Removal of Noise by Wavelet Method to Generate High Quality Temporal Data of Terrestrial MODIS Products

Xiaoliang Lu, Ronggao Liu, Jiyuan Liu, and Shunlin Liang

Abstract

Time-series terrestrial parameters derived from NOAA/AVHRR, SPOT/VEGETATION, TERRA, or AQUA/MODIS data, such as Normalized Difference Vegetation Index (NDVI), Leaf Index Area (LAI), and Albedo, have been extensively applied to global climate change. However, the noise impedes these data from being further analyzed and used. In this paper, a wavelet-based method is used to remove the contaminated data from time-series observations, which can effectively maintain the temporal pattern and approximate the "true" signals. The method is composed of two steps: (a), time-series values are linearly interpolated with the help of quality flags and the blue band, and (b), time series are decomposed into different scales and the highest correlation among several adjacent scales is used, which is more robust and objective than the threshold-based method. Our objective was to reduce noise in MODIS NDVI, LAI, and Albedo time-series data and to compare this technique with the BISE algorithm, Fourier-based fitting method, and the Savitzky-Golay filter method. The results indicate that our newly developed method enhances the ability to remove noise in all three time-series data products.

Introduction

Time-series data for some land surface parameters, such as Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and albedo, have been successfully used a wide range of fields. Many analysis methods have been developed from NDVI time-series data to (a) detect land-cover changes (Zhan *et al.*, 2002; Friedl *et al.*, 2002; Roy *et al.*, 2002), (b) derive biophysical parameters for other models (Sellers *et al.*, 1994; Moody and Johnson, 2001; Lu *et al.*, 2003), and (c) monitor vegetation dynamics (Sakamoto *et al.*, 2005; Beck *et al.*, 2006). Observing the change of LAI in time and space plays a significant part in understanding and modeling the land surface processes in the entire climate system (Running *et al.*, 1988; Potter *et al.*, 1993; Chase *et al.*, 1996). Studying albedo time series plays a central role in global energy budget and climate forcing issues (Dirmeyer and Shukla, 1994; Dickinson, 1995; Roesch *et al.*, 2002).

After National Aeronautics and Space Administration (NASA) launched the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard both the Terra and Aqua satellites, respectively, in December 1999 and May 2002, researchers were given an unprecedented way to get a variety of time-series data. However, these time-series data inevitably contain disturbances caused by cloud presence (Gutman, 1991), atmospheric variability (Huete and Liu, 1994), and aerosol scattering (Xiao *et al.*, 2003). Noise degrades data and hinders analysis. To reduce noise, the Maximum Value Composite (MVC) method (Holben *et al.*, 1986) is usually composited to get a higher percentage of clear-sky data. However, if the composite period is long, the land surface does not remain static; and if it is too short, the atmospheric disturbance cannot be removed effectively, especially in cloudy regions. For example, there exist many low quality pixels in 8- or 16-day composite MODIS products (Moody *et al.*, 2005). Several methods, based on interpolation of time series data, have been proposed to remove such noise and to reconstruct high-quality NDVI time-series data. These methods can be generally categorized into two general types. The first methods include removing noise in the time domain, such as the best index slope extraction (BISE) algorithm (Viovy *et al.*, 1992), the asymmetric Gaussian function fitting approach (Jonsson and Eklundh, 2002), the weighted least-squares linear regression approach (Swets *et al.*, 1999), the Savitzky-Golay filter approach (Chen *et al.*, 2005), and the ecosystem-dependent temporal interpolation technique (Moody *et al.*, 2005). The second type includes noise-removal methods in the frequency domain, such as Fourier-based fitting methods (Sellers *et al.*, 1994; Roerink *et al.*, 2000). Each of the approaches has advantages. Before time-series data can be utilized in advanced research applications, one of these data smoothing approaches must be applied. The BISE algorithm has been used to classify vegetation and forest types (Xiao *et al.*, 2002). The Fourier-based fitting approach has been employed to derive terrestrial biophysical parameters (Moody and Johnson, 2001), and to classify land-cover types (Anders, 1994). Asymmetric Gaussian

Xiaoliang Lu is with the Department of Earth & Atmospheric Sciences, Purdue University, CIVIL 550 Stadium Mall Drive, West Lafayette, IN 47907, and formerly with the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China (lxiaolia@purdue.edu).

Ronggao Liu, and Jiyuan Liu are with the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, No.11A, Datun Road, Chaoyang District, Beijing, 100101 China.

Shunlin Liang is with the Department of Geography, University of Maryland, 2181 LeFrak Hall, College Park, MD 20742.

function fitting methods have been used to extract vegetation seasonality information (Jonsson and Eklundh, 2002). The ecosystem-dependent, temporal-interpolation technique can provide researchers with a snow-free, land-surface albedo product (Moody et al., 2005).

All these methods also have disadvantages. The effectiveness of the BISE algorithm is subjective and dependent on the researchers' experience; it is difficult to determine a reasonable length of sliding window for the smoothing as well as a threshold for acceptable percentage change. Finding one optimal sliding window and threshold that can be used in any circumstance is impossible. Furthermore, these parameters are often obtained by a tedious trial and error procedure. The result processed by the Fourier-based fitting approach is smooth, but often shows large biases with the original time-series data. It operates well on time-series data as long as that data can be approximately expressed as the total of several sine or cosine functions, otherwise it may generate spurious oscillations. As for the Savitzky-Golay filter approach, it also required empirical analyses to determine the width of the smoothing window and the degree of the smoothing polynomial. In addition, to prevent the adverse edge-effect associated with filtering, original time series should be padded with some data points to constitute a complete filter window. But padding may skew/distort the accuracy of values in the head and tail of post-processed, time-series data. And, due to its design of keeping the upper time-series envelope, some useful information may be lost in the filtering process. The asymmetric Gaussian function fitting approach works well under some favorable conditions, but it requires users to determine a set of maxima and minima to which the local functions can be fitted. There are difficulties in identifying the two parameters from noisy time-series data. Finally, for the ecosystem-dependent, temporal-interpolation technique, pixels of the same ecosystem classification may exhibit different phenological or temporal behavior within a broad region. Imposing the behavior onto retrieved pixel data may lead to errors in the resulting value-added products.

