**Global evaluation of gap-filling approaches for seasonal NDVI with
considering vegetation growth trajectory, protection of key point,
noise resistance and curve stability**

Ronggao Liu^{1*} Rong Shang^{1,2} Yang Liu¹ Xiaoliang Lu³

1. State key Laboratory of Resources and Environmental Information system, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing100101, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts, 02543, USA.

Corresponding Author

Ronggao Liu
No11A, Datun Road, Beijing, 100101, China
Email: liurg@igsnrr.ac.cn

To be re-submitted

Remote Sensing of Environment
November 18, 2016

Abstract

A variety of approaches are available to fill the gaps in the time series of vegetation parameters estimated from satellite observations. In this paper, a scheme considering vegetation growth trajectory, protection of key point, noise resistance and curve stability was proposed to evaluate the gap-filling approaches. Six approaches for gap filling were globally evaluated pixel-by-pixel based on a reference NDVI generated from MODIS observations during the past 15 years. The evaluated approaches include the Fourier-based approach (Fourier), the double logistic model (DL), the iterative interpolation for data reconstruction (IDR), the Whittaker smoother (Whit), the Savitzky-Golay filter (SG) and the locally adjusted cubic spline capping approach (LACC). Considering the five aspects, the ranks of the overall performance are $LACC > Fourier > IDR > DL > SG > Whit$. The six approaches are similar in filling the gaps and remaining the curve stability but there are large difference in protection of key points and noise resistance. The SG is sensitive to noises and the Whit is poor in protection of key points. In the monsoon regions of India, all evaluated approaches don't work well. This paper provides some new views for evaluating the gap filling approaches that will be helpful in selecting the optimal approach to reconstruct the time series of parameters for data applications.

Keywords: MODIS, NDVI time series, gap filling, seasonal patterns, vegetation phenology

1. Introduction

Time series of vegetation parameters derived from remote sensing imagery, such as LAI and NDVI, have been widely applied to investigate the land surface change (Alcantara et al., 2012; Wessels et al., 2012), vegetation photosynthetic activity (Chen et al., 2014; Kawabata et al., 2001) and vegetation phenology (Atkinson et al., 2012; Walker et al., 2012). However, the continuity of these derived parameters is usually interrupted by residual atmospheric contamination, clouds, snow and system failure (Atkinson et al., 2012; Chen et al., 2006), which greatly hinder the monitoring of terrestrial ecosystems. The gaps of vegetation parameters should be filled before data applications.

The approaches for gap filling make use of the valid observations to estimate the invalid observations. They assume that the seasonal trajectory of vegetation growth can be fitted by a mathematical model. The model parameters were first determined from the valid observations, and then the invalid observations were inferred from the model. A variety of approaches for gap filling have been proposed. They can be categorized into four types: (1) Threshold-based approaches, which set a threshold to control the smoothness of the reconstructed time series, such as the best index slope extraction algorithm (BISE) (Viovy et al., 1992), the modified BISE (Lovell and Graetz, 2001), the mean-value iteration (MVI) (Ma and Veroustraete, 2006) and the iterative interpolation for data reconstruction (IDR) (Julien and Sobrino, 2010). (2) Filter-based approaches, which define a filter to fill the gaps with the valid observations in a moving window, such as the ARMD3-ARMA5 filter

(Filipova-Racheva and Hall-Beyer, 2000), the 4253H twice filter (Velleman, 1980), the Savitzky-Golay (SG) filter (Chen et al., 2004), the changing-weight filter (Zhu et al., 2012) and the adapted local regression filter (Moreno et al., 2014). (3) Curve-fitting approaches, which fit the seasonal trajectory of vegetation growth with a mathematical curve, such as the double logistic function (DL) (Beck et al., 2006), the asymmetric Gaussian function (AG) (Jonsson and Eklundh, 2002), the Fourier-based approach (hereafter referred to as Fourier) (Dash et al., 2010; Zhou et al., 2015) and the locally adjusted cubic spline capping approach (LACC) (Chen et al., 2006). (4) Other approaches, such as the wavelet transform (Lu et al., 2007), the data assimilation (Gu et al., 2009), the CACAO approach (Verger et al., 2013) and the Whittaker smoother (hereafter referred to as Whit) (Atzberger and Eilers, 2011).

Many studies have been conducted to evaluate the gap filling approaches. Hird and McDermid (2009) evaluated the noise-reducing ability of six approaches in a local area of west-central Alberta, demonstrating that the DL and AG were better than the MVI and SG for reducing the negatively-biased noise. Julien and Sobrino (2010) evaluated three approaches for estimating the upper envelope of the time series by randomly selecting 17 points in the globe scale for each IGBP land cover class. They showed that the DL was unable to identify the upper envelope for some vegetation types, such as tall grass, evergreen shrubs, deciduous shrubs, bogs and marshes. Atkinson (2012) compared four models for smoothing remotely-sensed images at sample sites in India, which showed the Whit was better than the AG and DL for the real image but the DL was better than the Whit and AG in resisting simulated noise.

Kandasamy (2013) compared eight approaches for gap filling at sites broadly distributed on the globe, with emphasis on the different fractions of invalid observations and the gap lengths. When the gap fraction was more than 20%, the AG failed to work and the reconstructed accuracy for the Whit and SG degraded rapidly. Michishita (2014) evaluated seven approaches for noise reduction on 12 land cover cluster types in the Poyang Lake area of China, demonstrating the SG was the best approach for noise reduction.

Despite so much work, current evaluations are still limited. First, these evaluations were made only on a small region or at several sampling sites, which were far from the real global evaluations. The effect of an approach for gap filling relies on the seasonal trajectory of vegetation growth and the gap conditions (e.g. the amount and period of invalid observations). The seasonal trajectories of vegetation growth are usually different even in the same regions or for the same land cover type due to the complex vegetation components mixture. The amounts and periods of invalid observations vary from region to region depending on the climate conditions, which influence the reconstruction of invalid observations. For example, the invalid observations during the dormant stage of vegetation should be filled more reliably than those during the rapid growth stage. The global pixel-by-pixel evaluation, which includes all the possible seasonal trajectories and gap conditions, should be more comprehensive and convincing than the regional or the site-based evaluations. Second, previous evaluations were mainly focused on evaluating the abilities of curve smoothing and noise reducing. However, the requirements of application for the gap

filling are various. For those aiming to extract the vegetation phenology, the recovery/ protection of the key status of vegetation growth is the priority. For those estimating the vegetation productivity in the whole year, the smoothing to reduce noise is important. For those comparing the inter-annual change, the stability is a prerequisite. Therefore, the goals of data applications should be considered for evaluations of the gap filling approaches.

