Improved algorithm for reconstructing vegetation index image time series based on Fourier Harmonic Analysis

ZHANG Xia[1], LI Ru[2], YUE Yuemin[3], LIU Bo[1], LIU Haixia[1]

1. *Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing* 100101, *China*;
2. *Institute of Space and Earth Information Science, Chinese University of Hong Kong, Hong Kong, China*;
3. *Institute of Subtropical Agriculture, Chinese Academy of Sciences, Hunan Changsha* 410125, *China*

Abstract: This paper, based on Fourier-transform-based Harmonic Analysis Algorithm, proposes an improved algorithm to overcome the drawback of artificially setting key parameters and to reconstruct high-quality time-series data. Firstly, outlier detection algorithm, instead of threshold setting method, is used to find unreasonable data points before curve fitting. Then, regarding the inherent phenological regulation for each land cover, Numbers of Frequency (NOF) is calculated pixel by pixel by polynomial fitting, which is more reasonable than setting a unique global NOF for the whole scene which contains complex land cover types. Fitting-effect Index, instead of manually setting one fitting tolerance, is employed to decide automatically when to terminate the iteration. The improved method is validated by the MODIS_EVI time series of Huabei plain of 2003. The widely used Harmonic Analysis of Time Series (HANTS) is chosen as a comparison. The result shows that both of the two methods can reflect the phenological regulations of land covers, the reconstructed EVI temporal profile of single-season cropland (e.g. cotton) takes on one peak pattern, and those of double-season cropland and inland water take on double peak pattern and low-steady curve. But the improved method performs better in tracking the change tendency of the original curve. Moreover, the peak time and peak value of croplands are mostly consistent with the original curve, which will be useful for VI-based crop yield prediction.

Key words: vegetation index image time series, filter, Fourier Harmonic Analysis, outlier detection, fitting-effect Index

CLC number: TP751.1 **Document code:** A

Citation format: Zhang X, Li R, Yue Y M, Liu B and Liu H X. 2010. Improved algorithm for reconstructing vegetation index image time series based on Fourier Harmonic Analysis. *Journal of Remote Sensing*. **14**(3): 437—447

1 INTRODUCTION

As one useful indicator of vegetation greenness and intensity of photosynthesis, vegetation index (VI) could not characterize land cover types directly. But VI time series data can represent certain phenological regulations, thereby enabling it to monitor regional even global phenological dynamics (Moody & Johnson, 2001; Zhang *et al*., 2004). Up to now, VI time series have been used successfully to improve the accuracy of land cover classification (Bartalev *et al*., 2003; Loveland *et al*., 2000) and to study effects of climatic change (Roerink *et al*., 2003). However, VI time series still suffer from residual noise of cloud contamination, atmospheric variability and so on; therefore its further applications are limited.

Fourier-transform-based Harmonic Analysis Algorithm is one of the most widely used methods for VI time series data reconstruction. The temporal VI curve of one pixel is transformed as a sum of a series of harmonic curves. Each harmonic curve is determined by its unique amplitude and phase. When they are assembled together by a certain regulation, a smooth and complicated curve can be reconstructed. The typical algorithms of Fourier-transform-based harmonic analysis include Sellers algorithm (Sellers *et al*., 1996) and HANTS (Huete *et al*., 2002; Immerzeel *et al*., 2005). Both of them have the advantage that they can reveal the inherent phenological regulation implied by VI time series data. However, this kind of approach still has some drawbacks (Li *et al*., 2009): (1) Some key parameters (e.g. NOF) need to be determined by manual trials and experience, there is no definite rule to follow. It is easy to introduce new subjective error and interference in the original data. (2) A unique set of parameters are used for the whole scene that may contain complex land cover types, which is unreasonable for the real situation. (3) For the weight criterion, Sellers (1996) used distance-based rule. Lin *el al.* (2006) in their study revealed that this rule was suitable for tropical broadleaf forest but not helpful for VI time series composed by short interval data. HANTS has its own weight criterion which integrated in its software package; its criterion cannot be modified based on different application situations, thus this would influence its

Received: 2009-04-02; **Accepted**: 2009-07-17

Foundation: National Natural Science Foundation of China (No. 40971205), National Key Technology R&D Program (No. 2007BAH15B01), Open foundation of State Key Lab of Remote Sensing Science (No. 03Q00300649).

First author biography: ZHANG Xia (1972—), female, Associate Professor. her main research interests include image-based atmospheric correction, and hyperspectral image classification, and biochemical and biophysical parameter mapping by using hyperspectral data. E-mail address: zx@irsa.ac.cn

application effects.

In this paper, our study is based on Sellers algorithm. The main improvements include outlier detection, automatic estimation for NOF and fitting-effect Index to automatically terminate iteration. Some new parameters and strategies are introduced to avoid artificial influences and over-interference in the data. Finally a more objective VI time series reconstruction result can be obtained.

2 IMPROVED ALGORITHM AND ITS IMPLEMENT

2.1 Rules for image time series reconstruction

To ensure that the improved algorithm reveal phenological regulation accurately, four rules are proposed for reconstructing the VI time series:

(1) Reduce the disturbances to the original data as few as possible. Because reconstruction algorithms usually disturb original data a lot; the more the disturbances are, the lower reliability the result is.

(2) Use auxiliary data as little as possible. Although auxiliary data can provide some diagnostic information, for example, aid in distinguishing cloud-contaminated pixel, it can also introduce uncertainty to some extent to data reconstruction because of error of the auxiliary data itself.

(3) Take full consideration of the physical meanings of the data. VI time series imply specific phenological regulations of different land covers, so it will lead to a reasonable result when full consideration is given to the physical meanings of the data during data processing.

(4) Keep a balance between effects and efficiency. Automatically choosing NOF on pixel base may be time-consuming in spite of considering complex land covers. The employment of Fitting-effect index may improve the processing efficiency.

2.2 Implement of the improved algorithm

2.2.1 Outlier detection

Due to residual noise, it is hard to identify the real value and outlier. Most traditional methods employ setting a threshold to find outliers, but they could not work well all the time. Therefore it is necessary to analyze the inherent regulations of data to identify the outliers before data reconstruction. The improved method introduces outlier detection algorithm and identify if there are outliers in the data set before further process. This new strategy can also avoid artificial error and new uncertainty caused by distance-based weight criterion. In this study, we utilize Grubbs Test (Baksalary & Puntanen, 1990) to detect outliers. The theoretical basis is described as follows.

Suppose a set residue of repeated observations is sorted by its values: $V(1)\leq V(2)\leq\cdots\leq V(n)$. Grubbs statistics could be calculated by Eq. (1) and Eq. (2):

$$g_n=\frac{V_{(n)}-\overline{V}}{\sigma} \tag{1}$$

$$g_n'=\frac{\overline{V}-V_{(1)}}{\sigma} \tag{2}$$

where, $\overline{V}=\frac{1}{n}\sum_{i=1}^{n}V_{(i)}$, σ is standard deviation, g_n is a key value to judge if the maximum was an outlier, g_n' is used to judge whether the minimum is an outlier. g_n and g_n' have the same probability distribution. $g_0(n,\alpha)$ shows the critical value under the significant level α, which can be obtained from the look-up table.

