Identifying Outliers of the MODIS Leaf Area Index Data by Including Temporal Patterns in Post-Processing

Baibing Ma [1,2,3] and Ming Xu [1,2,3,*]

[1] Synthesis Research Centre of Chinese Ecosystem Research Network, Key Laboratory of Ecosystem Network Observation and Modelling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; mabb.18b@igsnrr.ac.cn
[2] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100190, China
[3] Jiangmen Laboratory of Carbon Science and Technology, Hong Kong University of Science and Technology, Jiangmen 529199, China
* Correspondence: xuming@hkustgz-jcl.ac.cn; Tel.: +86-13552593273

Information S1: MCD12Q2 product information about sample points and ground measurements LAI stations

The Terra and Aqua combined MODIS Land Cover Dynamics Version 6.1 (MCD12Q2v6.1) data product provides yearly global land surface phenology metrics. The MCD12Q2v6.1 data product is derived from the time series of the 2-band enhanced vegetation index (EVI2) calculated from MODIS nadir bidirectional reflectance distribution function (BRDF) adjusted reflectance (NBAR). Vegetation phenology metrics at a 500 m spatial resolution are identified for up to two detected growing cycles per year. The data represent the two cycles with the largest NBAR-EVI2 amplitudes for pixels with more than two valid vegetation cycles. Figure S1a – c represent the DOY of greenup (when EVI2 first crossed 15% of the segment EVI2 amplitude) for single-growth-cycle vegetation, the first and the second cycles of double-growth-cycle vegetation, respectively. The DOY was calculated from 1 January 2003 as the actual date. A number with a DOY less than 0 indicates the days before 1 January 2003, whereas a number with a DOY of more than 365 indicates the days after 31 December 2003. In the same way, d, e, and f represent the DOY of the peak (when EVI2 reached the segment maximum) for single-growth-cycle vegetation and the first and second cycle of double-growth-cycle vegetation. Moreover, g, h, and i represent the DOY of maturity (when EVI2 first crossed 90% of the segment EVI2 amplitude) for single-growth-cycle vegetation and the first and the second cycles of double-growth-cycle vegetation. In addition, j, k, and l represent the DOY of dormancy (when EVI2 last crossed 15% of the segment EVI2 amplitude) for single-growth-cycle vegetation and the first and the second cycles of double-growth-cycle vegetation.

