To Referee #1

Dear reviewer,

We are very pleased to finish a revised version of the manuscript essd-2023-1 entitled "**Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022**". In preparing this revision we have considered all your comments and incorporated most of the suggestions. **Temporal coverage of PKU GIMMS NDVI has been extended from 2020 to 2022.** We greatly appreciate your time and effort spent in reviewing this manuscript, which have improved the revised version of the manuscript.

Substantial improvements have been made based on your comments, including:

(1) Random forests regressor with a moving window and extra explanatory variables (longitude and latitude) was employed to consolidate PKU GIMMS NDVI (1982−2015) with MODIS NDVI (2003−2022).

(2) We explained the reason why only the previous version of GIMMS NDVI (GIMMS NDVI3g) was used for comparison.

(3) We have re-structured the PKU GIMMS NDVI published in ZENODO.

Below we provide point-to-point responses, each following the specific comment from the reviewer. All the changes have been marked in red in the revised manuscript.

Sincerely yours,

Zaichun Zhu, Ph. D. (on behalf of the author team)
School of Urban Planning and Design
Peking University
Tel: 86 185 0042 6608
Email: zhu.zaichun@pku.edu.cn

[Comment 1] *For the past decades, NDVI has been one of the most frequently used and verified remote sensing indices to monitor vegetation dynamics at regional and global scales, with its unique advantages including the simple form, a low sensor requirement, robustness to changes in plant canopies, insensitiveness to atmospheric effects and the sun-sensor geometry, etc. The developments of global long-term NDVI products have thus been critical. The manuscript presented a new version of GIMMS NDVI products (denominated as PKU GIMMS NDVI in the manuscript), (I) to address the issues of orbital drift and sensor degradation effects in AVHRR data using massive Landsat samples and (II) to extend the temporal coverage to recent using MODIS NDVI. All parts of the manuscript were well organized and written. The methodology is feasible and reliable to me, and the results provide relatively comprehensive evidence on the spatial and temporal accuracies of the products. The product could potentially benefit many future vegetation studies. I have an overall positive attitude on the product and basically recommend the manuscript for publication, as long as the authors address the following concerns.*

[Response 1] We thank the reviewer for this positive comment. One of the most significant features of PKU GIMMS NDVI is that it successfully eliminated the effects of NOAA satellite orbital drift and AVHRR sensor degradation. In the revised manuscript, we have further improved the data consolidation of PKU GIMMS NDVI with MODIS NDVI so as to better extend the temporal coverage. We hope all the concerns from the reviewer have been addressed.

[Comment 2] *My major concern is on the data consolidation with MODIS NDVI. I am happy that a pixel-scale linear fusion method was adopted, as the relationship between AVHRR NDVI and MODIS NDVI could be spatially different. However, the relationship might also be driven by some other factors such as the plant function type, phenological cycle, and sensor design, which I believe are beyond the capacity of the simple linear function used in the study (Mao et al., 2012). This could be confirmed by the failure of the method in some EBF (Figure 10). Also, the phrases 'before consolidation' and 'after consolidation' have frequently confused me when reading. To me, the data before* *consolidation are an intermediate product that certainly needs a discussion but is improper and unnecessary to compare with other NDVI products, for example, in trend analysis. It is not an individual product.*

[Response 2] Thank you for the suggestions. We agree that the relationship between GIMMS NDVI and MODIS NDVI could hardly be explained by the linear regression model with the only explanatory variable of NDVI. As such, we have replaced the simple linear function with the Random Forests (RF) regression model and have further incorporated the location information (longitude and latitude) as extra explanatory variables. To meet the sample size requirement of the RF model, an 11 × 11 moving window (approximately 1° equivalent) around each pixel has been used. The significance of each RF model has also been tested.

Also, we understand that the PKU GIMMS NDVI dataset before consolidation could be sometimes confusing. However, we may still retain it in the manuscript because it represents the version of PKU GIMMS NDVI that was generated solely based on AVHRR data. Its comparison with GIMMS NDVI3g reflects the efficacy of using the machine learning model and massive high-quality Landsat NDVI samples to calibrate uncertainties in NOAA/AVHRR data. This has been clarified in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Consolidation method in Section 3.3:

"In this study, we used a pixel-wise method inspired by Mao et al. (2012) to splice the PKU GIMMS NDVI product (1982–2015) and MODIS NDVI product (2003–2022). The pixel-wised method has been demonstrated more accurate than the global models (Yang et al., 2021). Specifically, the MODIS NDVI was first resampled to have the same spatial resolution (1/12°) and temporal resolution (half a month) as the PKU GIMMS NDVI (see Section 2.3). Then, during the overlapping periods (2003–2015), an 11 × 11 moving window (approximately 1° equivalent) was placed around each pixel. All the neighbors that had the same vegetation biome type with the center pixel were identified and their NDVI values were extracted from both products. This resulted in at most 1573 GIMMS-MODIS NDVI sample pairs (11 × 11 pixels per year in 13 years) for each pixel location. The sample pairs were further screened based on the data quality of PKU GIMMS NDVI (quality information adopted from GIMMS NDVI3g; see Section 2.2) and MODIS NDVI (see Section 2.3). Based on the sample pairs, the Random Forests (RF) regression model was constructed (Breiman, 2001), with explanatory variables of the PKU GIMMS NDVI and the longitude and latitude of samples and target variable of the MODIS NDVI. This study found that the significance of the RF model largely depended on the data quality of PKU GIMMS NDVI and MODIS NDVI. As such, we used 90% of the sample pairs for RF establishment and 10% for validation. R^2 was calculated. The pixel-wise RF model was applied to the non-overlapping period only when $R^2 > 0.2$ with $p < 0.001$; otherwise, the PKU GIMMS NDVI was adjusted by aligning its mean value to that of the MODIS NDVI." (Page 9, Line 244-258)

(b) Evaluation of PKU GIMMS NDVI in Section 3.4:

"The PKU GIMMS NDVI before consolidation was included because it represents the version of our NDVI product that is solely based on AVHRR data, and it can provide a more direct evaluation on the efficacy of the BPNN model and Landsat NDVI samples." (Page 11, Line 293-295)

[Comment 3] *The second issue is related to NDVI products used for comparison. I have noticed that only GIMMS NDVI3g was involved; meanwhile there exists some other global long-term NDVI products such as the LTDR4 and VIP3 (as the authors also mentioned in Introduction). So why were they excluded in the data comparison? Please clarify.*

[Response 3] Thanks for this comment. It has been a serious consideration for us to use the LTDR4 and VIP3 for data comparison. However, we then realized that a simple comparison could be superficial and misleading as the global long-term NDVI products were developed from different perspectives; and a comprehensive comparison could be challenging and out of the scope of this study. The performance of the NDVI products is spatiotemporally varied in terms of trend, seasonality, and absolute accuracy (e.g., Marshall et al., 2016; Tian et al., 2015). We believed that individual studies are required to better evaluate the performance of current NDVI products. As such, we chose to compare PKU GIMMS NDVI to its predecessor (GIMMS NDVI3g) to evaluate the framework proposed in this study. This strategy allows for well demonstrating the effectiveness of using massive high-quality Landsat NDVI samples to calibrate the bias in GIMMS NDVI3g, especially for the effects of NOAA satellite orbital drift and AVHRR sensor degradation. This framework could be also applied to other products. In the revised manuscript, we have proposed future work for a thorough evaluation of current global long-term NDVI products.

The following changes are made in the revised manuscript:

(a) Discussion on improvements of PKU GIMMS NDVI in Section 5.1:

"While this is out of the scope of the current study, future evaluation work is suggested to comprehensively compare the PKU GIMMS NDVI to other global long-term NDVI products such as the LTDR4 and VIP3." (Page 23-24, Line 424-426)

[Comment 4] *Last, the data published in ZENODO (https://zenodo.org/record/7441559#.Y7J7y3ZByCo) have not been well structured and the data size is way larger than the analogs. I suggest the authors, if possible, compress the data and stack the monthly NDVI to preferably every 10 years before uploading.*

[Response 4] We thank the reviewer for pointing this out. We have updated the PKU GIMMS NDVI dataset published in ZENODO. The dataset is now zipped and compressed every 10 years during 1982−2022 (i.e., 1982−1990, 1991−2000, 2001−2010, and 2011−2022). It is now available at https://doi.org/10.5281/zenodo.7509116.

[Comment 5] *Specific comments*

line 17-19

Better specify how were R2, MAE, and MAPE calculated.

[Response 5] Thanks for the suggestion. We have added and explained the equations of R^2, MAE, and MAPE in the revised manuscript. RMSE has also been included in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Evaluation of PKU GIMMS NDVI in Section 3.4:

"R^2 measures the percentage of variations that models can explain, RMSE measures the variance of errors, and MAE and MAPE measure absolute and relative error values at the sample level.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(NDVI_{Landsat} - NDVI_{GIMMS})^2}{\sum_{i=1}^{N}(NDVI_{Landsat} - \overline{NDVI_{Landsat}})^2} \tag{1}$$

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n}(NDVI_{Landsat} - NDVI_{GIMMS})^2}{N}} \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|NDVI_{Landsat} - NDVI_{GIMMS}| \tag{3}$$

$$MAPE\ \% = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{NDVI_{Landsat} - NDVI_{GIMMS}}{NDVI_{Landsat}}\right| \times 100\% \tag{4}$$

" (Page 10, Line 265-267)

[Comment 6] *Line 44-46*

This is particularly severe when estimating long-term changes (e.g., Shen et al 2022). Shen et al 2022, Plant phenology changes and drivers on the Qinghai–Tibetan Plateau. Nature Reviews Earth & Environment, 3, 633–651.

