Intelligent Automation & Soft Computing

ISSN: 1079-8587 (Print) 2326-005X (Online) Journal homepage: http://www.tandfonline.com/loi/tasj20

Selecting the Optimal NDVI Time-Series Reconstruction Technique for Crop Phenology Detection

Wei Wei, Wenbin Wu, Zhengguo Li, Peng Yang & Qingbo Zhou

To cite this article: Wei Wei, Wenbin Wu, Zhengguo Li, Peng Yang & Qingbo Zhou (2016) Selecting the Optimal NDVI Time-Series Reconstruction Technique for Crop Phenology Detection, Intelligent Automation & Soft Computing, 22:2, 237-247, DOI: 10.1080/10798587.2015.1095482

To link to this article: http://dx.doi.org/10.1080/10798587.2015.1095482

Published online: 16 Nov 2015.

INTELLIGENT AUTOMATION & SOFT COMPUTING, 2016
VOL. 22, NO. 2, 237–247
http://dx.doi.org/10.1080/10798587.2015.1095482

Selecting the Optimal NDVI Time-Series Reconstruction Technique for Crop Phenology Detection

Wei Wei, Wenbin Wu, Zhengguo Li, Peng Yang and Qingbo Zhou

Key Laboratory of Agri-informatics, Ministry of Agriculture/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China

ABSTRACT
A new scored method has been proposed in this study to evaluate the performances of different NDVI time-series reconstruction techniques. By giving a synthetic score to each of the candidates techniques based on two quantified criteria the optimal one is selected for the purpose of phenology detection. Three widely used techniques including Asymmetric Gaussian function fitting (AG), Double Logistic function fitting (DL) and Savitzky-Golay filtering (SG) are compared using NDVI time-series products from Moderate Resolution Imaging Spectroradiometer (MODIS) on Terra satellite over cropland of Northeast China. The results show that AG approach outperforms the two others in our study area. Cropland NDVI values have been improved obviously after the reconstruction by AG. Spatial patterns of the crop phenology detected from the AG reconstructed NDVI time-series are reasonable. The errors of the derived crop phenology metrics are within an acceptable limit.

KEYWORDS
NDVI; time-series reconstruction; crop phenology; MODIS

1. Introduction

Normalized difference vegetation index (NDVI) is designed by enhancing the difference between the spectral reflectance of visible and infrared band to reflect vegetation status in remote sensing (Brown, Kastens, Coutinho, Victoria, & Bishop, 2013; Landmann, Schramm, Huettich, & Dech, 2013). Time-series of NDVI data from satellite sensors carry useful information about vegetation phenology, which can be used for characterizing growth process and seasonal dynamics of vegetation in support of research on climate change and ecosystem response (Fensholt & Proud, 2012; Forkel et al., 2013; Kariyeva & Van Leeuwen, 2011; Li, Zhao, & Yang, 2011; Verbesselt, Hyndman, Zeileis, & Culvenor, 2010). However, due to unfavourable atmospheric conditions and viewing geometries during the data acquisition, NDVI time-series curve often contains a lot of noise and fluctuates greatly (Motohka, Nasahara, Murakami, & Nagai, 2011; Park, 2013). So there is a strong need for the reconstruction of NDVI time-series before extracting information from the noisy data.

Although most of the NDVI data products are temporally composited through maximum value compositing (MVC) method (Holben, 1986) to retain relatively cloud-free data, residual noise still exists in the data, which will prevent the accurate detection of vegetation phenology. Several techniques have been presented to reduce noise and reconstruct NDVI time-series, which can be grouped into two general types: filtering methods and function fitting methods (Reed, Schwartz, & Xiao, 2009). Filtering methods consist of Savitzky-Golay filtering (SG) (Chen et al., 2004), the best index slope extraction (BISE) (Viovy, Arino, & Belward, 1992) and its modified version (Lovell & Graetz, 2001), moving median and mean filtering (Kogan & Sullivan, 1993), temporal window operation (TWO) (Park & Tateishi, 1998), maximum value iteration filtering (MVI) (Taddei, 1997), fast Fourier (Sellers et al., 1994) and wavelet (Lu, Liu, & Liang, 2007) transformation based frequency domain low pass filtering, 4253H Twice (4253HT) filtering (Velleman, 1980), ARMD3-ARMA5 (ARM3–5) filtering (Davis, 2002) and iterative interpolation for data reconstruction (IDR) (Julien & Sobrino, 2010), etc. Function fitting methods include Asymmetric Gaussian function fitting (AG) (Jönsson & Eklundh, 2002), Double Logistic function fitting (DL) (Beck, Atzberger, Høgda, Johansen, & Skidmore, 2006), Fourier function fitting such as harmonic analysis of time-series (HANTS) (Roerink, Menenti, & Verhoef, 2000) and Sellers algorithm (Sellers et al., 1996), piecewise Logistic function fitting (PL) (Zhang et al., 2003), etc.

Previous literatures have evaluated the performances of different noise reduction techniques for NDVI time-series either qualitatively or quantitatively (Bradley, Jacob, Hermance, & Mustard, 2007; Geng et al., 2014; Hird & McDermid, 2009; Julien & Sobrino, 2010; Michishita, Jin, Chen, & Xu, 2014). Although relatively easy, qualitative evaluations may only reveal the technique that produces the most visually pleasing result. Quantitative evaluations may be more accurate, but with more complex analysis. How to evaluate different noise reduction techniques both easily and accurately? Moreover, phenology detection is an important application of NDVI time-series. With an increasingly emphasis of food secure, comprehensive understanding of crops is critical and crop phenology is a very important element in it. Accuracy information about crop phenology will be useful in irrigation scheduling, fertilizer management and yield estimating (Shan & Xu, 2013; Xu, Yang, Long, & Wang, 2013). An appropriate reconstruction technique will restore the 'true' NDVI time-series as much as possible and facilitate crop phenology detection.

