Available online at www.sciencedirect.com

SCIENCE DIRECT®

Remote Sensing of Environment 91 (2004) 332–344

Remote Sensing of Environment

www.elsevier.com/locate/rse

A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter

Jin Chen[a,b,*], Per. Jönsson[c], Masayuki Tamura[b], Zhihui Gu[a], Bunkei Matsushita[b], Lars Eklundh[d]

[a] *Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Ministry of Education of China Beijing 100875, China*
[b] *Social and Information System Division, Japan National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, 305-8506, Japan*
[c] *Division of Mathematics, Natural Sciences and Language, Malmö University, Malmö, Sweden, and the Department of Physics, Lund University, Lund, Sweden*
[d] *Department of Physical Geography and Ecosystems Analysis, Lund University, Lund, Sweden*

Received 11 February 2003; received in revised form 5 November 2003; accepted 14 March 2004

Abstract

Although the Normalized Difference Vegetation Index (NDVI) time-series data, derived from NOAA/AVHRR, SPOT/VEGETATION, TERRA or AQUA/MODIS, has been successfully used in research regarding global environmental change, residual noise in the NDVI time-series data, even after applying strict pre-processing, impedes further analysis and risks generating erroneous results. Based on the assumptions that NDVI time-series follow annual cycles of growth and decline of vegetation, and that clouds or poor atmospheric conditions usually depress NDVI values, we have developed in the present study a simple but robust method based on the Savitzky–Golay filter to smooth out noise in NDVI time-series, specifically that caused primarily by cloud contamination and atmospheric variability. Our method was developed to make data approach the upper NDVI envelope and to reflect the changes in NDVI patterns via an iteration process. From the results obtained by applying the newly developed method to a 10-day MVC SPOT VGT-S product, we provide optimized parameters for the new method and compare this technique with the BISE algorithm and Fourier-based fitting method. Our results indicate that the new method is more effective in obtaining high-quality NDVI time-series.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Savitzky–Golay filter; NDVI; Time-series data set; SPOT vegetation

1. Introduction

Time-series data for the Normalized Difference Vegetation Index (NDVI) derived from NOAA/AVHRR, SPOT/VEGETATION, TERRA, or AQUA/MODIS have proven to be appropriate for detecting long-term land-use/cover changes and for modeling terrestrial ecosystems on the global, continental, and regional scales, since NDVI carries valuable information regarding land-surface properties (IGBP, 1992; Justice et al., 1985; Myneni et al., 1997; Potter et al., 1993; Prince, 1991; Reed et al., 1994; Running & Nemani, 1988; Tucker & Sellers, 1986; Tucker et al., 1985 and others). Theoretically, NDVI, calculated from a normalized transform of the near-infrared (NIR) and red reflectance ratio, is an index of the absorptive and reflective characteristics of vegetation in the red and near-infrared portions of the electromagnetic spectrum. For this reason, changes in NDVI time-series indicate changes in vegetation conditions proportional to the absorption of photosynthetically active radiation (Sellers, 1985). However, there are nearly always disturbances in these time-series, caused by cloud contamination, atmospheric variability, and bi-directional effects. These disturbances greatly affect the monitoring of land cover and terrestrial ecosystems and show up as undesirable noise (Cihlar et al., 1997; Gutman, 1991). Although the most often-used NDVI data sets are 10-day Maximum Value Composite (MVC) products (Holben, 1986), such as the Pathfinder land data set, the GIMMS NDVI data set, and the SPOT VGT product, these still include a lot of such noise. For this reason, a number of methods for reducing noise and constructing high-quality NDVI time-series data sets for further analysis have been formulated, applied, and eval-

* Corresponding author. Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Ministry of Education of China Beijing 100875, China Tel.: +086-10-62207656; fax: +086-10-62208460.
E-mail address: chenjin@irs.bnu.edu.cn (J. Chen).

0034-4257/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2004.03.014

uated in the last two decades. These methods can be broadly grouped into three general types: (1) threshold-based methods, such as the best index slope extraction algorithm (BISE) (Viovy et al., 1992); (2) Fourier-based fitting methods (Cihlar, 1996; Roerink et al., 2000; Sellers et al., 1994); and (3) asymmetric function fitting methods such as the asymmetric Gaussian function fitting approach (Jonsson & Eklundh, 2002) and the weighted least-squares linear regression approach (Swets et al., 1999). Each abovementioned approach possesses its own advantages and has been successfully applied to NDVI time-series pre-processing for some applications. The BISE algorithm has been used to extract seasonal metrics of vegetation phenology (e.g. Reed et al., 1994), to classify vegetation or land cover types (e.g. Lovell & Graetz, 2001; Xiao et al., 2002) and to estimate gross primary productivity (GPP) and net primary productivity (NPP) (e.g. Ruimy et al., 1996). The Fourier-based fitting approach has been employed to derive terrestrial biophysical parameters (e.g. Sellers et al., 1994) and to evaluate NPP dynamics (e.g. Malmstrom et al., 1997). Asymmetric function fitting methods have been mainly used to extract seasonality information for phenological studies (Jonsson & Eklundh, 2002).

However, these methods also suffer several drawbacks that limit their use (Jonsson & Eklundh, 2002). For example, the BISE algorithm requires the determination of a sliding period and a threshold for acceptable percentage increase in NDVI for re-growth during a sliding period based on an empirical strategy that is usually subjective and depends on the skills and experience of the analyst. Therefore, like other threshold-based methods, the remaining noise after applying the BISE algorithm may make the extracted temporal information unreliable. Fourier-based fitting methods may be problematic when applied to irregular or asymmetric NDVI-data, since they depend critically on symmetric sine and cosine functions. In addition, they may generate spurious oscillations in the NDVI time series. Compared with the above methods, the asymmetric Gaussian function-fitting approach is more flexible and effective in obtaining a high-quality NDVI time-series. However, it may be difficult to identify a reasonable and consistent set of maxima and minima to which the local functions can be fitted, especially for noisy data or for data from areas where there is no clear seasonality. Additionally, the complexity of this approach makes it more time-consuming.

In light of the abovementioned drawbacks, this paper presents a simple but robust method based on the Savitzky–Golay filter, to more efficiently reduce contamination in the NDVI time-series that is caused primarily by cloud contamination and atmospheric variability. Our method was developed to make data approach the upper NDVI envelope and to portray patterns of NDVI change through an iteration process. The method was tested with a 10-day MVC SPOT VGT-S product generated by the VEGETATION Programme, which is developed jointly by France, the European Commission, Belgium, Italy and Sweden.

2. Methodology

Similarly to other strategies for reducing noise and constructing a high-quality NDVI time-series, our method is based on two assumptions: (1) that the NDVI data from a satellite sensor is primarily related to vegetation changes. As such, an NDVI time-series follows annual cycle of growth and decline; and that (2) clouds and poor atmospheric conditions usually depress NDVI values, requiring that sudden drops in NDVI, which are not compatible with the gradual process of vegetation change, be regarded as noise and removed. In line with these two assumptions, a new method based on the Savitzky–Golay filter was developed to make data approach the upper NDVI envelope and to best fit the NDVI variations during full vegetational season through an iteration process. The method can be applied to NDVI data sets sampled at different intervals, including daily data, 10-day, or monthly MVCs. In addition, there are no restrictions regarding the scaling of the NDVI and specific sensors. In the following, we first briefly introduce the Savitzky–Golay filter, and then describe the main steps for implementing the new method according to the flowchart shown in Fig. 1.