When $|g_n|\geqslant g_0(n,\alpha)$ or $|g_n'|\geqslant g_0(n,\alpha)$, we can regard the corresponding observation value of the residue includes gross error and the maximum (or minimum) is an outlier. The significant level α is 0.05 or 0.01 generally.

Because the input data set of Grubbs Test is required to obey normal distribution, in order to avoid large programming consuming, we employ the absolute value of the residue and then put a log-transformation so that the data set can adapt to the test algorithm approximately.

2.2.2 Automatically estimation for NOF

One significant advantage of Fourier-transform-based harmonic analysis is to consider the inherent phenological regulation implied by VI time series. This advantage is represented by choosing suitable number of frequency (NOF). Traditional methods apply a unique global NOF on the whole scene. It implies an assumption that the phenological regulation on the whole scene was the same. However, this assumption is not always correct especially for a large scale research. Applying a unique global NOF on the whole scene is unreasonable; it could disturb the accuracy and reliability of fitting curve.

To solve this problem, our improved algorithm has a prepositive estimation for NOF. In this process, NOF of each time point has been obtained. NOF is the number of harmonic frequency used to stimulate the real time series curve, and it is equal to the number of peaks of the real curve ideally (here zero NOF is excluded). Given the noises in time series, taking the number of local maxima as NOF would not be reliable. However, we could use the number of the peaks of pseudo-fitting curve as NOF. The so called “pseudo-fitting curve” is the curve obtained by simple fitting algorithm and may not be consistent with the real curve shape, but these curves have the same number of local peaks (NOF) approximately. NOF could be estimated by the following steps:

(1) After data availability judging, put the first element of the series append to the last element, so we generate a new series that has (n+1) elements to avoid peak lost at head and tail of the series.

(2) Perform three-order polynomial fitting on the new series obtained after step (1) to generate a new curve. According to past experience, three-order polynomial is enough. If the order is smaller than three, peaks would be lost; while the order is larger than 3, it would be more time-consuming and generate some artificial peaks.

(3) Get the number of maxima (peaks), namely NOF, by calculating second derivative of the new curve.

Of course, we can also obtain NOF by filtering then fitting.

2.2.3 Fitting-effect Index

HANTS sets fitting tolerance obtained by experience and experiment to determine when to terminate the iteration, so does Sellers algorithm. To avoid importing objective errors and to automatically stop the iteration, the improved algorithm introduces a new parameter, Fitting-effect Index (FEI) (Gu, 2003; Chen, 2004), instead of traditional threshold. Calculating FEI for each iteration, when this parameter reaches the minimum (or local minimum), the iteration stops automatically.

FEI can be defined as: the k^{th} FEI represents the similarity of the fitted value and its true value. When F_k reaches the minimum, it means that the fitted value has the largest similarity to its true value.

$$F_k = \frac{\sum_{j=1}^{L}(N_j^{k+1} - N_j^0)^2}{L} \tag{3}$$

where, N_j^{k+1} is the new series obtained after k times fitted, N_j^0 is the original series, L is the number of available data points that are valid and will take part in the $(k+1)^{th}$ fitting. When $F_{k-1} \geqslant F_k \leqslant F_{k+1}$, the fitting iteration stops automatically.

2.3 workflow of the improved algorithm

Fig.1 shows the work flow of the improved algorithm. For each time point, NOF is calculated once, while outlier detection is applied in input data validation and iteration separately.

3 VALIDATION AND ANALYSIS OF THE IMPROVED ALGORITHM

MODIS_EVI data is a 16d synthesized series of NASA USGS vegetation index product- MOD13Q (Huete *et al.*, 2002), whose spatial resolution is 250 m. The data chosen covers the whole Huabei plain, containing 23 time points (phase) data. Meanwhile, relevant local phenological information has also been collected from a lot of weather stations in Hebei Province and Shandong Province. This information was obtained by field investigation, so it has good representation of large scale phenology.

In this study, we both applied our improved algorithm and HANTS software to reconstruct the data series, and chose several classical curves to assess our improved algorithm by comparing the two results. Furthermore, we also located the important time points of vegetation on fitting curves and make comparison with phenological information according to the approach of Xin *et al.* (2001).

According to investigation data in four countries of Heibei province (Shenzhou, Nangong, Feixiang and Fucheng), we can find that at the end of April, cotton (single-season pattern) would be planted and begins to grow (the 8^{th} temporal phase in MODIS time series approximately). Cotton plant grows to enter three leaves stage and then five leaves stage at the end of May. In early June (about the 11^{th} temporal phase), cotton plant grows to show buds and split boll stage in the middle of August