CONTACT Wei Wei wei20031479@126.com

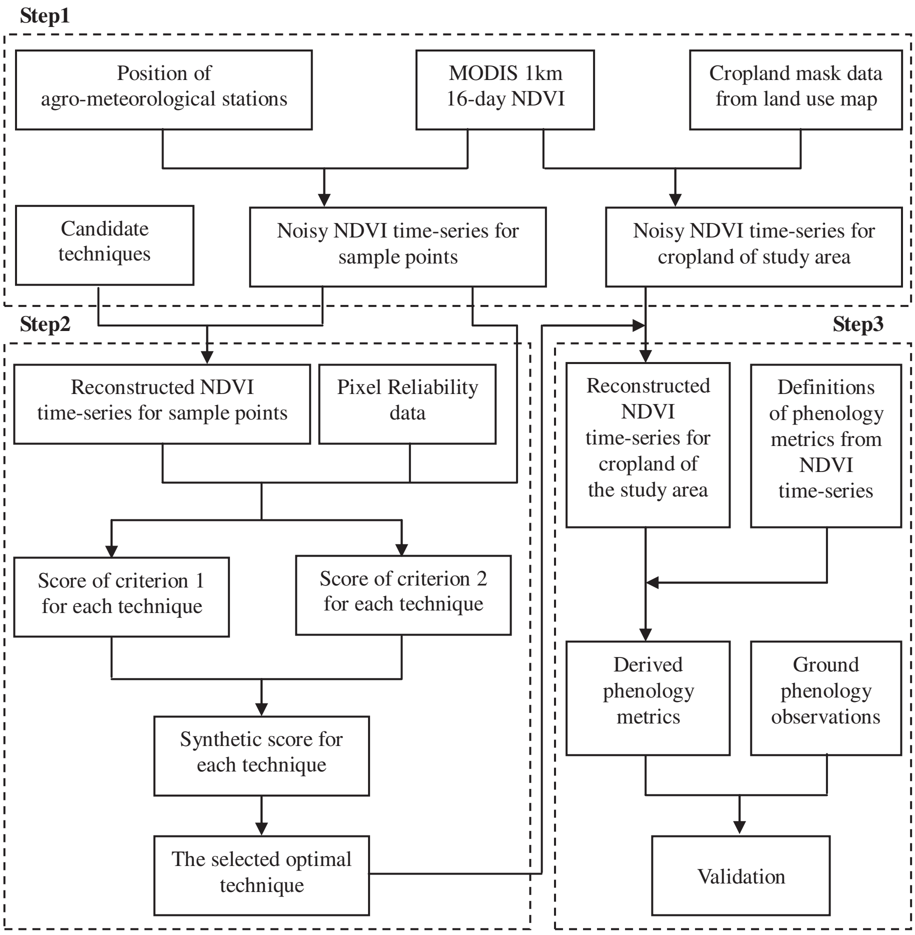


Figure 1. Flowchart of the Basic Steps and Procedures Followed in this Study.

TIMESAT (Jönsson & Eklundh, 2004) is a software package developed for analyzing time-series of remote sensing data from satellite sensor. It provides three techniques for the reconstruction of NDVI time-series including SG, AG and DL, which have been widely used by a lot of studies. Brown, de Beurs, and Vrieling (2010) used SG to smooth the NDVI time-series before extracting vegetation phenology of Africa continent for the year 1981–2008. Tuanmu et al. (2010) also processed time-series data by means of SG and then obtained phenological cycles for further analysis. Li et al. (2009) generated smooth time-series of NDVI using AG for the research of crop phenological characteristics and cropping system in North China. Wu et al. (2010) characterized spatial patterns of phenology in cropland of China based on AG reconstructed NDVI time-series. Butt, Turner, Singh, and Brottem (2011) estimated phenology metrics in western Africa using a DL fitted NDVI time-series from MODIS for the period 2000–2010. Cai, Zhang, and Yang (2012) investigated the dynamics of forest phenology in relation to climate change with selecting DL for the time-series data reconstruction.

There is a problem that various kinds of NDVI time-series reconstruction techniques often confuse us. It is difficult to determine which one should be chosen when we need to reconstruct NDVI time-series for phenology detection. A new scored method has been proposed in this paper aims to evaluate the performances of different reconstruction techniques and select the optimal one for the purpose of crop phenology detection. The objectives of this study are: (i) to determine which of the candidate techniques perform best under given conditions, and (ii) to detect crop phenology based on the reconstructed NDVI time-series produced by the selected technique. The paper is organized as follows: First, the significance of this study is investigated; second, relevant datasets and methods are introduced; then the experimentation results are showed and analyzed; at last the conclusions and future work are discussed.

2. Methodology

Figure 1 shows the flowchart of this study in which the overall procedures are divided into three steps (see the rectangle boxes in dashed line). The process starts with the data preparation to extract noisy NDVI time-series for sample points and the whole cropland of study area (Step 1). In the second step we have developed a new scored method to compare different NDVI time-series reconstruction techniques and selected the optimal one under given conditions (Step 2). Finally crop phenology metrics of the study area is derived based on the reconstructed NDVI time-series produced by the selected optimal technique and validated using ground phenology observations (Step 3).

2.1. Study area and datasets

The study area is located in Northeast China including three provinces: Heilongjiang, Jilin and Liaoning (Figure 2). It covers an area of 791,800 km2, in which 264,400 km2 is cropland, accounting for 16.5% of the total arable land in China. This region displays a boreal and temperate moist (humid) climate. Most parts of the region have an accumulated temperature (AT) ≥0 °C of 2000-4,200 °C.days, an active AT ≥ 0 °C of 2000-4,200 °C. days, average summer temperature of 20–25 °C, a