[Response 6] Thanks. We have modified the corresponding sentence to underscore the impact of uncertainties in NDVI products on long-term vegetation changes. The work from Shen et al. (2022) is now mentioned: "However, uncertainties in the NDVI products have also led to inconsistency not only between different products but also for the same product in different periods, placing many studies in a dilemma particularly when characterizing long-term changes (Wang et al., 2022; Zeng et al., 2022; Fensholt and Proud, 2012; Shen et al., 2022)." (Page 2, Line 46-49)

[Comment 7] *Line 126-127*

Could the authors provide more details here?

[Response 7] We performed a time-weighted aggregation method to convert the temporal resolution of the MODIS NDVI product (MOD13C1) from 16 days to half-month. The method was adopted from Zhu et al. (2013). Its central idea is to assign weights to all MOD13C1 scenes that could temporally intersect with a half-month interval, where the weight depends on the possibility of intersection. The half-month NDVI product was finally calculated as the weighted sum of the scenes. More details including schematic illustrations, can be found in Zhu et al. (2013).

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in Section 2.3:

"To match the temporal and spatial resolutions, we first performed a time-weighted aggregation method on MOD13C1 to produce an NDVI product at a temporal resolution of half-month. The method was adopted from Zhu et al. (2013). Its central idea is to assign weights to all MOD13C1 scenes that could temporally intersect with a particular half-month interval, where the weight depends on the possibility of intersection. The half-month NDVI product was finally calculated as the weighted sum of the scenes. We then performed nearest neighbor sampling to upscale the spatial resolution to 1/12°." (Page 5, Line 134-139)

[Comment 8] *Line 149*

How was the spatial aggregation used? By highest frequency?

[Response 8] We have investigated the manuscript and found out that the reviewer might refer to Line 139 (instead of Line 149) which was stating "The original spatial resolution of MCD12Q1 was 500 m, but it was spatially aggregated with Landsat NDVI samples to 1/12° to match the GIMMS NDVI3g product (Section 3.2)". After double-checking the MODIS Land-Cover Type data used in our study, we must clarify that two products were employed. The 500 m product (MCD12Q1) was employed to select sample locations in Landsat NDVI cross-calibration (Section 3.1.1), and the 0.05° product (MCD12C1) was employed to establish biome-specific BPNN models (Section 3.2.2). For the issue raised by the reviewer, the spatial aggregation from 0.05° to 1/12° (0.08°) in BPNN model establishment was achieved by the nearest neighbor sampling.

In the revised manuscript, Section 2.4 ("MODIS Land-Cover Type products (MCD12Q1 and MCD12C1, V6.1)") has been re-written to explain which and how MODIS Land-Cover Type products were used in the study.

The following changes are made in the revised manuscript:

(a) Introduction to MCD12Q1 in Section 2.4:

"The MODIS Land-Cover Type products provide global maps of land cover for each year between 2001–2019 (Friedl et al., 2002)." (Page 5, Line 141-142)

"This study employed two MODIS Land-Cover Type products with different spatial resolutions, i.e., 500 m (MCD12Q1) and 0.05° (MCD12C1). The MCD12Q1 was used to select sample locations for Landsat NDVI cross-calibration (Section 3.1.1). The MCD12C1 was used to establish biome-specific BPNN models with GIMMS NDVI3g after being spatially aggregated to 1/12° using the nearest neighbor resampling method (Section 3.2.2)." (Page 5, Line 150-153)

[Comment 9] *Line 196-197*

would 9 30m-resolution landsat pixels represent 8km GIMMS pixel? and was this done separately for each time step? How was the data temporally matched? Or did I miss something here?

[Response 9] We apologize for the confusion. In the Landsat NDVI sample selection, some details were missing because they were similar to those in Landsat NDVI cross-calibration (Section 3.1.1). Specifically, 40,000 random sample locations (1/12°) were first generated across the globe. Then at a time step of half-month, we identified sample locations with high-quality GIMMS NDVI3g data (QC=0), searched all available Landsat data, and uniformly placed 9 matrices of 20 × 20 Landsat pixels (30 m resolution) within each sample location. The quality of the Landsat pixel was determined in the same way as Section 3.1.1. We removed all matrices whose proportion of high-quality pixels < 90% (360 pixels). For those sample locations where more than half of the matrices were retained, we treated them as Landsat NDVI samples. The sample value was calculated as the average NDVI from high-quality Landsat pixels.

As for the reviewer's concern, (1) we used 9 matrices of 20 × 20 Landsat pixels, instead of 9 single Landsat pixels, to represent an 8 km GIMMS pixel. We believed that this sample size shall be a good compromise between representativeness and computation. (2) It is true that the sampling was done separately for each time step. (3) The Landsat pixels were selected as long as they fell in a particular time span of half-month. We understand that this might produce a maximum time difference of 14 days. However, we can hardly, if ever possible, infer the exact acquisition time of the AVHRR scene as the GIMMS NDVI3g is a maximum-value composite of half-month. Our practice might be an acceptable choice at the time.

The following changes are made in the revised manuscript:

(a) Introduction to Landsat NDVI sample selection in Section 3.2.1:

"A total of 40,000 sample locations were randomly selected from the GIMMS NDVI3g product with a spatial resolution of 1/12°. Then at a time step of half-month, we identified sample locations with high-quality GIMMS NDVI3g data (QC=0) and uniformly placed 9 matrices of 20 × 20 Landsat pixels within each location (1/12°). Landsat pixel values were extracted from all available scenes. Their quality was examined in the same way as Section 3.1.1. We removed all matrices whose proportion of high-quality pixels < 90% (360 pixels). The sample locations at a particular time were treated as Landsat NDVI samples if more than half (i.e. >= 5) of 9 matrices remained. The sample value was calculated as the average NDVI from high-quality Landsat pixels." (Page 8, Line 211-217)

[Comment 10] *Line 201 section 3.2.2*

was there any procedure performed to ensure the data quality of the two NDVI datasets?

[Response 10] The data quality of the GIMMS NDVI3g and Landsat NDVI samples has been ensured at sample locations. In this study, the quality of GIMMS NDVI3g was identified using its QC layer (QC=0: good quality) and the quality of Landsat NDVI was determined by a series of criteria including clouds, cloud shadows, water, snow, aerosol, radiation performance, etc. More details can be found in Section 3.2.1 of the revised manuscript (please also refer to **Response 9**).

[Comment 11] *Line 203-204*

were all the currently available landsat data included here? or just for the sample locations?

[Response 11] The BPNN models were established just for the sample locations. We acknowledged that better BPNN models could be established if all currently available Landsat data were included. However, that would produce a data volume too large to process in an efficient way. As such, we had to use samples instead. From a statistical perspective, we deemed that our 3.6 million high-quality Landsat NDVI samples could be a good representation of NDVI variations across time, space, and biome. This has also been clarified in updated Section 3.2.1 (please refer to **Response 9**).

[Comment 12] *Line 207-208*

Please consider if it is necessary to build more than one models for a given biome due to large spatial variability within the same biome (not mandatory).

[Response 12] This is a good suggestion. Indeed, the spatial information (latitude and longitude of the sample) has been incorporated as important explanatory variables in the BPNN model (Section 3.2.2). It accounts for the spatial variability of NDVI within the biome, as suggested by the reviewer. The results showed that the spatial information has greatly improved the performance of all BPNN models (Figure 4).

[Comment 13] *Line 219*

2016 or 2015?

[Response 13] Thanks for pointing out this mistake. The availability of GIMMS NDVI3g is until 2015 (Pinzon and Tucker, 2014). This has been corrected in the revised manuscript. More details on the status of GIMMS NDVI3g can be found at https://climatedataguide.ucar.edu/climate-data/ndvi-normalized-difference-vegetation-index-3rd-generation-nasagfsc-gimms. We have also gone through the manuscript to ensure there is no such error anymore.

The following changes are made in the revised manuscript:

(a) Consolidation method in Section 3.3:

"However, the latest data in GIMMS NDVI3g is until 2015 and no further upgrades will be provided." (Page 9, Line 236-237)

[Comment 14] *Line 220-222*

But some studies have shown that modis VI also has quality issues due to, such as, sensor degradation.

[Response 14] We apologize for the confusion. It is true that the MODIS Vegetation Index Product has long suffered from the sensor degradation effect. However, the effect has been argued well eliminated since collection 6.0 (Didan et al., 2015; Zhang et al., 2017). We have clarified this in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in Section 2.3:

"Compared to old versions, the latest MOD13C1 version 6.1 provides several algorithmic improvements and well corrects for the sensor degradation effect (Didan et al., 2015)." (Page 5, Line 128-129)

[Comment 15] *Line 226*

Which is the dependent variable here? MODIS NDVI or GIMMS NDVI?