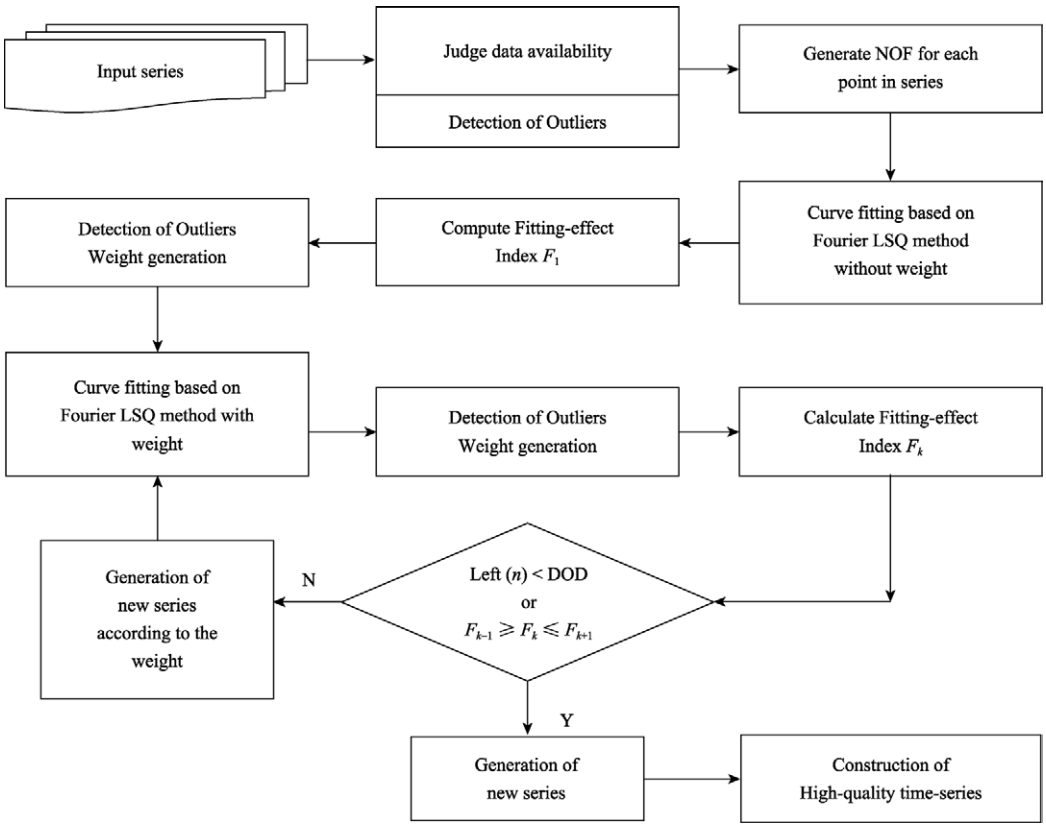

Fig. 1 Flow chart of the improved Fourier Harmonic analysis algorithm

(about the 15th temporal phase). At about the middle and the end of October (the 19th or 20th time phase), cotton enters harvest time, totally 180d (about 11.25 temporal phases). The reconstructed data curve has only one peak, which shows the local phenological regulation of single-season cropland. Fig.2(a) shows the MODIS_EVI curve which represents single-seasonal pattern of cotton in Shenzhou, Hebei. Data reconstructed by our improved algorithm shows that cotton begins to grow at the 8th temporal phase (May 2nd at seeding stage); cotton flourished at the 15th temporal phase (August 29th at splitting boll stage); cotton were harvested at the 19th phase (the observed date, October 23rd, at this phase). All the information revealed by the reconstructed data is consistent with the data obtained from observation stations. While data reconstructed by HANTS shows that between the 1st and the 4th temporal phases, the curve goes down then up. It shows the growth stage is at the 7th temporal stage which is not consistent well with the observed data. Moreover, the peak appears one temporal phase later than the observed one.

Investigation in the four countries of Shandong province (Dezhou, Huimin, Hanting and Liaocheng) shows that for WinterWheat-SummerCorn pattern, winter wheat would be planted around early October (the 18th temporal phase), wheat begins to return green around the end of February the next year (the 4th temporal phase), then heading around early May (the 8th temporal phase) and harvest in early June (the 10th temporal phase), totally 243d approximately (15.2 phases). Summer corn would be planted in early and middle June (the 10th and 11th temporal phase), around early and middle August (the 14th and 15th temporal phase) this crop begins heading and silking, then harvest time is around end September and early October (the 17th and 18th temporal phase), totally 117d approximately (7 temporal phases). Thus this reconstructed curve of cropland pattern would have two peaks, the first peak represents flourish of winter wheat from the 4th to 11th temporal phases, while the second one shows summer corn's from the 11th to 18th temporal phases. Fig.2(b) is the classic temporal curve of Winter Wheat-SummerCorn pattern in Liaocheng. Data reconstructed by improved algorithm and HANTS both show that winter wheat begins to return green from the 5th temporal phase which is one phase later than the observed date (February 20). At the 8th temporal phase (observed data shows May 3 which belongs to the 8th phase) the crop began heading and flourishing; winter wheat were harvested around the 10th temporal phase which is consistent with the investigated date, June 8. Summer corn would grow to flourish around the 15th phase (observed date, August 18, located in this phase) and were harvested at the 19th phase (consistent with observed date, October 16). The data reconstructed by the improved method could maintain the inherent pattern of the original data.

For single-seasonal pattern of cotton and winter wheat, because of none coverage or wheat suffering growth retardation, the curve at this period would be gently. In Fig.2(a), HANTS data shows that EVI value goes down then up between 1st and 6th temporal phase. While in Fig.2(b), both data reconstructed by the improved algorithm and HANTS are inconsistent with actual situation. This problem is caused by harmonic analysis's own drawback and spectral mix effects. Harmonic analysis

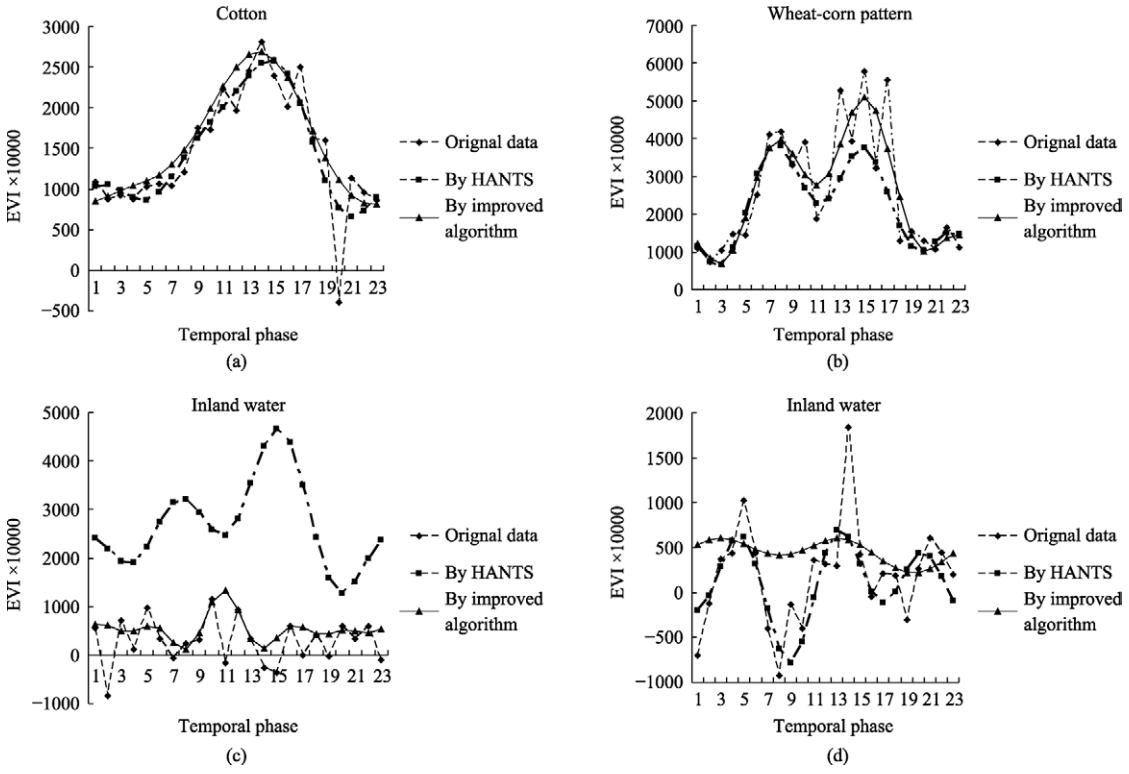

Fig. 2 MODIS_EVI temporal profiles of typical cropping systems and inland water in Huabei Plain
(a) Cotton single-season cropland (Shenzhou, Hebei); (b) Wheat-corn double-season cropland (Liaocheng, Shandong); (c) Inland water (119°19'E, 32°55'N); (d) Inland water (116°11'E, 35°57'N)

applies absolutely symmetrical sine wave to fit real curve. This problem could not be avoided but choosing suitable number of harmonic can relieve this problem. The difference between data by the improved algorithm and HANTS respectively is due that these two methods use different NOF. HANTS employs a global number, but the improved algorithm automatically chooses a NOF. MODIS_EVI data has a very low spatial resolution, leading to spectral mix effect, which makes it hard to represent local single crop-pattern phenological regulation. Moreover, the data are 16d synthesized, if the key period of growing was up to only a little part of the range, the synthesized data could not represent the true situation of local area and the useful information would be merged.