2.1. The Savitzky–Golay filter

Savitzky and Golay (1964) proposed a simplified least-squares-fit convolution for smoothing and computing derivatives of a set of consecutive values (a spectrum). The convolution can be understood as a weighted moving average filter with weighting given as a polynomial of a certain degree. The weight coefficients (referred to below as coefficients), when applied to a signal, perform a polynomial least-squares fit within the filter window. This polynomial is designed to preserve higher moments within the data and to reduce the bias introduced by the filter. This filter can be applied to any consecutive data when the points of the data are at a fixed and uniform interval along the chosen abscissa, and the curves formed by graphing the points must be continuous and more or less smooth. NDVI time-series clearly satisfy these conditions. The general equation of the simplified least-squares convolution for NDVI time-series smoothing can be given as follows:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N} \quad (1)$$

where Y is the original NDVI value, Y^* is the resultant NDVI value, C_i is the coefficient for the ith NDVI value of

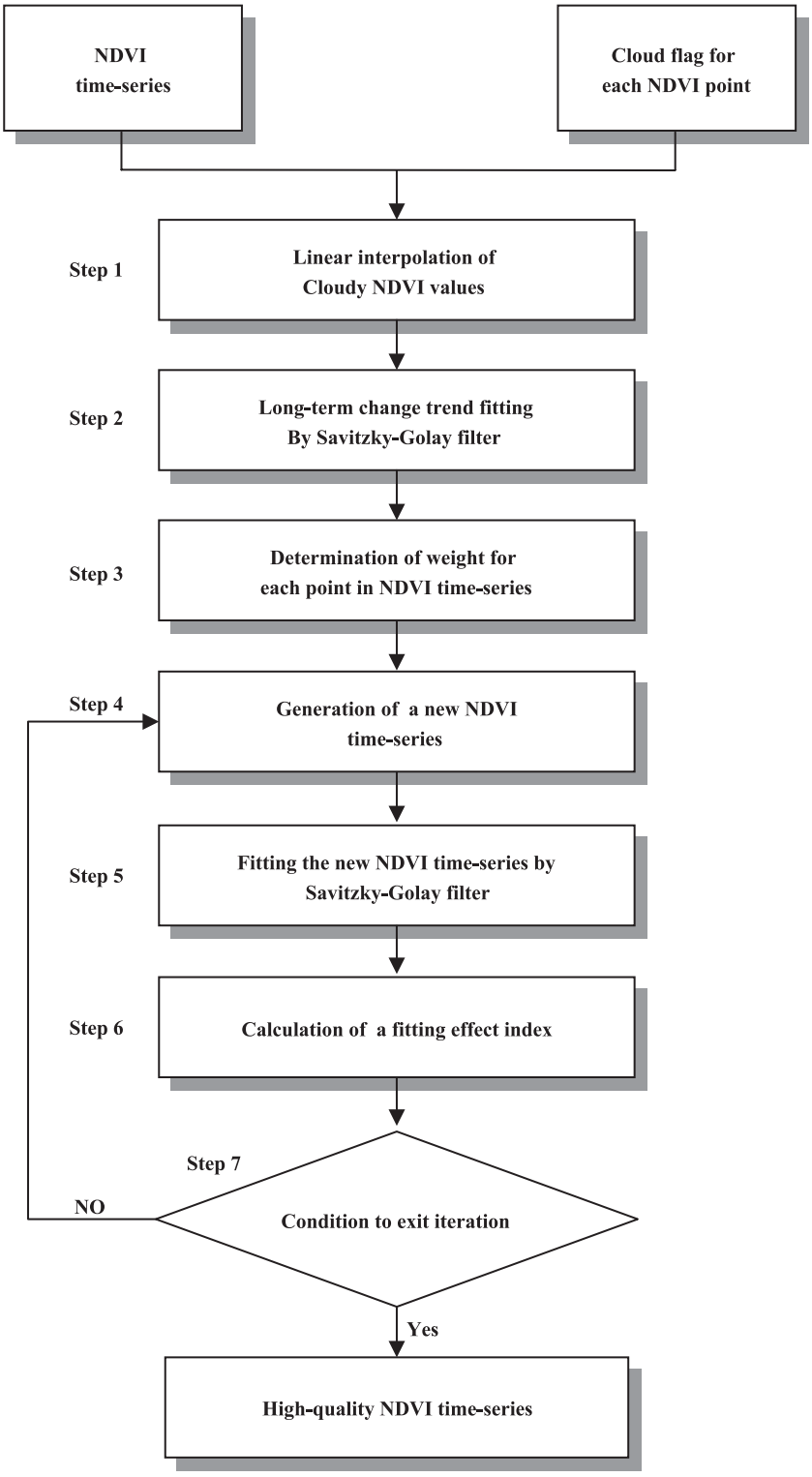


Fig. 1. Flowchart of the newly developed method.

the filter (smoothing window), and N is the number of convoluting integers and is equal to the smoothing window size ($2m+1$). The index j is the running index of the original ordinate data table. The smoothing array (filter size) consists of $2m+1$ points, where m is the half-width of the smoothing window. The coefficients of a Savitzky–Golay filter (C_i) can be obtained directly from Steinier et al. (1972) as a corrected version of Savitzky and Golay's work (1964), or calculated from the equations presented by Madden (1978).

When the filter is applied to NDVI time-series smoothing, there are two parameters that must be determined according to the NDVI observations. The first parameter is m, the half-width of the smoothing window. Usually, a larger value of m produces a smoother result at the expense of flattening sharp peaks. The second parameter is an integer (d) specifying the degree of the smoothing polynomial, which is typically set in a range from 2 to 4. A smaller value of d will produce a smoother result but may introduce bias; a higher value of d will reduce the filter bias, but may "over fit" the data and give a noisier result.

2.2. Implementation of the new method

2.2.1. Step 1: linear interpolation of cloudy NDVI values

Most NDVI time-series data sets such as the Pathfinder land data set or the SPOT VGT product have included the cloud flag band as ancillary data (Stowe et al., 1991); these data provide a valuable indicator of the cloud status of each data point in the time-series although these data do not include all cases in which NDVI data points were affected by cloud and poor atmospheric conditions. It is important to take advantage of such flag data to estimate the uncertainty of the NDVI value (Jonsson & Eklundh, 2002). In this study, cloud flag data were used to improve the NDVI time-series by linear interpolation of the cloudy NDVI values. Specifically, assuming that there is a NDVI time series of data points (t_i,N_i,F_i), $i=1, 2, 3\ldots n$, where t_i is the date, N_i is the NDVI value, and F_i is the cloud flag, if F_j of the jth point is identified as a cloudy point, then the N_j will be replaced by a linearly interpolated value using adjacent points that are not identified as cloudy points. In addition, points with a random NDVI increase greater than 0.4 during 20 days are also rejected and replaced by linearly interpolated values using the adjacent points, as such increases cannot be caused by natural vegetation changes. As a result, we obtain a new NDVI time series of data points (t_i,N_i^0), $i=1, 2, 3\ldots n$, where t_i is the date and N_i^0 is the new NDVI value after the linear interpolation. Fig. 2a shows examples of the linear interpolation of cloudy NDVI values according to the cloud flags, marked by circles. It can be seen that two points identified by cloud flags show sudden drops in the NDVI, while some other sudden drops in the NDVI are not indicated by cloud flags, which implies difficulties in categorizing cloud flags based on applying certain thresholds to visible and near infrared reflectance (VNIR) bands and thermal bands.