As MODIS data have been globally accumulated over many years, it is possible to generate a reference dataset to represent the basic seasonal patterns of vegetation growth for each pixel. In this paper, a reference NDVI was generated from the historical archived MODIS data during the past 15 years and taken as a baseline to evaluate the six approaches for gap filling. The six approaches include the IDR, SG, DL, Fourier, LACC and Whit. The evaluations include the filling abilities for the simulated gap and for the gaps from real observations, the recovery of the maximum curvature point, the resistance to noise on the valid observations and the stability of the reconstructed curve.

2. Data and Methods

2.1 Data

The six approaches for gap filling were evaluated using the MODIS land surface reflectance products (MOD09A1 and MYD09A1). The MOD09A1 and MYD09A1 provide 8-day composited surface reflectance at a 500-m resolution, which were generated from the Terra and Aqua MODIS observations by correcting the atmospheric effects (Vermote and Kotchenova, 2008). The NDVI was calculated from the NIR and red band of the MODIS 8-day reflectance products. The cloud and snow

were detected using a refined cloud mask method (Liu & Liu, 2013), which could provide a better separation of cloudy observations from clear-sky observations than the MODIS cloud mask does (Liu & Liu, 2013). The valid NDVI derived from the MOD09A1 and MYD09A1 during 2000 to 2014 were averaged to generate the reference NDVI. There are a total of 46 composites in each year.

2.2 Generation of the reference NDVI

The reference NDVI was taken as the benchmark to evaluate the approaches for gap filling. The reference NDVI was generated as follows: (1) the cloudy or cloud-contaminated observations in the MODIS surface reflectance were screened by a refined cloud mask (Liu and Liu, 2013); (2) the NDVIs for the valid observations were calculated with the NIR and red band reflectance; and (3) all valid NDVIs in the same day of year (DOY) during 2000-2014 were averaged to obtain the reference NDVI for the DOY. If the number of valid NDVIs in the same DOY is less than 4, the reference NDVI was set as invalid. These invalid data were labelled in their quality control (QC) values and would be temporarily filled by linear interpolation of the nearest valid values. Through these procedures, most noises, angular effects and inter-annual fluctuations in the original observations could be diminished. Fig. 1 is an example showing the generation of the reference NDVI for the time series of observations. The reference NDVI is regarded as the truth data for the evaluations, and their characteristics are described in chapter 3.2.

Fig. 1 Example of the reference NDVI.

2.3 Six approaches for evaluations

The evaluated approaches include the IDR, SG, Fourier, DL, LACC and Whit, which were selected to represent the four types of gap filling approaches. The IDR fills the gaps from neighboring values, which represents the threshold-based approach. The SG is widely used for noise reduction, gap filling and data smoothing, which represents the filter-based approach. The Fourier, DL and LACC were selected to represent the curve-fitting approach. The Fourier was selected for its few model inputs. The DL is widely used for the extraction of vegetation phenology because of its good simulation of the seasonal trajectories of vegetation growth with six parameters (Beck et al., 2006). The LACC was chosen for its flexibility to simulate any seasonal vegetation growth patterns and the ability to protect the maximum curvature (Chen et al., 2006). For other approaches, the Whit was selected for its balanced mechanism between the fidelity and the roughness of the time series (Atzberger and Eilers, 2011).

2.3.1 The Fourier-based approach

This approach uses the sum of a series of cosine and sine waves to fit the vegetation growth curve (Jakubauskas et al., 2001). The vegetation growth curve could be simulated by individual sinusoids and their frequencies for which noise and gaps could be reconstructed (Dash et al., 2010). The number of harmonics is required to reconstruct the time series parameters (Zhou et al., 2015). The values of 2, 4 and 6 for the input harmonics were tested for the global evaluations. The harmonics of 4 was shown the best and set as the default input parameter, though the Fourier with high harmonics would show better performance for the special regions with complicated seasonal trajectories (Zhou et al., 2015).

2.3.2 The Whittaker smoother

This approach can be regarded as the penalized least squares, which puts a penalty of the fidelity on the roughness of the smooth curve (Atzberger and Eilers, 2011). The fidelity was expressed as the quadratic difference between estimates and actual observations. The roughness was defined as the quadratic difference between successive estimates. The cost function balances the fidelity and the roughness in which a smoothing parameter regulates these two parts. The larger the parameter, the smoother would be the reconstructed time series. Based on the work of Atkinson (2012) for running the Whittaker smoother with two smoothing parameter (2 and 15), the default smoothing parameter was set to 2 because the parameter of 15 would over-smooth the maximum curvature points.

2.3.3 The IDR approach

The IDR approach was developed to reconstruct high-quality time-series NDVI with an iterative mechanism. The threshold controls the times of iteration. An

operator is defined as the absolute difference between the given NDVI and the average value of pre- and post-NDVI. If the operator is larger than the threshold value, the given NDVI would be replaced by the average NDVI. The smaller the threshold value, the smoother the reconstructed NDVI time series. The original threshold was set to 0.02. However, it tends to overestimate the low NDVI values (Julien and Sobrino, 2010) and the maximum curvature point. The threshold of 0.02 and 0.1 were tested for the global evaluations. The threshold of 0.1 shows better performance than that of 0.02 so the default threshold was set as 0.1.

2.3.4 The double logistic function

This approach can fit the unimodal shape of the vegetation growth curve (Beck et al., 2006). It originally simulates the time-series data as a function of six parameters: the minimum NDVI, the maximum NDVI, the curve rise point or drop point and their rate of increase or decrease. The DL performs well for fitting the symmetric seasonal trajectory. For the asymmetric seasonal trajectories, the DL was unsuitable and should be divided into two separate parts: the curve rising part and the curve decreasing part. Each part was fitted with one logistic function. The four parameter logistic model was demonstrated to be the best in the parameterizing test of the logistic model from two parameters to five parameters (White et al., 2014), which was selected in this paper.

2.3.5 The LACC approach

This approach was developed to produce flexible and mathematically smooth cubic-spline capping curves to fit rapid seasonal changes (Chen et al., 2006). The time series data were fitted with iterative procedures. The first curve fitting was conducted with a fixed global smoothing parameter of 0.5. For all data points in the time series, a

local curve smoothness matrix (Chen et al., 2006) was used to automatically scale the local smoothing parameter based on the curvature of the fitted curve. Optimal weights of all points for the final curve fitting were regulated by the matrix. Three iterations was set for this approach as suggested in the original paper (Chen et al., 2006). Based on the local curve smoothness matrix, the first fitting determines the weights of the second fitting and the second curve fitting regulates the weights. The final fitting reconstructs the smoothed time series data by applying the regulated weights. This approach shows the flexible abilities for fitting any seasonal patterns and variable curvatures of a surface parameter in all seasons (Chen et al., 2006).