Fig.2 (c) and Fig.2 (d) shows EVI annual change curve of inland water. Because of phytoplankton, EVI of water shows fluctuation rather than flatness. Ideally, the curve of water should be flat with low EVI value. And phytoplankton in water has a slight influence on water EVI, thus the change of curve inland water would be small, and none drastic fluctuations would happen. The figure shows that data reconstructed by the improved algorithm is more similar to the actual situation.

4 RESULT

This paper, based on Fourier-transform-based harmonic analysis algorithm, proposes an improved algorithm which includes outlier detection algorithm, automatically NOF estimation and fitting-effect index. These new parameters or strategies help the improved algorithm overcome the drawback of artificially setting key parameters and to reconstruct high-quality time-series data. Firstly, it introduces an outlier detection algorithm Grubbs Test instead of threshold setting method, in order to find unreasonable data points by its own inner characteristics before curve fitting. Then, considering the inherent phenological regulation for each land cover, Numbers of Frequency (NOF) is calculated pixel by pixel through polynomial fitting; it is more reasonable than setting a unique global NOF for the whole scene which may contain complex land cover types. To avoid importing external subjective disturbances and uncertainties to original data, Fitting-effect Index, instead of manually setting one fitting tolerance, is employed to decide automatically when to terminate the iteration.

The improved method is validated by the MODIS_EVI time series which were 16d synthesized. The widely used Harmonic Analysis of Time Series (HANTS) is chosen as a comparison. The result shows that both of the two methods can reflect phenological regulations of different land covers, the reconstructed EVI temporal profile of single-season cropland (e.g. cotton) takes on one peak pattern, and those of double-season cropland and inland water take on double peak pattern and low-steady curve. But the improved method performs better in tracking the change tendency of the original curve. Moreover, the peak time and peak value of croplands are mostly consistent with the original curve, which will be useful for VI-based crop yield prediction.

Both result curves of the two methods are abnormal between the 1st and the 6th temporal phases. They show wrong time of crop returning green both for single-season crops and double-season crops. The reason is that the two methods use absolutely symmetrical sine wave to fit the real curves. Choosing a suitable NOF can relieve this problem. Another reason is that low spatial and temporal resolution of MODIS_EVI is inadequate to show the inherent phenological characteristics of a single land cover.

REFERENCES

Baksalary J K and Puntanen S. 1990. A complete solution to the problem of robustness of Grubbs's test. *The Canadian Journal of Statistics* , **18**: 285—287

Bartalev S A, Belward A S, Erchov D V and Isaev A S. 2003. A new SPOT4-vegetation derived land cover map of Northern Eurasia. *Int. J. Remote Sensing*, **24**(9): 1977—1982

Chen J, Per J, Masayuki T, Gu Z H, Bunkei M and Lars E. 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, **91**: 332—344

Gu Z H. 2003. A Study of Calculating Multiple Cropping Index of Crop in China Using SPOT/VGT Multi-temporal NDVI Data. Beijing: Beijing Normal Universiity

Huete A, Didan K, Miura T, Rodriguez E P, Gao X and Ferreira L G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. (special issue). *Remote Sensing of Environment*, 2002, **83**(1—2): 195—213

Immerzeel W W, Quiroz R A and Jong S M. 2005. Understanding precipitation patterns and land use interaction in tibet using Harmonic analysis of SPOT VGT-S10 NDVI time series. *International Journal of Remote Sensing*, **26**(11): 2281—2296

Li R, Zhang X, Liu B and Zhang B. 2009. Review on methods of remote sensing time-series data reconstruction. *Journal of Remote Sensing*, **13**(2): 246—252

Lin Z H and Mo X G. 2006. Phenologies from Harmonics Analysis of AVHRR NDVI Time Series. *Transactions of the Chinese Society of Agricultural Engineering*, **22**(12): 138—144

Loveland T R, Reed B C, Brown J F, Ohlen D O, Zhu Z, Yang L and Merchant J W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. *Int. J. Remote Sens*., **32**:1303—1330

Moody A and Johnson D M. 2001. Land-surface phenologies from AVHRR using the discrete fourier transform. *Remote Sensing of Environment*, **75**(3): 305—323

Roerink G J, Menenti M, Soepboer W and Su Z. 2003. Assessment of climate impact on vegetation dynamics by using remote sensing. *Physics and Chemistry of the Earth*, **28**: 103—109

Sellers P J, Los S O, Tucker C J, Dazlich D A, Collatz G J and Randall D A. 1996. A reversed land surface parameterization (SiB_2) for atmospheric GCMs, Part 2: The generation of global fields of terrestrial biophtsical parameters from satellite data. *Journal of Climate*, **9**: 706—737

Xin J F,Yu Z R and Driessen P. M. 2001. Monitoring phenological key stages of winter wheat with NOAA NDVI data. *Journal of Remote Sensing*, **5**(6): 442—448

Zhang F, Wu B F, Liu C L and Luo Z M. 2004. Methods of monitoring crop phonological stages using time series of vegetation indicator. *Transactions of the Chinese Society of Agricultural Engineering*, **20**(1): 155—159

谐波改进的植被指数时间序列重建算法

张 霞[1], 李 儒[2], 岳跃民[3], 刘 波[1], 刘海霞[1]

(1. 遥感科学国家重点实验室, 中国科学院 遥感应用研究所 北京 100101;
2. 香港中文大学 太空与地球信息科学研究所, 香港;
3. 中国科学院 亚热带农业生态研究所, 湖南 长沙 410125)

摘 要: 提出一种基于傅里叶谐波分析的改进算法, 引入异常值检测算法, 检测拟合过程中的异常值, 增加数据拟合的真实性; 迭代前动态估算出待处理序列点的峰值个数(即频数), 解决整个区域预设单一频数的不合理性; 引入拟合影响因子, 自动控制迭代终止条件, 避免传统方法中人为设置阈值导致的不确定性。利用 2003 年华北平原 MODIS_EVI 时间序列图像验证表明, 较之 HANTS 算法, 改进算法能够有效修正噪声污染像元值, 修正后的 EVI 时序曲线更能反映地物内在的物候变化规律, 并能够更好地保真原始曲线上的特征(点), 如作物 EVI 最大值、最小值出现的时间和大小关系。