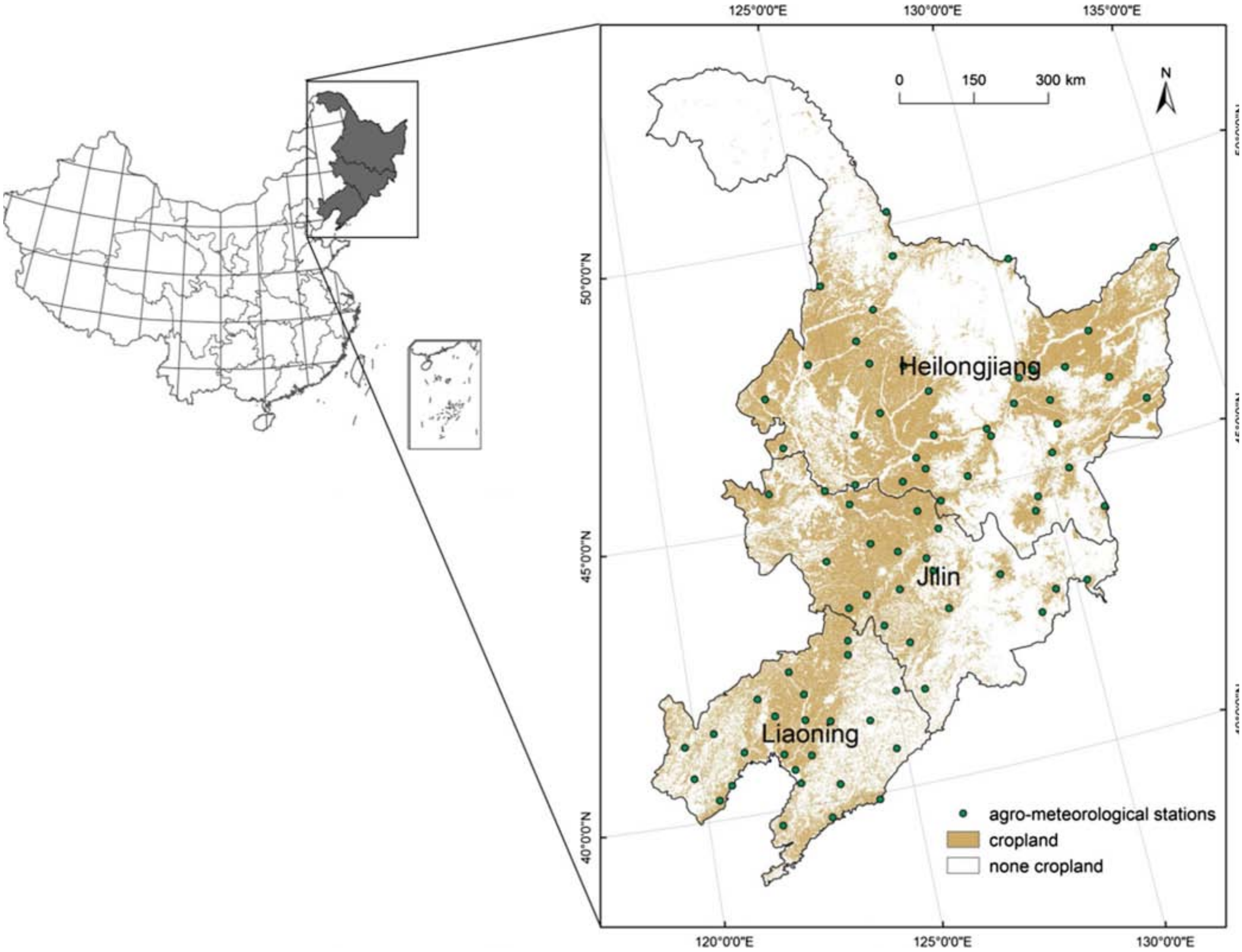


Figure 2. Study Area as Well as Geographical Distribution of Cropland and Agro-meteorological Stations.

frost-free period of 140–170 days and precipitation of 500–800 mm (60% of the rainfall concentrates between July and September) (Guo, Zhang, & Wang, 2010). The major crops of this region are corn, rice, wheat and soybean, and the region exhibits considerable internal differences in the planting structure.

The MOD13A2 (version 005) NDVI and Data Pixel Reliability products from Moderate Resolution Imaging Spectroradiometer (MODIS) on Terra are used for our analysis. The data are produced by a modified MVC method to minimize the view angle effects and consequently BRDF related issues. The composite periods are 16 days with a spatial resolution of 1 km. We have acquired the data-set from the website of Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov/products/modis_products_table). The download data covered the complete year of 2005, which consist of 23 images. The data are HDF (Hierarchical Data Format) files and contain several data layers. MODIS Reprojection Tool (MRT) is used to extract the desired data layers and re-project them into Albers Conical Equal Area Projection. NDVI in the data-set ranged from -3,000 to 10,000, in which -3,000 is the fill value and a factor of 0.0001 is used to convert the DN value to standard NDVI.

The land use/cover map of China in 2005 has been employed to build a cropland mask for the study area. This data-set, which has a spatial resolution of 30 m with a scale of 1: 250,000, is from a national land use survey performed via remote sensing and provided by the Chinese Academy of Sciences. We have resampled the data-set to 1 km to match the NDVI data and used the cropland coverage data layer for masking analysis. There are 84 agro-meteorological stations distributed uniformly within the study area (see also in Figure 2) from which ground phenology observations are collected and provided by the China Meteorological Administration.

2.2. Methods

We have taken pixels in which agro-meteorological stations located as sample points and extracted the NDVI time-series from these points for the evaluation of different reconstruction techniques.

2.2.1 Candidate reconstruction techniques

Comparisons are conducted among three widely used reconstruction techniques provided by TIMESAT including AG, DL and SG.

AG approach fits the time-series curve from local to global and the process can be divided into three parts: interval extraction, local fitting and functions merging. First, the peak and valley values of an original NDVI time-series curve are extracted. Then the left and right parts of the curve are separately fitted using a Gaussian function. The local fitting function is given as

$$f(t) = f(t; c_1, c_2, a_1, \dots, a_5) = c_1 + c_2 g(t; a_1, \dots, a_5) \quad (1)$$

where c_1 and c_2 are linear parameters that determine the base level and the amplitude; the non-linear parameters $a1$ to a_5 determine the shape of the Gaussian function $g(t,a_1,\ldots,a_5)$. These parameters can be generated via the calculation of the optimizing function. The global function is given as

$$F(t) = \begin{cases} \alpha(t)f_L(t) + (1-\alpha(t))f_L(t), t_L < t < t_C \\ \beta(t)f_C(t) + (1-\alpha(t))f_R(t), t_C < t < t_R \end{cases} \quad (2)$$

Here $\alpha(t)$ and $\beta(t)$ are cut-off functions that in small intervals around $(t^L + t^C)/2$ and $(t^C + t^R)/2$, respectively, smoothly drop from 0 to 1. The global function assumes the character of $f^L(t)$, $f^C(t)$ and $f^R(t)$ in, respectively, the left, central and right part of the interval $[t^L, t^R]$. The merging of the local functions to a global function is a key feature of the method. It increases the flexibility and allows the fitted function to follow a complex behavior of the time-series (Jönsson & Eklundh, 2002).

DL approach is very similar to AG and the only difference is local fitting function. For DL there is one parameter less than AG in the local fitting function and the basis function $g(t,a^1,\ldots,a^4)$ is in double logistic form as

$$f(t) = f(t;c_1,c_2,a_1,\ldots,a_4) = c_1 + c_2 g(t;a_1,\ldots,a_4) \quad (3)$$

Similarly parameters a_1 to a_4 determine the position of the inflection point and the rate of change at this point for the left and right parts. Also for this function the parameters are restricted in range to ensure a smooth shape (Jönsson & Eklundh, 2004).

SG approach is a simplified least square fit convolution firstly proposed by Savitzky and Golay for smoothing and computing derivatives of a set of consecutive values. The convolution can be understood as a weighted moving average filter with weighting given as a polynomial of a certain degree. The general equation of the simplified least square convolution for NDVI time-series smoothing is given as:

$$Y'_j = \frac{\sum_{i=-n}^{n} C_i Y_{j+1}}{N} \quad (4)$$

where Y is the original NDVI value, Y' is the resultant NDVI value, Ci is the coefficient for the ith NDVI value of the filter, and $N = 2n+1$ is the smoothing window size. The index j is the running index of the original ordinate data table. The smoothing window size and the degree of the smoothing polynomial will influence the results (Chen et al., 2004).

2.2.2 Best reconstruction assessment

NDVI time-series can reflect the annual cycle of vegetation growth and decline, but the observations are often contaminated by unfavourable atmospheric conditions. For the most strategies to reduce noise and reconstruct NDVI time-series, uncontaminated observations are regarded as 'true' and retained while contaminated observations are regarded as noise and modified. That is to say a proper reconstruction technique should perform best both on the fidelity of the reconstructed time-series to the uncontaminated observations and the correct modification of the contaminated observations. Pixel Reliability data layer from MOD13A2 data-set are used to identify whether one observation had been contaminated. This layer contains values that have been ranked into five categories to describe the overall pixel quality (Table 1) (Ramon, Kamel, Andree, & Huete, 2010). In this study, observations with rank value of 0 or 1 are identified as uncontaminated while observations with rank value of 2 or 3 are considered to be contaminated.