2.3.6 The Savitzky-Golay filter

This approach could be interpreted as a weighted moving average window, and the weighting was given as a polynomial of window width (Chen et al., 2004). It was designed to fit the upper envelope and to describe the changing patterns of the parameter through an iteration process (Chen et al., 2004). The coefficients of the smoothing window (Madden, 1978) were calculated by the half-width of the smoothing window and the degree of the smoothing polynomial. The larger the half-width of the smoothing window, the smoother the reconstructed NDVI time series. A linear interpolation preprocess was used for the gaps in this approach (Chen et al., 2004). The threshold of 4 for smoothing polynomial and 6 for the half-width of the smoothing window was suggested in the original paper (Chen et al., 2004). But the threshold of 5 for the half-width of the smoothing window shows better performance

in the global scale. Therefore, the smoothing polynomial was set as 4 and the half-width of the smoothing window was set as 5 in this study.

2.4 Evaluation method

The approaches for gap filling were globally evaluated pixel-by-pixel. Fig. 2 shows the flowchart of evaluation. The evaluations include the filling effect for the simulated gaps (Simgaps evaluation) and for the gaps from real observations (Realgaps evaluation), the recovery of the maximum curvature point (Keypoint evaluation), the resistance to noise on valid observations (Noise evaluation) and the stability of the reconstructed curve (Stability evaluation). The evaluations were performed based on the reference NDVI, except for the stability evaluation. First, the test series were generated by setting gaps or adding noises into the reference NDVI according to the evaluation purpose. Then, the gaps in the test series were filled by the evaluated approaches to generate the gap-filled series. Finally, the gap-filled series were compared with the original reference NDVI series. The mean bias between the reference NDVI and the gap-filled series was taken as the index to evaluate the performance of the approaches. The smaller the bias, the better the approach performed. The biases were only calculated between the reconstructed values ($NDVI_{reconstructed}$) and the counterpart values in the reference NDVI ($NDVI_{reference}$):

$$Bias = \frac{\sum_{i=1}^n |NDVI_{reference}(i) - NDVI_{reconstructed}(i)|}{n}$$

Fig. 2 Data process flowchart of the global evaluations.

The first step was to generate the test series for each evaluation. The test series for Simgaps evaluation were generated by equidistantly setting the first point in each three points as a gap in the reference NDVI series (for example, the values in the positions of 1, 4, 7, 10, 13 and so on were set as invalid). For the Realgaps evaluation, the positions of invalid observations in 2008 were first detected, and then the test series for Realgaps evaluation were generated by setting the counterpart values in the reference NDVI as the gaps. That is, if the observation in the DOY 193 of 2008 was invalid, the reference NDVI in this position was set as invalid. For the Keypoint evaluation, the maximum curvature point in the reference NDVI series was first detected, and then this point was set as invalid to generate the test series. For the Noise evaluation, random noises were added to the 10% randomly selected position of

the reference NDVI. The noises were negative and random, and the maximum magnitude was constrained to 40% of the selected values. For the Stability evaluation, the MOD series and the MYD series were generated from the Terra and Aqua MODIS observations in 2008, respectively. The gaps in these two series were filled by the evaluated approaches. The absolute biases between the MOD gap-filled series and the MYD gap-filled series were taken as the indicator of stability. Theoretically, the series from Terra and Aqua should be similar if the BRDF effect is neglected. A lower difference between two series indicates a higher stability of the gap-filling approach.

Some pixels were excluded from the evaluations. The noises in the low NDVI series would have large error on the evaluations. The reconstruction would be unreliable if the gap percentage was larger than 60% (Kandasamy et al., 2013). Therefore, those pixels with the valid reference NDVI less than 20 or the maximum NDVI less than 0.4 were excluded. For the Keypoint evaluation, those pixels with the maximum NDVI less than 0.5 and those with the difference between the maximum and minimum NDVI less than 0.3 were excluded to guarantee the obvious seasonality and the existence of the maximum curvature point.

3. Results

3.1 Global distributions of the invalid observations

The invalid observations include the cloud- and snow-covered observations and the missing observations by system failure. The cloud and snow cover are spatially and temporally heterogeneous, and are constrained by certain geographic rules. The invalid percentage for each pixel was calculated by the number of invalid

observations divided by the total possible observations from 2000 to 2014. The global invalid percentage is shown in Fig. 3a. There are 69.84% of pixels (23,308,435 km²) with an invalid percentage < 15% and 1.26% of pixels (420,689 km²) with an invalid percentage > 40%. The invalid percentage is typically less than 15% in the middle latitudes of the Northern Hemisphere (30°N-50°N), Mexico and the middle and low latitudes in the Southern Hemisphere (10°S-50°S). In the northern high latitudes (above 50°N), the invalid percentage is high due to the snow cover and the high solar zenith in the high latitude region in the winter. In the tropical zone, the invalid percentage is above 20% due to the frequent cloud cover. The invalid percentage is above 40% around the Gulf of Guinea.

Fig. 3 (a) Global map of the gap percentage for the 15-year observations in the MOD09A1 and (b) Global map of the merged Koppen climate classification.

The temporal distribution of invalid observations was demonstrated to be varied largely in different zones using several temporal profiles of the NDVI series for different climate zones (Fig. 4). The periods of invalid observations are similar in the same climate zone. For the subarctic climate, the invalid observations are mainly concentrated in the dormancy period with continuous snow cover in winter (Fig. 4a).

For the tundra climate in the Tibetan Plateau, the invalid observations are attributed to the snow cover in the vegetation dormant stage (Fig. 4b). For the tropical monsoon climate in India, the invalid observations are concentrated in the rapid growth stage of vegetation during the wet season (Fig. 4c). For the humid subtropical climate, the high rainfalls in the wet season make the valid observations usually unavailable (Fig. 4d). For the tropical rainforest climate around the Gulf of Guinea (Fig. 4e), the clouds almost cover the whole year, so that the valid observations are quite sparse. For the tropical savanna climate in the north of Australia (Fig. 4f), the invalid observations are mainly concentrated during the wet season.

Fig. 4 Temporal profiles of the MODIS NDVI with cloudy or snowy observations: (a) the subarctic climate (54.66°N , 155.73°E); (b) the tundra climate in the Tibetan Plateau (30.38°N , 92.76°E); (c) the tropical monsoon climate in India (22.16°N , 75.14°E); (d) the humid subtropical climate (24.97°N , 102.18°E); (e) the tropical rainforest climate around the Gulf of Guinea (6.37°N , 5.12°E); (f) the tropical savanna climate in the north of Australia (18.15°S , 140.13°E). The positions were showed in figure 1a.