关键词: 植被指数图像时间序列, 滤波, 傅里叶谐波, 异常值检测, 拟合影响因子

中图分类号: TP751.1 文献标识码: A

引用格式: 张 霞, 李 儒, 岳跃民, 刘 波, 刘海霞. 2010. 谐波改进的植被指数时间序列重建算法. 遥感学报, **14**(3): 437—447
Zhang X, Li R, Yue Y M, Liu B and Liu H X. 2010. Improved algorithm for reconstructing vegetation index image time series based on Fourier Harmonic Analysis. *Journal of Remote Sensing*. **14**(3): 437—447

1 引 言

一般而言, 植被指数只能表征植被绿度、光合作用强度, 而不能直接表征土地覆盖类型, 但是植被指数时间序列可以反映生物物候变化规律, 因此植被指数时间序列分析使得研究区域乃至全球范围的物候现象(如返青, 生长区间, 生长衰退期)成为可能(Moody Johnson, 2001; 张峰等, 2004), 并在改进土地覆盖分类效果(Bartalev 等, 2003; Loveland 等, 2000)、研究气候变化效应(Roerink 等, 2003)等方面发挥着重要作用。但植被指数图像时间序列受到云等噪声的影响, 这在一定程度上制约了其应用的精度。

傅里叶谐波分析法是在植被指数时间序列去噪重建方面应用最广泛的算法, 它将每个独立数据点的时间曲线表达为一系列正(余)弦波加性项的和, 每一个余弦波由唯一振幅和相位确定, 将这些连续的谐波项叠加起来就生成一条复杂而平滑的曲线。傅里叶谐波分析最为典型的是 Sellers 方法(Sellers 等, 1996)和 Hants 方法(Huete 等, 2002; Immerzeel 等, 2005), 其优势是借助谐波分析能够尽最大可能揭示植被指数曲线蕴涵的物候规律。但傅里叶谐波分析方法也存在一些缺陷(李儒等, 2009): (1)最突出的是关键参数(如有效值范围, 频率个数、拟合容许误差和超定度等)需要人为凭经验和不断试验确定, 没有明确的规则可循, 因而易引入主观误差、导致对原始数据的干扰。(2)整个测区使用同一个(组)参数值也多与实际不符, 如华北平原存在多种土地覆盖类型, 单季和双季作物的植被指数时间序列分别呈单峰和双峰态势, 设定唯一的频数是不合理的。(3)关于定权准则, Sellers 采用的是按距离定权法, 林忠辉等(2006)研究指出 Sellers 定权准则主要适用于热带阔叶林, 而对短时间间隔合成的时间序列效果欠佳; Hants 软件的定权准则是集成在软件里面的, 属于黑箱操作, 无法根据现实条件及时调整, 从而影响其具体应用效果。

本文在 Sellers 傅里叶谐波分析法基础上, 从异

收稿日期: 2009-04-02; 修订日期: 2009-07-17
基金项目: 国家自然科学基金项目(编号: 40971205), 国家科技支撑课题(编号: 2007BAH15B01)及遥感科学国家重点实验室开放基金(编号: 03Q00300649)
第一作者简介: 张霞(1972—)中国科学院遥感应用研究所副研, 博士。主要从事高光谱图像处理和信息提取研究, 发表论文 30 余篇。

常值检测方法、自动选频和拟合影响因子的自动迭代计算等方面进行改进，引入新的参量或策略，试图减少人为影响及其对原始数据的过度扰动，使算法运行结果更为客观，同时增强算法普适性。

2　算法原理和实现

2.1　时间序列数据滤波重建原则

为使滤波算法在有效去噪同时，揭示植被指数时间曲线蕴涵的物候规律，滤波重建算法将遵循如下 4 条原则:

(1) 尽可能减少对数据过度扰动。滤波重建算法对原始数据扰动越大，其结果可信度越低，某些重要规律甚至会因此被抹杀;

(2) 尽可能少地使用辅助数据。辅助数据的使用一方面为滤波重建提供了一定的指示，如 MODIS 云图像，但由于数据获取、生产等过程中的不确定性因素，使辅助数据本身就存在类型复杂的误差，它的使用一定程度上增加程序开销、还引入新的误差;

(3) 尽可能多的考虑参量物理意义。植被指数时间序列的优势在于可以通过对其分析提取一定的季相物候等规律，因此滤波重建过程能考虑此问题将使数据处理更有意义，这也是傅里叶谐波分析系列算法的优势所在;

(4) 算法效率与效果的兼顾。引入拟合影响因子，控制拟合迭代的自动终止；提出使用粗拟合的方法，动态估计频数(NOF)。

2.2　改进算法具体实现

2.2.1　异常值**(outlier)**检测

由于残存误差的存在，真实值和异常值难以通过表象得以区分。传统方法通过设置阈值处理突变值(疑似异常值)虽然也有效，但难免出现“误杀误放”的情况，因此有必要从数据内部入手，通过分析数据内在规律探测序列中异常值的存在。改进算法引入异常值检测算法，检测拟合迭代过程中的异常值，减少因传统方法根据距离定权剔除无效点而引入的人为误差和新的不确定性。具体使用 Grubbs Test(Baksalary & Puntanen,1990)进行异常检测，其原理如下:

设一组重复观测值的残差，按大小顺序排列为: $V_{(1)}\leq V_{(2)}\leq\cdots\leq V_{(n)}$. 计算 Grubbs 统计量:

$$g_n=\frac{V_{(n)}-\overline{V}}{\sigma} \tag{1}$$

$$g_n'=\frac{\overline{V}-V_{(1)}}{\sigma} \tag{2}$$

式中，$\overline{V}=\frac{1}{n}\sum_{i=1}^{n}V_{(i)}$，$\sigma$为样本标准差，$g_n$ 用来判别最大值是否为异常值，g_n'用来判断最小值是否为异常值，g_n和 g_n'服从相同的概率分布。$g_0(n,\alpha)$为显著水平α下的临界值，可通过查表获得。

当$|g_n|\geqslant g_0(n,\alpha)|$或$|g_n'|\geqslant g_0(n,\alpha)|$时,可认为该残差相应的观测值含有粗差，判定最大值(或最小值)是异常值。α为显著度，一般取 0.05 或 0.01。

由于 Grubbs Test 要求数据服从正态分布，为减少算法开销，将观测值的残差取绝对值再取对数，作近似处理以适应检测算法。

2.2.2　频数前置自动估算

傅里叶谐波分析最大的优势是考虑了植被物候规律，这种优势通过谐波个数(频数 NOF)来体现。以前的方法对时间序列上所有点使用同一频数，这显然暗含一条假设，即: 在作业区域内，所有地方物候规律是相同的。这个假设并不是普遍成立的。即便是同一地区，耕地也可以一年数茬，即在曲线上反映出数个波峰，但是林地则不会有这种现象，显然代表耕地和林地的谐波数是不同的。所以若不加区分，使用同一谐波数模拟植被物候明显不合理，它的存在影响曲线拟合精度和可靠性。