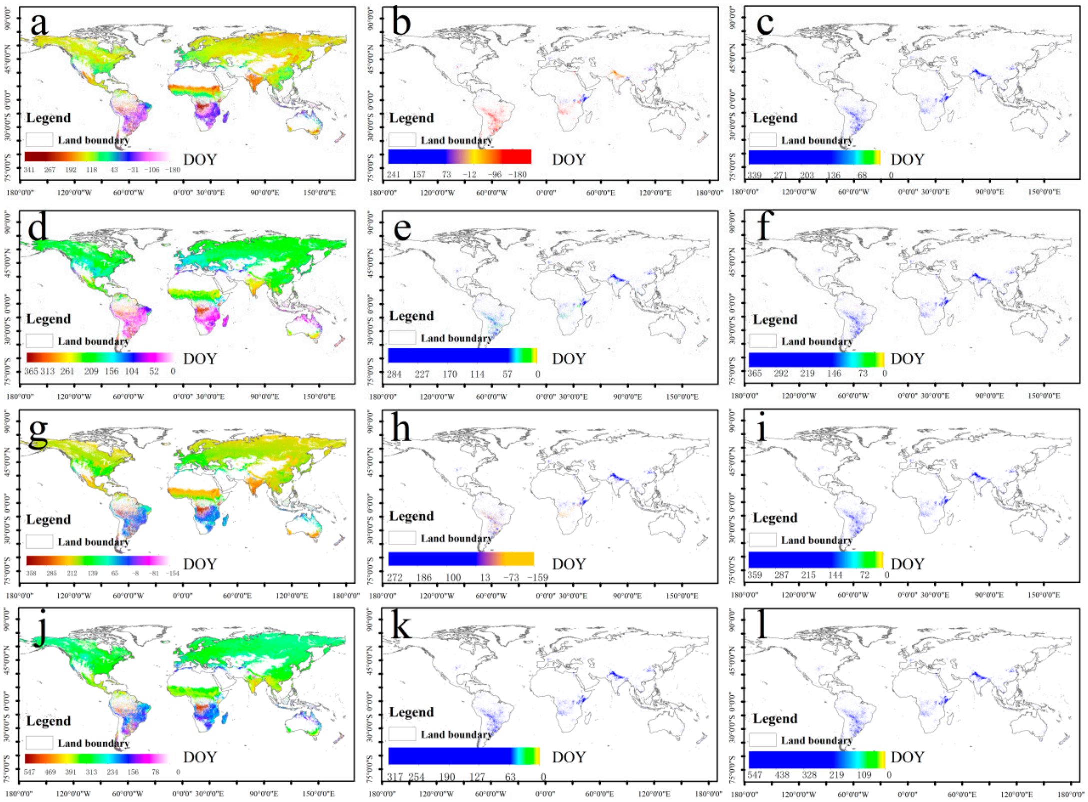

Figure S1. Vegetation phonological information. Panels (a – c) represent the DOY of greenup for single-growth-cycle vegetation and the first and second cycle of double-growth-cycle vegetation, respectively. Panels (d – f) represent the DOY of the peak for single-growth-cycle vegetation and the first and second cycles of double-growth-cycle vegetation. Panels (g – i) represent the DOY of maturity for single-growth-cycle vegetation and the first and second cycles of double-growth-cycle vegetation. Panels (j – l) represent the DOY of dormancy for single-growth-cycle vegetation and the first and second cycles of double-growth-cycle vegetation. The figure shows the phenological information for the year 2003 .

Table S1 shows the basic information of single growth-cycle sample points. The value of MODIS land use cover change (LUCC) means the land use cover, referring to Table S4.

Table S2 indicates the basic information of double growth-cycle sample points. The value of MODIS LUCC means the land use cover, referring to Table S4.

Table S3 suggests the phonological information of ground measurements LAI stations.

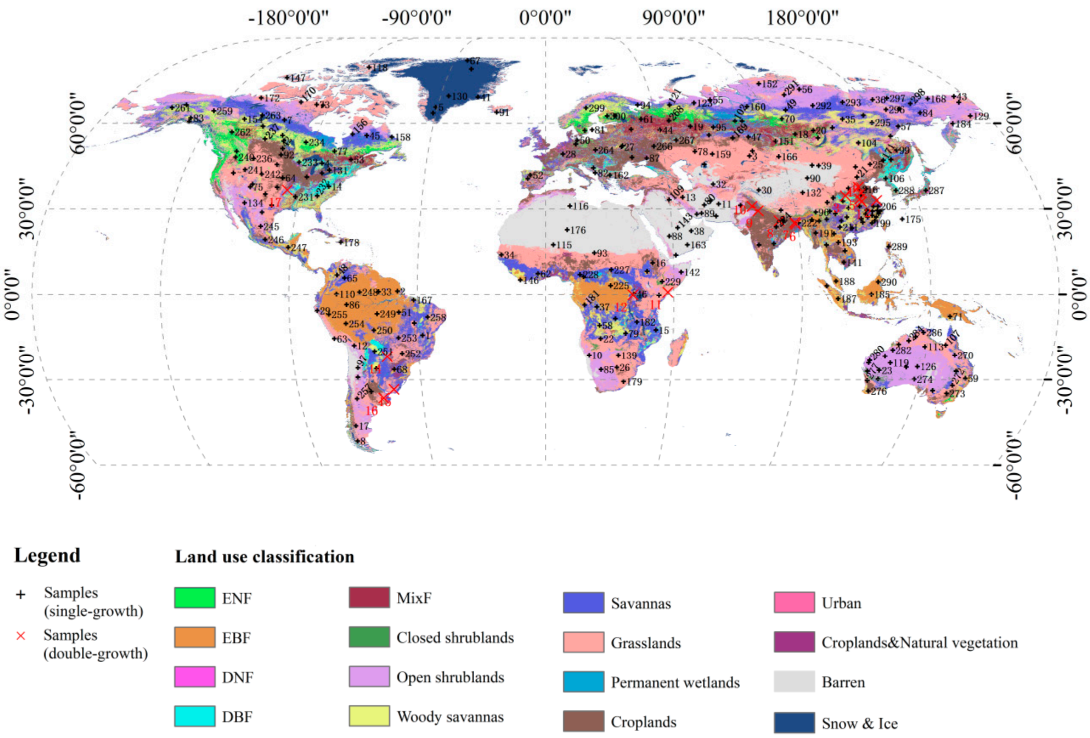

Figure S2. The distribution of random samplings of MODIS LAI (the black cross symbol and adjacent black number represent the random samples of MODIS LAI in a single−growth cycle and the sample number codes of random samples, respectively; red cross symbol and adjacent red number represent the random samples of MODIS LAI in double−growth cycle and the sample number codes of random samples, respectively).