2.2.2. Step 2: long-term change trend fitting using the Savitzky–Golay filter

According to the assumptions mentioned above, a NDVI time-series should follow the gradual process of the annual vegetation cycle, so sudden falls in the NDVI time-series that are not compatible with the process can be regarded as noisy points affected by clouds or poor atmospheric conditions. Therefore, if we can obtain a long-term change trend curve representing the gradual process of annual vegetation cycle,

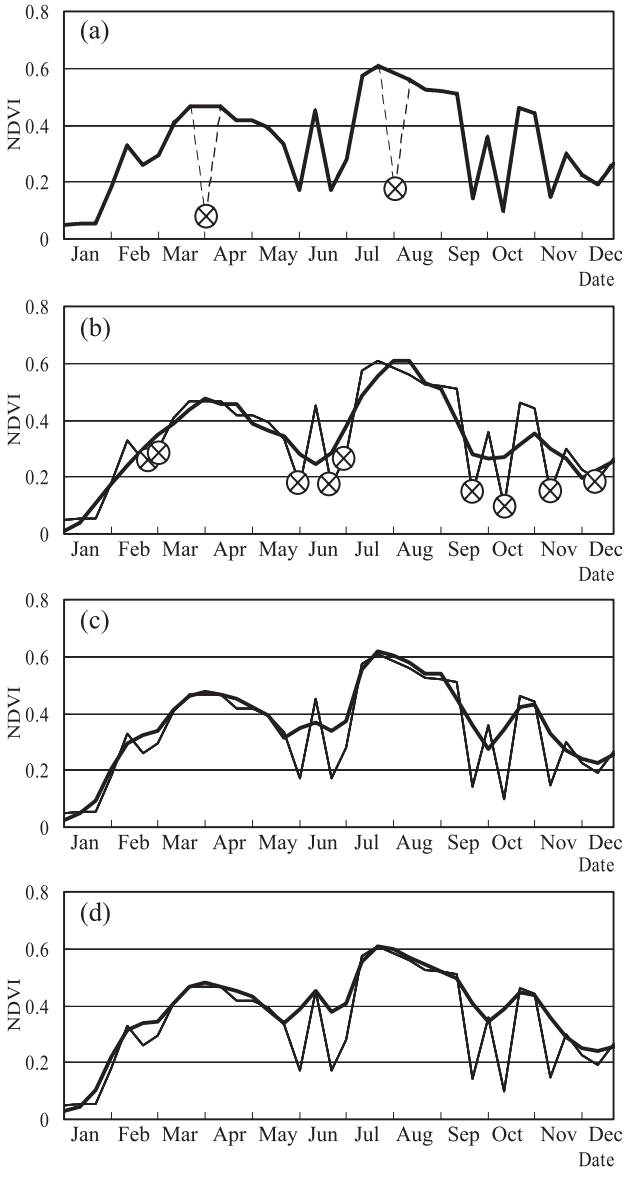


Fig. 2. An example showing the NDVI time-series in different steps of the newly developed method (NDVI data from No. 8 test pixel): (a) Original NDVI time-series and cloud flag points (circled in the figure). (b) Long-term change trend curve (thick solid line) fitted by the Savitzky–Golay filter. Noisy NDVI points are circled. (c) The first fitted NDVI time-series by the Savitzky–Golay filter. The first fitted NDVI time-series is plotted as a thick solid line, and the linear interpolated NDVI time-series (t_i,N_i^0) is plotted as a solid line. (d) The final NDVI time-series using the Savitzky–Golay filter (thick solid line).

it is helpful to identify these noisy data points and treat them as being less important in further fittings (Jonsson & Eklundh, 2002). To obtain a satisfactory long-term change trend curve, two criteria need to be considered: (1) the long-term change trend curve should follow the gradual process of annual vegetation cycle without too great a loss of temporal detail in the NDVI time-series; (2) most noisy points should be below the long-term change trend curve, since noise caused by clouds and poor atmospheric conditions is negatively biased. Based on these two criteria, the Savitzky–

Golay filter is used to smooth the NDVI variation and obtain the long-term change trend. From the characteristics of Savitzky–Golay filter, it can be found that setting too small a value of m (half-width of the smoothing window) may "over fit" the data points and cause difficulties in capturing the long-term change trend, while setting too large a value of m may neglect some important variations in the NDVI time-series. Therefore, middle values of m in the range of 4–7 can be considered as appropriate parameters for generating the long-term change trend curve. Consequently, there are 4×3 combinations of m and d (degree of polynomial) when d is typically set in a range from 2 to 4. Considering the importance of temporal detail in NDVI time-series for phenological studies, a combination of m and d corresponding to the best fit using the least-squares fitting method of all the combinations is automatically selected as the optimal set of parameters for generating the long-term change trend curve. According to our experimental results, shown below, this type of parameter selection can provide a good trade-off between preserving temporal detail in NDVI time-series and identifying noisy points. Fig. 2b shows a long-term change trend curve for NDVI time-series (t_i,N_i^0) obtained using the Savitzky–Golay filter. It is clear that the long-term change trend curve preserves temporal detail in NDVI time-series and all sudden drops are seen to fall below the long-term change trend curve.

2.2.3. Step 3: determination of weight for each point in NDVI time-series

After the NDVI time-series (t_i,N_i^0) is smoothed, a new time-series representing the NDVI long-term change trend is obtained as (t_i,N_i^{tr}). This new time-series is used to determine the weight of each point in the NDVI time-series by comparing it with the time-series (t_i,N_i^0). The weights of NDVI points will be used to calculate the fitting-effect index below such that the final NDVI time-series is the best description of NDVI variations during the full vegetation season and approach to the upper NDVI envelope. As shown in Fig. 2b, the points that are above or more and which approach the long-term change trend curve are more likely reflect the actual vegetation cycle and can be regarded as acceptable points, so we give them relatively higher weights. The local dropped points that are below the long-term change trend curve (circled points in Fig. 2b), on the other hand, are most likely to be the result of disturbance factors in the data rather than actual physical phenomena. Therefore, we give these locally dropped points relatively lower weights than the acceptable points. Based on the above idea, the weight (W_i) for each NDVI point can be calculated according to its distance from the long-term change trend curve as:

$$W_i = \begin{cases} 1 & \text{when } N_i^0 \geq N_i^{\mathrm{tr}} \\ 1 - d_i/d_{\max} & \text{when } N_i^0 < N_i^{\mathrm{tr}} \end{cases} \quad (2)$$

where $d_i = |N_i^0 - N_i^{\mathrm{tr}}|$ and $d_{\max}$ is the maximum of the absolute difference value of N_i^0 and N_i^{tr}.

2.2.4. Step 4: generation of a new NDVI time-series

In Fig. 2b, it can be seen that the time-series (t_i,N_i^{tr}) representing the NDVI long-term change trend displays larger NDVI values than the time-series (t_i,N_i^0) at the noisy NDVI points. The above suggests that if we can generate a new NDVI time-series by replacing the noisy NDVI points in the time-series (t_i,N_i^0) with the corresponding points in the time-series (t_i,N_i^{tr}), then refitting the new NDVI time-series will more closely approach the upper envelope of the original NDVI data. It is obvious that approaching the upper NDVI envelope is a gradual process and should be designed as an iteration process in the new method. As such, a new time-series (t_i,N_i^1) is generated by

$$N_i^1 = \begin{cases} N_i^0 & \text{when } N_i^0 \geq N_i^{\mathrm{tr}} \\ N_i^{\mathrm{tr}} & \text{when } N_i^0 < N_i^{\mathrm{tr}} \end{cases} \quad (3)$$

2.2.5. Step 5: fitting the new NDVI time-series using the Savitzky–Golay filter

Based on the new time-series (t_i,N_i^1), the Savitzky–Golay filter is used again to fit variations in the new time-series (t_i,N_i^1) rather than smoothing the time-series to obtain the long-term change trend. Therefore, a smaller value of m and a larger value for the degree (d) of the polynomial are set in the Savitzky–Golay filter to better fit the new time-series. From our experiments discussed below, m set to 4 and d set to 6 are recommended to fit the variations in the new time-series. As a result of this fitting, a new time-series (t_i,N_i^{k+1}) is generated, where $k=1$ for the first fitting. Fig. 2c shows the initially fitted curve (t_i,N_i^2) and time-series (t_i,N_i^0). It is clear that the initially fitted curve is closer to the upper NDVI envelope.