3.2 The characteristics of the reference NDVI

The integrity of the reference NDVI was described by the invalid percentage and the maximum length of gaps, and its continuity was described by the roughness. The global invalid percentage of the reference NDVI was shown in Fig. 5a. The invalid distribution of the reference NDVI shows a significant latitudinal variation. When compared with the invalid percentage of real observations (Fig. 3a), the invalid percentages of the reference NDVI are very low ($\leq 5\%$) at the middle and lower latitudes of the Northern Hemisphere and the whole Southern Hemisphere, especially for the tropical monsoon forest of India, the southeast of Asia and the vegetated areas around the Gulf of Guinea. The invalid percentage is above 30% in the high latitudes of the Northern Hemisphere due to the snow cover in the whole winter (Fig. 4a and 4c). In the winter, the vegetation greenness remains nearly unchanged in this zone, so that the invalid reference NDVI could be filled by a linear interpolation using the neighboring valid observations. The maximum gap length in the reference NDVI series is shown in Fig. 5b. Its distribution is similar to the invalid percentage map. The maximum gap length in most of the regions is less than 32 days. It is usually above 90 days in the high latitudes due to the continuous snow cover in the winter. The temporal roughness of the reference NDVI was expressed as the quadratic difference between successive valid NDVI values (Fig. 5c). The roughness in most of the regions (94.52% of pixels, 31,546,193 km²) is less than 0.05, suggesting that the reference NDVI shows smoothing seasonal cycles for the majority of the areas. The roughness is slightly large (0.06 to 0.08) in the agricultural zones in the North China Plain, where double crops would decrease the smoothness of the NDVI seasonal cycles. The

roughness is above 0.07 in the regions around the Gulf of Guinea, where there are not enough valid observations to build the reference NDVI.

Fig. 5 Global maps of the characteristics of the reference NDVI: (a) the invalid percentage; (b) the maximum gap length; (c) the roughness.

3.3 Site examples for gap filling

The effectiveness of an approach for gap filling is highly determined by the agreement between the fitting model and the seasonal trajectory of vegetation growth. Six sites, including the boreal forest (43.50°N, 117.45°E), tropical savannas (8.90°N, 68.15°W), forest in the Southern Hemisphere (4.29°S, 11.66°E), grassland (49.33°N, 120.05°E), mixture of forest and grassland (55.49°N, 11.65°E) and the double rice or savanna (26.73°N, 115.07°E), were taken as examples to demonstrate the effect of gap filling for different vegetation conditions (Fig. 6). The NDVI of deciduous vegetation is the minimum in the dormancy stage, increases when the vegetation

began to grow and reaches the maximum in the mature stage. Then, the NDVI decreases while the vegetation withered. Finally, the vegetation remains dormant until the start of the next growth cycle. These four processes constitute a positive unimodal shape for the boreal forest (Fig. 6a) and for the grass (Fig. 6d) in the Northern Hemisphere and a negative unimodal shape in the Southern Hemisphere (Fig. 6c). The vegetation greenness of the tropical monsoon forest fluctuates during the wet and dry seasons. The vegetation greenness gradually decreases when the dry season came, but it recovers rapidly when the rainy season came (Fig. 6b). For the mixed vegetation of forest and grass, the growth and senescence of forest is faster than that of grass. The greenness of the forest reaches a maximum and remains nearly stable during its maturity stage but that of grass always changes. When the grass began to wither, the forest is still in its mature stage that leads to another stable NDVI for one or two months. The trend of NDVI in the mixed area is mainly decided by the forest growth conditions. This constitutes a complex seasonal shape (Fig. 6e). For the double rice or savanna, the double seasonal pattern constitutes the bimodal shape (Fig. 6f).

Fig. 6 Site examples of typical shapes for the gap filling: (a) the boreal forest; (b) the tropical savanna; (c) the forest in the Southern Hemisphere; (d) the grassland; (e) the mixed area of forest and grassland; (f) the double rice or savanna.

The six approaches show the different abilities to capture the seasonal cycles of vegetation greenness. The LACC and SG perform the best for all six seasonal patterns and could both fill the gaps and preserve the fidelity of the original time series. However, its performance deteriorates when the seasonal shapes of the vegetation are

complex (Fig. 6a and Fig. 6c), especially over the mixed area of forest and grass (Fig. 6e), the double rice and the savanna vegetation (Fig. 6f) (Beck et al., 2006). The Fourier with four harmonics presents unreasonable spurious oscillations for the temperate vegetation in the dormancy stage and mature stage (Fig. 6a), but it fit the shape of the grassland quite well (Fig. 6d). This could be explained by the symmetric shape of the grassland (Fig. 6d) and the complicated fluctuations appeared in the beginning of dormancy stage of the boreal forest (Fig. 6a). When compared with the DL, the Whit and IDR are more suitable to maintain the integrity of the time series. However, the Whit and IDR are too smoothing around the inflexion points of the boreal forest (Fig. 6a), the peak of the grassland (Fig. 6d) and the valley between the double peaks of double rice or savanna (Fig. 6f).

3.4 Overall performance

To evaluate the approaches quantitatively, the percentage of pixels with bias < 0.03 was calculated for different evaluations (Tab. 1). The ranks for the six approaches are based on the statistics of the pixel percentage. The ranks for the Simgaps evaluation are LACC, SG $>$ IDR $>$ Whit $>$ Fourier $>$ DL. The ranks for the Realgaps evaluation are SG $>$ LACC $>$ IDR $>$ DL $>$ Fourier $>$ Whit. The ranks for the Keypoint evaluation are SG $>$ LACC $>$ IDR $>$ DL $>$ Fourier $>$ Whit. The ranks for the Noise evaluation are LACC $>$ Fourier $>$ IDR $>$ DL $>$ SG $>$ Whit. The ranks for the Stability evaluation are Whit $>$ Fourier $>$ LACC $>$ DL $>$ SG $>$ IDR. The ranks of the overall performance considering all five of the evaluations are LACC $>$ Fourier $>$

IDR > DL > SG > Whit, which were ranked with their average percentage of the five conditions. The performances of all six approaches on the Simgaps and Realgaps evaluations are similar, but the biases for the Realgaps evaluation are larger. The LACC is the best approach for global gap filling when considering all of the evaluations. The IDR performs well in all of the evaluations except for the Stability evaluation and the Noise evaluation. The SG performs well in all of the evaluations except for the Noise evaluation, indicating that the SG is very good for gap filling if the cloud, snow and atmospheric residual contamination could be well detected.