Our objective is to present a new method to reduce contamination in time-series data. Our method was tested on the NDVI eight-day composite data generated by MOD09A1 surface reflectance data, MODIS LAI eight-day composite data and MODIS Albedo 16-day composite data. The time series parameters processed by our method were also compared with the existing noise reducing methods. The results indicate that our method can reduce noise in time-series data and maintain its real pattern effectively.

Method Description

The wavelet transform (WT) can analyze signals in time-frequency space. Its strength is the feasibility of identifying and reducing noise while maintaining useful information in time-series data. In the past two decades, WT has been developed as a powerful tool in signal processing. At present, the WT has been widely used in remote sensing applications such as image fusion (Ranchin et al., 2003), hyperspectral data feature extraction (Pu and Gong, 2004), phenology detection (Sakamoto et al., 2005), morphotectonic lineament investigation (Jordan and Schott, 2005). We use the wavelet transform to remove noise in time-series data.

The Denoising Method by Wavelet Technique

Wavelets are groups of functions $\Psi_{a,b}(x)$ generated from a mother wavelet $\Psi(x)$ by dilations and translations:

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{x-b}{a}\right) \tag{1}$$

where a is the dilation parameter, and b is the translation parameter. Since $\Psi(x)$ is compact support, which indicates that the duration of the function is very limited, wavelet analysis can accurately capture the local characteristics of non-stationary signals. The WT of function $f(x)$ can be expressed by:

$$wf(a,b) = \int_{-\infty}^{+\infty} f(x) * \Psi_{a,b}(x)dx \tag{2}$$

From Equation 2, the wavelet transform decomposes signals at various scales or resolution and shifts. For most practical applications, the discrete wavelet transform (DWT), which analyzes signals over a discrete set of scales usually sampled at dyadic sequence ($2^j, j = 1,2,3, \ldots$), is accurate enough and can recover signals perfectly (Mallat, 1989). In the dyadic form, the wavelet function can be given as:

$$\Psi_{j,k}(x) = 2^{\frac{-j}{2}} \Psi\left(2^{-j}x - k\right) \tag{3}$$

where j is the jth decomposition level or scale and k is the kth wavelet coefficient. In contrast with Equation 1, we have $a = 2,4,8, \ldots 2^j$. The WT can decompose a signal both the large-scale components that represent the optimal approximation or low frequency parts of the original signal and small-scale components that represent the detailed information or high frequency parts of the signal.

The high frequency part of the original data is dominant at fine scales. Larger amplitude variations and slowly changing features of the original signal are mainly represented at coarse scales. On the basis of the distinct characteristics of decomposition coefficients at different scales, the traditional scheme of directly applying a certain threshold to wavelet coefficients is most widely used. Specifically, significant wavelet coefficients above a certain threshold at fine scales are discarded as noise and the signal is restored by the remaining coefficients. However, it may be difficult for the traditional scheme to set a reasonable threshold. The variation of decomposition coefficients at several adjacent scales, however, may help us resolve this dilemma (Xu et al., 1994). The coefficients corresponding to low frequency signal have high correlation at the neighboring scales and their absolute values increase from the finest scale to the coarsest scale. In contrast, noise has an inverse relationship with scales, as its decomposition coefficients diminish swiftly with increasing scale and do not propagate up to the next scale. Direct multiplying of coefficients at the adjacent two levels may enhance important variations while diluting high-frequency components. We can use the direct spatial correlation coefficients at adjacent two levels, which is defined (Xu et al., 1994) as:

$$cor(m,n) = w(m,n) \times w(m+1,n). \tag{4}$$

To build comparability between the direct spatial correlation and wavelet decomposition coefficients, we normalize the direct spatial correlation coefficients (Xu et al., 1994):

$$ncor(m,n) = cor(m,n) \times \sqrt{pw(m)/pcor(m)}$$
$$pw(m) = \sum_n w(m,n)^2 \tag{5}$$
$$pcor(m) = \sum_n cor(m,n)^2$$

where $ncor(m,n)$ is the normalization of correlation coefficients, $pw(m)$ and $pcor(m)$ represent the power of wavelet decomposition coefficients and correlation coefficients at the level of m, respectively.

For each level, we compare the absolute values of $ncor(m,n)$ with $w(m,n)$ at all positions. Thus, we implement our method at the first scale. If $|ncor(1,n)|$ is smaller than $|w(1,n)|$ at the position of n, which implies that correlation

coefficient at this position becomes small through the first scale to the second scale and noise is dominant at that position, the $w(1,n)$ is reset to zero. This comparison can be made at all levels. After we identify and modify the values, we refer to $w(m,n)$ as $w_{new}(m,n)$.

In the flowchart, we make a comparison between the normalized data $|ncor(m,n)|$ and $|w(m,n)|$ to extract the important information from $w(m,n)$ at the mth scale. By repeating the procedure at every scale, we can acquire filtered wavelet transform data $w_{new}(m,n)$. Finally, the reconstruction from $w_{new}(m,n)$ through the inverse wavelet transform will yield the filtered time series.