2.2.6. Step 6: calculation of a fitting-effect index

A fitting-effect index is defined to evaluate the degree to which the fitted NDVI values approach the more highly weighted NDVI points during one fitting process. The fitting-effect index (F_k) for a kth times fitting is calculated as:

$$F_k = \sum_{i=1}^{n} (|N_i^{k+1} - N_i^0| \times W_i) \quad (4)$$

where N_i^{k+1} is the ith NDVI value of the kth fitted time-series, N_i^0 is the ith NDVI value of the original NDVI time-series after linear interpolation of cloudy NDVI values, and W_i is the weight of ith NDVI point determined in Step 3. From this definition, it can be seen that with decreasing F_k, the fitted curve more closely approaches the higher-weighted NDVI points (the upper envelope of the original NDVI data).

After fitting the NDVI time-series, a fitting-effect index (F_k) can be obtained. If F_k does not achieve its minimum, a new fitting process is iterated from Step 4 to Step 6. Here, it

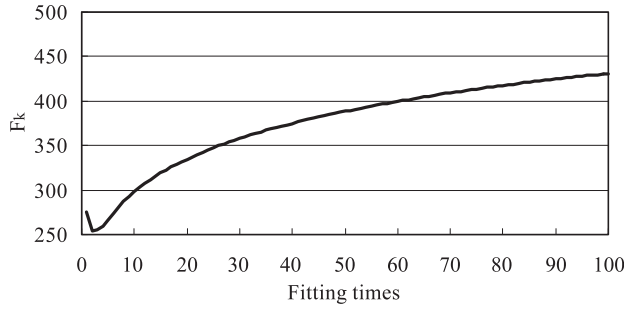


Fig. 3. The fitting-effect index (F_k) change pattern with increasing fitting times.

should be noted that a new time-series generated by the following formula will be fitted for the new iteration.

$$N_i^{\text{new}} = \begin{cases} N_i^0 & \text{when } N_i^0 \geq N_i^{k+1} \\ N_i^{k+1} & \text{when } N_i^0 < N_i^{k+1} \end{cases} \quad (5)$$

2.2.7. Step 7: conditions for exiting iteration

Fig. 3 shows the F_k change pattern with increasing fitting times. It can be seen that F_k decreases very quickly in the first few fittings, and that after it reaches its minimum, F_k begins to increase. Although the figure shows only the first 100 fittings, our experimental results indicate that this increasing trend continues. This pattern of change suggests that the exit condition of the iteration can be defined as:

$$F_{k-1} \geq F_k \leq F_{k+1} \quad (6)$$

where F_{k-1} is the fitting-effect index for the $(k-1)$th times fitting, F_k is for the kth times fitting, and F_{k+1} is for the $(k+1)$th times fitting. This condition ensures that the fitting-effect index reaches its minimum. After the above iterating process satisfies this condition, it will terminate and we can obtain the final NDVI time-series, as shown in Fig. 2d. The figure demonstrates a good fit to the upper NDVI envelope.

Here, it should be noted that the edge effect of the Savitzky–Golay filter has an adverse effect on the fitting quality of the first four points and the last four points, because such points can not constitute a complete filter window. To address the problem, we can assume that the NDVI series is cyclic (Sellers et al., 1994), hence the neighbors for the last four points are the first four points, and the neighbors for the first four points are the last four points. Naturally, if a longer NDVI time series than fitting NDVI time series is available, the edge effects of Savitzky–Golay filter can be avoided.

3. Results

As test data for evaluating the new method's performance, SPOT VGT-S product (1 × 1 km, 10-day MVC Product) for Southeast Asia (68–147°E, 5–55°N) during the period from January 2000 to December 2000 was used to construct a high-quality NDVI time-series data set. Considering the edge effect seen in the Savitzky–Golay filter, the same SPOT VGT-S product of Nov., Dec. 1999 and Jan., Feb. 2001 was employed to constitute a complete filter window for the first four points and the last four points of the NDVI time series. During the data preparation stage, NDVI values from the NDV band were recalculated to the range of −1 to 1 according to the formula (VEGETATION Programme, 1998): NDVI = 0.004 × DN − 0.1. The cloud flag for each NDVI data point was extracted from Bit NR 0 and Bit NR 1 of the Status Map (SM) band. As is well known, the SPOT VGT-S product has been preprocessed using a consistent processing algorithm including geometric, radiometric, and atmospheric corrections. The atmospheric correction is based on SMAC to correct water vapor, ozone, and aerosol effects (Rahman & Dedieu, 1994; VEGETATION Programme, 1998). As such, further pre-processing of the SPOT VGT-S product was not implemented.

To determine the parameters of the new method and compare the new method with existing methods, 438 test pixels were selected from the SPOT VGT-S product using the randomly stratified sampling method, in which the China vegetation map (Wu, 1980) was used for stratification. Fig. 4 shows the vegetation type of China (Wu, 1980) and the distribution of the 438 test pixels. For each of the selected 438 test pixels, NDVI time-series during the period from Nov. 1999 to Feb. 2001 were plotted, respectively, and then the possibly noisy points of each NDVI time-series were identified by visual interpretation. These interpretation results were used as the reference data for assessing the new method's performance.

3.1. Determination of m and d for the Savitzky–Golay filter in the fitting iteration

To obtain the final NDVI time-series that best approached the upper envelope of original NDVI data and reflecting the NDVI change pattern, it is necessary to determine optimal m and d of the Savitzky–Golay filter in the fitting iteration process. These two parameters were determined through a series of experiments using data from the 438 test pixels. Table 1 shows the fitting-effect index (F_k) of the No. 8 test pixel using different m and d combinations. It is seen that the fitting-effect decreases quickly (F_k increases), with m increasing when the degree (d) of polynomial is fixed. It can also be seen that increasing d can improve the fitting effect when m is fixed, but the extent of this improvement is not so evident as the changes in m. From the results of experiments similarly carried out for 438 test pixels, it was found that smaller values of m and larger values of d result in a better fitting effect; therefore the (m,d) combination of (4,6) was determined to be optimal for the Savitzky–Golay filter in the fitting iteration process because it can provide the best-fitting effect in most cases.

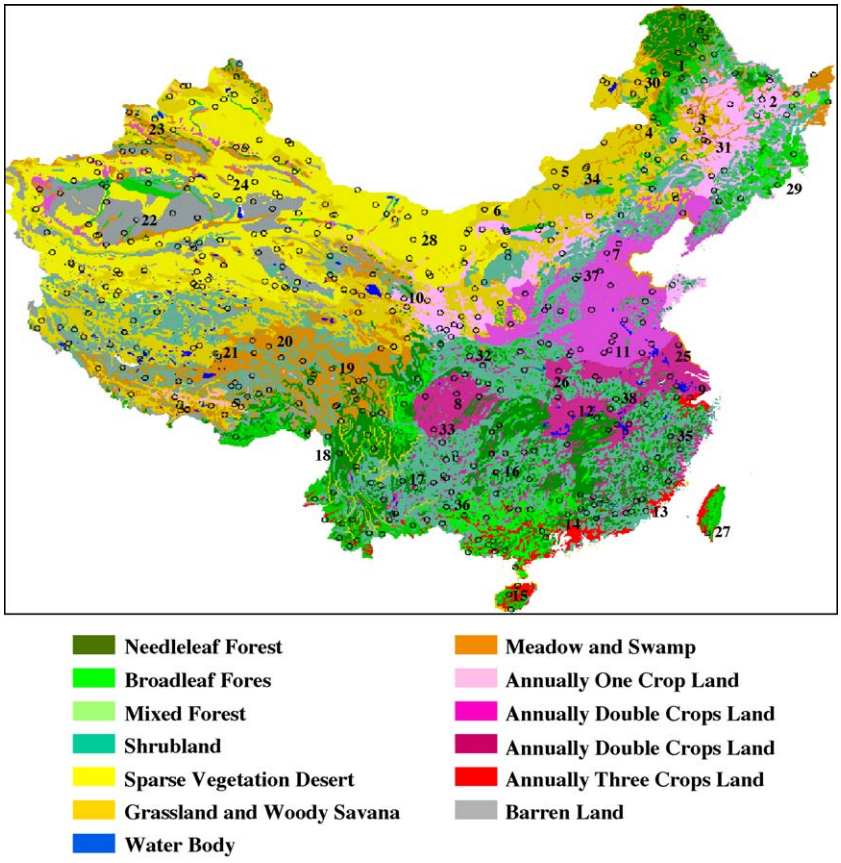


Fig. 4. Vegetation type in China and distribution of 438 test pixels.

