2015, we can observe that there is relatively much higher precipitation than the surrounding land area both in Europe and the South of China, as well as regions around the equator. As a result, the remote sensing observations from the two sensors (AVHRR and MODIS) are hard to be consistently caused by the frequent noises induced by the cloudy climate, and lead to a relatively poor *r* in the fusion result.

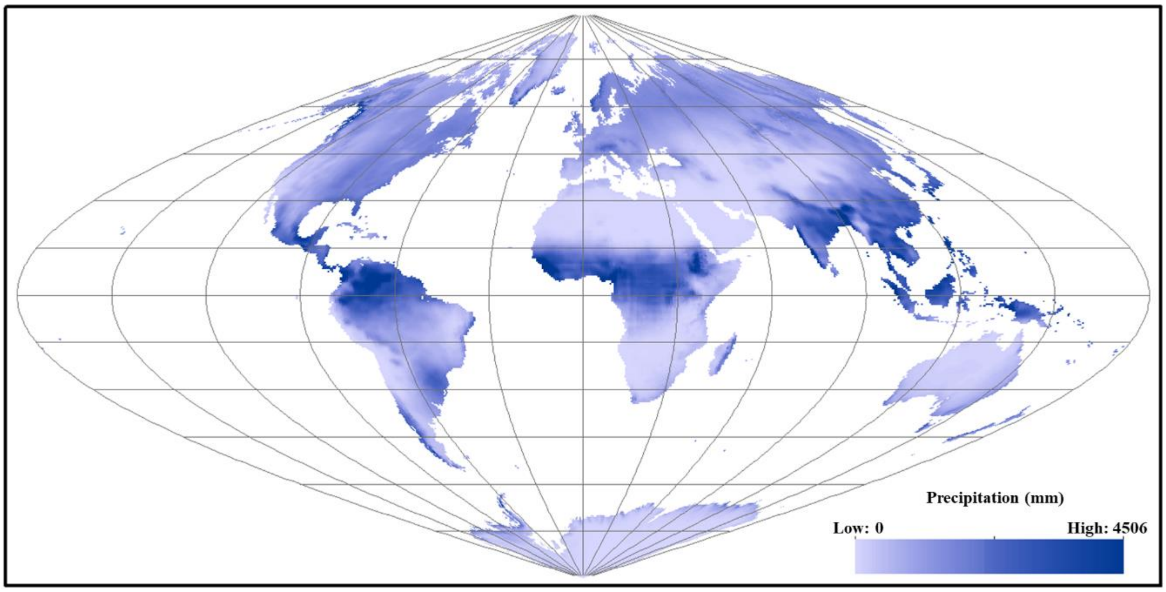

Figure 3. Global distribution of mean precipitation calculated from the MERRA-2 reanalysis data during the growing season in the Northern Hemisphere from 1980 to 2015.

5. **Fig 5, Ln 357: It would be better to give the statistics of rand bias for each climatic biome in Fig 5. The patterns in Fig 4 show low consistency in some regions, but this information has been hidden in global statistics.**

Response: Thanks for the comments. In order to further investigate the rand bias of the fusion result, we calculate the statistics of rand bias in 13 climatic biomes according to the Ko¨ppen-Geiger climate classification [8]. The Ko¨ppen-Geiger climate classification divides the global climate into seven zones and 13 types (Figure 4), respectively are Af: Tropical rainforest climate; Am: Tropical monsoon climate; Aw: Tropical sparse forest climate; Bs: Steppe climate; Bw: Desert Climate; Cf: Normally humid and warm climate; Cs: Dry summer and warm climate; Cw: Dry winter and warm climate; Df: Normal humidity and cold temperature climate; Ds: Subarctic continental climate; Dw: Sub-frigid monsoon climate; EF: Tundra climate; ET: Ice Field Climate.