[Response 15] The dependent variable is MODIS NDVI. We have explicitly indicated this in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Consolidation method in Section 3.3:

"Based on the sample pairs, the Random Forests (RF) regression model was constructed (Breiman, 2001), with explanatory variables of the PKU GIMMS NDVI and the longitude and latitude of samples and target variable of the MODIS NDVI." (Page 9, Line 252-254)

[Comment 16] *Line 228-229*

May be a little more detail about the spatial and temporal matching between modis and pku gimms ndvi would help. And was there any procedure performed to ensure the data *quality of the two NDVI datasets?*

[Response 16] Thanks for this suggestion. The temporal matching between MODIS NDVI and PKU GIMMS NDVI adopted the method from Zhu et al. (2013). Basically, the method assigns weights to all 16-day MODIS scenes (MOD13C1 in this study) that could temporally intersect with a particular half-month interval, where the weight depends on the possibility of intersection. Then, the half-month MODIS NDVI was calculated as the weighted sum of the scenes. The spatial matching between MODIS NDVI and PKU GIMMS NDVI directly upscaled the resolution of MODIS NDVI from 0.05° to 1/12° (or ~0.008°) using nearest neighbor sampling. This has been clarified in Section 2.3 (please refer to **Response 7**).

The data quality of both PKU GIMMS NDVI and MODIS NDVI has been ensured via their quality layer. The MODIS NDVI provides a layer of pixel reliability and the PKU GIMMS NDVI adopted quality information from GIMMS NDVI3g. We further tested the significance of the pixel-wise model built between the two NDVI datasets using the 10% validation sample pairs. The model was considered significant and applied to the non-overlapping periods only when $R^2 > 0.2$ with $p < 0.001$; otherwise, the PKU GIMMS NDVI was adjusted by aligning its mean value to that of the MODIS NDVI.

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in Section 2.3:

"MOD13C1 provides a pixel reliability layer that distinguishes good-quality data from no data, marginal data, snow/ice, and cloudy and estimated data." (Page 5, Line 133-134)

(b) Consolidation method in Section 3.3:

"The sample pairs were further screened based on the data quality of PKU GIMMS NDVI (quality information adopted from GIMMS NDVI3g; see Section 2.2) and MODIS NDVI (see Section 2.3). Based on the sample pairs, the Random Forests (RF) regression model was constructed (Breiman, 2001), with explanatory variables of the PKU GIMMS NDVI and the longitude and latitude of samples and target variable of the MODIS NDVI. This study found that the significance of the RF model largely depended on the data quality of PKU GIMMS NDVI and MODIS NDVI. As such, we used 90% of the sample pairs for RF establishment and 10% for validation. R^2 was calculated. The pixel-wise RF model was applied to the non-overlapping period only when $R^2 > 0.2$ with $p < 0.001$; otherwise, the PKU GIMMS NDVI was adjusted by aligning its mean value to that of the MODIS NDVI." (Page 9, Line 251-258)

[Comment 17] *Line 229*

2004-2015 or 2001-2015?

[Response 17] We apologize for this confusion. We have double-checked the overlapping period between PKU GIMMS NDVI and MODIS NDVI that was used for data consolidation. We now confirm that the overlapping period is 2013−2015. It should be noted that although the MODIS NDVI based on Terra (MOD13C1) has been available since 2000, we could only find year-round global MOD13C1 since 2003 in both USGS Earth Explorer (http://earthexplorer.usgs.gov/) and NASA Earth Data (https://search.earthdata.nasa.gov/search). This has been clarified in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Last paragraph in Introduction Section:

"In this context, this study uses the long-term Landsat data to develop a new version of the GIMMS NDVI product (PKU GIMMS NDVI) (1982–2022) from the GIMMS NDVI3g (current version) (1982–2015) and MODIS NDVI products (2003–2022)." (Page 3, Line 88-90)

(b) Introduction to MOD13C1:

"In this study, the year-round global MOD13C1 during 2003−2022 was employed." (Page 5, Line 132-133)

(c) Consolidation method in Section 3.3:

"In this study, we used a pixel-wise method inspired by Mao et al. (2012) to splice the PKU GIMMS NDVI product (1982–2015) and MODIS NDVI product (2003–2022)." (Page 9, Line 244-245)

(d) Figure 6 caption:

“**Figure 6.** Accuracies of the GIMMS NDVI3g and PKU GIMMS NDVI products measured by R^2 for pre-MODIS (1982–2000) and MODIS (2001–2015) period. The R^2 was calculated between the NDVI products and Landsat NDVI samples. (a) to (d) shows the spatial distributions of R^2 in 2°× 2° grids. Non-vegetated grids and grids with less than 20 validation samples are marked in white. (e) and (f) shows the probability distribution of R^2 differences between the two periods (before 2000 and after 2000) and between the two products (GIMMS NDVI3g and PKU GIMMS NDVI), respectively.”

[Comment 18] *Line 252*

does landsat have an orbit shift problem?

[Response 18] Yes, the Landsat, like many other satellites, has the orbital shift problem though the problem has been minimized (Storey et al., 2014). However, Landsat provides the longest space-based record of Earth's land and has been one of the most reliable data sources with great accuracies and consistencies in geometric and radiometric properties (Wulder et al., 2019; Wulder et al., 2016). It could best serve as reference data for our study. This has been clarified in the Introduction section.

The following changes are made in the revised manuscript:

(a) Last paragraph in Introduction Section:

“Landsat sensors have a high spatial resolution, low frequencies of sensor change, and in particular, high accuracy and consistency in geometric and radiometric properties (Zhang et al., 2021; Weng et al., 2014; Dong et al., 2020; Storey et al., 2014).” (Page 3, Line 78-80)

[Comment 19] *Figure 5*

Please delete ‘unitless’ in the figure.

[Response 19] Thank for pointing it out. The Figure 5 has been updated and ‘unitless’ has been deleted.

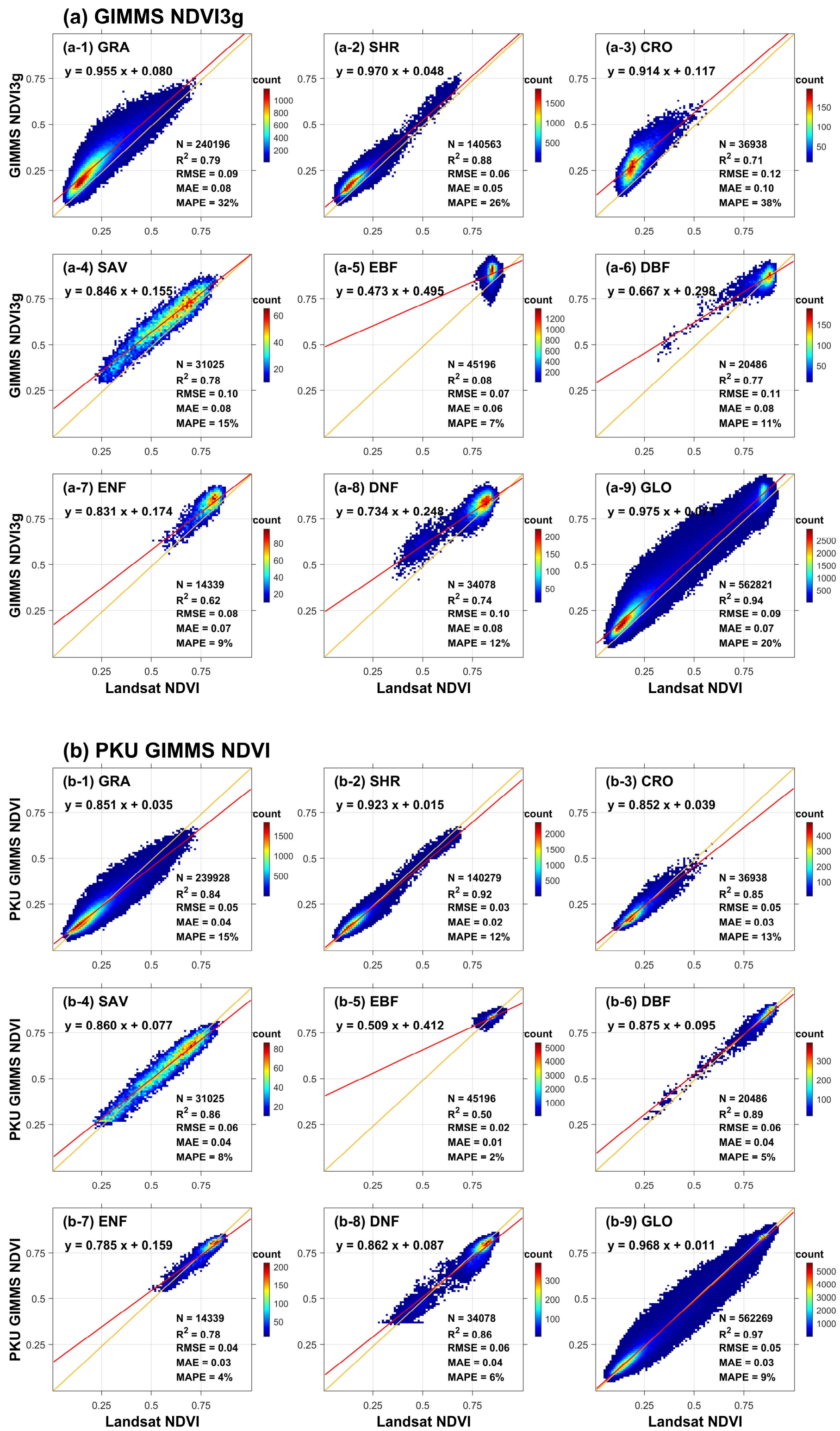

Figure 5. Direct validation of the (a) GIMMS NDVI3g and (b) PKU GIMMS NDVI products. Individual Landsat NDVI samples from 1984 to 2015 were used in the validation at a 1/12° resolution. Orange lines represent a 1:1 line. GLO represents the global vegetation biome.