3.2. Comparison of new and existing methods

To test the performance of the new method, we carried out a comparison among three methods using 438 test pixels: (1) the new method developed in the study; (2) a BISE algorithm (Viovy et al., 1992) and (3) a Fourier-based fitting method (Sellers et al., 1994). The latter two methods are widely used for reducing noise in the NDVI time-series. In contrast with the new method, the BISE algorithm requires determination of a sliding period and a threshold of acceptable percentage increase in NDVI for re-growth during a sliding period according to an empirical strategy. It then reconstructs an NDVI time-series by scanning the time period, ignoring low values, and selecting high values based on the threshold in a sliding period. On the other hand, a Fourier-based fitting method is similar to some extent to the new method. It is also a weighted fitting method with the exception of selecting symmetric sine and cosine as fitting function. It constructs a new NDVI time series by means of a robust least-squares optimizing method using the first three harmonics of the Fourier series and taking their weights into account. In this comparison, for the new method, the (m,d) combination of (4,6) was used in the Savitzky–Golay filter. For the BISE algorithm, the sliding period was set at 1 month and the threshold of acceptable percentage increase was determined by a manual trial-and-error procedure to obtain the best result for each pixel.

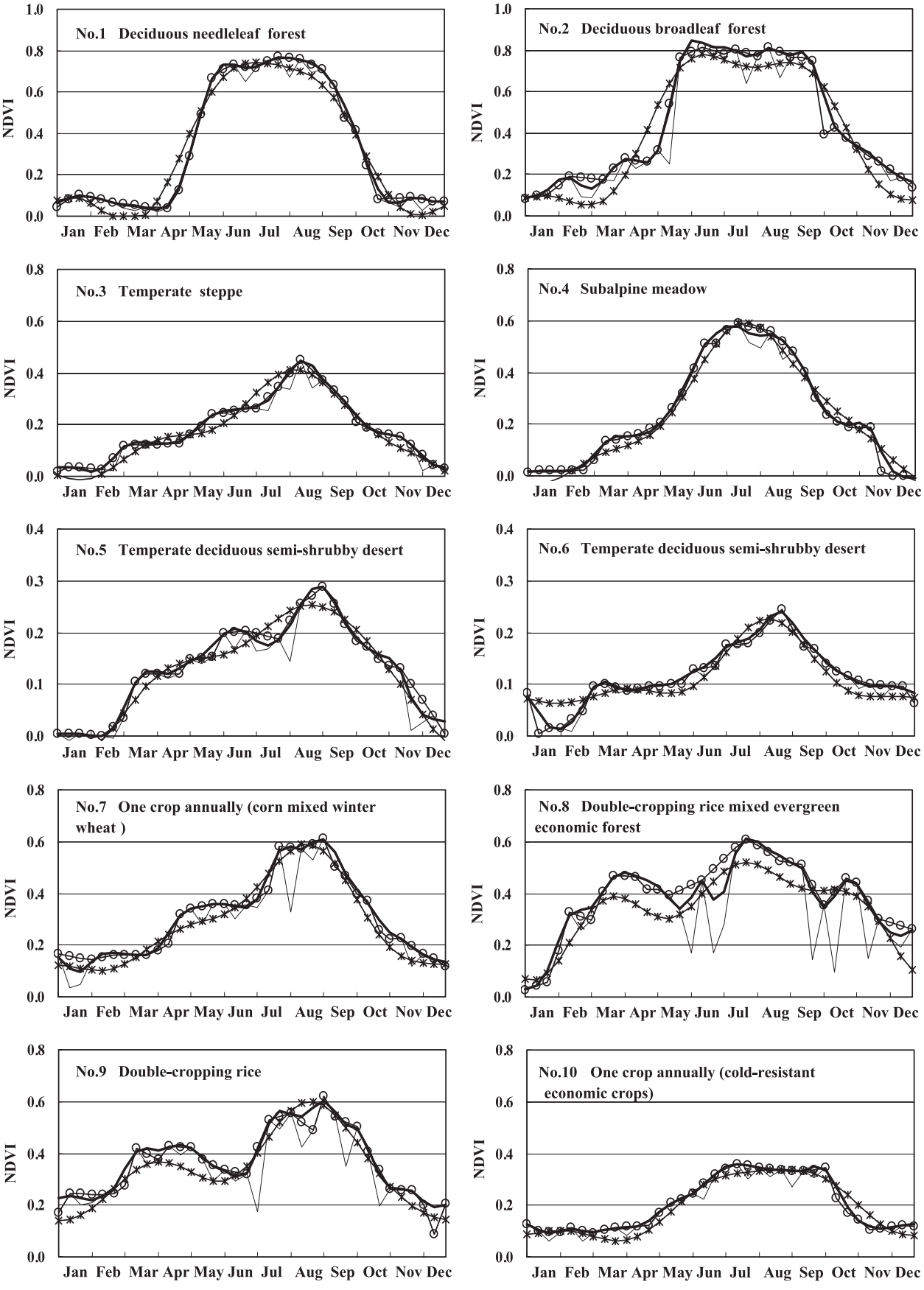
Table 1
Fitting effect assessment of m and d combinations in fitting iteration