Information S2: the identification results of MODIS LAI noise points in random sample points

According to LAI outlier identifying methods in section 2.2.2, this study selected upper and lower limits interval parameters as x_1=1.0 and x_2=1.0 for multi-years mean, and multi-years detrend denoising processes. The number of outliers judged by multi-years detrend denoising and multi-years mean denoising was 1383 and 1546, respectively, including 1040 outliers identified by either method. In contrast, there were 1889 outliers identified by jointed multi-years detrend denoising and multi-years mean denoising. Outliers identified by multi-years mean and trend denoising effects throughout the year. We selected the denoising results from main land cover types: grassland, DBF, ENF, shrublands, croplands, EBF, and woody savanna. Specifically, a total of 64 outliers can be determined by the multi-years mean denoising for grassland ecotype sample of No.0 point, including the date of 2003-01-09, 2003-02-22, 2003-06-26, 2003-08-21et al. Besides, 66 outliers can be detected by the multi-years detrend denoising in sample No.0 point, including the dates 2003-01-09, 2003-02-18, 2003-06-26, 2003-07-04 et al. (Figure S3a). For DBF land cover type, the outliers determined by the multi-years mean denoising include the date of 2003-03-14, 2003-04-07, 2003-07-04, 2003-08-17 et al., a total of 126 outliers in sample No.74 point. Besides, the multi-years detrend denoising identify outliers, including the date of 2003-03-14, 2003-06-30, 2003-08-17, 2003-09-10 et al., a total of 112 outliers in sample No.74 point (Figure S3c). From the perspective of denoising in ENF land type, the multi-years mean denoising identifies 122 outliers in sample No.162 point, including the dates 2003-03-14, 2003-04-07, 2003-07-04, 2003-08-17 et al. And the multi-years detrend denoising identifies 104 outliers in sample No.162 point, including the dates 2003-03-14, 2003-06-30, 2003-08-17, 2003-09-10 et al. (Figure S3d). For shrublands cover type, the multi-years mean denoising can identify 86 outliers, including the date of 2003-03-14, 2003-04-07, 2003-07-04, 2003-08-17 et al., and the multi-years detrend denoising can identify 88 outliers in sample No.260 point, include the date of 2003-03-14, 2003-06-30, 2003-08-17, 2003-09-10 et al (Figure S3e). For the land cover type of croplands, the outliers identified by the multi-years mean denoising including the date of 2003-03-14, 2003-04-07, 2003-07-04, 2003-08-17 et al., a total of 112 outliers, and the multi-years detrend denoising can identify 114 outliers including the date of 2003-03-14, 2003-06-30, 2003-08-17, 2003-09-10 et al., in sample No.10 (double-growth) point (Figure S3f). For EBF land cover type, 170 and 150 outliers were identified by multi-years mean and multi-years detrend denoising for samples No.68 (Figure S3b). Moreover, the 282 and 286 outliers were identified by multi-years mean and multi-years detrend denoising for another 55 sample points of grassland types in 2003 yr. Besides, 21 and 28 outliers were identified through the denoising of multi-years mean, and multi-years detrend for another 4 sample points of DBF types in 2003yr, respectively. In addition, the denoising of multi-years mean and multi-years detrend identify 23 and 18 outliers of another 4 sample points of ENF types in 2003, respectively. Moreover, the denoising of multi-years mean and multi-years detrend identified 191 and 211 outliers in 2003yr of another 36 sample points of shrubland types, respectively. Additionally, 91 and 64 outliers were identified by the multi-years mean and multi-years detrend denoising of another 16 croplands samples in 2003yr, respectively. Furthermore, the denoising of multi-years mean and multi-years detrend can identify 99 and 100 outliers of another 22 EBF type sample points in 2003yr. Moreover, the multi-years mean and multi-years detrend denoising can identify 272 and 218 outliers for 36 woody savanna type sample points in 2003yr (see supplementary material Figure S2).