Implementation of the New Wavelet-based Method

Weighted Inverse Distance Interpolation of Low Quality Values

Most remote sensing data include quality control (QC) information. For time-series data, each observation is assigned a QC flag during processing. This information is encoded in a standard format and stored in a special dataset so that any specific condition can be easily extracted and used. The ancillary data are valuable for us to indicate and remove low-quality data points (Jonsson and Eklundh, 2002; Chen *et al.*, 2005). In this paper, we use weighted inverse distance interpolation to improve the quality of time-series data. Concretely, supposing that there is a time series of data points (t_i, V_i, F_i), $i = 1,2,3 \ldots n$, where t_i is the date, V_i is the value of time-series data, and F_i is the quality status flag. If the ith point is flagged as a cloud or snow point, then V_i will be replaced by the result from the weighted inverse distance interpolation, which uses adjacent clear points that are not identified as cloud or snow points. The distance is the time interval between points to be replaced and adjacent clear points used in the interpolation. But the QC flag is not reliable in all cases; the cloud mask algorithm is adequate for identifying larger, cooler clouds but problematic in detecting small clouds and cloud edges (Giglio *et al.*, 2003). We use the blue band of MODIS as an additive indicator of cloud status. The observations whose reflectance of the blue band is more than 0.1 are rejected as abnormal data and replaced by the interpolated results. In Figure 1a, some drops in the NDVI time series are identified by cloud flags or mix flags, while those flags do not indicate the sudden drop on the 151st day, which shows that the cloud mask algorithm based on setting certain thresholds is not very reliable. The linear interpolation of this NDVI time series, using only quality control flags, is limited, as is illustrated by the fact that the abnormal value on the 151st day remains in the interpolation result (Figure 1b). Fortunately, the information provided by the MODIS blue band helps overcome the drawback and the observation on the 151st day is properly captured using the weighted inverse distance interpolation (Figure 1c).

Removal of Noise In Time Series by Wavelet

After the process of step one, we assume that the sharp variations remaining result from intense changes in the land surface. Some residual noise also remains from the first processing step, but this residual noise is shown as small fluctuations. We assume that the frequency of remaining noise components in time-series data is higher than that of the normal surface change. Removing this noise allows us to highlight actual surface-related changes while suppressing noise with relatively high frequency.

There are many different types of mother wavelets, such as Daubechies family (dbN), Biorthogonal family, Coiflets family, and Symlets family. Daubechies family (dbN) is a series of compactly supported orthogonal wavelets. N specifies the order of the mother wavelet and is related to the number of coefficients necessary to
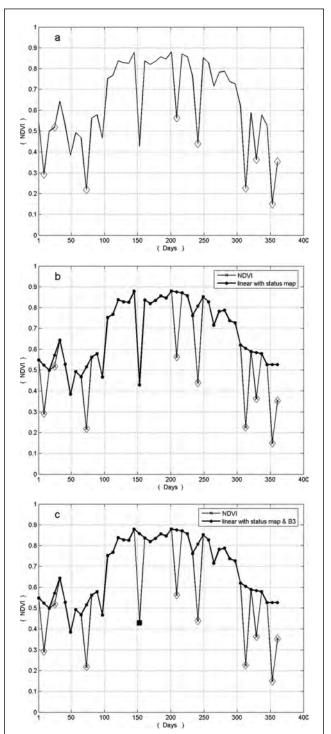

Figure 1. An example showing the NDVI time series and its linear interpolation results by two methods: (a) Original NDVI time series and cloud and mixed points (diamonds in the figure), (b) The linear interpolated NDVI time series (thick solid line) on the basis of quality control flags, and (c) The linear interpolated NDVI time series (thick solid line) on the basis of quality control flags and the information of MODIS blue band (point with more than 0.1 in the blue band is squared in the figure).

represent the associated low-pass and high-pass filters in the dyadic filter tree implementation. The "db" is the "surname" of the wavelet. The high order means we get smooth results but have a poor ability to localize the small features. In this study, we use "db3" as the mother wavelet, and the decomposition level is three. Unfortunately, the classical DWT suffers a drawback: the DWT is not a time-invariant transform. This means the DWT of a translated version of a signal X is not, in general, the translated version of the DWT of X even with periodic signal extension. But in some studies such as phenology, the position of the variations may correspond to important changes in time series and thus need to be accurately reserved; time distortions may lead to a large number of artifacts. In order to overcome this difficulty and to get more complete characteristics of the analyzed signal, we use the stationary wavelet transform (SWT) (Pesquet *et al.*, 2003) to decompose the time series since it has the following characteristics: (a) the time series processed by DWT may have overshoots at discontinuities, while SWT will not bring the phenomenon into the results; (b) the number of decomposition coefficients at different scales must be equal, which is a prerequisite to multiplying coefficients at adjacent scales. The numbers of decomposition coefficients acquired by SWT are equal on all scales, while these numbers decrease from fine to coarse scales.

Results and Analysis

Description of the Study Data
We used the eight-day composite NDVI product as the test data set for evaluating the performance of our method. Because the standard NDVI products from the EOS Data Gateway are 16-day composite, MODIS eight-day composite surface reflectance products (MOD09A1) were downloaded to create an NDVI data set for the experiments. The surface reflectance product name is "MODIS/TERRA surface reflectance eight-day Global 500 m SIN GRID v004." In the production of MOD09A1, atmospheric corrections for gases, thin cirrus clouds and aerosols are implemented. Each eight-day composite product includes the measurement of surface spectral reflectance of the seven spectral bands at 500 m spatial resolution as it would be measured at ground level in the absence of atmospheric scattering or absorption. MOD09A1 also includes quality control flags to account for various image artifacts (e.g., clouds, cloud shadow). The MODIS products are organized in a tile system with the Sinusoidal (SIN) projection grid, and each tile covers an area of 1,200 km by 1,200 km (approximately 10′ latitude by 10′ longitude at equator). In this study, we downloaded MOD09A1 data during the period from January 2004 to December 2004 for eastern China from the EOS data center. The tile H27V05 covers the study area. For each eight-day composite, we calculated NDVI by using surface reflectance values from the red and NIR (841 to 875 nm) bands:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}. \qquad (6)$$