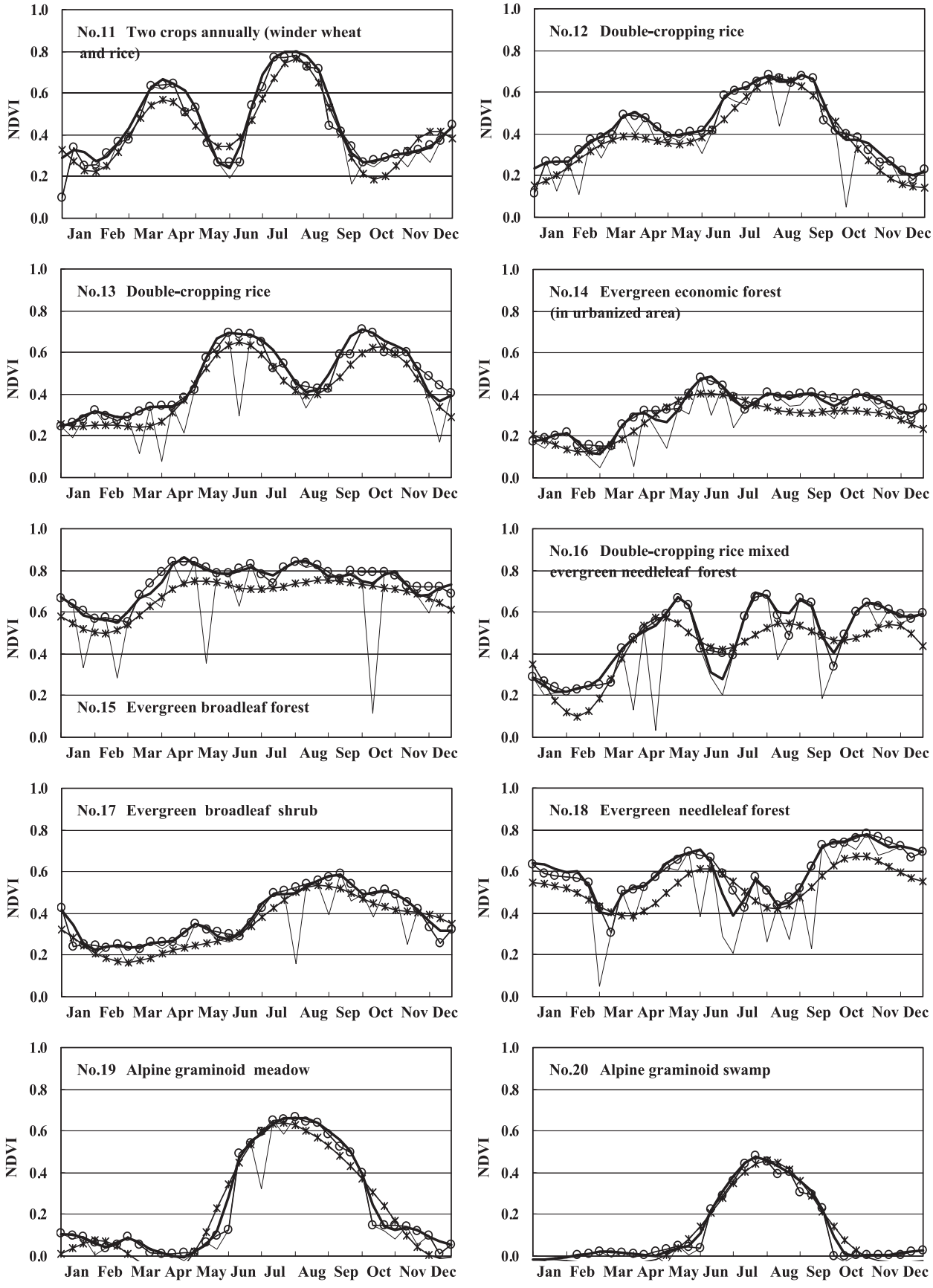
d / m	2	3	4	5	6
2	278.6	278.6			
3	353.5	353.5	255.4	255.4	
4	383.4	383.4	316.6	316.6	243.4
5	404.5	404.5	348.6	348.6	303.7
6	414.9	414.9	376.7	376.7	332.4
7	424.8	424.8	406	406	363.8
8	451.4	451.4	408	408	393.1

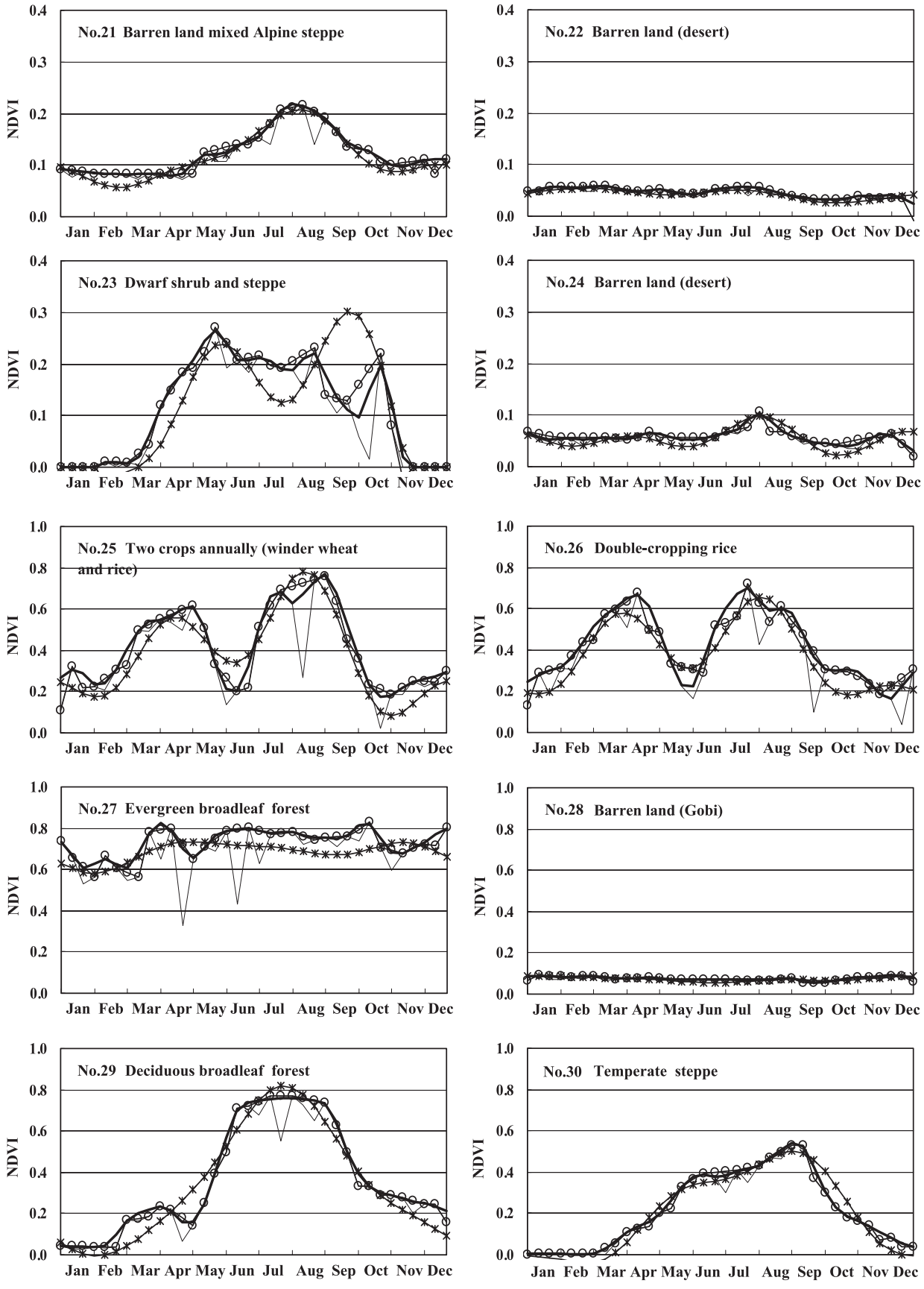
The figures in the table is fitting-effect index (F_k).

Fig. 5 gives the final NDVI time-series obtained using the three methods. Due to limitations on space, only the results of 38 test pixels are shown in the figure. The location of 38 test pixels can be seen in Fig. 4. The general pattern of changes of NDVI time series is apparent from these methods, all three of which identify and correct most noisy points. The results suggest that the three methods are effective for constructing high-quality NDVI time-series data sets.