针对这一问题，本文设计了前置动态估算方法，即通过迭代前预处理，动态估算出待处理序列点的峰值个数，即频数，参与下步拟合。谐波用来叠加拟合研究点的时间谱曲线，时间谱曲线峰值的个数与谐波个数有对应的关系，即时间谱曲线有几个峰，就应该选用几个谐波(此处谐波不含 0 谐波)。考虑到时间序列噪声影响，如果直接进行局部最大值统计确定峰数作频数，其结果不可靠。但可以转化为统计粗拟合曲线的波峰数来解决。所谓“粗拟合”是指不苛求拟合曲线与真实曲线符合，只要拟合曲线的波峰与实际相同即可。实现方法:

(1) 将进行输入值有效性判断后的序列首尾相连，形成含 n+1 个值的新序列，保证不因首尾截断丢失波峰。

(2) 对新序列采用三次多项式拟合生成新曲线。通常使用 3 次就足够了，低于 3 次，可能会造成波峰丢失，高于 3 次，拟合时间长，容易产生虚假波峰。

(3) 对新曲线求一次导数，获得零点；再求二次导数，判断零点正负性，最终获得极大值个数，亦即曲线波峰个数(频数)。也可以先使用简单滤波处理然后估算峰值个数。

2.2.3 迭代自动终止

Hants 通过人为设置拟合允许误差(FET)，在曲线迭代过程中，当所有点的估值与观测值绝对差小于FET时，曲线拟合迭代终止; Sellers则采用估测值与原始值的最近 4 个邻域比较预设限差，来判断拟合是否终止。为实现迭代自动终止，避免主观性，本文借鉴基于 Savitzky-Golay 滤波的拟合法中的拟合影响因子(辜智慧,2003; Chen 等, 2004)，计算每次拟合的影响因子，在该因子达到最小值(或局部最小值)时自动终止迭代，放弃使用传统方法中人为设置的阈值限差。

拟合影响因子 F_k 定义为：第 k 次拟合的影响因子，是对拟合值与真值间接近程度的刻画。当 F_k 达到最小时，标志着这一序列点的拟合值与真值达到了最接近的程度。计算公式如下:

$$F_k = \frac{\sum_{j=1}^{L}(N_j^{k+1} - N_j^0)^2}{L} \tag{3}$$

式中，N_j^{k+1} 是指 NDVI 经过 k 次拟合后得到的数列，N_j^0 原始的 NDVI 值，L 值是经过“输入值有效性”判别后，被认定为有效输入值、参加下步拟合的点的个数。当 F_k 符合 $F_{k-1} \geqslant F_k \leqslant F_{k+1}$ 条件时，拟合终止。

2.3 改进算法流程

图 1 为改进算法流程。对于每一个像元点，进行一次选频频数(NOF)计算，异常值检测则根据需要，在输入值有效性判断、迭代计算中分别使用。输入值有效性判断时也可以不选用异常值检测，而只给出一个宽泛的有效值范围，将异常检测留到拟合迭代过程中。

3 算法验证与分析

MODIS_EVI图像数据来源于NASA USGS的植被指数产品—MOD13Q(Huete 等, 2002)，是 250 米空间分辨率，16 天合成的数据产品，涵盖 2003 年华北平原一整年共 23 个时相的数据。同时，为便于结合物候时相分析，收集整理了河北和山东各 4 县市标准气象台站的农气观测报表资料数据，该资料是

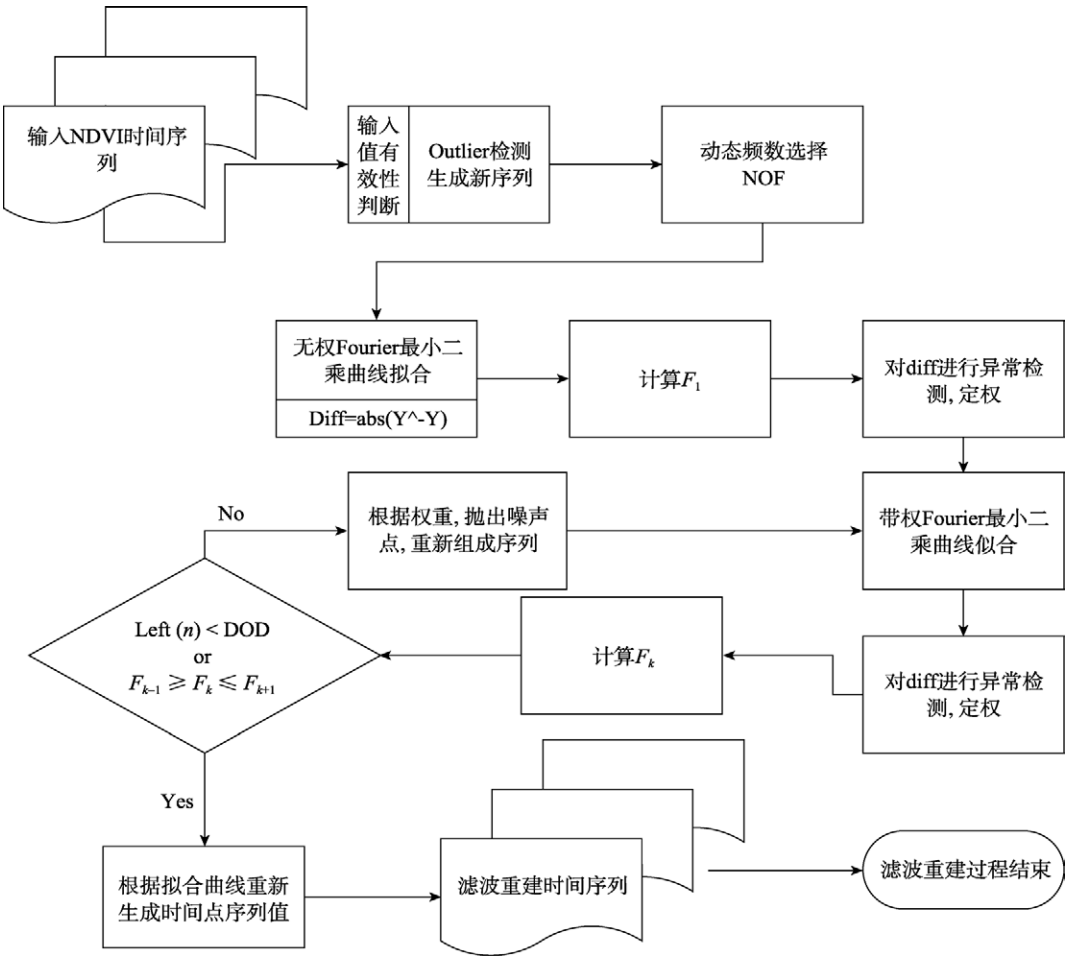

图 1 改进的植被指数时间序列 Fourier 谐波分析算法流程

通过调查大田作物获得的，因此对大面积的作物物候具有很好的面上代表性。

利用本文提出的傅里叶谐波改进算法和HANTS算法，对华北平原植被指数时间序列图像进行处理，提取几种典型作物种植模式和内陆水体的MODIS_EVI 时序曲线进行分析，对比二种方法对物候规律的揭示效果及对原始数据的忠实性。同时，根据辛景峰等(2001)方法从 EVI 时间曲线上确定小麦返青期和其他作物的生育开始期、作物成熟期，与物候观测资料比较。