Tab. 1 Global percentage of pixels with bias < 0.03 for the five schemes.

Evaluations	Fourier	DL	IDR	SG	LACC	Whit
Simgaps	91.28%	88.93%	96.84%	96.91%	96.91%	94.06%
Realgaps	79.34%	80.01%	80.10%	80.41%	80.37%	76.59%
Keypoint	52.59%	61.61%	64.72%	70.32%	67.87%	32.14%
Noise	59.80%	30.10%	38.69%	13.55%	75.28%	9.47%
Stability	50.63%	43.64%	30.21%	40.75%	44.61%	54.83%

Regional quantitative evaluations of each approach were also conducted based on the Koppen climate classification in Fig. 3b (Rubel and Kottek, 2010). Fig. 7 describes the pixel percentage with bias < 0.03 for the four largest climate regions: the humid continental climate, the humid subtropical climate, the subarctic climate and the tropical savanna climate. For the Simgaps evaluation, the Realgaps evaluation and the Keypoint evaluation, the regional performances of the LACC, SG and IDR accord

with their global performance. The Fourier performed well for the Simgaps and Realgap evaluations in both the humid subtropical climate regions and the tropical savanna climate regions. The DL showed the best performance in the recovery of the maximum curvature point in both the subarctic climate regions and the humid continental climate regions. For the Noise evaluation, the LACC and Fourier show great resistance to noise despite the climate regions. For the Stability evaluation, all six of the approaches show relatively poor stability in the subarctic climate regions, and the IDR has the poorest stability considering the four climate regions. The Fourier and Whit are relatively stable in both the humid subtropical climate regions and the tropical savanna climate regions.

Fig. 7 Histogram of the pixel percentage with bias < 0.03 for the four largest climate regions (EV1 is for the Simgaps evaluation, EV2 is for the Realgaps evaluation,

EV3 is for the Keypoint evaluation, EV4 is for the Noise evaluation and EV5 is for the Stability evaluation).

The overall performances for each climate zone considering all five of the evaluations were also sorted with their average percentages. The ranks of the overall performances considering all of the evaluations in the humid continental climate region are LACC > DL > Fourier > IDR > SG > Whit. The ranks in the humid subtropical climate region are LACC > Fourier > IDR > DL > Whit > SG. The ranks in the subarctic climate region are LACC > Fourier > IDR > DL > SG > Whit. The ranks in the tropical savanna climate region are LACC > Fourier > DL > IDR > Whit > SG. The LACC performs the best for all four climate regions. The Fourier performs relatively poor in the humid continental climate region (ranks the third), but the performance in other three climate regions ranks the second. The DL has better performance in the humid continental climate region than other three climate regions. The performance of IDR is relatively stable which ranks the third or fourth for the four climate regions.

3.5 The filling effect for simulated gaps

Fig. 8 shows the global distributions of the gap filling biases for the Simgaps evaluation. Overall, the performances of the IDR, SG and LACC are better than the Whit, Fourier and DL (Tab. 1). Most biases of the IDR (96.84%), SG (96.91%) and LACC (96.91%) show quite a small range under 0.03. When compared with IDR, SG and LACC, the performance of the Whit seems generally good, with biases < 0.03 for

94.06% of the vegetated areas, while it is slightly worse in the boreal forest, with gap filling biases of approximately 0.05 due to the over-smoothing around the maximum curvature points (Fig. 6a). The Fourier shows good performance in the low and middle latitudes of Northern Hemisphere and the Southern Hemisphere, with gap filling biases lower than 0.03 for 91.28% of the vegetated areas, but the biases increases in the boreal forest of Russia, which is caused by the spurious oscillations happened in the dormancy stage and mature stage (Fig. 6a). The DL shows a good performance, with bias less than 0.05 for the surface with single seasonal pattern of vegetation growth. However, the biases for some pixels are high (>0.10) in the tropical rainforest and agricultural zones. For the tropical rainforest such as Malaya and the Indonesia Archipelago, Congo Basin and Amazon Basin, it was hard for the DL to fit the small fluctuation greenness of the observations. Similarly, the biases are also greater than 0.10 for some evergreen vegetation in southern Brazil. For the agricultural zones with double or triple cropping seasons, such as the plains of India, Ganges and China, the DL is unsuitable to describe a double or triple annual growing season (Beck et al., 2006).

Fig. 8 Global distributions of the gap filling biases under the Simgaps evaluation.

3.6 The filling effect for the gaps from real observations

The global bias maps for the Realgaps evaluation were presented in Fig. 9. The spatial patterns of biases are similar to that in the Simgaps evaluation, while the biases were somewhat larger due to larger gap sizes in some series. The biases for all of the approaches are somewhat larger in India, and it might be derived from the frequent cloudy cover in the wet season when the vegetation grew rapidly and the seasonal trajectories of the vegetation disagreed with the fitting models. Except for the DL, all other five approaches showed large biases in the east of Europe, which might be caused by the long successive gaps appeared in winter.

Fig. 9 Global distributions of the gap filling biases under the Realgaps evaluation.

The performances of the IDR, SG, LACC and Fourier are similar. Most of the biases are less than 0.04. The Fourier performs poorly in part of the boreal forest due to the spurious oscillations happened in the dormancy stage or mature stage (Fig. 6a). The Whit presents large biases in the agricultural zones of China with bimodal or multimodal cropping patterns due to the over-smoothing around the turning points (Fig. 6f). The DL showed the best performance in the middle and high latitudes of Northern Hemisphere ($> 40^{\circ}\text{N}$), but large biases were appeared in the tropical rainforest due to the vague seasonality (Kandasamy et al., 2013), and in the agricultural zones due to the inapplicability of the double logistic function to describe bimodal and multimodal seasonal patterns (Beck et al., 2006).

3.7 The recovery of the maximum curvature point

The biases in the Keypoint evaluation are larger than those in the Simgaps evaluation for all of the approaches (Fig. 10). Larger biases occur in India and the vegetated areas around the Gulf of Guinea and the southeast of Asia for all of the approaches. In India, the temporal shapes of the vegetation growth vary spatially and are difficult to fit in the gap filling model. The maximum difference between the model and the reference NDVI occur in the maximum curvature position. In areas around the Gulf of Guinea and the southeast of Asia, the large invalid observations from the frequent cloud cover (Fig. 3a) increase the uncertainties to fit the maximum curvature point.

Fig. 10 Global distributions of the gap filling bias under the Keypoint evaluation.