For eight-day composite datasets, the time series contains 46 observations per pixel per year. The number of elements in the input data array should be 2^N for SWT, where N stands for the decomposition level. With this in mind, the original input data array was extended by the symmetric boundary value replication to constitute the input array with 48 observations. During the data preparation stage, NDVI values were scaled to the range of −1 to 1.
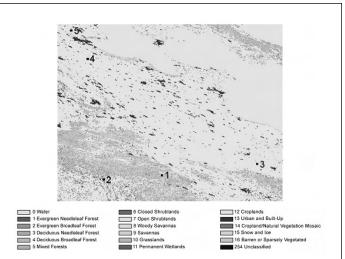

Figure 2. Land-cover map in study area and distribution of test pixels. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Comparison of New and Existing Methods
Comparisons at the Pixel Level
For a detailed assessment of the new method, comparisons between the new method and other existing methods were carried out for three primary vegetation types at pixel level. The three vegetation types are: deciduous broadleaf forest (No.1 test pixel), evergreen broadleaf forest (No.2 test pixel), and double-cropping rice (No.3 test pixel). One pixel is selected for each cover type, respectively. Figure 2 shows a MODIS land-cover (Friedl *et al.*, 2002) map of our study area and the distribution of the three test pixels. The attributes of these pixels are given in Table 1. The methods in our comparison are: (a) the BISE algorithm; (b) the Savitzky–Golay filter method; (c) the weighted Fourier series fitting method, and (d) our wavelet-based method. For the BISE algorithm, the slide window is three and the threshold is 0.2. In the Savitzky–Golay (S-G) filter method, the half-width of the smoothing window is four and the degree of the smoothing polynomial is five. For the weighted Fourier series fitting method, we reconstruct a new NDVI time series by using the first three harmonics of the Fourier series. The mother wavelet used in the new method is db3 and the decomposition level is three.

Figure 3 demonstrates the reconstructed NDVI time series for the type of deciduous broadleaf forest. From Figure 3a, the BISE algorithm can remove a large amount of noise in the time series. On the other hand, because the rising does not exceed the pre-established threshold, the abnormal peak in data at

TABLE 1. ATTRIBUTES OF PIXELS SELECTED FOR TESTING THE NEW AND EXISTING METHODS. THE SPATIAL DISTRIBUTION OF THESE PIXELS IS SHOWN IN FIGURE 2

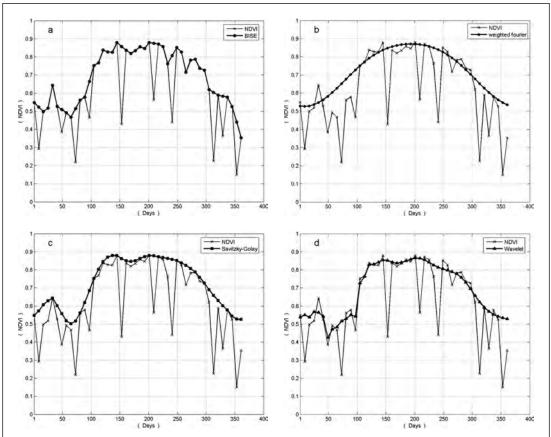| Test Pixels	Land-cover	Latitude (N)	Longitude (E)
1	Deciduous Broadleaf Forest	31.475234°	110.968835°
2	Evergreen Broadleaf forest	31.225241°	107.643634°
3	Cropland	32.108550°	116.719538°
4	Cropland	38.658370°	117.751021°
5	Cropland	39.391683°	117.016452°

Figure 3. The denoised NDVI time series of deciduous broadleaf generated by use of (a) the BISE algorithm, (b) the weighted Fourier series fitting method, (c) the S-G filter method, and (d) the wavelet method.

day 33 is not smoothed. The weighted Fourier series fitting method can make the most smoothed profile of NDVI time series. Since negative deviations receive low weights during the fitting process and positive deviations are assigned high weights, the weighted Fourier series fitting has better de-noising performance than traditional Fourier-based ones. Because the original NDVI time series in Figure 3 can be well approximated by the summation of several sine or cosine functions, the result of the weighted Fourier series fitting is good for most observations except that the overestimation was made at several head points, as showed in Figure 3b. From Figure 3c, the S-G method remains the value at data day 33 and overestimates several points near the peak observation. By comparison, our method can also remove noise shown as abrupt rises or drops and the unreasonable peak value at data day 33 is also partly smoothed. Moreover, the right observations that are near the contaminated points are not confused as noise (Figure 3d).

Figure 4 shows the reconstructed NDVI time series for an evergreen broadleaf forest. In Figure 4a, there is still much noise in the time series after the BISE algorithm is applied. To interpret this poor result, it is important to note that there is cloud cover during the three consecutive eight-day periods (Julian days: 153 to 160, 161 to 168, and 169 to 176 days) and the slide window of the BISE algorithm is only 3. We found that the BISE algorithm has no ability to remove the data contaminated by clouds more than for two consecutive eight-day periods. The weighted Fourier series fitting method obtained the smoothed result and the most-noisy observations were remained, as shown in Figure 4b.

Because the values of the blue band for Julian days 41 and 49 are small (0.021 and 0.022), we assume the decline in these two data points is due to sudden land surface changes. The two observations should remain unchanged in the noise removal, but the values for the two data days are overestimated by the weighted Fourier series fitting method. The situation can be explained by the idea of punishing negative deviations in the weighted fitting. The weights of the two observations are low in the fitting operation and their values are falsely exaggerated in the de-noising. Examining Figure 4c, we find that the Savizky-Golay (S-G) method can effectively remove noise. But the values for Julian days 41 and 49 are affected by the nearby high value points in the polynomial fitting and are overestimated by about 0.1. The S-G method is biased toward high values. In Figure 4d, the wavelet-based method indicates the good performance in denoising time series even in the face of consecutive cloud cover days, including Julian days 41 and 49.