根据河北 4 县市(深州、南宫、肥乡、阜城)调查资料，棉花单作在 4 月下旬(对应 MODIS 第 8 时相)播种出苗，5 月下旬依次进入三真叶和五真叶期，6 月上旬(时相号 11)现蕾，8 月中下旬裂铃(MODIS 第 15 时相), 10 月中下旬(时相号 19 或 20)拔秆收获，全生育期天数平均在 180 d (约 11.25 个时相)左右。重建后的 EVI 曲线呈现单峰态势，能够很好地体现单季作物地一年只有一个生长峰期的规律。图 2(a)为河北深州棉花单作种植模式的 MODIS_EVI 曲线。根据本文改进方法处理后的曲线，可确定在第 8 时相开始生长(对应出苗期 5 月 2 日，属于该时相)，在第 15 时相达到生长旺盛期(对应棉花裂铃期 8 月 29 日，落在该时相)，在第 19 时相拔秆收获(观测时间为 10 月 23 日，落在该时相)，与观测台站资料具较好一致性。而 HANTS 处理结果在 1—4 时相内，呈现先走低再回升态势，导致确定的生育开始期为第 7 时相，与观测资料不符，另外峰值出现时间比原始曲线错后一个时相。

根据山东省 4 县市(德州、惠民、寒亭、聊城)，对于冬小麦－夏玉米种植方式，冬小麦 10 月上旬(时相号 18)播种，次年 2 月底(时相号 4)返青, 5 月初(时相号 8)抽穗, 6 月上旬(时相号 10)成熟收获，全生育期天数平均 243d(15.2 个时相)左右；下茬夏玉米 6 月上中旬(时相号 10, 11)播种, 8 月上中旬(时相号 14, 15)抽雄与吐丝, 9 月底 10 月初(时相号 17, 18)成熟收获，全生育期天数平均 112d(7 个时相)左右。去噪重建后的 EVI 时间曲线均呈现前双峰模态，前茬冬小麦在其次年生长期(第 4—11 时相)形成第一个峰，后茬夏玉米在第 11—18 时相形成第二个峰。图 2(b)为聊城冬小麦-夏玉米种植模式的 EVI 时间曲线，利用 HANTS 和改进算法处理后的曲线均确定，冬小麦在第 5 时相开始生长，与观测时间(2 月 20 日)所属

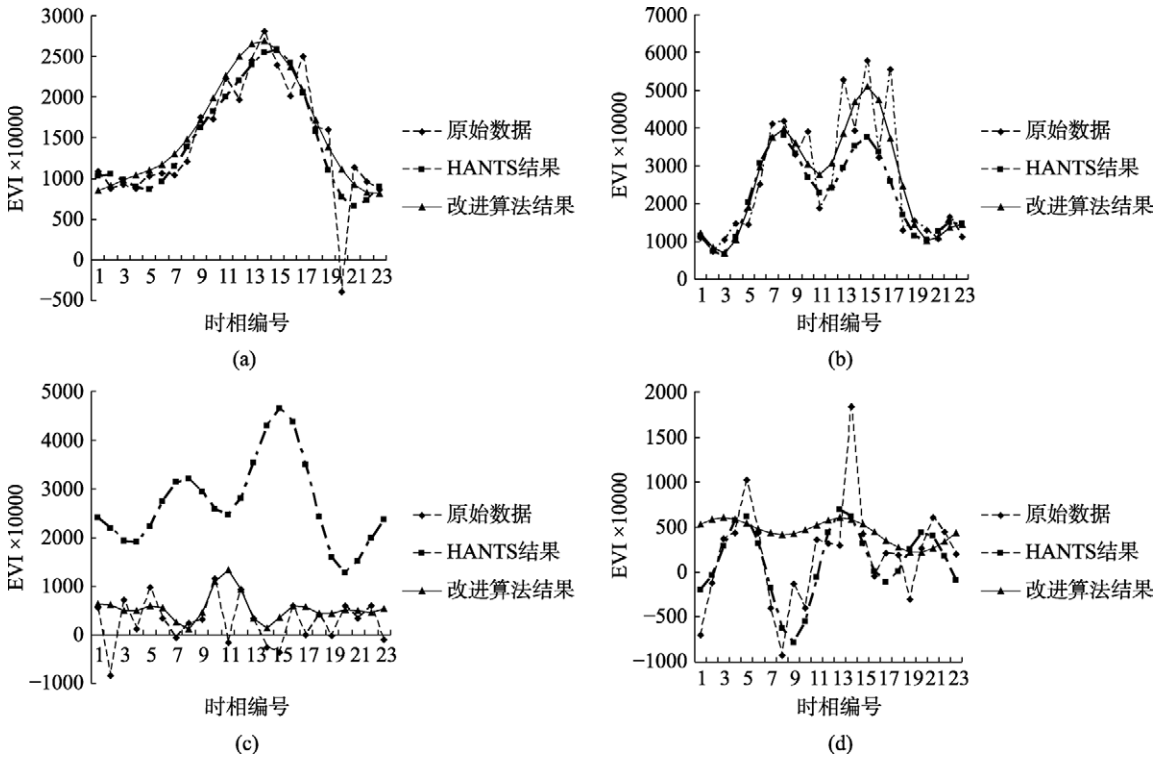

图 2　华北平原典型作物种植模式与内陆水体的 MODIS_EVI 时间曲线

(a). 棉花单作(河北深州); (b). 小麦-玉米种植模式(山东聊城); (c). 内陆水体(119°19′E, 32°55′N); (d). 内陆水体(35°57′0N, 116°11′E)

时相推后一个时相，在第 8 时相(5 月 3 日，落在该时相)抽穗达到生长旺盛期，在第 10 时相(观测时间为 6 月 8 日，落入该时相)成熟收获，后茬夏玉米在 15 时相吐丝达到生长盛期(观测时间为 8 月 18 日，落入该时相)，在 19 时相成熟收获(观测时间 10 月 16 日，落入该时相)。改进算法能够较好保持原始曲线先低峰后高峰的模式。