Except for these areas, the SG and LACC show great performance in the recovery of the maximum curvature point. In the agricultural zones of China with bimodal or multimodal seasonal patterns, the biases for the IDR, Fourier and Whit are above 0.08, 0.08 and 0.10, respectively, meaning that these approaches would lead to a large bias in this local area for the extraction of vegetation phenology. The over-smoothing around the turning point led to the large biases for the Whit in the boreal forest (Fig. 6a). The performance is poor for the Fourier in the middle latitudes of the Northern Hemisphere, with biases above 0.08. The DL performed quite well, except for those inapplicable areas, such as areas with double rice or savanna and the mixed area of forest and grass.

3.8 The resistance to noise on valid observations

Global maps of the biases for Noise evaluation are presented in Fig. 11. Generally, the curve-fitting approaches (Fourier and LACC) shows stronger resistance than the others, especially for the SG and Whit. Compared with other evaluated aspects, high noise level (up to 40% of the reference NDVI value) used in the Noise evaluation resulted in much larger global overall biases and lower statistical results with biases lesser than 0.03. Most of the biases for Fourier (59.80%) and LACC (75.28%) are less than 0.03, which is better than those of the SG, Whit IDR and DL (Tab. 1), indicating that these two approaches are more resistant to noise contamination. The iterative mechanism makes the IDR more resistant to the noises

than the SG, Whit and DL. A slight fluctuation in the decrease or increase period would change the shape of logistic function, making the DL less resistant to the noises. All six approaches present poor resistance to the noises in the boreal forest of east Russia and local areas in the central of North America. The LACC exhibits the best resistance to the noises in the Southern Hemisphere compared with other five approaches. The DL, SG and Whit perform relatively poorer to the noises in most part of the Southern Hemisphere. The sensitivity of the SG to noisy time series parameters supports the conclusions drawn by Jonsson and Eklundh (2002) and Hird and McDermid (2009).

Fig. 11 Global distributions of the gap filling biases under the Noise evaluation.

3.9 The stability of the reconstructed curve

The stability of an approach refers to the ability to preserve the shape of the seasonal trajectories of vegetation growth. High stability means that the uncertainties in the valid observations would lead to little change on the vegetation growth curve. The stability of gap filling is a prerequisite for the inter-annual comparison of vegetation status. Fig. 12 shows the global biases distribution of the six approaches for the Stability evaluation. The six approaches show great stability for the middle latitudes of the Northern Hemisphere (30°N-50°N) and the Southern Hemisphere (10°S-50°S), with biases less than 0.04. Large biases (>0.08) are found in the tropical rainforest and the boreal forest. The ranks of the stability for the gap filling approaches are Whit > Fourier > LACC > DL > SG > IDR (Tab. 1). Compared with the other five approaches, the Whit has the best global stability. The Whit, Fourier and LACC are more stable than the DL, IDR and SG in the Amazon Basin and the Congo Basin. The Fourier with two harmonics shows the best overall stability (61.32% of pixels with biases smaller than 0.03) compared with other five approaches or the Fourier with four harmonics. It indicates that the Fourier with higher harmonics had lesser stability.

Fig. 12 Global distributions of the final biases under the Stability evaluation.

4. Discussion

The main contribution of this paper is the global pixel-by-pixel evaluations of the gap-filling approaches from five aspects. The selection of an approach for gap filling depends on the purpose of the data application (Keypoint and Stability evaluations), the seasonal trajectory of the vegetation growth (global pixel-by-pixel evaluations), the amount and period of the invalid observations (Simgaps and Realgaps evaluations) and the quality of valid observations (Noise evaluation). To test various trajectories of vegetation growth, the evaluations should be made pixel-by-pixel in the global scale rather than in some sampling sites. The Simgaps and Realgaps evaluations test the filling abilities for different amount and period of gaps. As the maximum curvature

point of seasonal trajectory is usually taken as an indicator of the vegetation phenology, the protection of the maximum curvature point is regarded as an indicator for gap filling. The residual contamination on valid observations would have the ill-effect on the filling results. The resistance to noise was also evaluated. The stability of an approach was evaluated because it is important for the inter-annual comparison of vegetation status. But the smoothing ability was not tested because the real observations are not always smoothing. Global bias maps with different evaluations describe the ability of each approach for gap filling, which could be a guide to select the optimal approach to reconstruct the time series of vegetation parameters for a certain data application.

The six approaches were evaluated with the same parameters in the global scale without modification for local conditions. The tunable parameters in some evaluated approaches are adaptable to be suitable for the local conditions, such as the harmonics of the Fourier, the smoothing parameter of the Whit, the threshold of the IDR and the moving window width of the SG. It is impossible to set input parameters which are optimal for different regions in the global scale. The parameters were tried to set as optimal as possible for global performance. The Fourier with four harmonics shows better performance than that with two harmonics (Fig. 13a and Fig. 13b). But for higher input harmonics, such as six, except for the agricultural zones with double or triple cropping seasons, the performance became worse than the four harmonics.

Fig. 13 Comparisons between different tunable parameters for the approaches under the Simgaps evaluation.

For the IDR, small threshold makes the reconstructed value approximate to the mean value of its pre- and post-value. The threshold of 0.02 was used in original study (Julien and Sobrino, 2010), but it tended to overestimate low NDVI values. Several candidate thresholds were tested and the 0.1 was found the best (Fig. 13c and Fig. 13d), therefore the default threshold was set as 0.1. For the SG, the threshold of 5 for the half-width of smoothing window showed better performance than the suggested threshold of 6 in the literature (Fig. 13e and Fig. 13f). For the Whit, Atzberger and Eilers (2011) suggested the lambda parameter as 2 or 15. The lambda of 15 would

over-smooth the maximum curvature points, and the lambda of 2 showed obvious better performance than the lambda of 15 for almost all the global scale (Fig. 13g and Fig. 13h). Therefore, the lambda of 2 was set in this study.

There was no special modification for the double logistic function (DL) approach. The DL approach was proposed to extract the vegetation phenology at high latitudes in the first place (Beck et al., 2006) because of its similar shape to the vegetation growth trajectory in this region. Five global evaluations about this approach was conducted to guide its global usage and future modification. The DL performed well for the pure temperate or boreal vegetation with two ideal “S”-shaped vegetation growth trajectories. But it was inapplicable for the complex seasonal trajectories which departed far from the ideal “S” shape, such as the grassland with obvious two-stage greenness increase (Cao et al., 2015) and the mixed vegetation with two-stage senescence (Fig. 6e). In addition, the DL were unable to describe the vegetation pixels with more than one annual growing season, and more logistic functions (e.g. four or six) or the piecewise procedure (Zhang et al., 2003) could be used to simulate these pixels.