Table 1. MOD13A2 Pixel Reliability.

Rank Key	Summary QA	Description
−1	Fill/No Data	Not Processed
0	Good Data	Use with confidence
1	Marginal Data	Useful, but look at other QA information
2	Snow/Ice	Target covered with snow/ice
3	Cloudy	Target not visible, covered with cloud

Because clouds and poor atmospheric conditions usually depress NDVI values, most of the reconstruction techniques try to approach the upper envelope of the NDVI time-series curve. We have introduced 2 criteria presented by Julien and Sobrino (2010) and evaluated the performance of different reconstruction techniques by 1) distance to original data and 2) proximity to the upper envelope of the raw time-series. For the uncontaminated observations the average of the distance between raw and reconstructed time-series are used to represent distance to original data. For the contaminated observations the frequencies of reconstructed observations with lower NDVI than the raw observations are calculated to reflect proximity to the upper envelope of the raw time-series. Lower values of the preceding criteria correspond to shorter distance to original data and better proximity to the upper envelope of the raw time-series. A synthetic scored method has been proposed to determine the best reconstruction technique of the three candidates. For each of the two criteria to every sample point the reconstruction technique with lowest value gets one score while others get zero (when there are more than one lowest value each one gets one score). The best reconstruction technique is identified as the one with the highest average score of the two criteria's sum for all sample points.

2.2.3 Phenology detection from NDVI time-series

After the reconstruction the NDVI time-series curves can characterize the annual dynamic features of crop growth much better. Therefore, it is feasible to derive the phenology metrics of crops such as the dates of seedling emergence, heading and maturity from the smoothed NDVI time-series. Many studies have used threshold method to achieve this goal. Justice, Townshend, Holben, and Tucker (1985) took 0.099 as the NDVI threshold of vegetation growth start. Fischer (1994) and Markon, Fleming, and Binnian (1995) used 0.17 and 0.09 for the start of growth based on NDVI time-series, respectively. However, a significant limitation is that varied land surfaces require the use of different thresholds. Thus, a dynamic threshold as a percentage of the distance between minimum and maximum at the increase and decrease parts of the NDVI time-series curve is presented. Because 20% has been the most often used threshold (Heumann, Seaquist, Eklundh, & Jönsson, 2007; Li et al., 2009; Wu et al., 2010), here we also adopt it to determine the onset-of-growth and end-of-growth. In such, the onset-of-growth is determined to be the date when the reconstructed NDVI time-series curve increases to 20% of the overall level. Similarly, the end-of-growth is determined to be the date when the reconstructed NDVI time-series curve reduces to 20% of the overall level. For the peak-of-growth it is defined as the date corresponding to the maximum of the reconstructed NDVI time-series curve. To validate the derived crop phenology metrics ground phenology observations from agro-meteorological stations are used. In our study area of

Table 2. Definitions of Derived Phenology Metrics From NDVI Time-series and Corresponding Ground Phenology Observations for Crops in Northeast China.

Derived phenology metrics	Definition	Corresponding ground phenology observations
Onset-of-growth	Date when the reconstructed NDVI time-series curve increases to 20% of the overall level	Seedling stage
Peak-of-growth	Date when the reconstructed NDVI time-series curve reaches the maximum	Heading stage
End-of-growth	Date when the reconstructed NDVI time-series curve reduces to 20% of the overall level	Maturity stage

Northeast China where crops only have one harvest per year, the onset-of-growth date corresponds to the observed seedling stage; the peak-of-growth date represents the observed heading stage (tasseling stage for corn and pod-bearing stage for soybean); the end-of-growth date may correspond to the observed maturity stage (Table 2) (Li et al., 2012).

3. Experimentation and Discussion

3.1 Results

The experimentation results are consisting of five parts. The reconstructed NDVI time-series curves, the scores for selection of the optimal technique, the effect of reconstruction for NDVI images, spatial patterns and validation results of crop phenology metrics are showed.

3.1.1 Qualitative comparisons of candidate techniques

The original and reconstructed NDVI time-series with AG, DL and SG approaches for major crop types in the study area are shown in Figure 3. As seen in the figure, peak values of corn and wheat are higher than rice and soybean, which suggests differences exist in the NDVI time-series for various kinds of crops within the cropland. However, the fitted curves tend to approach the upper envelope of the original data for all the crop types. AG has obtained the persistent results and most of the noisy points in the NDVI time-series are modified correctly except the 13th observation for wheat as well as the 5th and 22nd observation for soybean. DL results are very similar to AG for corn and soybean, but it seemed to overestimate NDVI peak values for rice and wheat. SG results come much closer to the original data at the starts and ends of the NDVI time-series and overestimated the peak value for corn and soybean. In general, AG, DL and SG approaches are all effective for reducing noise and reconstructing NDVI time-series. It is difficult to determine which one perform best only by visual examination.

3.1.2 Selection of the optimal technique using scored method

Figure 4a shows the synthetic scores for three candidate reconstruction techniques. The two function fitting techniques—AG and DL perform similarly with synthetic score of 1.24 and 1.15, respectively. The filtering method - SG performs not so well with a lower synthetic score of 0.89. Take AG as reference, the largest difference of synthetic scores is 28% (between AG and SG). However, more complex patterns emerge when the scores