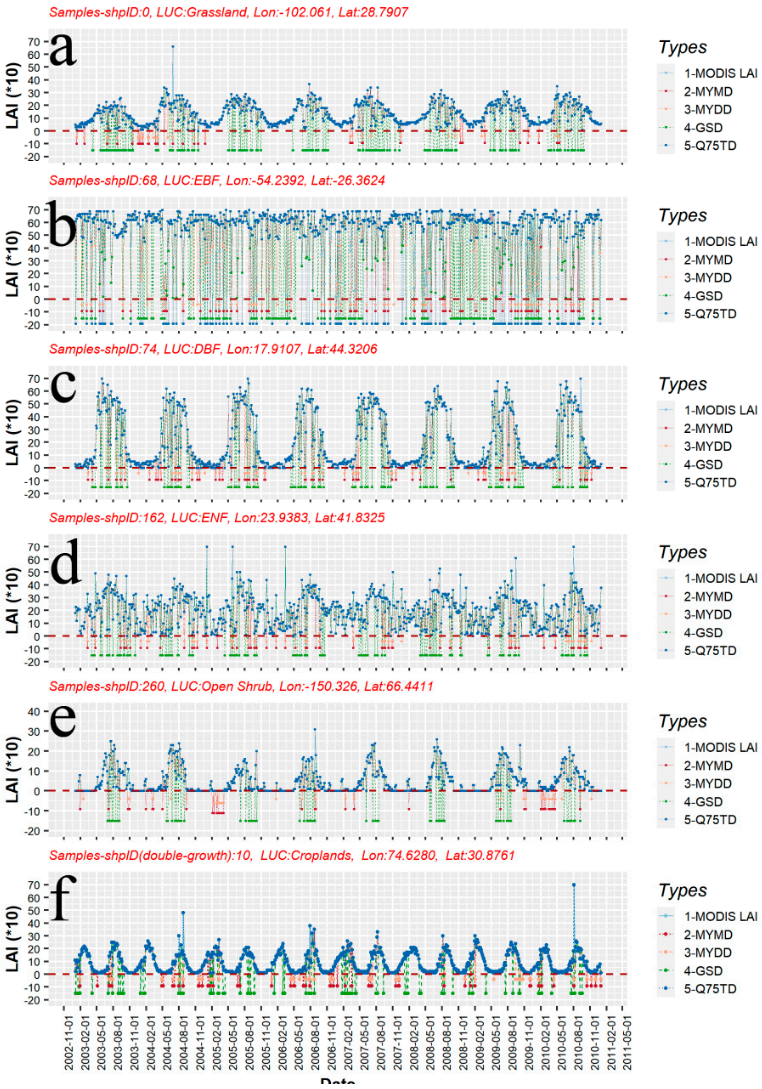

Figure S3. Identification of outliers of sample points (from the years 2003 to 2010). The outliers identified by multi-year detrend, multi-year mean, season denoise, and the threshold were marked with −5, −10, −15, and −20 (corresponding to the light blue, aurantiacus, dark red, green, and blue solid dots in the figures), respectively. Panels (a − e) represent the effect of denoise for a single growth cycle point. In addition, (f) stands for the effect of denoise for points of double growing cycles.

According to LAI outlier identifying methods in section 2.2.2, we selected x_1=1.5 and x_2=0.3 as upper and lower limit interval parameters for the season denoising process of a single GS. Besides, we selected x_1=1.5 and x_2=0.2 as the upper and lower limit parameters for the double growth cycle denoising process. In the condition of the optimal interval, we obtain the denoising results of all samples. The result shows a total of 3297 outliers were identified by season denoising, among which 275 outliers could also be determined by multi-years mean and multi-years detrend denoising in 2003yr. For grassland, 346 outliers can be detected by season denoising in sample No.0 point from 2003 to 2010yr, including the date of 2003-04-07, 2003-04-11, 2003-05-21 et al (Figure S3a). In addition, the outliers that can be determined by season denoising include the date of 2003-04-07, 2003-04-11, 2003-04-15 et al, a total of 272 outliers in sample No.74 point from 2003 to 2010yr (Figure S3c). Besides, the outliers that can be determined by season denoising include the date of 2003-04-07, 2003-04-11, 2003-04-15 et al, a total of 254 outliers in sample No.162 points from 2003 to 2010yr (Figure S3d). Besides, the outliers that can be determined by season denoising have the date of 2003-04-07, 2003-04-11, 2003-04-15 et al, a total of 140 outliers in sample No.260 point from 2003 to 2010yr (Figure S3e). In addition, the outliers that can be determined by season denoising include the date of 2003-04-07, 2003-04-11, 2003-04-15 et al, a total of 168 outliers in the sample No.10 (double-growth) from 2003 to 2010yr (Figure S3f). Correspondingly, the number of 151 outliers was identified by season denoising dynamic information for samples No.68 (Figure S3b).