The reference NDVI, which was generated from observations during 15 years, is not perfect. In some regions, the valid observations in a certain period are always unavailable, so that the reference NDVI is invalid. In this case, the reference NDVI was interpolated linearly from the neighboring valid values. In some cases, the linear interpolation could not represent the true values, especially the missing values in the minimum or maximum points. But these cases are quite few in the global scale.

Besides, the reference NDVI was generated under the assumption that the trajectory of the vegetation growth in the same site is similar. This assumption would not be appropriate for some vegetation types, such as savanna or cropping. The growth of the savanna vegetation is mainly determined by the precipitation but the date of precipitation is very different for year to year. The shape of crop growth is mainly determined by the agricultural patterns. At the same time, the ephemeral water bodies would decrease the observed NDVI, which leads the reference NDVI to diverge from the truth. Additionally, others factors, such as the frequent atmospheric contamination and the atmospheric over-correction, would introduce uncertainties to the reference NDVI. Although these uncertainties existed, the reference NDVI could represent the basic trajectory of the vegetation growth for most of the sites.

The angular effect exists in MOD09A1 and MYD09A1 reflectance. The MODIS NBAR product (MCD43A4) has been corrected the directional effect but there were too many missing values in it. It was found that the angular effect was not so obvious in the reference NDVI. The ratio used in the definition of NDVI and the averaging procedure for generating the reference NDVI could reduce most of the angular effect in the MOD09A1 and MYD09A1 directional reflectance. Fig. 14 shows several examples for the comparison between the NDVI series and BRDF-corrected NDVI series (named as NDVI_NBAR) in 2006. These examples show that, although there is a little difference between the NDVI from directional reflectance and that from NBAR, the shape of time series is nearly same. The large view zenith angles would introduce noises into NDVIs calculated from directional spectral reflectance. But the

observations with large view zenith angle were quite few in the global scale. 98.96% of the pixels were less than five observations ($\approx 10\%$) with view zenith angle $> 55^\circ$ in 2008.

Fig. 14 Several temporal examples for the comparison between the NDVI series and BRDF-corrected NDVI series (named as NDVI_NBAR) in 2006 from 70°N to 20°N . The cloud and snow in MOD09A1 was masked by the inflexion-based cloud detection method, and was set as 0 in the calculated NDVI values. For the NDVI_NBAR, only the BRDF parameters from full inversion were used for correcting the BRDF effects of the MOD09A1 and then calculated the BRDF-corrected NDVI. The bracket in the upper left of each subfigure was the pixel position in the MODIS Tile.

It is impossible to evaluate all of approaches in this study. Other temporal approaches, such as the iterative caterpillar singular spectrum analysis approach (ICSSA) (Golyandina and Osipov, 2007), would perform well for filling the gaps with various fractions of invalid observations (Kandasamy et al., 2013). The spatial approaches were not evaluated due to their different assumptions. All of the evaluated approaches show relatively poor performance in the tropical monsoon forest in India, where the trajectory of the vegetation growth largely departs from the model curve. Some new approaches should be used for these vegetation types.

The NDVI was selected to represent the vegetation parameters to evaluate the approaches for gap fillings. Other vegetation parameters, such as EVI, LAI and FAPAR, are similar seasonal patterns with the NDVI. Therefore, the evaluations for the NDVI were also appropriate for other parameters.

5. Conclusion

The six approaches for gap filling, including the Fourier, DL, IDR, Whit, SG and LACC, were globally evaluated pixel-by-pixel based on a reference NDVI generated from observations during 2000-2014. The evaluations of these approaches were conducted comprehensively, including the filling effect for the simulated gap and for the gaps from real observations in 2008, the recovery of the maximum curvature point, the resistance to noise on the valid observations and the stability of the reconstructed curve. The pixel-by-pixel evaluation considers all the possible seasonal trajectories of the vegetation growth and the gap conditions.

Global bias maps from five evaluation schemes describe the global applicability of each approach. The ranks for considering all of the evaluation aspects are LACC > Fourier > IDR > DL > SG > Whit. The SG performs well in all aspects, except for the resistance to the noise. The LACC was the best approach for global gap filling. The stability of LACC also made it suitable for the inter-annual comparison of the vegetation status. These bias maps could guide to select the optimal approach to regionally or globally reconstruct the time series parameters for certain applications. This work provides a new way for the global evaluations of all temporal gap filling approaches.

Acknowledgement

This research was funded by the key research and development programs for global change and adaptation (2016YFA0600201), the National Natural Science Foundation from China (41171285) and the carbon project of the Chinese Academy of Sciences (XDA05090303). We would like to thank the MODIS science team for providing excellent and accessible data products that made this paper possible. MODIS data were obtained from the Land Processes Distributed Active Archive Center (<https://lpdaac.usgs.gov/>).

References

Alcantara, C., Kuemmerle, T., Prishchepov, A.V., & Radeloff, V.C. (2012). Mapping abandoned agriculture with multi-temporal MODIS satellite data. *Remote Sensing of Environment*, 124, 334-347

- Atkinson, P.M., Jegathan, C., Dash, J., & Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123, 400-417
- Atzberger, C., & Eilers, P.H. (2011). A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. *International Journal of Digital Earth*, 4, 365-386
- Beck, P.S., Atzberger, C., Høgda, K.A., Johansen, B., & Skidmore, A.K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, 100, 321-334
- Cao, R.Y., Chen, J., Shen, M.G., & Tang, Y.H. (2015). An improved logistic method for detecting spring vegetation phenology in grasslands from MODIS EVI time-series data. *Agricultural and Forest Meteorology*, 200, 9-20
- Chen, B.Z., Xu, G., Coops, N.C., Ciais, P., Innes, J.L., Wang, G.Y., Myneni, R.B., Wang, T.L., Krzyzanowski, J., Li, Q.L., Cao, L., & Liu, Y. (2014). Changes in vegetation photosynthetic activity trends across the Asia-Pacific region over the last three decades. *Remote Sensing of Environment*, 144, 28-41
- Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91, 332-344
- Chen, J.M., Deng, F., & Chen, M. (2006). Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter. *Geoscience and Remote Sensing, IEEE Transactions on*, 44, 2230-2238