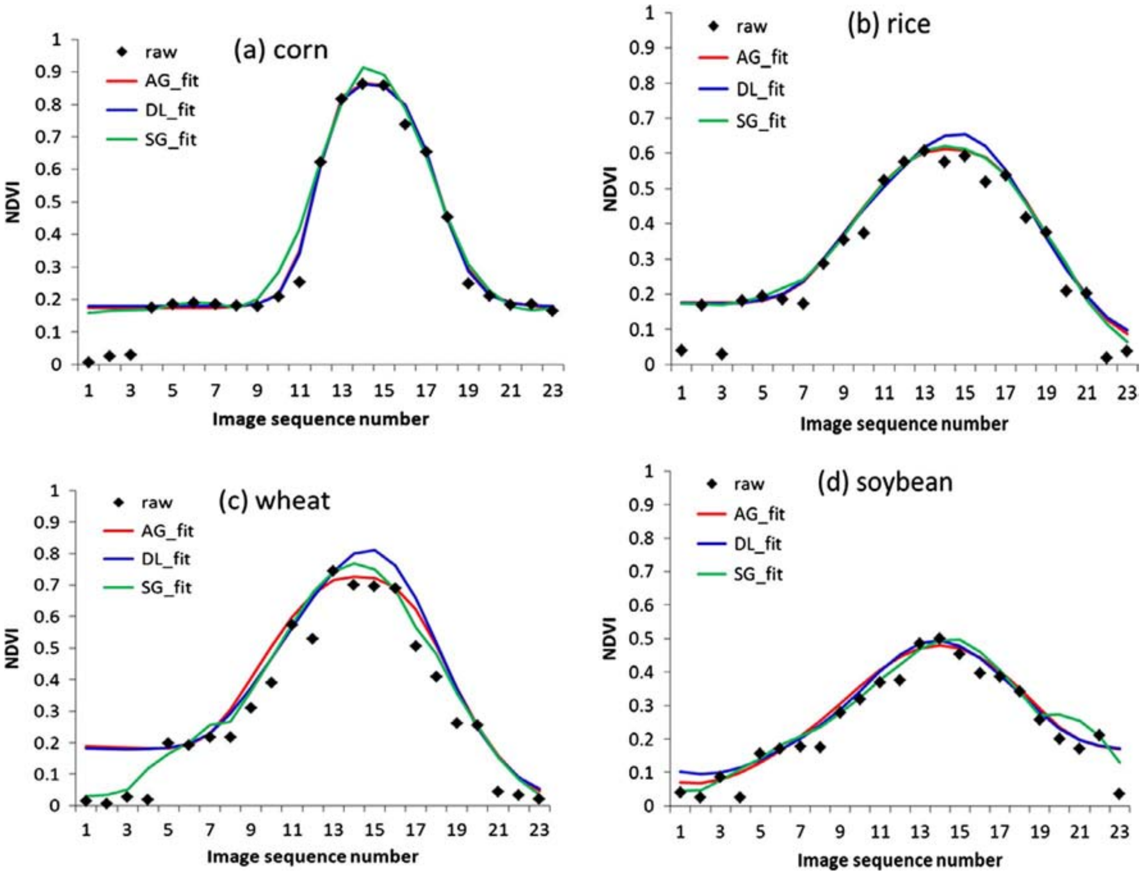


Figure 3. Comparison of the AG, DL and SG Approaches with the Original NDVI Time-series for Major Crop Types in the Study Area: Corn (a), Rice (b), Wheat (c) and Soybean (d).

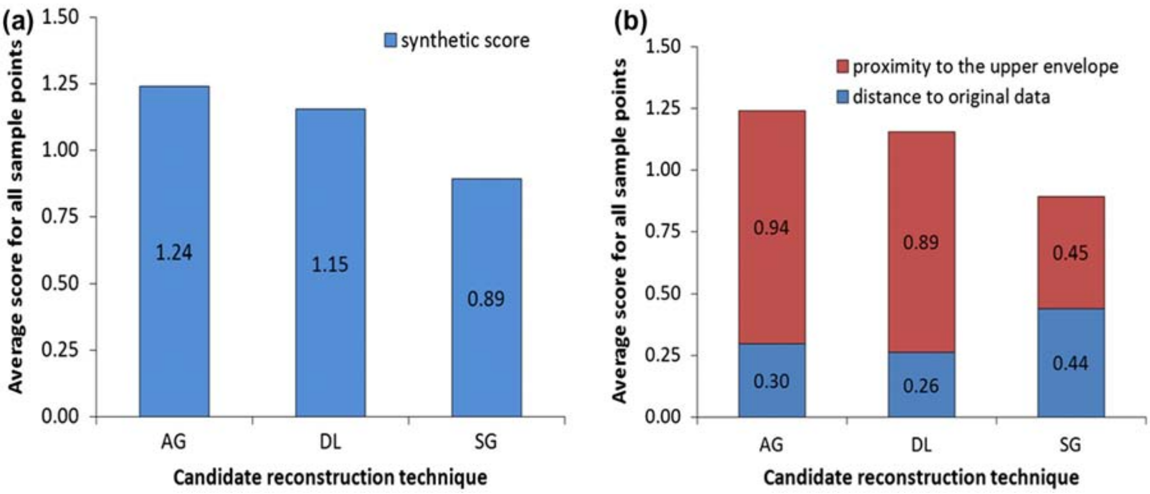


Figure 4. The Synthetic Scores (a) and the Scores of Distance to Original Data and Proximity to the Upper Envelope (b).