References:

Breiman, L.: Random Forests, Mach. Learn., 45(1), 5−32, https://doi.org/10.1023/A:1010933404324, 2001.

Didan, K., Munoz, A., Solano, R., and Huete, A.: MODIS Vegetation Index User's Guide, Version 3.00, Collection 6, 2015.

Marshall, M., Okuto, E., Kang, Y., Opiyo, E., and Ahmed, M.: Global assessment of Vegetation Index and Phenology Lab (VIP) and Global Inventory Modeling and Mapping Studies (GIMMS) version 3 products, Biogeosciences, 13, 625–639, https://doi.org/10.5194/bg-13-625-2016, 2016.

Pinzon, J. E. and Tucker, C. J.: A Non-Stationary 1981-2012 AVHRR NDVI3g Time Series, Remote Sens, 6, 6929-6960, https://doi.org/10.3390/rs6086929, 2014.

Storey, J., Choate, M., and Lee, K.: Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance, Remote Sens, 6, 11127–11152. https://doi.org/10.3390/rs61111127, 2014

Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., and Wang, Y.: Evaluating temporal consistency of long-term global NDVI datasets for trend analysis, Remote Sens Environ, 163, 326–340. https://doi.org/10.1016/j.rse.2015.03.031, 2015.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., and Roy, D. P.: The global Landsat archive: Status, consolidation, and direction, Remote Sens Environ, 185, 271-283, https://doi.org/10.1016/j.rse.2015.11.032, 2016.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosillo, T., Hipple, J. D., Hostert, P., Hughes, M. J., Huntington, J., Johnson, D. M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T. A., Schaaf, C., Schott, J. R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J. C., Wynne, R. H., and Zhu, Z.: Current status of Landsat program, science, and applications, Remote

Sens Environ, 225, 127-147, https://doi.org/10.1016/j.rse.2019.02.015, 2019.

Zhang, Y., Song, C., Band, L.E., Sun, G., and Li, J.: Reanalysis of global terrestrial vegetation trends from MODIS products: Browning or greening?, Remote Sens Environ, 191, 145–155, https://doi.org/10.1016/j.rse.2016.12.018, 2017.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011, Remote Sens, 5, 927-948, https://doi.org/10.3390/rs5020927, 2013.

To Referee #2

Dear reviewer,

We are very pleased to finish a revised version of the manuscript essd-2023-1 entitled "**Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022**". In preparing this revision we have considered all your comments and incorporated most of the suggestions. **Temporal coverage of PKU GIMMS NDVI has been extended from 2020 to 2022.** We greatly appreciate your time and effort spent in reviewing this manuscript, which have improved the revised version of the manuscript.

Substantial improvements have been made based on your comments, including:

(1) We have provided more details on how we performed the time-weighted aggregation method to convert the temporal resolution of the MODIS NDVI product (MOD13C1) from 16 days to half-month.

(2) We have also elaborated on the method used to splice the PKU GIMMS NDVI and MODIS NDVI.

Below we provide point-to-point responses, each following the specific comment from the reviewer. All the changes have been marked in red in the revised manuscript.

Sincerely yours,

Zaichun Zhu, Ph. D. (on behalf of the author team)
School of Urban Planning and Design
Peking University
Tel: 86 185 0042 6608
Email: zhu.zaichun@pku.edu.cn

[Comment 1] *This study proposes PKU GIMMS NDVI, a new global long-term NDVI* *time series data that covers 1982 to 2020 based on AVHRR and MODIS sensors onboard satellite platforms. The PKU GIMMS NDVI extends the GIMMS NDVI3g data and has better data quality. It has better agreement with Landsat NDVI compared to GIMMS NDVI3g, alleviating the orbital drift problem in the AVHRR sensors. The method proposed in this study could be used to generate consistent global NDVI data in the future, which would help study global terrestrial biosphere dynamics.*

[Response 1] We thank the reviewer for helping improve our manuscript. We believe that the framework proposed in this study, which employs massive high-quality Landsat NDVI samples and a data consolidation method, could benefit future work that aims to generate long-term remote sensing-based earth data. We hope that our PKU GIMMS NDVI could provide more accurate vegetation monitoring in the context of global environmental changes.

[Comment 2] *Detailed comments:*

Line 14: "global-wide" may be simplified as "global"

[Response 2] We thank the reviewer for pointing it out. The term "global-wide" has been simplified to "global" in the new version of the manuscript.

The following changes are made in the revised manuscript:

(a) Abstract Section:

"In this study, we presented a machine learning model that employed massive high-quality global Landsat NDVI samples and a data consolidation method" (Page 1, Line 14-15)

[Comment 3] *Line 26: When introducing NDVI, please cite the original NDVI paper: Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ, 351(1), p.309.*

[Response 3] Thank you for providing the original NDIV paper from Rouse et al. (1974). It is now cited in the manuscript.

The following changes are made in the revised manuscript:

(a) First paragraph of the Introduction Section:

“The Normalized Difference Vegetation Index (NDVI) characterizes the biophysical, biochemical, and physiological conditions of vegetation (Rouse et al., 1974; Rondeaux et al., 1996; Gao et al., 2000; Yin et al., 2022).” (Page 1, Line 28-29)

[Comment 4] *Line 97: Please provide literature references to the data sources.*

[Response 4] Thanks for the suggestion. The literature references were originally available under sub-sections of specific data. Now, they have also been added in the data overview.

The following changes are made in the revised manuscript:

(a) First paragraph of the Data Section:

“Four global satellite products were used in this study: Landsat Surface Reflectance data (Collection 1 Tier 1) (Masek et al., 2006; Vermote et al., 2016), MODIS Land-Cover Type product (V6.1) (Friedl et al., 2002), GIMMS NDVI3g product (V1.0) (Pinzon and Tucker, 2014), and MODIS NDVI product (V6.1) (Didan, 2021).” (Page 4, Line 101-103)

[Comment 5] *Line 126: How is the time-weighted aggregation performed? Please explain in detail.*

[Response 5] In the revised manuscript, we have provided more details on the time-weighted aggregation method, which converted the temporal resolution of the MODIS NDVI product (MOD13C1) from 16-day to half-month. The method was adopted from Zhu et al. (2013). Its central idea is to assign weights to all MOD13C1 scenes that could temporally intersect with a particular half-month interval, where the weight depends on the possibility of intersection. The half-month NDVI product was finally calculated as the weighted sum of the scenes. More details including schematic illustrations can be found in Zhu et al. (2013).

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in Section 2.3:

“To match the temporal and spatial resolutions, we first performed a time-weighted aggregation method on MOD13C1 to produce an NDVI product at a temporal resolution of half-month. The method was adopted from Zhu et al. (2013). Its central idea is to assign weights to all MOD13C1 scenes that could temporally intersect with a particular half-month interval, where the weight depends on the possibility of intersection. The half-month NDVI product was finally calculated as the weighted sum of the scenes. We then performed nearest neighbor sampling to upscale the spatial resolution to 1/12°." (Page 5, Line 134-139)

[Comment 6] *Line 127: Maybe the authors want to say "upscale" instead of "downscale."*

[Response 6] The reviewer is right. The spatial resampling from 0.05° to 1/12° (0.083°) should be an upscaling process. Thanks for this and we have modified the wording in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in Section 2.3:

"We then performed nearest neighbor sampling to upscale the spatial resolution to 1/12°." (Page 5, Line 138-139)

[Comment 7] *Line 139: How were the Landsat NDVI samples aggregated to 1/12°? Please explain in detail.*

[Response 7] The reviewer referred to this sentence "The original spatial resolution of MCD12Q1 was 500 m, but it was spatially aggregated with Landsat NDVI samples to 1/12° to match the GIMMS NDVI3g product (Section 3.2)". Here we apologize for the confusion as two MODIS Land-Cover Type products (MCD12Q1 in 500 m and MCD12C1 in 0.05°), rather than only MCD12Q1, were employed in this study. The MCD12Q1 was employed to select sample locations in Landsat NDVI cross-calibration (Section 3.1.1), and the MCD12C1 was employed to establish biome-specific BPNN models (Section 3.2.2).

As for the issue raised by the reviewer, the Landsat NDVI samples were aggregated to 1/12° for training BPNN models with GIMMS NDVI3g (1/12°). Specifically, 40,000 random sample locations (1/12°) were first generated across the globe. Then at a time step of half-month, we identified sample locations with high-quality GIMMS NDVI3g data (QC=0), searched all available Landsat data, and uniformly placed 9 matrices of 20 × 20 Landsat pixels (30 m resolution) within each sample location. We have added many details in Section 2.4 and Section 3.2.1 to clarify this.

The following changes are made in the revised manuscript:

(a) Introduction to MCD12Q1 and MCD12C1 in Section 2.4:

"The MODIS Land-Cover Type products provide global maps of land cover for each year between 2001–2019 (Friedl et al., 2002)."