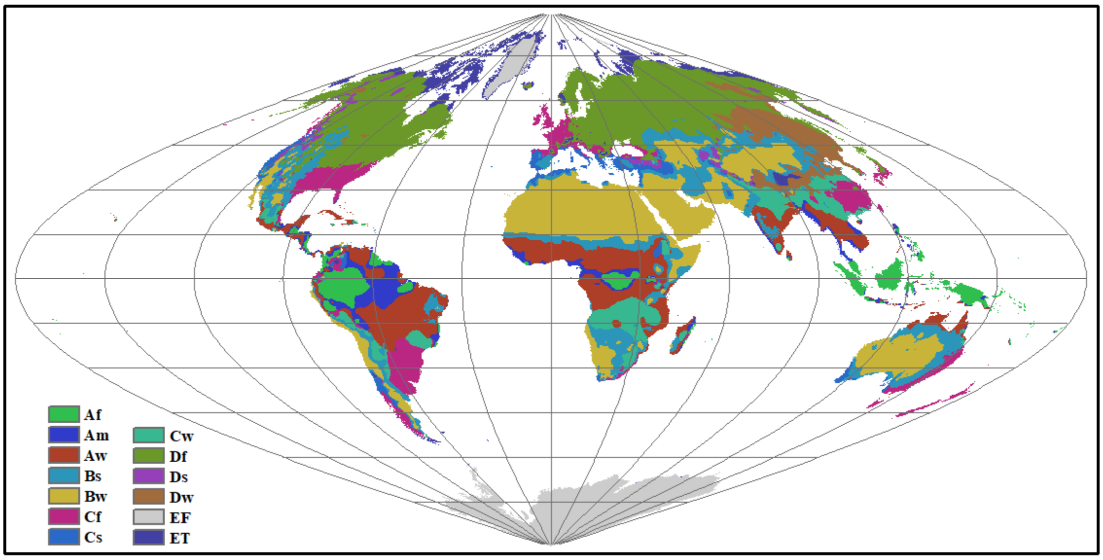

Figure 4. Global distribution of the Ko¨ppen-Geiger climate classification. Af: Tropical rainforest climate; Am: Tropical monsoon climate; Aw: Tropical sparse forest climate; Bs: Steppe climate; Bw: Desert Climate; Cf: Normally humid and warm climate; Cs: Dry summer and warm climate; Cw: Dry winter and warm climate; Df: Normal humidity and cold temperature climate; Ds: Subarctic continental climate; Dw: Sub-frigid monsoon climate; EF: Tundra climate; ET: Ice Field Climate.

As shown in Figure 5, the distribution of rand bias across the 13 climate types is quite similar to the overall distribution showed in Fig. 5 in the manuscript, except for the Tundra climate and Ice Field Climate area. The fusion results in the Tundra climate and Ice Field Climate area showed a relatively positive bias when compared to the true MODIS data, and there are some pixels showing a bias higher than 0.1. This is mostly due to these two climate types distributed in the high latitude area with very low vegetation cover. As a result, they may be meaningless when evaluating the accuracy of the vegetation index. Besides, in the other 11 climate types, the bias of the fusion data and true MODIS data is very low, with ideal distribution concentrated around zeros, and most of the pixels showing bias within the range of −0.1 to 0.1, or even −0.05 to 0.05. It indicates the fusion results are similar to the actual MODIS data with minor bias in almost all the climatic biomes. We believe the climate will impact the fusion results, mainly because the rainy climate will induce uncertainties in the original MODIS and AVHRR data. However, the division of climate zones is not totally related to the amount of precipitation or the rainy days, so it is hard to find the regions showing low consistency between fusion results and actual MODIS data in this way. The results may be more significant if calculating the statistics of rand bias for the regions with different degrees of total precipitation or rainly days.

We have added these two figures in the supplementary materials, and added some more discussion on the distribution of bias across climatic biomes in the section of Results in our revised manuscript.