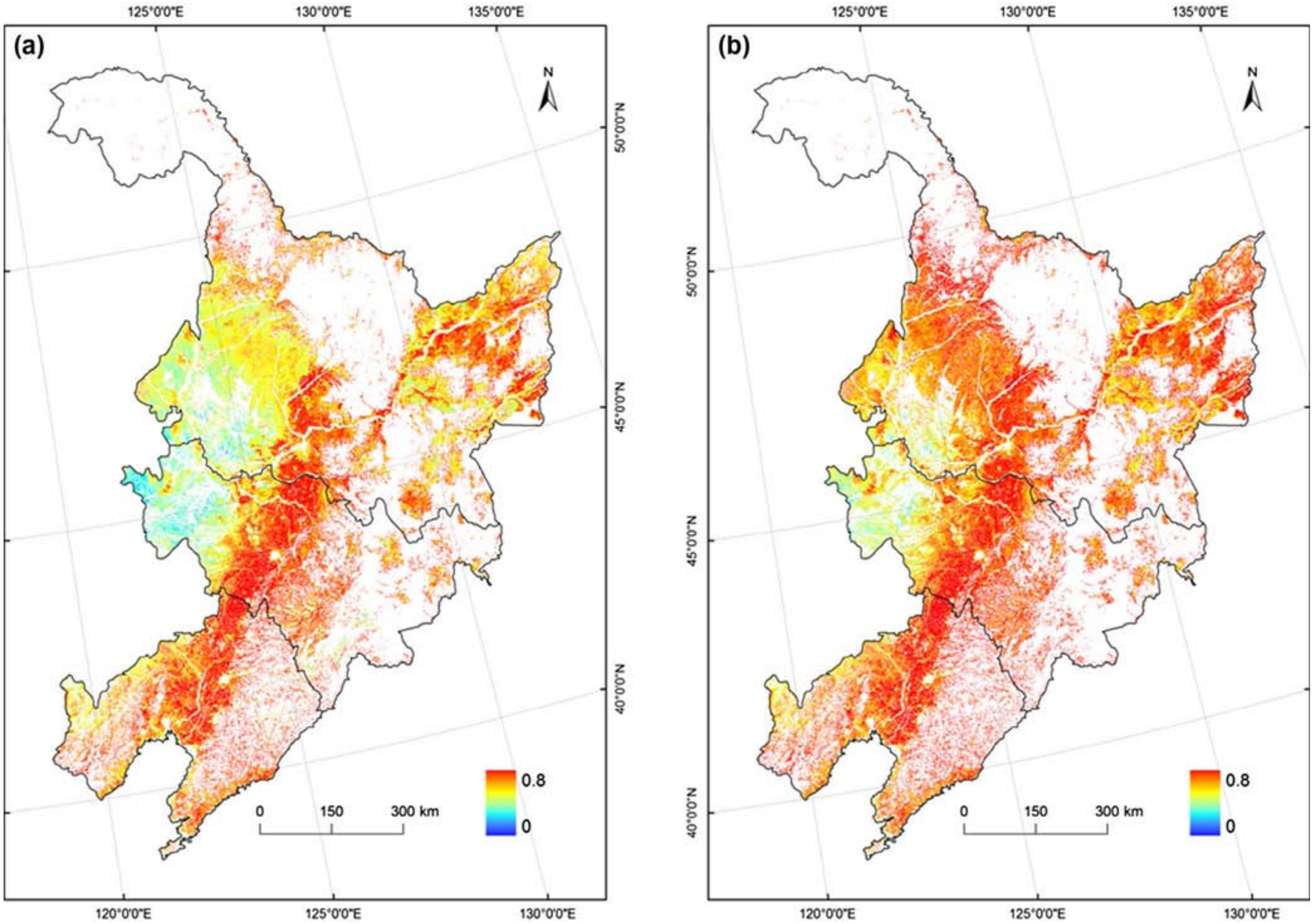


Figure 5. MODIS NDVI Image Produced through MVC Method of Cropland in Northeast China for the Period of 26-June to 11-July 2005 Before (a) and After (b) the Application of AG to the NDVI Time-series.

of two criteria we used to evaluate the performance of different reconstruction techniques are compared separately (Figure 4b). For distance to original data SG out-performs the other two techniques with the highest score of 0.44. As regards proximity to the upper envelope SG approach gets the lowest score of 0.45 and the technique with the highest score is AG (0.94). Although the pattern of proximity to the upper envelope score is same as the synthetic score, the largest difference up to 52% (between AG and SG) when take AG as reference. From these scores we can infer that SG is better able to retain the original data of the NDVI time-series, the two function fitting techniques, AG and DL prove better able to approach the upper envelope of the raw time-series. Overall, AG performs best among three candidate techniques under given conditions in this study.

3.1.3 Improvement of NDVI image by selected Technique

Examples of the original MODIS NDVI image produced through MVC method and the corresponding AG reconstructed results for cropland of the study area are shown in Figure 5. The acquisition dates of the NDVI values are between 26-June and 11-July 2005 during which the major crops of this region are growing vigorously. As such the NDVI values of the images should be on a high level. However, some regions still appear middle or low level of NDVI values in the original image probably due to atmospheric contamination (Figure 5a). Fortunately, Figure 5b shows that higher NDVI values can be reconstructed from contaminated data over the northeast and southwest of Heilongjiang province. In addition, although less obvious, improvement of the NDVI values can also be found over the west of Jilin province. Compared with the original NDVI image the corresponding AG reconstructed results are clearly more homogeneous over the regions mentioned above.

3.1.4 Spatial patterns of the detected crop phenology

Figure 6 shows the spatial distribution of the derived phenology parameters for the year 2005 in Northeast China. Earlier dates of onset-of-growth appear over the east part of Sanjiang Plain and north part of Songnen Plain in Heilongjiang province, predominantly from early April to early May. In comparison, most of the left regions have experienced onset-of-growth dates from middle May to early June. The spatial pattern of the peak-of-growth date indicates that for most areas it mainly occurs in early August. A few parts of north Songnen Plain and southeast Liaoning province have earlier peak-of-growth date in late July. Later peak-of-growth dates to middle August can be found in some parts of Sanjiang Plain, Songnen Plain, west Jilin province and southeast coast of Liaoning province. The end-of-growth dates are almost concentrated to middle September, when the major crops of this region are in the maturity stage. The spatial distribution of end-of-growth dates seems to be more homogeneous which suggests that autumn-harvesting crops might be dominant in the region.

3.1.5 Validation of the derived phenology metrics

Ground observed phenology metrics are used to verify the accuracy of the crop phenology detection. Because the data from agro-meteorological stations are recorded by manual work, there might have certain phenology metrics lost for some stations. Furthermore, a lot of stations often have records for more than one crop types. In our study ground observations are collected from stations having complete records of seedling emergence, heading and maturity dates for only one kind of crop. For the above requirements 16 stations are available. Scatterplot of the derived phenological metrics and ground observations from these stations are showed in Figure 7. The RMSEs between them for the onset-of-growth date, peak-of-growth date and end-of-growth date are also calculated. From the figure we can see that the majority of the points lie around the 1:1 line within the interval of ± 16 days. The RMSE between the derived phenological metrics and ground observations for the onset-of-growth date, peak-of-growth date and end-of-growth date is 18.21, 7.93 and 13.71 days, respectively. These results indicate that the accuracy of the detected cropland phenology metrics is pleasant and the errors are within a reasonable limit.

3.2 Discussion

Comparison of the AG, DL and SG approaches with the original data for major crop types were made using data from pixels where agro-meteorological stations with only one kind of crop located. The reconstructed NDVI time-series curves of three candidate reconstruction techniques are so close to each other that qualitative evaluation is difficult to conduct. The synthetic scored method presented in this study aims to choose a NDVI time-series reconstruction technique with the best performance through quantitative evaluation. Using the criteria simple yes or no result is given and one or zero score is got correspondingly to enlarge the difference of the reconstructed NDVI time-series curves by different reconstruction techniques. Pixel Reliability data are used to identify whether a pixel has been contaminated and different criterion is applied depending on the result of the identification. Previous studies using similar criteria to evaluate different techniques have not involve the quality assessment data might give evaluation results less accurately.

The scores for three candidate reconstruction techniques showed in the results have been averaged from all sample points. For the distance to original data the total score of AG, DL and SG equals one while for the proximity to the upper envelope is much more than one. This is because for the distance to original data there is only one technique with lowest value and gets one score. But for the proximity to the upper envelope there are usually more than one lowest value and each one get one score. Also the scores of proximity to the upper envelope are larger than the distance to original data (see especially for AG and DL in Figure 4b) can be explained. As commented in Section 3.2, the SG approach is much closer to the original data and better able to minimize the overall noise in NDVI time-series. The size of temporal window is very important when using SG for NDVI time-series reconstruction. In this study a recommend value of 4 by the TIMESAT software is used. which should be suitable for most common conditions. The AG and DL approach perform well in maintaining the integrity of the time-series for the purpose of phenology metrics detection. For most vegetation the left and right part of the NDVI time-series curves are not symmetrical. Asymmetric Gaussian function might describe this character better than Double Logistic function and that is why AG approach performs a little better than DL.