"This study employed two MODIS Land-Cover Type products with different spatial resolutions, i.e., 500 m (MCD12Q1) and 0.05° (MCD12C1). The MCD12Q1 was used to select sample locations for Landsat NDVI cross-calibration (Section 3.1.1). The MCD12C1 was used to establish biome-specific BPNN models with GIMMS NDVI3g after being spatially aggregated to 1/12° using the nearest neighbor resampling method (Section 3.2.2)." (Page 5, Line 150-153)

(b) Landsat NDVI sample selection in Section 2.4:

"A total of 40,000 sample locations were randomly selected from the GIMMS NDVI3g product with a spatial resolution of 1/12°. Then at a time step of half-month, we identified sample locations with high-quality GIMMS NDVI3g data (QC=0) and uniformly placed 9 matrices of 20 × 20 Landsat pixels within each location (1/12°). Landsat pixel values were extracted from all available scenes." (Page 8, Line 211-214)

[Comment 8] *Line 144: "temporal" should be "temporally."*

[Response 8] Thanks for pointing out this grammar error. It has been fixed in the revised manuscript.

The following changes are made in the revised manuscript:

(a) First paragraph of the Methodology Section:

"1) Landsat sensor cross-calibration to create temporally consistent Landsat data as a benchmark;" (Page 6, Line 157-158)

[Comment 9] *Line 226: could the authors elaborate more on how they spliced the PKU GIMMS NDVI and MODIS NDVI?*

[Response 9] In the revised manuscript, more details have been provided on the consolidation of PKU GIMMS NDVI and MODIS NDVI, including how we used an 11 × 11 moving window to establish the pixel-wise Random Forests (RF) model during the overlapping periods of 2003–2015, how we tested the significance of the RF model, and how we applied the RF model to the non-overlapping period.

The following changes are made in the revised manuscript:

(a) Consolidation method in Section 3.3:

"In this study, we used a pixel-wise method inspired by Mao et al. (2012) to splice the PKU GIMMS NDVI product (1982–2015) and MODIS NDVI product (2003–2022). The pixel-wised method has been demonstrated more accurate than the global models (Yang et al., 2021). Specifically, the MODIS NDVI was first resampled to have the same spatial resolution (1/12°) and temporal resolution (half a month) as the PKU GIMMS NDVI (see Section 2.3). Then, during the overlapping periods (2003–2015), an 11 × 11 moving window (approximately 1° equivalent) was placed around each pixel. All the neighbors that had the same vegetation biome type with the center pixel were identified and their NDVI values were extracted from both products. This resulted in at most 1573 GIMMS-MODIS NDVI sample pairs (11 × 11 pixels per year in 13 years) for each pixel location. The sample pairs were further screened based on the data quality of PKU GIMMS NDVI (quality information adopted from GIMMS NDVI3g; see Section 2.2) and MODIS NDVI (see Section 2.3). Based on the sample pairs, the Random Forests (RF) regression model was constructed (Breiman, 2001), with explanatory variables of the PKU GIMMS NDVI and the longitude and latitude of samples and target variable of the MODIS NDVI. This study found that the significance of the RF model largely depended on the data quality of PKU GIMMS NDVI and MODIS NDVI. As such, we used 90% of the sample pairs for RF establishment and 10% for validation. R^2 was calculated. The pixel-wise RF model was applied to the non-overlapping period only when $R^2 > 0.2$ with $p < 0.001$; otherwise, the PKU GIMMS NDVI was adjusted by aligning its mean value to that of the MODIS NDVI."

(Page 9, Line 244-258)

[Comment 10] *Line 254: How can seasonal fluctuations in the time series of NDVI bias be removed via the multi-year averaging method? Please explain.*

[Response 10] The seasonal fluctuations were removed by subtracting the multi-year average at a particular time of the year from the original data, i.e.,

$$bias\_deseason_{i,j} = bias\_origin_{i,j} - mean(bias\_origin_j) \quad (1)$$

Where $bias\_origin_{i,j}$ is the original NDVI bias at the time j of the year i (e.g., the first half-month of January in 2005); $mean(bias_origin_j)$is the multi-year average at the time j (e.g., average on the first half-month of January for all years); and $bias\_deseason_{i,j}$ is the NDVI bias after removing the seasonal fluctuation. We have clarified this in the revised manuscript.

The following changes are made in the revised manuscript:

(a) PKU GIMMS NDVI evaluation in Section 3.3:

"Seasonal fluctuations in the time series of NDVI bias were first removed by subtracting the multi-year average at a particular time of the year, i.e.,

$$bias\_deseason_{y,t} = bias\_origin_{y,t} - mean(bias\_origin_t) \quad (5)$$

Where $bias\_origin_{y,t}$ is the original NDVI bias at the time t of the year y (e.g., the first half-month of January in 2005); $mean(bias\_origin_t)$ is the multi-year average at the time t (e.g., the first half-month of January for all years); and $bias\_deseason_{y,t}$ is the NDVI bias after removing the seasonal fluctuation." (Page 10, Line 279-284)

[Comment 11] *Line 257: Maybe it should be "... was evaluated at 1,000 random points ..."*

[Response 11] Thank you. The sentence has been re-written in the revised manuscript.

The following changes are made in the revised manuscript:

(a) PKU GIMMS NDVI evaluation in Section 3.3:

"The consolidation of PKU GIMMS NDVI with MODIS NDVI was evaluated at 1,000 random points for each vegetation biome type." (Page 11, Line 286-287)

[Comment 12] *Figure 2: maybe the authors could also show the regression line and equation in each panel?*

[Response 12] Figure 2 has been updated to show the regression lines and equations for all panels.

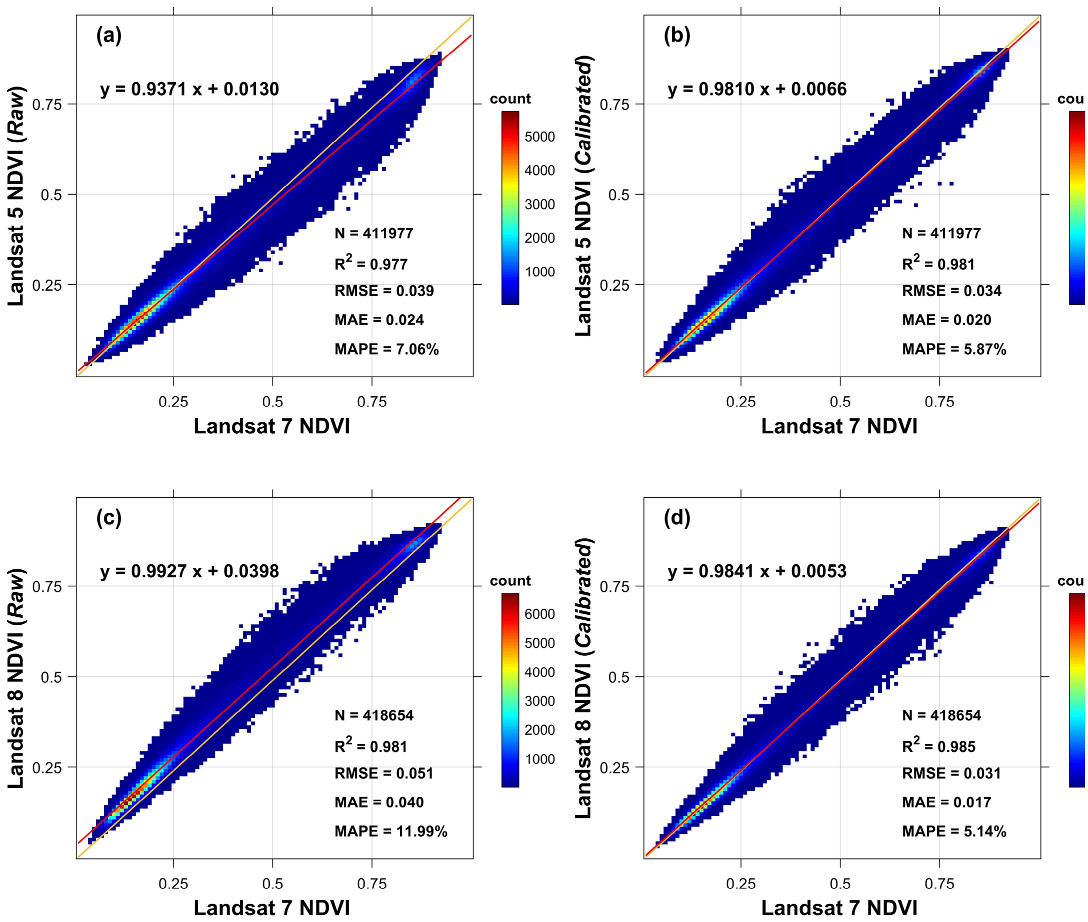

Figure 2. The efficiency of NDVI cross-calibration between Landsat sensors. (a) Landsat 7 NDVI vs. uncalibrated Landsat 5 NDVI. (b) Landsat 7 NDVI vs. calibrated Landsat 5 NDVI. (c) Landsat 7 NDVI vs. uncalibrated Landsat 8 NDVI. (d) Landsat 7 NDVI vs. calibrated Landsat 8 NDVI. The red line is the regression line and the orange diagonal line represents a 1:1 relationship. The size of the NDVI interval in the maps is 0.01. NDVI intervals with sample number < 10 were omitted.

[Comment 13] *Line 301: The section title could be "Validation of PKU GIMMS NDVI and GIMMS NDVI3g"*

[Response 13] Thanks for this comment. We used the term "Direct validation" instead of "Validation" for the section title because we would like to distinguish our validation analysis based on reference samples from those based on inter-product comparison analysis. After serious consideration, we decide to keep the term "Direct validation". We sincerely appreciate the suggestion and understanding from the reviewer.