Of the three methods, the Fourier-based fitting method obtained the smoothest fitted curve, but shows a large displacement away from the original NDVI values. Such


No.1 Deciduous needleleaf forest
No.2 Deciduous broadleaf forest
No.3 Temperate steppe
No.4 Subalpine meadow
No.5 Temperate deciduous semi-shrubby desert
No.6 Temperate deciduous semi-shrubby desert
No.7 One crop annually (corn mixed winter wheat)
No.8 Double-cropping rice mixed evergreen economic forest
No.9 Double-cropping rice
No.10 One crop annually (cold-resistant economic crops)
NDVI
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec


No.11 Two crops annually (winder wheat and rice)
No.12 Double-cropping rice
No.13 Double-cropping rice
No.14 Evergreen economic forest (in urbanized area)
No.15 Evergreen broadleaf forest
No.16 Double-cropping rice mixed evergreen needleleaf forest
No.17 Evergreen broadleaf shrub
No.18 Evergreen needleleaf forest
No.19 Alpine graminoid meadow
No.20 Alpine graminoid swamp
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec


No.21 Barren land mixed Alpine steppe
No.22 Barren land (desert)
No.23 Dwarf shrub and steppe
No.24 Barren land (desert)
No.25 Two crops annually (winder wheat and rice)
No.26 Double-cropping rice
No.27 Evergreen broadleaf forest
No.28 Barren land (Gobi)
No.29 Deciduous broadleaf forest
No.30 Temperate steppe
NDVI
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

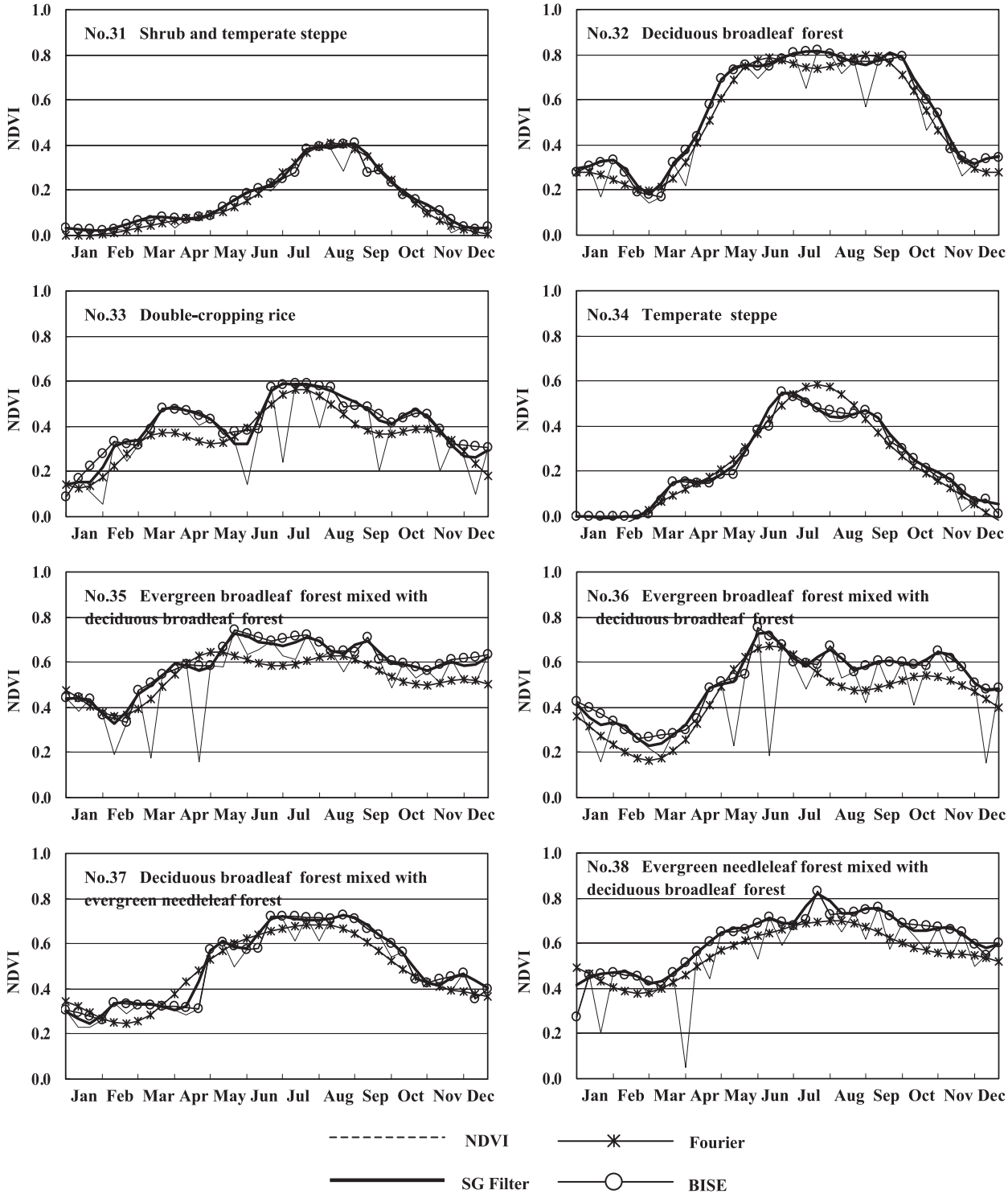


Fig. 5. The final NDVI time-series generated using the new method, the BISE algorithm and the Fourier-based fitting method (dashed line is the linearly interpolated NDVI time-series (t_i,N_i^0); the thick solid line is the smoothed NDVI time-series using the new method; the solid line with circles is the smoothed NDVI time-series using the BISE algorithm; the solid line with asterisks is the smoothed NDVI time-series using the Fourier-based fitting method).

displacements are more obvious when Fourier-based fitting method was used with irregular or asymmetric NDVI time series such as in No. 8, No. 9, No. 16, No. 18, No. 23, No. 27, No. 35 and No. 36. These displacements can be explained by the fact that the method depends critically on symmetric sine and cosine functions. In addition, a Fourier-based fitting method also generates spurious oscillations in the NDVI time series. This can be observed in No. 23. In contrast with the Fourier-based fitting method, the new method and BISE algorithm obtained almost identical

results for most of these test pixels. The final NDVI time series comes close to the upper NDVI envelope, and most of the noisy points in the NDVI time series were successfully identified and corrected. However, it is notable that the threshold of the BISE algorithm for each test pixel needed to be set using a manual trial-and-error procedure to obtain the best result, and the parameters of the new method were only set as constant combinations (4,6). In short, the new method is able to reconstruct a comparably high-quality NDVI time-series as the BISE algorithm, but without the difficulty of finding an optimal threshold such as that in the BISE algorithm. Here, in addition, as a result of this manual trial-and-error procedure, the optimal threshold for each test pixel displays several differences from pixel to pixel, although most of the optimal thresholds were in the range of 0.1–0.4. These results suggest that the BISE algorithm needs some improvement, such as the development of an automatic or semi-automatic method for effectively determining the threshold for each pixel in NDVI time-series data sets.

To test the applicability of the new method to large images, we applied the new method to SPOT VGT-S product for Southeast Asia during the period from November 1999 to February 2001, which includes 8849 × 5601 pixels and 48 layers. Using a DELL Dimension 8100 desktop computer system (4 CPU, 1800 MHz, 1 GB RAM), the time for processing the product was 22 h using the new method. For comparison with the new method, we also applied the BISE algorithm and the Fourier-based fitting method to the same product. The processing times for the two methods were 17 and 29 h, respectively, showing that the computing efficiency of the new method is higher than the Fourier-based fitting approach but lower than the BISE algorithm. For such larger images, we think that the computation time cost of the new method is acceptable. There is clear potential to greatly shorten it by using more powerful computer systems and improving the program used for the new method. Fig. 6 illustrates the results of implementing the new method on an NDVI image in July 2000. It should be noted that there are large numbers of low NDVI values (depicted in yellow) in China's Sichuan Basin and in India due to atmospheric perturbations during this 10-day period. The new method effectively replaces these low values with higher NDVI values that are more typical of the cropland found in these areas.