Figure 5 compares the reconstructed NDVI time series for the land-cover type, double-crop rice. The BISE algorithm effectively removes noise and does not change the original low NDVI value points after harvesting, as shown in Figure 5a. In Figure 5b, the observations for Julian days 137, 145, and 153 corresponding to the periods after cropping are overestimated by the weighted Fourier series fitting method. The reconstructed time series does not correctly show the pattern of double-cropping. The distortion in de-noising is due to both the high priorities assigned to positive deviations and observations of low value are replaced by the relatively high-value results. Figure 5c shows that the S-G
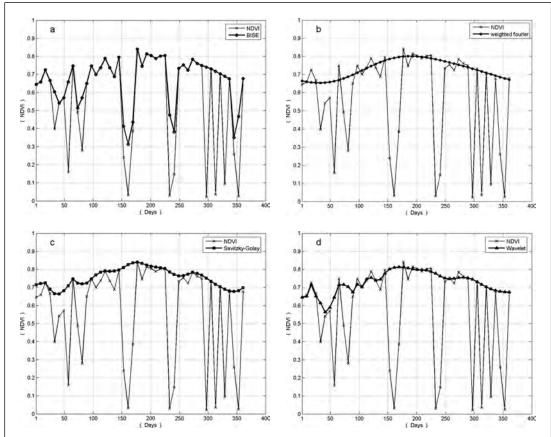
Figure 4. The denoised NDVI time series of evergreen broadleaf generated by use of (a) the BISE algorithm, (b) the weighted Fourier series fitting method, (c) the S-G filter method, and (d) the wavelet method.

filter method has steady de-noising ability. But due to its bias towards high values, the data at Julian days 137, 145, and 153, corresponding to the periods after cropping, are overestimated. In Figure 5d, the wavelet-based method removes noisy points accurately and preserves key features of the time series. The low NDVI value points after cropping are not confused in the filtered time series. The reconstructed time series represents the actual process of the bi-annual vegetation cycle.

Comparisons at the Image Level
In order to compare the new and existing methods at the regional scale, we used the same NDVI data at the image level. Figure 6a and Figure 6b are the NDVI images for Julian days 233 and 241 of 2004, respectively. We found from the two images that cloudy days lasted more than half of one month and thus many areas were masked out. Since the threshold in the BISE algorithm is not suitable for eliminating the continuous noise, Figure 6c still shows much noise in the southern area. Little obvious noise remains in the NDVI image after the weighted Fourier series fitting method is used. The high NDVI data in the original image have also been modified to give lower values (Figure 6d). To interpret the result, it is important to note that the first three harmonics can effectively characterize the basic behavior of the surface, while these high-value observations are mainly expressed by high-frequency harmonics. Figure 6e and Figure 6f are the results of implementing the S-G method and our method, respectively. Both methods effectively remove noise and the correct

parts of the original data are also retained perfectly. It indicates that the two methods work more effectively than other methods in the face of long-term noise.

Figure 7a and Figure 7b are the NDVI images for Julian days 297 and 305 in 2004, respectively. Due to cropping, many areas in the two images have low NDVI values, and it is clear that there is still some noise in the image for Julian day 297. Figure 7c shows a reconstructed NDVI image for Julian day 297 by using the BISE method, and we can see from the image that the noise is almost completely removed, and the NDVI values in the cropped areas are not modified. The result of implementing the weighted Fourier series fitting method for Julian day 297 shows less noise, but the NDVI values in the cropped area are also replaced by higher NDVI values, shown in Figure 7d. The concept of ignoring low-value observations in time series causes undesirable distortion. Figure 7e and Figure 7f demonstrate that the S-G method and our method are effective for constructing noise-free, NDVI time-series data sets. But after cautious comparison with Figure 7a, it can be found that some NDVI values in cropping areas in Figure 7e are overestimated to some extent. We assume that the S-G filter method is designed to be biased toward relatively high values, thus explaining the results for this method.

Applications
To demonstrate the performance of the proposed method when applied to other time-series data, it is also used to reduce noise in MODIS LAI and albedo time-series data.
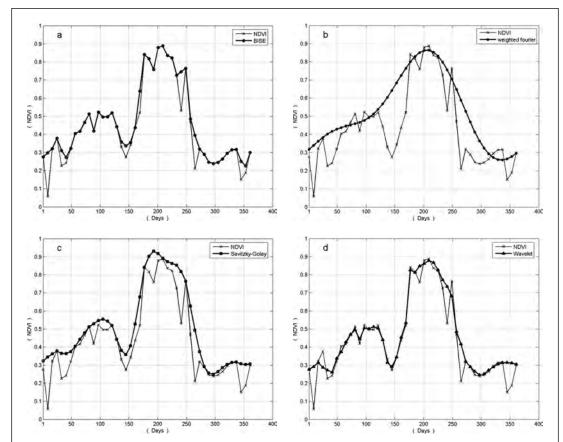
Figure 5. The denoised NDVI time-series of double-cropping rice generated by use of (a) the BISE algorithm, (b) the weighted Fourier series fitting method, (c) the S-G filter method, and (d) the wavelet method.
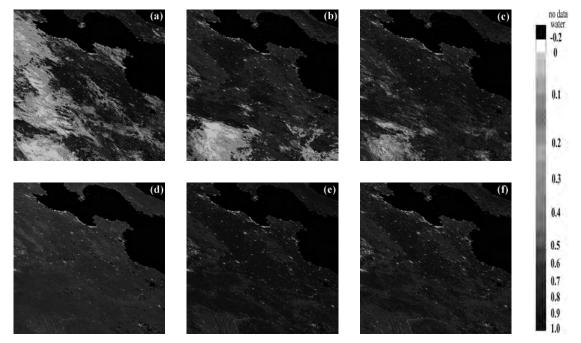