[Comment 14] *Figure 6: please explain in the figure caption how the R2 was computed in detail.*

[Response 14] The R^2 was calculated between Landsat NDVI samples and the GIMMS NDVI3g/PKU GIMMS NDVI products (Section 3.4). We recognized that the current figure caption could be misleading ("Comparison of R^2 between the GIMMS NDVI3g and PKU GIMMS NDVI products in pre-MODIS (1982–2000) and MODIS (2001–2015) period."). In the revised manuscript, the caption has been clarified and the calculation of R^2 has been explained.

The following changes are made in the revised manuscript:

(a) Figure 6 caption:

"Accuracies of the GIMMS NDVI3g and PKU GIMMS NDVI products measured by R^2 for pre-MODIS (1982–2000) and MODIS (2001–2015) period. The R^2 was calculated between the NDVI products and Landsat NDVI samples…"

[Comment 15] *Line 337: The section title could be "Comparison with MODIS NDVI"?*

[Response 15] Thanks for this suggestion. The title of Section 4.3.1 has been changed to "Comparison with MODIS NDVI" in the revised manuscript.

References:

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011, Remote Sens, 5, 927-948, https://doi.org/10.3390/rs5020927, 2013.

To Referee #3

Dear reviewer,

We are very pleased to finish a revised version of the manuscript essd-2023-1 entitled "**Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022**". In preparing this revision we have considered all your comments and incorporated most of the suggestions. **Temporal coverage of PKU GIMMS NDVI has been extended from 2020 to 2022.** We greatly appreciate your time and effort spent in reviewing this manuscript, which have improved the revised version of the manuscript.

Substantial improvements have been made based on your comments, including:

(1) We have clarified the rationality to use Landsat NDVI samples.

(2) Figures have been implemented and updated to show the annual anomalies and trends of all NDVI products involved in this study for a full time span of 1982−2022.

(3) The greening trend in most tropical regions during 2003−2015.

Below we provide point-to-point responses, each following the specific comment from the reviewer. All the changes have been marked in red in the revised manuscript.

Sincerely yours,

Zaichun Zhu, Ph. D. (on behalf of the author team)
School of Urban Planning and Design
Peking University
Tel: 86 185 0042 6608
Email: zhu.zaichun@pku.edu.cn

[Comment 1] *Overall Comments:*

A reliable long-term vegetation time series is critical to understand the dynamic of *vegetation and its feedback to the climate. This study by Li et al reconstructs a spatiotemporally consistent global NDVI dataset for 1982-2020 integrating Back Propagation Neural Network and a total of 3.6 million Landsat NDVI samples that well spread across the globe as input. This product, along with its predecessor (the GIMMS NDVI3g dataset), has been evaluated with the Landsat NDVI samples, showing substantial improvement. This study has originality and significance in uniqueness and usefulness.*

[Response 1] We appreciate the reviewer for confirming the uniqueness and significance of the study. The idea of using massive high-quality Landsat NDVI samples ensures the spatiotemporal consistency and reliability of the PKU GIMMS NDVI dataset and eliminates the effects of NOAA satellite orbital drift and AVHRR sensor degradation. We hope that the PKU GIMMS NDVI could provide a more solid data basis for global vegetation dynamic studies.

[Comment 2] *Below are some comments that may help to further improve the manuscript. First, it seems that the golden truth of NDVI is Landsat NDVI samples. I recommend providing details in section 3.2.1 of Landsat NDVI samples to illustrate: 1) Why Landsat NDVI is more accurate than other products? 2) Does Landsat NDVI have any limitations (e.g. the influence of clouds)? 3) Adding a plot to show the distribution of these samples (time and space).*

[Response 2] We thank the reviewer for providing these recommendations. More details have been added in Section 3.2.1 of the revised manuscript. We hope the following questions have been answered:

1) Why Landsat NDVI is more accurate than other products?

The reason is that Landsat sensors have unparalleled radiometric and geometric accuracy and stability with the longest continuity, global coverage, and relatively high spatial resolution (Wulder et al., 2019; Wulder et al., 2016). This has been partially explained in the Introduction section. Now we further state it in the beginning of Section 3.2.1.

2) Does Landsat NDVI have any limitations (e.g. the influence of clouds)?

Landsat NDVI did have limitations. Because of the relatively long revisit time and small field-of-view (see Introduction section), it suffers from influences from clouds, cloud shadows, water, snow, aerosols, etc. These limitations make Landsat NDVI unsuitable for global long-term vegetation trend analysis but could be an ideal candidate as reference data. This study used a rigorous process that considered the abovementioned factors as well as others such as the eruption of Mount Pinatubo and radiation performance to retain high-quality Landsat pixels only (see Section 3.1.1).

3) Adding a plot to show the distribution of these samples (time and space).

A new sub-figure has been added to Figure 3. Now the spatial, seasonal, and interannual distribution of the Landsat NDVI samples are presented.

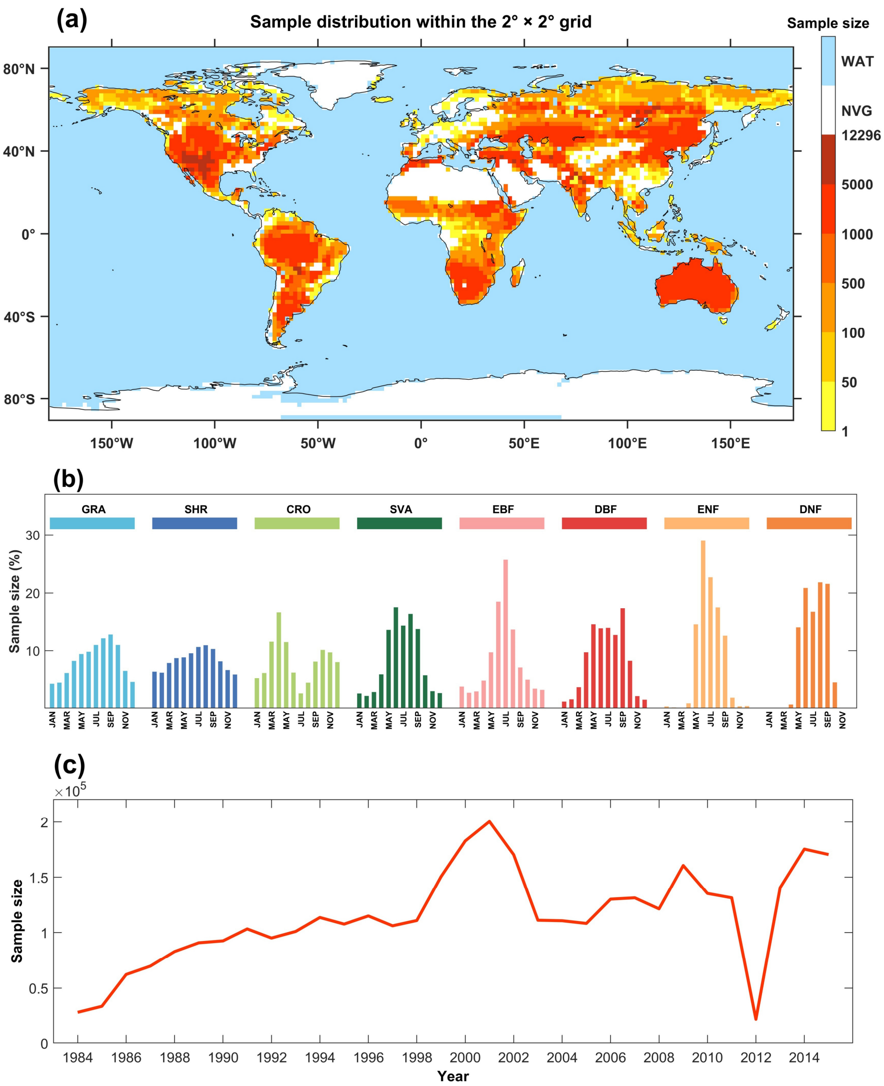

Figure 3. Spatial and temporal distribution of refined Landsat NDVI samples (3.6 million). (a) Distribution of Landsat NDVI samples within the 2° × 2° grid. (b)

Percentage of samples among the eight vegetation biome types in each month. (c) Annual variation of Landsat NDVI sample size.

The following changes are made in the revised manuscript:

(a) Landsat NDVI sample selection in Section 3.2.1:

"Landsat data is known for its unparalleled radiometric and geometric accuracy and stability, as well as the longest continuity, global coverage, and relatively high spatial resolution (Wulder et al., 2019; Wulder et al., 2016)." (Page 8, Line 209-211)

(b) Spatiotemporal variations of Landsat NDVI samples in Section 4.2.1:

"During 1984−2015, the Landsat NDVI sample size generally increased from Landsat 5 to Landsat 7 and Landsat 8 except for two periods. Between 1999 and 2003, the sample size was significantly larger as both Landsat 5 and Landsat 7 were available; and between November 2011 and May 2012, very few images were acquired when Landsat 5 was decommissioning (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-4-5-thematic-mapper-tm-level-1-data) and Landsat 8 was not available yet (Figure 3c)." (Page 13, Line 320-324)

[Comment 3] *Second, this dataset extends the time-span of 1982-2015 for its predecessor to 1982-2020, but there is no figure or analysis of this extension. It would be great to show more details of this extension (e.g. a long time series spanning from 1982 to 2015). In addition, it would be great to explain the results of consolidating the PKU dataset with MODIS NDVI data for the years after 2016. A comparison between MODIS and PKU NDVI datasets from 2017 to 2020 would be helpful (e.g. Figure 12).*

[Response 3] We thank the reviewer for this suggestion. In the revised manuscript, we have updated Figure 12 to show the full time span of 1982−2022 and to draw the time-series of all NDVI products involved in this study, i.e., PKU GIMMS NDVI before consolidation (1982−2015), PKU GIMMS NDVI after consolidation (1982−2022), MODIS NDVI (2003−2022), and GIMMS NDVI3g (1982−2015). Biome-specific version of updated Figure 12 has also been implemented in the Appendix. Anomalies and trends of the NDVI products have been compared and analyzed.