The supplementary material provides the results of season denoising of another 55 sample points of grassland types. A total of 663 outliers in grassland type were identified in 2003yr, respectively. Besides, a total of 58 outliers were identified by season denoising of another 4 sample points of DBF types in 2003 yr. In addition, the results of season denoising of another 4 sample points of ENF types, which identify 55 outliers in 2003yr. Additionally, 452 outliers were identified by season denoising of another 36 sample points of shrubland types in 2003yr. Moreover, a total of 198 outliers were identified by season denoising of another 16 cropland samples in 2003yr. Furthermore, the results of season denoising of another 22 EBF type sample points can identify 299 outliers in 2003yr. What's more, the results of season denoising identify 510 outliers for 36 woody savanna type sample points in 2003yr.

Although the multi-years mean, multi-years detrend and season denoising performed identify LAI outliers for most land cover types, they need to be more comprehensive for some tropical EBF and woody savanna types. In this study, the Q_{75} threshold of annual LAI was set each three months to diagnose the seasonal variation of LAI. The results showed that there were 22 sample points with LAI values in each season greater than third quantiles values of the year, including 19 samples of EBFs and 3 samples of woody savanna. Such as, the total number of 168 outliers was identified by setting a P_{75} threshold for samples No.68 from 2003 to 2010yr, with 12 outliers of those in 2003yr (Figure S3b). In addition, the supplementary material provides the results of setting a Q_{75} threshold denoising that can identify 670 outliers of another 18 sample points of EBF types and 138 outliers of 3 sample points of woody savanna. A total of 820 outliers were identified by including seasonal in the variation of LAI in 2003yr.

There are 2003 outliers (1889 for single-growth cycle sample points and 114 for double-growth cycle sample points) in 2003yr were identified by including interannual dynamics. Besides, there are 4078 outliers (3880 for single-growth cycle sample points and 198 for double-growth cycle sample points) in 2003yr were identified by including seasonal dynamics. Among them, 531 outliers were repeatedly identified by including both interannual and seasonal dynamics. So, a total of 5550 outliers was identified by including both interannual and seasonal dynamics in 2003yr, accounting for 21.70% of the total samples. In land types of grassland, DBF, ENF, shrubland, cropland, EBF, savanna, woody savanna, and other land types can identify 930, 79, 74, 643, 269, 902, 683, 879, and 1091 outliers, respectively.

Due to the many sample sites, we use the compressed file format as supplementary material. The results in the **Figure S2** attachment supplement the other samples' denoising effect of sample points (from 2003yr to 2010yr) of Figure S3. Note the outliers identified by the style of multi-mean, detrend, season denoise, and setting threshold were marked with -5, -10, -15, and -20 (corresponding to the lightblue, dark red, aurantiacus, green, and blue solid dots in the figures), respectively. The figures of

Num1_No.*_2003_2010 represent the denoise effect for the single-growth cycles per year point, corresponding to the black cross identifier and samples No. number in Figure S2. For example, No.2 means sample No. number 2 denoise result in Figure S2 for samples. In addition, the figures of Num2_No.*_2003_2010 represent the denoise effect for the double-growth cycles per year point, corresponding to the red cross and No. in Figure S2 of the manuscript. In addtion, Num2_No.0_2003_2010 means the denoise result of sample No. 0 in Figure S2 of the manuscript for double-growth samples.

Information S3: verification results of other Ground LAI observation sites

The results in the **Figure S3** attachment supplement the results from the other sites of Figure 5 in the manuscript. Note, Time-series plots of the MODIS LAI (lightblue), detrend denoise (dark green), multimean denoise (aurantiacus), season denoise (blue), Q_{75} threshold (green), and GM LAI (dark red). (The No. number corresponds to the No. field in Table 1 of the manuscript). The outliers identified by the style of multi-mean, detrend, season denoise, and setting threshold were marked with -5, -10, -15, and -20.