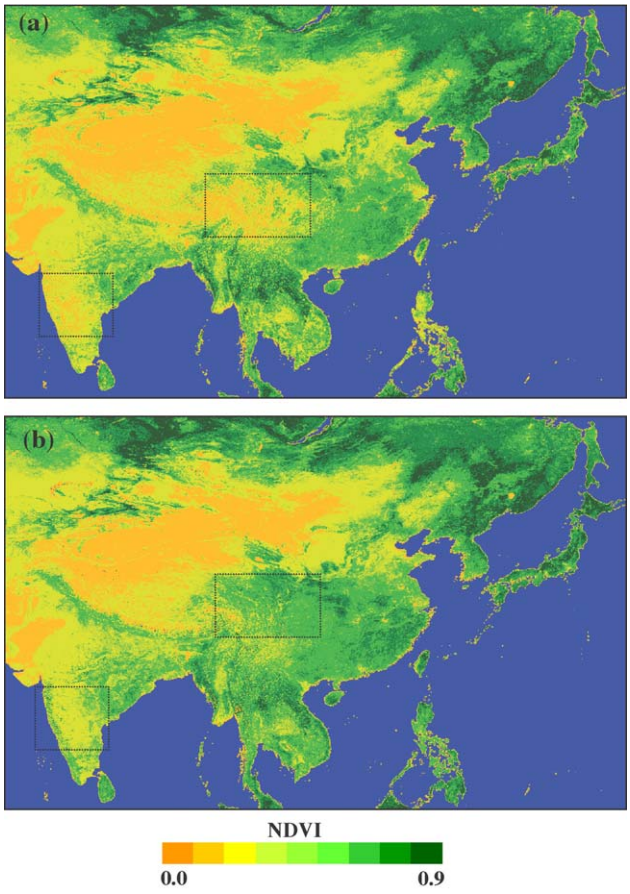


Fig. 6. NDVI data from July 2000 (a) and the results of implementing the new method (b).

4. Conclusion

The noise in NDVI time-series data, caused primarily by cloud contamination and atmospheric variability, is an important problem faced by global environmental change research when these NDVI data sets are used as input. However, existing methods for reducing noise in NDVI time-series are not sufficiently flexible or effective. Based on the assumptions that an NDVI time-series follows an annual cycle of growth and decline in vegetation, and that clouds or poor atmospheric conditions usually depress NDVI values, we have described herein a simple but robust method, based on the Savitzky–Golay filter, to smooth out the noise present in NDVI time-series. The method was developed to make data approach the upper NDVI envelope and to reflect the NDVI pattern of change through an iteration process. By applying the newly developed method to a 10-day MVC SPOT VGT-S product and by comparing the results with those from the BISE algorithm and Fourier-based fitting method, we have found that the new method shows the following four advantages over existing methods: (1) it takes advantage of ancillary data in the form of cloud flags; (2) it can reconstruct high-quality NDVI time-series by setting only two parameters in the Savitzky–Golay filter: the half-width of the smoothing window (m) and the degree of the smoothing polynomial (d). The results of our series of experiments suggest the optimal parameters to be (4,6) for the new method. The recommended parameters are robust in most cases. (3) It is very simple in theory and easy to implement because commercial software such as MATLAB and IDL include the Savitzky–Golay filter in their function library; and (4) it can be applied to NDVI data sets sampled

at different intervals such as daily data, 10-day, or monthly MVCs. In addition, there are no restrictions on the scaling of the NDVI and the specific sensor. For these reasons, we anticipate that the new method can be applied to reconstructing high-quality NDVI time-series data sets for use in global environmental change research, such as extracting seasonal metrics of vegetation phenology, land cover classification and change detection, deriving terrestrial biophysical parameters and modeling terrestrial ecosystems.

Acknowledgements

This research was partially supported by the Watershed Environments and Management Research Project at the National Institute for Environmental Studies, Japan, the project of Ecosystem Restoration Mechanism and Optimized Eco-productive Paradigm of the Grassland and Farming-pastoral Zone of North China (G2000018600) and the Outstanding Overseas Chinese Scholars Fund of Chinese Academy of Sciences. The authors would like to thank two anonymous reviewers for their insightful comments.

References

Cihlar, J. (1996). Identification of contaminated pixels in AVHRR composite images for studies of land biosphere. *Remote Sensing of Environment*, *56*, 149–153.

Cihlar, J., Ly, H., Li, Z. Q., Chen, J., Pokrant, H., & Huang, F. T. (1997). Multitemporal, multichannel AVHRR data sets for land biosphere studies—artifacts and corrections. *Remote Sensing of Environment*, *60*, 35–57.

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.

IGBP (1992). J. R. G. Townshend (Ed.), *Improved global data for land applications. IGBP Global Change Report, vol. 20*. Stockholm, Sweden: International Geosphere–Biosphere Programme.

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

Justice, C. O., Townshend, J. R. G., Holben, B. N., & Tucker, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, *6*(8), 1271–1318.

Lovell, J. L., & Graetz, R. D. (2001). Filtering pathfinder AVHRR land NDVI data for Australia. *International Journal of Remote Sensing*, *22*(13), 2649–2654.

Madden, H. (1978). Comments on the Savitzky–Golay convolution method for least-squares fit smoothing and differentiation of digital data. *Analytical Chemistry*, *50*(9), 1383–1386.

Malmstrom, C. M., Thompson, M. V., Juday, G. P., Los, S. O., Randerson, J. T., & Field, C. B. (1997). Interannual variation in global-scale net primary production: Testing model estimates. *Global Biogeochemical Cycles*, *11*(3), 367–392.

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., & Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, *386*(6626), 695–702.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., & Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles*, *7*(4), 811–841.

Prince, S. D. (1991). A model of regional primary production for use with coarse-resolution satellite data. *International Journal of Remote Sensing*, *12*(6), 1313–1330.

Rahman, H., & Dedieu, G. (1994). SMAC—a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing*, *15*(1), 123–143.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, *5*(5), 703–714.

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

Ruimy, A., Dedieu, G., & Saugier, B. (1996). TURC: A diagnostic model of continental gross primary productivity and net primary productivity. *Global Biogeochemical Cycles*, *10*(2), 269–285.

Running, S. W., & Nemani, R. R. (1988). Relating seasonal patterns of the AVHRR vegetation index to simulated photosynthesis and transpiration of forests in different climates. *Remote Sensing of Environment*, *24*, 347–367.

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

Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, *6*, 1335–1372.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich, D. A., & Randall, D. A. (1994). A global 1_1 NDVI data set for climate studies: Part II. The generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing*, *15*(17), 3519–3545.

Steinier, J., Termonia, Y., & Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least squares procedure. *Analytical Chemistry*, *44*(11), 1906–1909.

Stowe, L. L., McClain, E. P., Carey, R., Pellegrino, P., Gutman, G., Davis, P., Long, C., & Hart, S. (1991). Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Advances in Space Research*, *3*, 51–54.

Swets, D. L., Reed, B. C., Rowland, J. R., & Marko, S. E. (1999). A weighted least-squares approach to temporal smoothing of NDVI. 1999 ASPRS Annual Conference, From Image to Information, Portland, Oregon, May 17–21, 1999, Proceedings: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing.

Tucker, C. J., & Sellers, P. J. (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing*, *7*(11), 1395–1416.

Tucker, C. J., Townshend, J. R. G., & Goff, T. E. (1985). African land cover classification using satellite data. *Science*, *227*(4685), 369–375.

VEGETATION Programme, 1998. SPOT Vegetation User Guide, http://www.spotimage.fr/data/images/vege/vegetat/book_1/e_frame.htm.

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*(8), 1585–1590.

Wu, Z. (Ed.) (1980). *(Editor in chief) Vegetation of China*. Beijing: Science Press, in Chinese.

Xiao, X. M., Boles, S., Liu, J. Y., Zhuang, D. F., & Liu, M. L. (2002). Characterization of forest types in Northeastern China, using multitemporal SPOT-4 VEGETATION sensor data. *Remote Sensing of Environment*, *82*, 335–348.