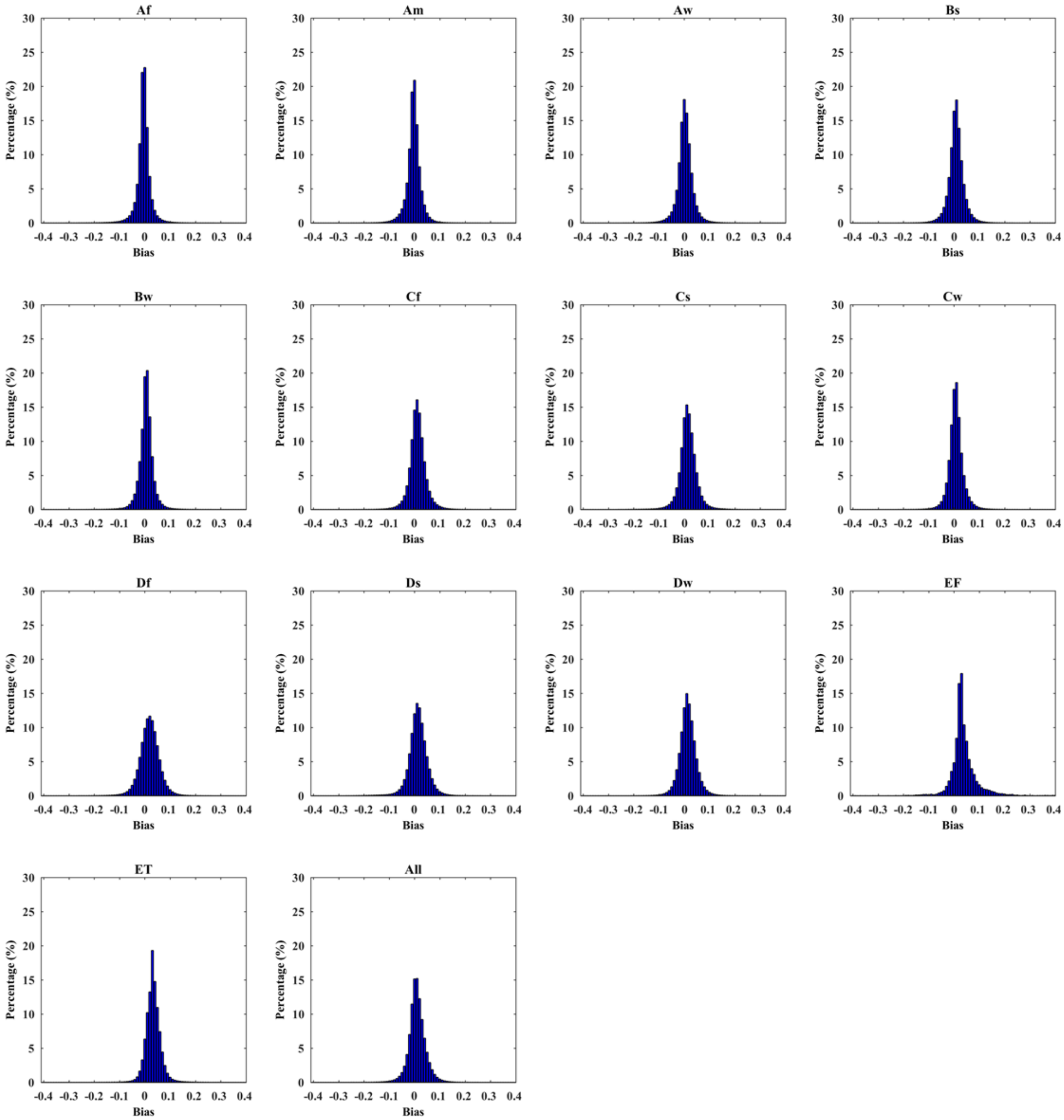

Figure 5. Statistics of rand bias for 13 climatic biomes according to the global Ko¨ppen-Geiger climate classification. Af: Tropical rainforest climate; Am: Tropical monsoon climate; Aw: Tropical sparse forest climate; Bs: Steppe climate; Bw: Desert Climate; Cf: Normally humid and warm climate; Cs: Dry summer and warm climate; Cw: Dry winter and warm climate; Df: Normal humidity and cold temperature climate; Ds: Subarctic continental climate; Dw: Sub-frigid monsoon climate; EF: Tundra climate; ET: Ice Field Climate.