The reconstructed NDVI image produced by AG (Figure 5b) shows the effective modification of NDVI values contaminated by clouds and poor atmospheric conditions. Although the overall level of NDVI values is high, regions with middle or low level of NDVI values still exist. This case can reflect the difference of crop types within the cropland from which we can probably infer the crop structure in Northeast China. For the spatial pattern of onset-of-growth dates, regions with earlier observations might be single-season rice, which requires a relatively longer growth period. It's earlier sowing dates result in an earlier seedling stage so as to the onset-of-growth dates derived from NDVI time-series. Moreover, the spatial resolution of the data used in this study is 1 km, which usually involves many mixed pixels. The onset-of-growth dates of these pixels, for example with forests and crops mixed, might be influenced by the earlier budding dates of forests as seen by remote sensing. The effects of different kinds of crops and mixed pixels are also displayed in the spatial distribution of peak-of-growth and end-of-growth dates.

Validation of the derived phenology metrics has showed a pleasant result. The RMSEs between the remote sensing and ground observations for the peak-of-growth date (7.93) and end-of-growth date (13.71) are less than the MODIS NDVI composite periods of 16 days. While the RMSE of onset-of-growth date (18.21) is a little more

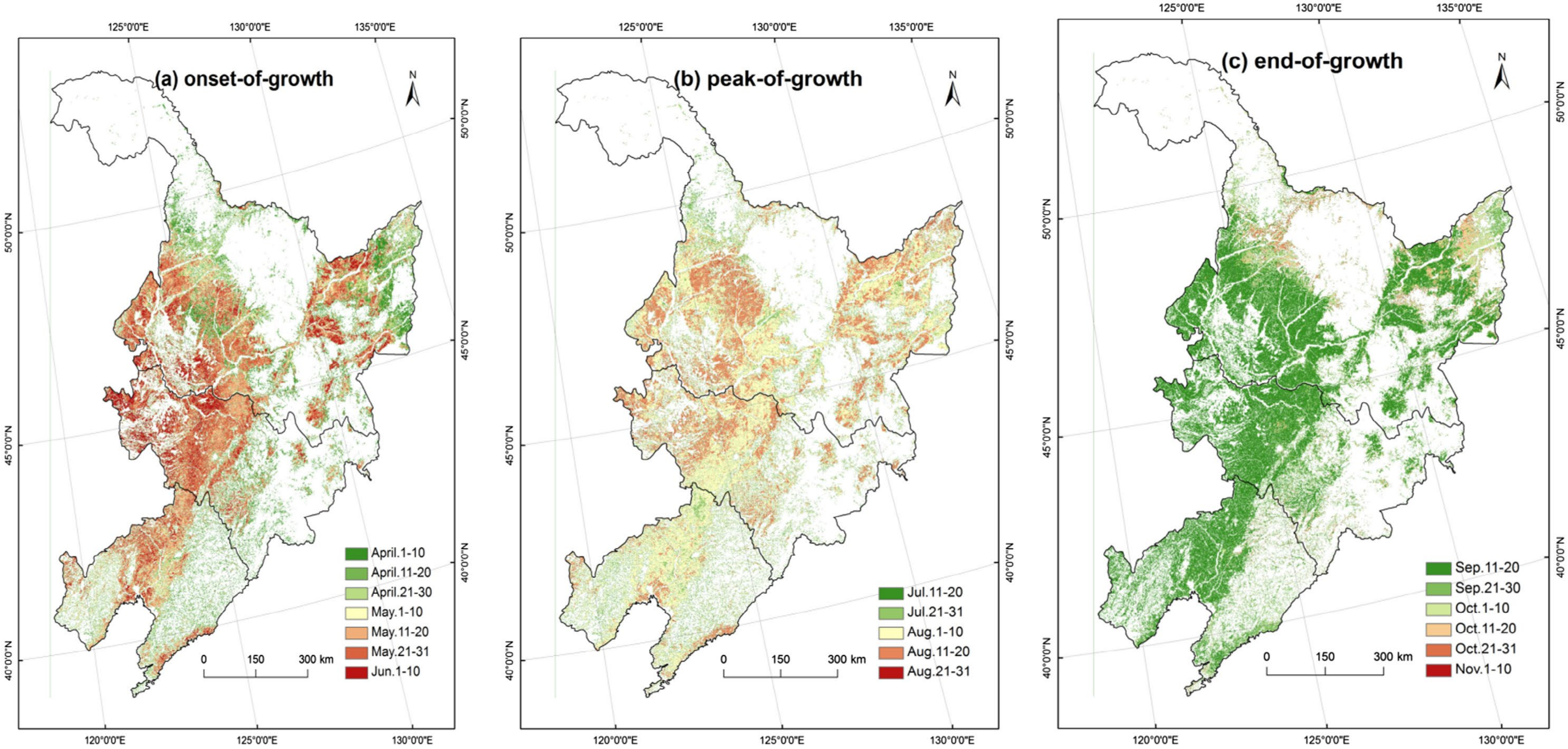


Figure 6. Spatial Patterns of the Derived Phenology Parameters Including the Onset-of-growth (a), Peak-of-growth (b) and End-of-growth (c) Date for the Year 2005 in Northeast China.

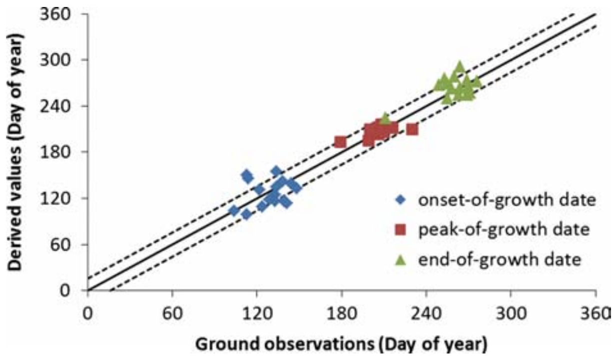


Figure 7. Scatterplot of the Derived Phenology Metrics and Ground Observations. The Solid Line is 1:1 Line and the Dashed Lines Show the Interval of ± 16 Days Which Equals the Composite Period of the MODIS NDVI Data We Used.

than 16 days. For our analysis we use composite data in which NDVI values are assigned to multiday periods rather than the specific date of image capture. The composite NDVI data is nominally temporal equidistant, in fact the NDVI values can be acquired on any day of the composite period. The actual temporal distance between adjacent NDVI values vary from 1 to 32 days, depending on the dates of acquisition of the data. Such an uncertainty in the temporal location of the NDVI values is much greater than the variations of phenology parameters. In addition, erroneous NDVI value positions affect the shape of the reconstructed time-series curve. Due to the loss of precision phenology metrics might be less precisely comparable with ground observations.

4. Conclusion and Future Work

Residual noise due to poor atmospheric conditions and other factors in the NDVI time-series has prevented the further application of the data. There is a strong need for the reconstruction of NDVI time-series before using them. An appropriate reconstruction technique can restore the 'true' NDVI time-series as much as possible. However, there are so many kinds of reconstruction techniques that we usually don't know which one should be chosen.