对于棉花单作和冬小麦，由于冬春土地空闲或小麦尚处在越冬期生长停滞，该时段地表覆盖变化小，在 EVI 时间序列上本应该反映为一段平直或者平缓曲线，或者略有波动，但在图 2(a)上 1—6 时相 HANTS 结果出现 EVI 降低再增大的情况，而在图 2(b)上 1—4 时相两种算法都出现了相同情况，这与实际物候规律不符。究其原因，一方面可能在于傅里叶谐波方法本身的问题，即使用绝对对称的正余弦曲线进行数据拟合，该问题无法避免，只能通过合适的波数尽量减小影响，HANTS 与改进方法之所以出现差异，是因为各自使用的频数(参与拟合的谐波个数)不同，HANTS 使用 3 谐波，而改进方法根据曲线峰值个数自动选频；另一方面，MODIS_EVI 产品空间分辨率较低，混合像元普遍存在使其难以完全体现单一地物目标的时间变化规律，另外，16d 最大值合成算法时间跨度大，如果关键生育期只占 16d 合成区间的很小一部分，且不是 16d 内 EVI 变化的最主要特征，则关键生育期的特征将被掩盖。

图 2(c)(d)为内陆水体的 EVI 年变化曲线。内陆水体中由于浮游植物的生长，EVI 曲线在一定程度上会出现波动。但是理想水体的 EVI 曲线是一水平直线，其值很低，且水中浮游植物对 EVI 值影响能力有限，年变化振幅很低，因而在内陆水体真实曲线上很难有大起大落的波动。从重建后的曲线上看，改进方法表现更好，更能从原始曲线中回归水体年变化的真实规律。

4 结论

本文从 Fourier 谐波分析方法原理入手，就异常值检测方法、自动选频和拟合影响因子的自动迭代计算等 3 个主要方面进行改进，使原方法由经验定参数向依准则定参数靠拢，减小主观干预造成的不确定。具体滤波前，首先对数值有效性进行规整，摒弃人为设置有效阈值的片面性，从分析数据内在规律入手，采用异常值检测算法 Grubbs Test，探测序列中的异常值并剔除不合理值；其次，以自动选频代替全局单一的频数设定，更符合复杂地类条件下物候空间变异较大的情况，通过粗拟合初步确定谐波的个数(即频数)，在后续拟合滤波中动态调整该值，实现针对不同物候选择不同谐波个数，体现 Fourier 分析的优势；为避免传统方法，如 HANTS 算法通过人为设置拟合允许误差(FET)导致的不确定性，引入拟合影响因子，自动控制迭代终止条件。

利用 MODIS 的 16d 合成植被指数(增强型植被指数，EVI)的年时间序列图像验证表明，改进算法和 HANTS 处理结果都能较好揭示 EVI 时间序列蕴涵的物候规律，棉花单作的 EVI 时间曲线呈现单峰模态，冬小麦-夏玉米种植模式则呈现双峰模态，水体呈低而平缓的曲线。但依改进算法重建后的曲线确定的作物关键期与物候观测资料具有更好的一致性， 特别是峰值的大小和位置于原始曲线具很好的保真性，对于基于植被指数时间序列的估产应用具有重要作用。改进算法处理后的水体曲线能够很好地逼近水体的低稳变化曲线态势。

改进算法和 HANTS 处理后曲线在 MODIS_EVI 的 1—6 时相出现先降低后增加的趋势，导致由其确定的单季作物和两季之前茬作物的生长开始期与实际观测不一致，原因主要在于两种算法均利用绝对对称的正余弦曲线进行数据拟合，这对于物候规律的揭示是把双刃剑，需要因像元设定合适的波数来减少此影响；另外，MODIS_EVI 数据产品的粗空间分辨率和低时间分辨率，不足以体现单一地物固有的物候特征，也是一个重要原因。

REFERENCES

Baksalary J K and Puntanen S. 1990. A complete solution to the problem of robustness of Grubbs's test. *The Canadian Journal of Statistics* , **18**: 285—287

Bartalev S A, Belward A S, Erchov D V and Isaev A S. 2003. A new SPOT4-vegetation derived land cover map of Northern Eurasia. *Int. J. Remote Sensing*, **24**(9): 1977—1982

Chen J, Per J, Masayuki T, Gu Z H, Bunkei M and Lars E. 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, **91**: 332—344

Gu Z H. 2003. A Study of Calculating Multiple Cropping Index of Crop in China Using SPOT/VGT Multi-temporal NDVI Data. Beijing: Beijing Normal Universiity

Huete A, Didan K, Miura T, Rodriguez E P, Gao X and Ferreira L G. Overview of the radiometric and biophysical performance of

the MODIS vegetation indices. (special issue). *Remote Sensing of Environment*, 2002, **83**(1—2): 195—213

Immerzeel W W, Quiroz R A and Jong S M. 2005. Understanding precipitation patterns and land use interaction in tibet using Harmonic analysis of SPOT VGT-S10 NDVI time series. *International Journal of Remote Sensing*, **26**(11): 2281—2296

Li R, Zhang X, Liu B and Zhang B. 2009. Review on methods of remote sensing time-series data reconstruction. *Journal of Remote Sensing*, **13**(2): 246—252

Lin Z H and Mo X G. 2006. Phenologies from Harmonics Analysis of AVHRR NDVI Time Series. *Transactions of the Chinese Society of Agricultural Engineering*, **22**(12): 138—144

Loveland T R, Reed B C, Brown J F, Ohlen D O, Zhu Z, Yang L and Merchant J W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. *Int. J. Remote Sens*., **32**:1303—1330

Moody A and Johnson D M. 2001. Land-surface phenologies from AVHRR using the discrete fourier transform. *Remote Sensing of Environment*, **75**(3): 305—323

Roerink G J, Menenti M, Soepboer W and Su Z. 2003. Assessment of climate impact on vegetation dynamics by using remote sensing. *Physics and Chemistry of the Earth*, **28**: 103—109

Sellers P J, Los S O, Tucker C J, Dazlich D A, Collatz G J and Randall D A. 1996. A reversed land surface parameterization (SiB_2) for atmospheric GCMs, Part 2: The generation of global fields of terrestrial biophtsical parameters from satellite data. *Journal of Climate*, **9**: 706—737

Xin J F,Yu Z R and Driessen P. M. 2001. Monitoring phenological key stages of winter wheat with NOAA NDVI data. *Journal of Remote Sensing*, **5**(6): 442—448

Zhang F, Wu B F, Liu C L and Luo Z M. 2004. Methods of monitoring crop phonological stages using time series of vegetation indicator. *Transactions of the Chinese Society of Agricultural Engineering*, **20**(1): 155—159

附中文参考文献

辜智慧. 2003. 中国农业复种指数的遥感估算方法研究. 北京: 北京师范大学

李儒, 张霞, 刘波, 张兵. 2009. 遥感时间序列数据滤波重建算法发展综述. 遥感学报, **13**(2): 246—252

林忠辉, 莫兴国. 2006. NDVI 时间序列谐波分析与地表物候信息获取.农业工程学报, **22**(12):138—144

辛景峰, 宇振荣, Driessen P.M. 2001. 利用 NOAA NDVI 数据集监测冬小麦生育期的研究. 遥感学报, **5**(6): 442—448

张峰, 吴炳方, 刘成林, 罗治敏. 2004. 利用时序植被指数监测作物物候的方法研究. 农业工程学报, 2004, **20**(1):155—159