- Dash, J., Jeganathan, C., & Atkinson, P.M. (2010). The use of MERIS Terrestrial Chlorophyll Index to study spatio-temporal variation in vegetation phenology over India. *Remote Sensing of Environment*, 114, 1388-1402
- Filipova-Racheva, D., & Hall-Beyer, M. (2000). Smoothing of NDVI time series curves for monitoring of vegetation changes in time. In, *Ecological Monitoring and Assessment Network National Science Meeting* (pp. 17-22)
- Golyandina, N., & Osipov, E. (2007). The "Caterpillar"-SSA method for analysis of time series with missing values. *Journal of Statistical Planning and Inference*, 137, 2642-2653
- Gu, J., Li, X., Huang, C.L., & Okin, G.S. (2009). A simplified data assimilation method for reconstructing time-series MODIS NDVI data. *Advances in Space Research*, 44, 501-509
- Hird, J.N., & McDermid, G.J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113, 248-258
- Jakubauskas, M.E., Legates, D.R., & Kastens, J.H. (2001). Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, 67, 461-470
- Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40, 1824-1832

- Julien, Y., & Sobrino, J.A. (2010). Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sensing of Environment*, 114, 618-625
- Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations - application to MODIS LAI products. *Biogeosciences*, 10, 4055-4071
- Kawabata, A., Ichii, K., & Yamaguchi, Y. (2001). Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *International Journal of Remote Sensing*, 22, 1377-1382
- Liu, R., & Liu, Y. (2013). Generation of new cloud masks from MODIS land surface reflectance products. *Remote Sensing of Environment*, 133, 21-37
- Lovell, J., & Graetz, R. (2001). Filtering pathfinder AVHRR land NDVI data for Australia. *International Journal of Remote Sensing*, 22, 2649-2654
- Lu, X.L., Liu, R.G., Liu, J.Y., & Liang, S.L. (2007). Removal of noise by wavelet method to generate high quality temporal data of terrestrial MODIS products. *Photogrammetric Engineering and Remote Sensing*, 73, 1129-1139
- Ma, M., & Veroustraete, F. (2006). Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Advances in Space Research*, 37, 835-840

- Madden, H.H. (1978). Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data. *Analytical Chemistry*, 50, 1383-1386
- Michishita, R., Jin, Z., Chen, J., & Xu, B. (2014). Empirical comparison of noise reduction techniques for NDVI time-series based on a new measure. *Isprs Journal of Photogrammetry and Remote Sensing*, 91, 17-28
- Moreno, A., Garcia-Haro, F.J., Martinez, B., & Gilabert, M.A. (2014). Noise Reduction and Gap Filling of fAPAR Time Series Using an Adapted Local Regression Filter. *Remote Sensing*, 6, 8238-8260
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19, 135-141
- Velleman, P.F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association*, 75, 609-615
- Verger, A., Baret, F., Weiss, M., Kandasamy, S., & Vermote, E. (2013). The CACAO Method for Smoothing, Gap Filling, and Characterizing Seasonal Anomalies in Satellite Time Series. *Ieee Transactions on Geoscience and Remote Sensing*, 51, 1963-1972
- Vermote, E.F., & Kotchenova, S. (2008). Atmospheric correction for the monitoring of land surfaces. *Journal of Geophysical Research-Atmospheres*, 113

- Viovy, N., Arino, O., & Belward, A. (1992). The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, 13, 1585-1590
- Walker, J.J., de Beurs, K.M., Wynne, R.H., & Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment*, 117, 381-393
- Wessels, K.J., van den Bergh, F., & Scholes, R.J. (2012). Limits to detectability of land degradation by trend analysis of vegetation index data. *Remote Sensing of Environment*, 125, 10-22
- White, K., Pontius, J., & Schaberg, P. (2014). Remote sensing of spring phenology in northeastern forests: A comparison of methods, field metrics and sources of uncertainty. *Remote Sensing of Environment*, 148, 97-107
- Zhang, X.Y., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84, 471-475
- Zhou, J., Jia, L., & Menenti, M. (2015). Reconstruction of global MODIS NDVI time series: Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sensing of Environment*, 163, 217-228
- Zhu, W.Q., Pan, Y.Z., He, H., Wang, L.L., Mou, M.J., & Liu, J.H. (2012). A Changing-Weight Filter Method for Reconstructing a High-Quality NDVI Time Series to Preserve the Integrity of Vegetation Phenology. *Ieee Transactions on Geoscience and Remote Sensing*, 50, 1085-1094

Figures List:

Fig. 1 Example of the reference NDVI.

Fig. 2 Data process flowchart of the global evaluations.

Fig. 3 (a) Global map of the gap percentage for the 15-year observations in the MOD09A1 and (b) Global map of the merged Koppen climate classification

Fig. 4 Temporal profiles of the MODIS NDVI with cloudy or snowy observations: (a) the subarctic climate (54.66°N, 155.73°E); (b) the tundra climate in the Tibetan Plateau (30.38°N, 92.76°E); (c) the tropical monsoon climate in India (22.16°N, 75.14°E); (d) the humid subtropical climate (24.97°N, 102.18°E); (e) the tropical rainforest climate around the Gulf of Guinea (6.37°N, 5.12°E); (f) the tropical savanna climate in the north of Australia (18.15°S, 140.13°E).

Fig. 5 Global maps of the characteristics of the reference NDVI: (a) the invalid percentage; (b) the maximum gap length; (c) the roughness.

Fig. 6 Site examples of typical shapes for the gap filling: (a) the boreal forest; (b) the tropical savanna; (c) the forest in the Southern Hemisphere; (d) the grassland; (e) the mixed area of forest and grassland; (f) the double rice or savanna.

Fig. 7 Histogram of the pixel percentage with bias < 0.03 for the four largest climate regions (EV1 is for the Simgaps evaluation, EV2 is for the Realgaps evaluation, EV3 is for the Keypoint evaluation, EV4 is for the Noise evaluation and EV5 is for the Stability evaluation).

Fig. 8 Global distributions of the gap filling biases under the Simgaps evaluation.

Fig. 9 Global distributions of the gap filling biases under the Realgaps evaluation.

Fig. 10 Global distributions of the gap filling bias under the Keypoint evaluation.

Fig. 11 Global distributions of the gap filling biases under the Noise evaluation.

Fig. 12 Global distributions of the final biases under the Stability evaluation.

Fig. 13 Comparisons between different tunable parameters for the approaches under the Simgaps evaluation.

Fig. 14 Several temporal examples for the comparison between the NDVI series and BRDF-corrected NDVI series (named as NDVI_NBAR) in 2006 from 70°N to 20°N. The cloud and snow in MOD09A1 was masked by the inflexion-based cloud detection method, and was set as 0 in the calculated NDVI values. For the NDVI_NBAR, only the BRDF parameters from full inversion were used for correcting the BRDF effects of the MOD09A1 and then calculated the BRDF-corrected NDVI. The bracket in the upper left of each subfigure was the pixel position in the MODIS Tile.