Figure 6. Composite NDVI images at (a) data day 233, (b) data day 241 and reconstructed time series at data day 233 generated by (c) the BISE method, (d) the weighted Fourier series fitting method, (e) the S-G filter method, and (f) the wavelet method. A color version of this figure is available at the ASPRS website: **www.asprs.org**.
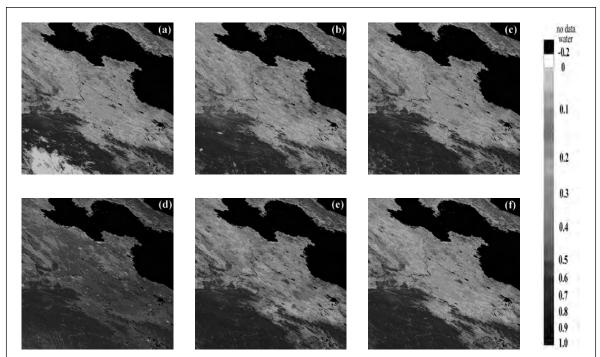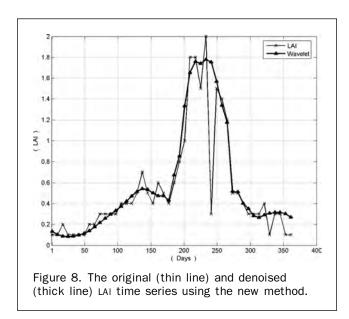
Figure 7. Composited NDVI images at (a) data day 297, (b) data day 305 and reconstructed time series at data day 297 generated by (c) the BISE method, (d) the weighted Fourier series fitting method, (e) the S-G filter method, and (f) the wavelet method. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Figure 8. The original (thin line) and denoised (thick line) LAI time series using the new method.

Denoising LAI Time Series

In the present study, MODIS LAI eight-day composite scenes from January 2004 to December 2004 were downloaded from the Earth Observing System data gateway as the experiment data set. LAI time series of a pixel (No. 4 test pixel in Figure 2) in the study area was plotted as test time series (Figure 8). The MODIS LAI product (MOD15A2; collection 4) with 1 km resolution is produced on eight-day compositing period, where the selected value in a compositing period is that with the highest corresponding fraction of absorbed photosynthetically active radiation (FPAR). But there are still many observations

under cloudy conditions in LAI time series, as is detected through the sudden drop or rebound of LAI in the time series. Before the removal of noise by the wavelet, the observations contaminated by cloud in the time series are also filtered by the weighted inverse distance interpolation. Some observations in the time series are not in the valid range and they are filled with fill values. These values are replaced by zeros before denoising by wavelet. Just like the wavelet used in NDVI time series, the mother wavelet used here is still db3, and the decomposition level is also 3.

There is an observation with abnormally sudden drop in the original LAI time series (Figure 8). The value is successfully removed in the de-noised time series, and some small fluctuations are also smoothed by the new method. Compared to the original LAI time series, the reconstructed LAI time series is more smooth and appropriate for further researches, such as assessing phenological events and estimating terrestrial biophysical parameters accurately when combined with advanced models.

Denoising Albedo Time Series

The MOD43B MODIS BRDF/Albedo algorithm provides four standard products in HDF-EOS format. The third operational product (MOD43B3) provides a standard suite of albedos. It provides users with black-sky and white-sky albedos for seven spectral bands (MODIS channels 1 through 7) and the three broadbands (0.3 to 0.7, 0.7 to 3.0, and 0.3 to 5.0 μm). All BRDF/Albedo products supply per-pixel quality flags, as do all MODIS land products. For the albedo quality dataset stored in BRDF/Albedo products, the first word of it is band-averaged quality information and the second is band-specific quality information. In addition to the mandatory QA flags, the snow cover flag for every pixel is also stored in the first word. The information on the number of observations for each pixel over 16-day period is provided by the second word.
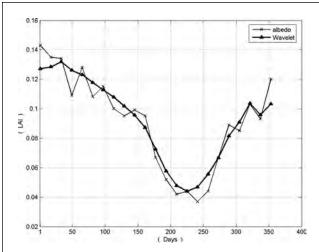
Figure 9. The original (thin line) and denoised (thick line) albedo time series using the new method.

contained in low values. The proposed wavelet-based method makes the most of quality flags in products. To make up the limitation of the flags, the blue band of MODIS is also used. When compared with other methods, the proposed method is more effective in identifying and removing noise. At the same time, it can properly maintain the characteristics of the original data and the correct data in original time series are not confused in the reconstruction.

It is difficult to provide an estimate of error for the temporal interpolation technique, as the validation data and the instruments used to validate the data both suffer from the same issues. Namely, when it is cloudy or poor in atmospheric condition, reliable retrievals cannot be made. However, we can use the pixels that have almost full clear temporal coverage as the validation data and these pixels can be acquired by the multi-year aggregation method. We use NDVI time-series data as an example to show the aggregation method and use it to evaluate our method. We downloaded eight-day composite reflectance data covered from 2001 to 2004 to create NDVI data in these four years. For each pixel, an average from the valid data in the four years is computed for generating the aggregation data. By the operation, the NDVI data in these four years can provide

The black-sky albedo for the MODIS red spectral band time series is used as the test time series in this paper. A test pixel (No. 5 test pixel in Figure 2) in the study area was selected and albedo time series for the pixel during January 2004 to December 2004 was plotted (Figure 9). The spatial distribution of the test pixel is shown in Figure 2. Both snow-covered and cloudy observations are replaced by the result of the weighted inverse distance interpolation. Because they are not enough observations to have confidence in the retrieved albedo values, data with less than three observations over the 16-day period were also replaced by the same interpolation method.