6. **Fig 7, Ln 392: As shown in Fig 3, the mean difference between fusion results and MODIS is less than 0.1. In Fig 7, the bins of colorbar are 0.2 or 0.1. It is possible that patterns of fusion results and MODIS have some differences, but these differences can't be shown because of colorbar setting. Could you please check this?**

Response: Thanks for the comments. We checked this issue by calculating the difference between the true MODIS and fused data in our revised manuscript, as shown in Figure 6 (q-t). The difference between the original AVHRR and MODIS are also calculated for comparison (Figure 6 (m-p)). It can be observed that the difference between MODIS and fusion results is slight, with most of the area within the range ±0.05, and only a few pixels showing an absolute difference greater than 0.05. While, the differences between MODIS and AVHRR data are pretty obvious, with lots of pixels showing an absolute difference greater than 0.1. Besides, the distribution of the difference between MODIS and AVHRR is quite similar to the distribution of MODIS data, which means that the spatial patterns in AVHRR are almost missing. In contrast, the distribution of the difference between MODIS and fusion results is almost random, which indicates the fusion results can well reproduce a similar spatial distribution with MODIS.

We updated Figure 7 in the revised manuscript version and added these comparisons in the section of Results.

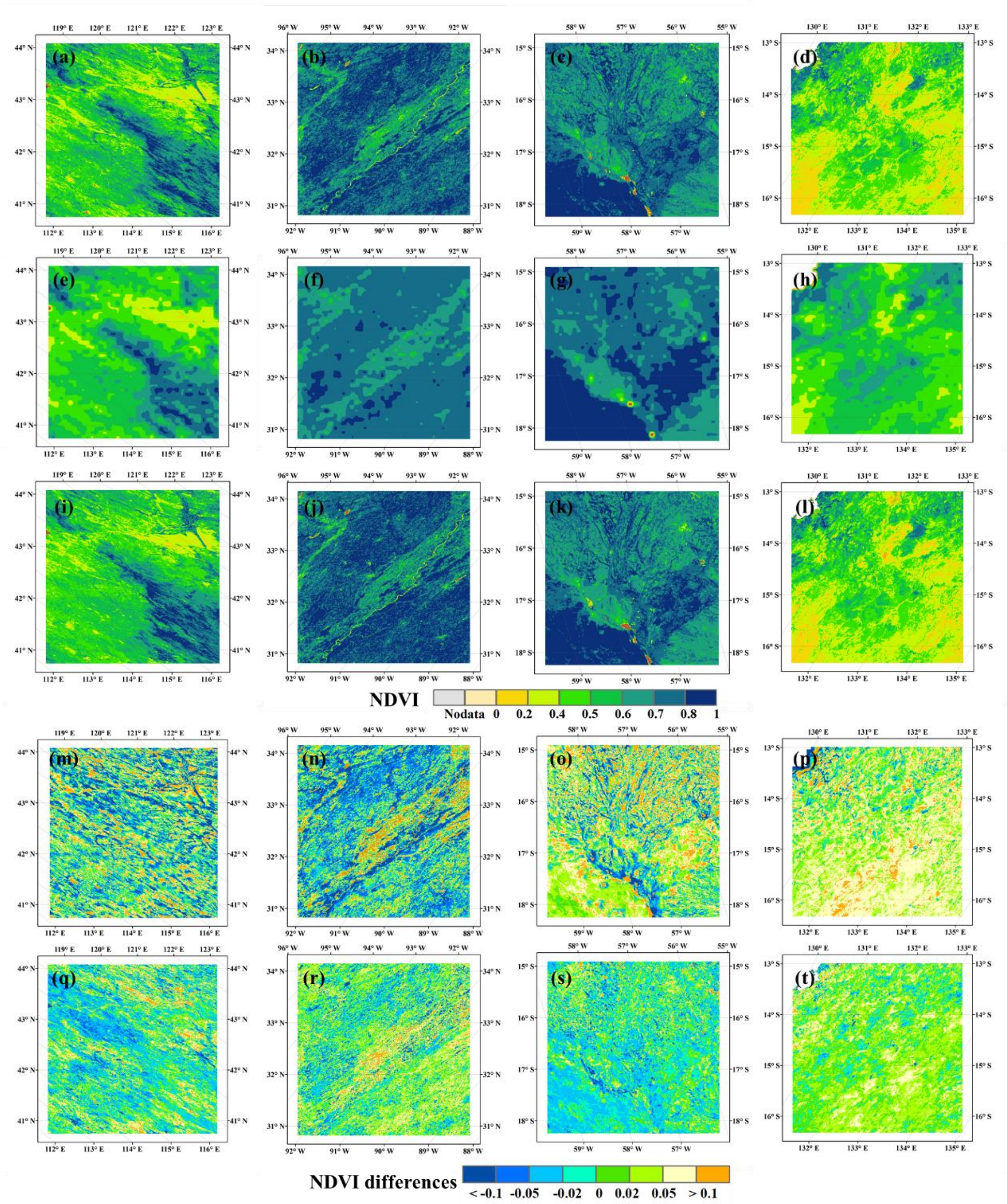

Figure 6. Subset spatial comparisons of the different data. (a)-(d) are the true MODIS data, (e)-(h) are the original AVHRR data, (i)-(l) are the fusion results, (m)-(p) are the differences between MODIS and AVHRR data, and (i)-(l) are the differences between MODIS and fusion results.

7. Other minor comments

Ln 36-37: The logic should be "NDVI can not only …vegetation coverage and growth status, which is associated with …". And it would be better to use "associated with" rather than "correlated with"