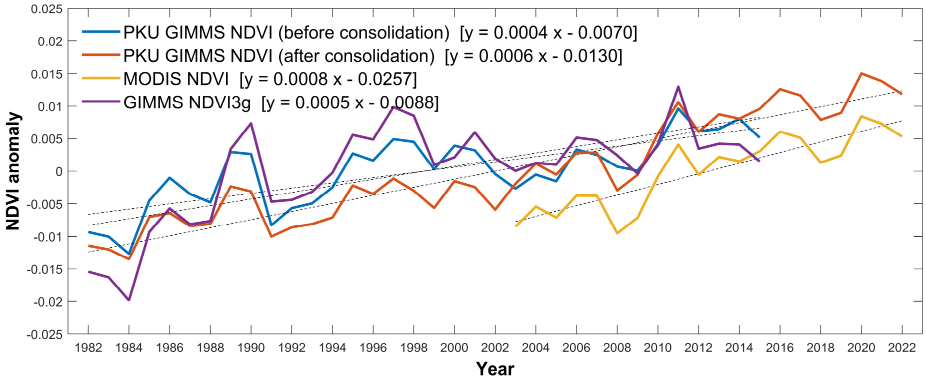

Figure 12. Annual anomalies and trends of PKU GIMMS NDVI (before consolidation), PKU GIMMS NDVI (after consolidation), MODIS NDVI, and GIMMS NDVI3g. The NDVI anomalies were calculated as area-weighted annual averages.

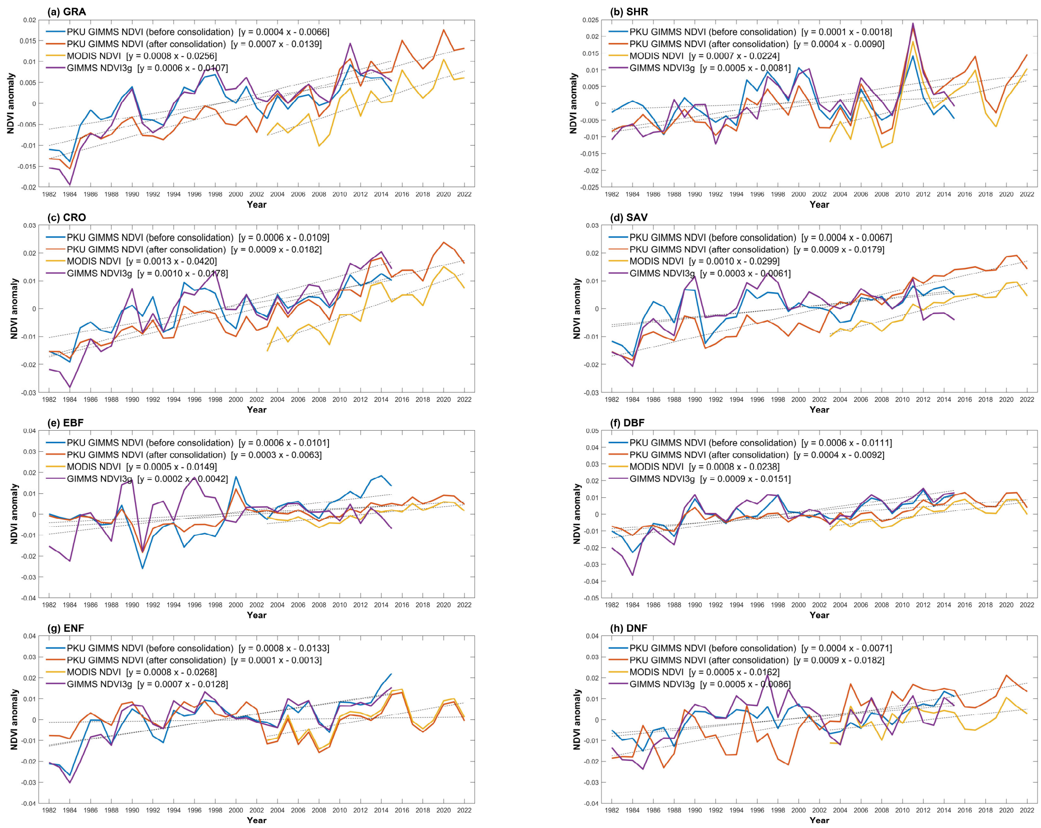

Figure A1. Annual anomalies and trends of PKU GIMMS NDVI (before consolidation), PKU GIMMS NDVI (after consolidation), MODIS NDVI, and GIMMS NDVI3g for different vegetation biome types. The NDVI anomalies were calculated as area-weighted annual averages.

The following changes are made in the revised manuscript:

(a) PKU GIMMS NDVI evaluation in Section 3.4:

"Trends from multiple NDVI products, i.e., GIMMS NDVI3g, MODIS NDVI, and PKU GIMMS NDVI (before and after consolidation), were compared over their overlapping period. The PKU GIMMS NDVI before consolidation was included because it represents the version of our NDVI product that is solely based on AVHRR data, and it can provide a more direct evaluation on the efficacy of the BPNN model and Landsat NDVI samples." (Page 11, Line 291-295)

(b) Vegetation trend analysis in Section 4.3.2:

"The time series of annual NDVI anomalies and trends from different products are shown in Figure 12. All products presented a similar shape of anomalies in their overlapping periods. During 1982−2015, PKU GIMMS NDVI before consolidation had a similar trend with GIMMS NDVI3g (0.4×10^{-3} yr^{-1} vs. 0.5×10^{-3} yr^{-1}). During 2003−2015 when all NDVI products were available, PKU GIMMS NDVI before consolidation (0.8×10^{-3} yr^{-1}), PKU GIMMS NDVI after consolidation (0.9×10^{-3} yr^{-1}), and MODIS NDVI (0.9×10^{-3} yr^{-1}) were similar in vegetation trend (trend values not shown in the Figure), higher than GIMMS NDVI3g (0.5×10^{-3} yr^{-1}). In the EBF area, GIMMS NDVI3g showed a browning trend since 2003 due to the impact of orbital drift and sensor degradation (Figure A1), which was consistent with the research by Wang et al. (2022). In PKU GIMMS NDVI products, the effect of orbital drift and sensor degradation has been alleviated. It showed a greening trend in EBF, consistent with MODIS NDVI (Figure A1)." (Page 22, Line 397-405)

[Comment 4] *Some minor comments: The Landsat NDVI samples were used both to train and validate the PKU GIMMS NDVI dataset. How the Landsat NDVI samples were separated into two groups?*

[Response 4] The Landsat NDVI samples were separated into 80% for model training and 20% for NDVI product evaluation (GIMMS NDVI3g and PKU GIMMS NDVI).

The following changes are made in the revised manuscript:

(a) Introduction to Landsat NDVI sample selection in Section 3.2.1:

"The samples were also divided into 80% for model training and 20% for NDVI product evaluation." (Page 8, Line 217-218)

[Comment 5] *The figure caption for Figure 6 is confusing: comparison of R2 between* *the GIMMS NDVI4g and PKU GIMMS NDVI products. Is it R2 between the GIMMS NDVI4g and PKU GIMMS NDVI products? Or R2 between the GIMMS NDVI4g and Landsat NDVI, and that between PKU GIMMS NDVI and Landsat NDVI?*

[Response 5] We apologize for the confusion. It is R^2 between the GIMMS NDVI4g and Landsat NDVI, and that between PKU GIMMS NDVI and Landsat NDVI. We have clarified this in the revised manuscript. The way of how R^2 was calculated has also been added.

The following changes are made in the revised manuscript:

(a) Figure 6 caption:

"**Figure 6.** Accuracies of the GIMMS NDVI3g and PKU GIMMS NDVI products measured by R^2 for pre-MODIS (1982–2000) and MODIS (2001–2015) period. The R^2 was calculated between the NDVI products and Landsat NDVI samples. (a) to (d) shows the spatial distributions of R^2 in 2°× 2° grids. Non-vegetated grids and grids with less than 20 validation samples are marked in white. (e) and (f) shows the probability distribution of R^2 differences between the two periods (before 2000 and after 2000) and between the two products (GIMMS NDVI3g and PKU GIMMS NDVI), respectively."

[Comment 6] *Why MOD13C1, not MOD13Q1, MOD13A3 or MOD13C2, is used in this study?*

[Response 6] In this study, the MODIS Vegetation Index product was used to consolidate with PKU GIMMS NDVI after the year 2003. The main differences between MOD13C1 and other MODIS Vegetation Index products are that it is (1) derived from MODIS Terra, (2) 16-day-based, and (3) of 0.05° resolution. MODIS Terra available since 2000 has a longer on-orbit period than MODIS Aqua (available since 2002). The temporal resolution of 16 days and the spatial resolution of 0.05° are close to those of PKU GIMMS NDVI (half-month; 1/12°). As such, MOD13C1 was selected. Note that we could only find year-round global MOD13C1 since 2003 in both USGS Earth Explorer (http://earthexplorer.usgs.gov/) and NASA Earth Data (https://search.earthdata.nasa.gov/search).