The mother wavelet and the decomposition level used here are db3 and 3, respectively. Because the composite period of MOD43B3 albedo data is 16 days, there are 23 observations for one year. To satisfy the qualification for input data length, the last observation in time series is duplicated and added to the tail of the albedo time-series data. Before denoising by wavelet, the observations with values outside of the valid range are reset to zero. The de-noising albedo time-series data is showed in Figure 9.

In Figure 9, we can see some drops and peaks in the original albedo data series. Using the newly developed method, the reconstructed albedo time series does not pick up short-term strong fluctuations and however shows smooth transitions closely related to the natural change. In comparison with the waved original time series, we can anticipate that the denoised time series is closer to the natural change pattern.

Discussion

From above comparison experiments, we learn that the weighted Fourier series fitting method obtains the smoothest profile of time series, but shows a large displacement away from the original observations with low values. Its indistinguishable discard of low-value observations may obliterate the useful information in time series. The BISE algorithm is simple in theory and its calculation is very fast, but its performance depends highly on the optimal threshold and the sliding window. The setting of the two parameters requires the tedious trial-and-error procedure, which may restrict its use in large-scale regions. The S-G filter method is an effective way to improve the quality of time series, but it always gives priority to high values and overlooks some useful information
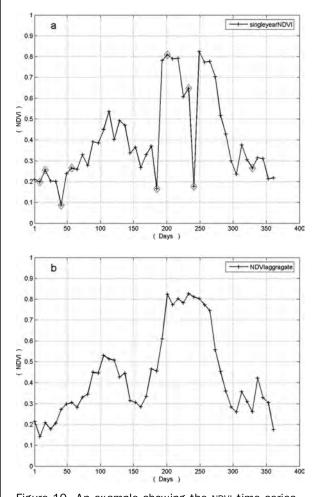

Figure 10. An example showing the NDVI time series of a single year and 4-year aggregation: (a) NDVI time series of a test pixel in 2003, cloud and mix flag points are shown as diamond flags, and (b) four-year aggregate NDVI time series of the same pixel.
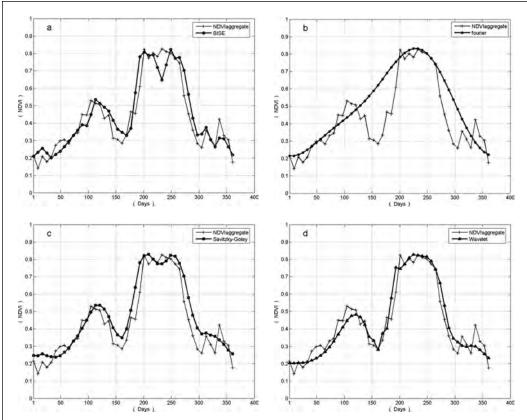
Figure 11. Comparison of the aggregate MODIS NDVI time series and the reconstructed results generated by different methods: aggregate NDVI time series and the final NDVI time-series using (a) the BISE algorithm, (b) the weighted Fourier series fitting method, (c) the Savitzky-Golay filter method, and (d) the new method.

us with aggregate NDVI product. It is clear that a pixel is unlikely to be screened by cloud in all the four years. So we assume that the aggregate product is almost noise-free and can express the real land condition to a great extent.

Figure 10 shows us the advantage of the aggregate production. In Figure 10a, there are 20 percent observations acquired on the condition of cloud or mix. On the other hand, the aggregate NDVI time series is completely clear for the test point (Figure 10b). The aggregate NDVI time series are used as reference data to evaluate the performance of our method. We use the proposed wavelet-based method to remove the noise in the NDVI time series, which is shown in Figure 10a and compare the reconstructed result with the aggregate time series. Other existing methods are also carried out to show their performances. The results are showed in Figure 11.

Of the four methods, almost all points obtained by the weighted Fourier series fitting method show large displacement away from the aggregate points (Figure 11b). The NDVI time series provided by the BISE and the S–G filter method are similar to the aggregate result in most of the points, which is clearly expressed in Figure 11a and Figure 11c. However, through carefully comparing them with the aggregate product, we can find that there is an abnormal decrease in the NDVI time series obtained by the BISE method and the S–G filter method gives higher outcome at the observations with low values in the aggregate time series. As for our wavelet-based method, visual examinations of Figure 11d reveal that the two time series are matched well. The results generated by the new wavelet-based method are in good agreement with most points of aggregate time series.

Conclusions

The noise in time series impedes the study of global environmental change when these time-series data sets are used as input. However, existing methods for reconstructing time series are not steady under various conditions or bring errors to original right data. We assume that the quality flags and the blue band of MODIS can effectively remove significant low-quality points in time series and wavelet can help us approximate the real profile of time-series data The remainder sudden changes in time series which have been processed by the first step indict the intense changes of land surface (e.g., due to cutting of meadows, removal of trees, agricultural harvesting, etc.). Because influences of these events on variations in the time series data are larger compared to the residual noise, these sudden changes are easily retained and are not degraded in the wavelet-based de-noising. By comparing the performance with the newly developed method, the BISE algorithm, the Fourier-based fitting method and the S–G filter method, it have been found that our method shows the following four advantages over existing methods: (a) it takes advantage of ancillary data in the form of cloud flags and the information of the blue band; (b) it does not depend on the subjective parameters and the threshold in the blue band is robust in most cases; (c) it can maintain the real patterns of time series change, while it removes noise; (d) it is easy to implement because commercial software such as MATLAB and IDL include the wavelets in their function library. For these reasons, the new method can be applied to reconstruct high-quality time-series data sets in many researches, such as the monitoring

inter-annual vegetation changes, deriving biophysical land surface variables and modeling terrestrial ecosystems.