Ln 295: freely downloaded => downloaded for free

Fig 8 and 9: Colorbars are missing!

Response: Thanks for your careful reading. According to the comments, We have corrected the corresponding places, and added the color bars in Fig. 8 and 9. We also have checked through the manuscript to avoid other similar minor issues.

References:

[1] Zhou, Junxiong, et al. "Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction." Remote Sensing of Environment 252 (2021): 112130.

[2] Boyte, Stephen P., et al. "Fusing MODIS with Landsat 8 data to downscale weekly normalized difference vegetation index estimates for central Great Basin rangelands, USA." GIScience & Remote Sensing 55.3 (2018): 376-399.

[3] Hongtao, Jiang, et al. "Extending the SMAP 9-km soil moisture product using a spatio-temporal fusion model." Remote Sensing of Environment 231 (2019): 111224.

[4] Yang, Wenze, et al. "A novel re-compositing approach to create continuous and consistent cross-sensor/cross-production global NDVI datasets." International Journal of Remote Sensing 42.16 (2021): 6025-6049.

[5] Brown, Molly E., et al. "Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS." International Journal of Remote Sensing 29.24 (2008): 7141-7158.

[6] Mao, Dehua, et al. "Integrating AVHRR and MODIS data to monitor NDVI changes and their relationships with climatic parameters in Northeast China." International Journal of Applied Earth Observation and Geoinformation 18 (2012): 528-536.

[7] Fensholt, Rasmus, et al. "Evaluation of earth observation based long term vegetation trends—Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data." Remote sensing of environment 113.9 (2009): 1886-1898.

[8] Peel, Murray C., Brian L. Finlayson, and Thomas A. McMahon. "Updated world map of the Köppen-Geiger climate classification." Hydrology and earth system sciences 11.5 (2007): 1633-1644.