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in section 2.3:

"As the MODIS NDVI product was used to consolidate with PKU GIMMS NDVI, we chose MOD13C1 over other MODIS Vegetation Index products because it was derived from MODIS Terra which has been available since 2000 and it has a close temporal (16 days) and spatial resolution (0.05°) to those of PKU GIMMS NDVI (half-month and 1/12°)." (Page 5, Line 130-132)

[Comment 7] *Figure 11: In the PKU dataset, most tropical regions show greening during 2004-2015, which may be unreasonable.*

[Response 7] This is a good comment. The vegetation trend in tropical regions after 2000 has been an arguable topic. In terms of NDVI, there is a substantial difference between existing GIMMS-based (from AVHRR) and MODIS-based products. Current findings from multiple studies suggested that in tropical regions GIMMS-based NDVI presented a decreasing trend while MODIS-based NDVI presented an increasing trend (Fensholt and Proud, 2012; Tian et al., 2015; Wang et al., 2022). Possible reasons for the trend discrepancy in tropical/humid areas could be a lack of high-quality data (Fensholt and Proud, 2012), orbital drift effects for GIMMS NDVI (Tian et al., 2015), and NDVI saturation (Wang et al., 2022). In the generation of PKU GIMMS NDVI, we attempted to account for all these factors. For example, we used global long-term massive high-quality Landsat NDVI as reference data to calibrate GIMMS NDVI3g via BPNN models. The BPNN models employed multiple types of explanatory variables to further explain NDVI variations in time, space, and satellite. The effects of NOAA satellite orbital drift and AVHRR sensor degradation have been efficiently eliminated in PKU GIMMS NDVI. After the efforts, PKU GIMMS NDVI (before or after consolidation with MODIS NDVI) demonstrated an increasing trend in tropical regions, as the reviewer has figured out in Figure 11. Whereas this phenomenon in PKU GIMMS NDVI is different from old versions of GIMMS-based NDVI products, we believe that PKU GIMMS NDVI could be a more reliable reference in studying long-term vegetation trends, especially in tropical regions. This point of view has been discussed in the revised manuscript.

The following changes are made in the revised manuscript:

(a) Introduction to MOD13C1 in section 2.3:

"The improvements in PKU GIMMS NDVI may help to clarify some discrepancies between existing NDVI products, for instance, the vegetation trend in humid tropical regions after 2000. In these regions, current findings from multiple studies suggested that GIMMS-based NDVI presented a decreasing trend while MODIS-based NDVI presented an increasing trend (Fensholt and Proud, 2012; Tian et al., 2015; Wang et al., 2022). Possible reasons could be the uncertainties from NDVI saturation or lack of high-quality data (Fensholt and Proud, 2012; Wang et al., 2022) and orbital drift effects for GIMMS NDVI (Tian et al., 2015). In the generation of PKU GIMMS NDVI, these uncertainties have been well accounted for and we found an increasing NDVI trend in tropical regions after 2000, both before and after data consolidation with MODIS NDVI." (Page 24, Line 427-433)

[Comment 8] *Figure 4: S5 shows no significant improvement compared to S4.*

[Response 8] We agree with the reviewer that S5 shows no significant improvement compared to S4. Indeed, we were also struggling between S4 and S5 when establishing BPNN models. But after analyzing all four error metrics (R^2, RMSE, MAE, and MAPE), we determined that S5 did provide some improvement over S4 in all vegetation biomes despite the improvement was generally subtle. In the end, we chose to use S5 where NDVI, spatial information, time information, NOAA satellite number, and years since its launch were included in BPNN models.

[Comment 9] *Figure 8: A color bar is missing.*

[Response 9] Thanks. Color bars have been added to sub-figures in the revised manuscript.

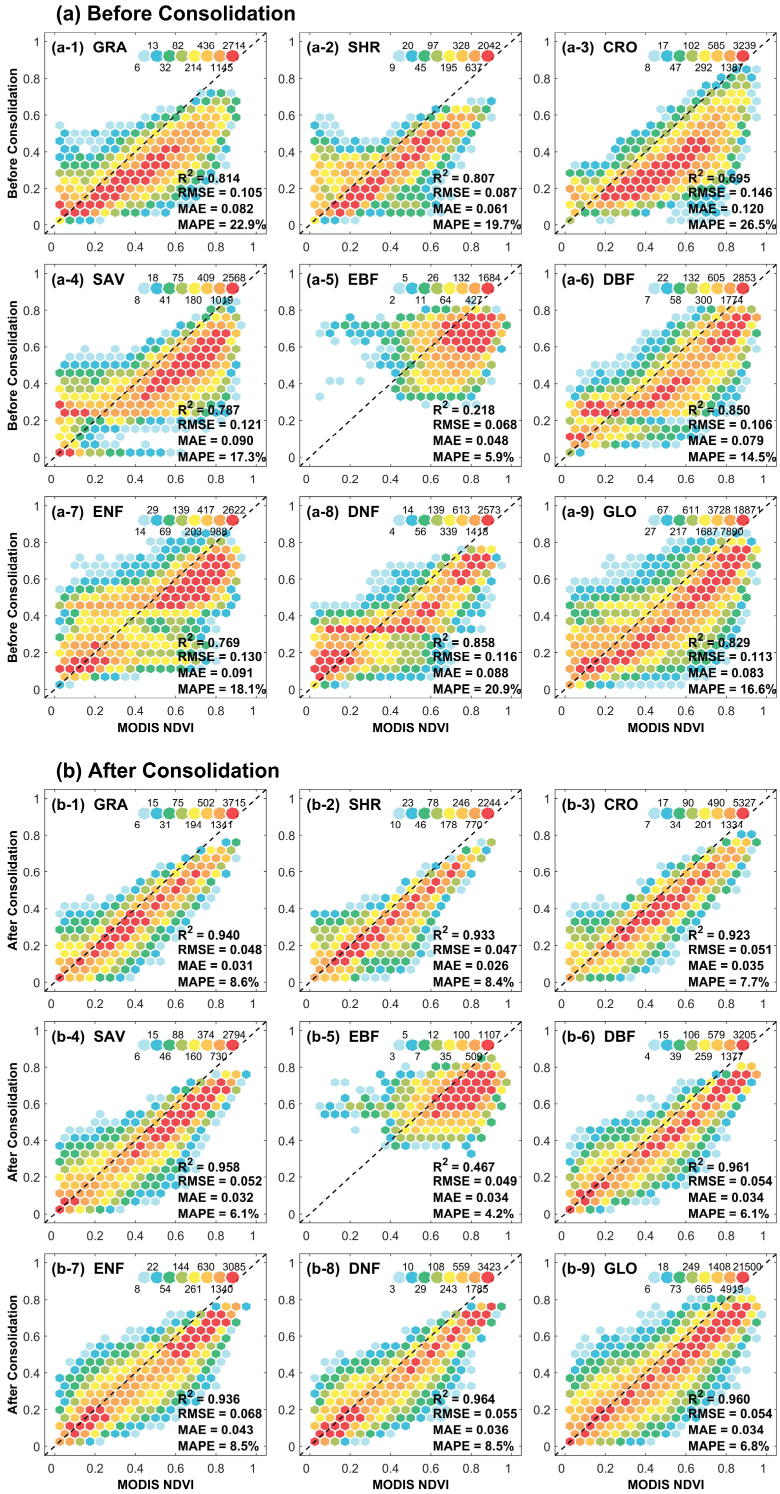

Figure 8. Direct validation of the PKU GIMMS NDVI product (a) before and (b) after consolidation. The validation was performed using 1,000 MODIS NDVI samples at a 1/12° resolution for each vegetation biome type from 2004 to 2015.

References:

Fensholt, R., and Proud, S.R.: Evaluation of Earth Observation based global long term vegetation trends — Comparing GIMMS and MODIS global NDVI time series, Remote Sens Environ, 119, 131–147, https://doi.org/10.1016/j.rse.2011.12.015, 2012.

Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., and Wang, Y. J.: Evaluating temporal consistency of long-term global NDVI datasets for trend analysis, Remote Sens Environ, 163, 326-340, https://doi.org/10.1016/j.rse.2015.03.031, 2015.

Wang, Z., Wang, H., Wang, T., Wang, L., Liu, X., Zheng, K., and Huang, X.: Large discrepancies of global greening: Indication of multi-source remote sensing data, Global Ecol. Conserv., 34, e02016. https://doi.org/10.1016/j.gecco.2022.e02016, 2022

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., and Roy, D. P.: The global Landsat archive: Status, consolidation, and direction, Remote Sens Environ, 185, 271-283, https://doi.org/10.1016/j.rse.2015.11.032, 2016.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosillo, T., Hipple, J. D., Hostert, P., Hughes, M. J., Huntington, J., Johnson, D. M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T. A., Schaaf, C., Schott, J. R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J. C., Wynne, R. H., and Zhu, Z.: Current status of Landsat program, science, and applications, Remote Sens Environ, 225, 127-147, https://doi.org/10.1016/j.rse.2019.02.015, 2019.