Acknowledgments

References

Anders, L., W.A. Salas, and D. Skole, 1994. Fourier analysis of multi-temporal AVHRR data applied to a land-cover classification, *International Journal of Remote Sensing*, 15(5):1115–1121.

Beck, P.S.A., C. Atzberger, and K.A. Høgda, 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI, *Remote Sensing of Environment*, 100:321–334.

Chase, T.N., R.A. Pielke, T.G.F. Kittel, and R. Nemani, 1996. Sensitivity of a general circulation model to global changes in leaf area index, *Journal of Geophysical Research*, 101:7393–7408.

Chen, J., P. Jonsson, M. Tamura, Z.H. Gu, B. Matsushita, and L. Eklundh, 2005. 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.

Dickinson, R.E., 1995. Land processes in climate models, *Remote Sensing of Environment*, 51:27–38.

Dirmeyer, P.A., and J. Shukla, 1994. Albedo as a modulator of climate response to tropical deforestation, *Journal of Geophysical Research*, 99:20863–20878.

Friedl, M.A., D.K. McIver, and J.C.F. Hodges, 2002. Global land-cover mapping from MODIS: algorithms and early results, *Remote Sensing of Environment*, 83:287–302.

Giglio, L., J. Descloitresa, C.O. Justicec, and Y.J. Kaufman, 2003. An enhanced contextual fire detection algorithm for MODIS, *Remote Sensing of Environment*, 87:273–282.

Gutman, G.G., 1991. Vegetation indices from AVHRR: An update and future prospects, *Remote Sensing of Environment*, 35:121–136.

Holben, B.N., 1986. Characteristic of maximum value composite images for temporal AVHRR data, *International Journal of Remote Sensing*, 7(11):1417–1434.

Huete, A.R., and H.Q. Liu, 1994. An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS, *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):897–905.

Jonsson, P., and L. Eklundh, 2002. Seasonality extraction by function fitting to time-series of satellite sensor data, *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1824–1832.

Jordan, G., and B, Schott, 2005. Application of wavelet analysis to the study of spatial pattern of morphotectonic lineaments in digital terrain models, A case study, *Remote Sensing of Environment*, 94:31–38.

Lu, H., M.R. Raupach, and T.R. McVicar, 2003. Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series, *Remote Sensing of Environment*, 86:1–18.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Moody, A., and D. Johnson, 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform, *Remote Sensing of Environment*, 65:305–323.

Moody, E.G., M.D. King, S. Platnick, C.B. Schaaf, and F. Gao, 2005. Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products, *IEEE Transactions on Geoscience and Remote Sensing*, 43:144–158.

Potter, C.S., J.T. Randerson, C.B. Field, and P.A. Matson, 1993. Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochemistry Cycles*, 7:811–841.

Pu, R.L., and P. Gong, 2004. Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping, *Remote Sensing of Environment*, 91:212–224.

Ranchina, T., B. Aiazzib, L. Alparonec, S. Barontib, and L. Wald, 2003. Image fusion-The ARSIS concept and some successful implementation schemes, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58:4–18.

Roerink, G.J., M. Menenti, and W. Verhoef, 2000. Reconstructing cloud-free NDVI composites using Fourier analysis of time series, *International Journal of Remote Sensing*, 21(9):1911–1917.

Roesch, A., M. Wild, R. Pinker, and A. Ohmura, 2002. Comparison of spectral surface albedos and their impact on the general circulation model simulated surface climate, *Journal of Geophysical Research*, 107(4221):doi:10.1029/2001JD000809.

Roy, D.P., P. Lewis, and C.O. Justice, 2002. Burned area mapping using multi-temporal moderate spatial resolution data-A bi-directional reflectance model-based expectation approach, *Remote Sensing of Environment*, 83:263–286.

Running, S.W., and J.C. Coughlan, 1988. A general model of forest ecosystem processes for regional applications, *Ecological Modeling*, 42:124–154.

Sakamoto, T., M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno, 2005. A crop phenology detection method using time-series MODIS data, *Remote Sensing of Environment*, 96:366–374.

Savitzky, A., and M.J.E. Golay, 1964. Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry*, 36:1627–1639.

Sellers, P.J., C.J. Tucker, G.J. Collatz, S.O. Los, C.O. Justice, and D.A. Dazlich, 1994. A global 1 by 1 NDVI data set for climate studies: 2. The generation of global fields of terrestrial biophysical parameters from the NDVI, *International Journal of Remote Sensing*, 151:3519–3545.

Swets, D.L., B.C. Reed, J.D. Rowland, and S.E. Marko, 1999. A weighted least squares approach to temporal NDVI smoothing, *Proceedings of the ASPRS Annual Conference*, Washington, D.C., 17–21 May, pp. 526–536.

Viovy, N., O. Arino, and A. Belward, 1992. The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series, *International Journal of Remote Sensing*, 13(8):1585–1590.

Xiao, X.M., B. Braswell, Q.Y. Zhang, S. Boles, S. Frolking, and B. Moore, 2003. Sensitivity of vegetation indices to atmospheric aerosols: Continental-scale observations in Northern Asia, *Remote Sensing of Environment*, 84:385–392.

Xiao, X.M., B. Stephen, J.Y. Liu, and D.F. Zhang, 2002. Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 Vegetation sensor data, *Remote Sensing of Environment*, 82:335–348.

Xu, Y., J.B. Weaver, D.M. Healy, and J. Lu, 1994. Wavelet transform domain filters: A spatially selective noise filtration technique, *IEEE Transactions on Image Processing*, 3:747–758.

Zhan, X., R.A. Sohlberg, J.R.G. Townshend, C. DiMiceli, M.L. Carroll, and J.C. Eastman, 2002. Detection of land-cover changes using MODIS 250 m data, *Remote Sensing of Environment*, 83:336–350.