This study has presented a synthetic scored method for evaluating performances of different NDVI time-series reconstruction techniques in order to choose the optimal one for crop phenology detection. Three widely used techniques provided by the TIMESAT software package including Asymmetric Gaussian function fitting (AG), Double Logistic function fitting (DL) and Savitzky-Golay filtering (SG) are compared based on MODIS NDVI time-series. The results show that AG approach has got the highest score which means it outperform the two other techniques. The cropland NDVI values contaminated by unfavorable atmospheric conditions in the study area have been improved obviously after the reconstruction through AG. Spatial patterns of the cropland phenology metrics based on the reconstructed NDVI time-series produced by AG are reasonable. The errors of the detected crop phenology metrics are within an acceptable limit.

There are also some limitations in this study for improvement by future work. We have only evaluated three NDVI time-series reconstruction techniques over a regional scale of Northeast China. In the future work we should involve more techniques and compare them over a larger scale. For the detection of crop phenology a 20% threshold is used both for the onset-of-growth and end-of-growth. Sometimes a different value needs to be given for the end-of-growth date, because it is more influenced by human activities rather than natural factors. Moreover, NDVI time-series data from MODIS with coarse temporal and spatial resolution (16 days and 1 km) are used in this study. Datasets from other satellite sensors with relatively higher temporal and spatial resolution are encouraged to use for further studies.

Acknowledgments

This study is financially supported by the National High Technology Research and Development Program of China (2012AA12A304) and the National Natural Science Foundation of China (41201184 and 41271112). We are very grateful to Dr. Lvwen Huang from Northwest Agriculture and Forestry University of China for helpful advices to revise the paper. The authors also wish to thank the editor and the anonymous reviewers for their insightful comments and constructive suggestions.

Disclosure statement

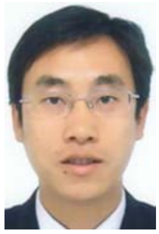
No potential conflict of interest was reported by the authors.

Notes on contributors

Wei Wei received a B.Ec. in Mechatronics Engineering from Beijing institute of technology, China, in 2007 and an M.Sc. degree in Agriculture Remote Sensing from Chinese Academy of Agricultural Sciences, Beijing, China, in 2012. He is currently a Ph.D. student in Institute of Agricultural Resources and Regional Planning at Chinese Academy of Agricultural Sciences, Beijing, China. His study interests include multi-sensor data collaborative applications remote sensing image processing and time-series analysis.

Wenbin Wu received a B.Sc. in Geography from The Central China Normal University, Wuhan, China, in 1998, an M.Sc. degree in Environmental Engineering from the College of Graduate of Chinese Academy of Agricultural Sciences, Beijing, China, in 2005 and a Ph.D. degree in Spatial Information from the University of Tokyo, Japan, in 2008. Currently he is the deputy director of the division of Key Laboratory of Agri-informatics, Ministry of Agriculture. His research is mainly focused on global/climate change impacts on agriculture using remote sensing and GIS, etc.

Zhengguo Li received a B.Sc. in Physical Geography from Beijing University, China, in 2002. Then he had been taking a successive postgraduate and doctoral program of study for doctoral degree and finally received a Ph.D. degree in Physical Geography from Beijing University, China, in 2007. He is currently an associate researcher in Institute of Agricultural Resources and Regional Planning at Chinese Academy of Agricultural Sciences, Beijing, China. His research interests focus on temporal-spatial patterns modeling.



Peng Yang received a B.Sc. degree in Environment Science from Wuhan University, China, in 1996 and an M.Sc. degree in Environmental Engineering from the College of Graduate of Chinese Academy of Agricultural Sciences, Beijing, China, in 2000 and a Ph.D. degree in Spatial information from Center for Spatial Information Science, University of Tokyo, Japan, in 2005. Currently he is the director of the Scientific Research Office in Institute of Agricultural Resources and Regional Planning at Chinese Academy of Agricultural Sciences, Beijing, China. His research interests are crop model and data assimilation.

Qingbo Zhou received a B.Ec. degree in Applied Meteorology from China Agricultural University, Beijing, China, in 1987, an M.Sc. degree in Physical Geography from Beijing University, China, in 1990 and a Ph.D. degree in Applied remote sensing from the Institute of Geography, Chinese Academy Sciences, Beijing, China, in 1993. He is currently the deputy director of Institute of Agricultural Resources and Regional Planning, Beijing, China. His research area mainly focuses on crop monitoring using remote sensing.

References

Beck, P. S. A., Atzberger, C., Høgda, K. A., Johansen, B., & Skidmore, A. K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment, 100*, 321–334. doi:10.1016/j.rse.2005.10.021.

Bradley, B. A., Jacob, R. W., Hermance, J. F., & Mustard, J. F. (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment, 106*, 137–145. doi:10.1016/j.rse.2006.08.002.

Brown, J. C., Kastens, J. H., Coutinho, A. C., Victoria, D. D. C., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment, 130*, 39–50. doi:10.1016/j.rse.2012.11.009.

Brown, M. E., de Beurs, K., & Vrieling, A. (2010). The response of African land surface phenology to large scale climate oscillations. *Remote Sensing of Environment, 114*, 2286–2296. doi:10.1016/j.rse.2010.05.005.

Butt, B., Turner, M. D., Singh, A., & Brottem, L. (2011). Use of MODIS NDVI to evaluate changing latitudinal gradients of rangeland phenology in Sudano-Sahelian West Africa. *Remote Sensing of Environment, 115*, 3367–3376. doi:10.1016/j.rse.2011.08.001.

Cai, H., Zhang, S., & Yang, X. (2012). Forest dynamics and their phenological response to climate warming in the khingan mountains, Northeastern China. *International Journal of Environmental Research and Public Health, 9*, 3943–3953. doi:10.3390/ijerph9113943.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment, 91*, 332–344. doi:10.1016/j.rse.2004.03.014.

Davis, J. C. (2002). *Statistics and Data Analysis in Geology* (3rd ed.). New York, NY: Wiley.

Fensholt, R. & Proud, S. R. (2012). Evaluation of earth observation based global long term vegetation trends - comparing GIMMS and MODIS global NDVI time series. *Remote Sensing of Environment, 119*, 131–147. doi:10.1016/j.rse.2011.12.015.

Fischer, A. (1994). A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Remote Sensing of Environment, 48*, 220–230. doi:10.1016/0034-4257(94)90143-0.

Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M. D., Neigh, C. S., & Reichstein, M. (2013). Trend change detection in ndvi time series: effects of inter-annual variability and methodology. *Remote Sensing, 5*, 2113–2144. doi:10.3390/rs5052113.

Geng, L., Ma, M., Wang, X., Yu, W., Jia, S., & Wang, H. (2014). Comparison of eight techniques for reconstructing multi-satellite sensor time-series ndvi data sets in the heihe river basin, China. *Remote Sensing, 6*, 2024–2049. doi:10.3390/rs6032024.

Guo, Z., Zhang, X., & Wang, Z. (2010). Responses of vegetation phenology in Northeast China to climate change. *Chinese Journal of Ecology*, 29, 578–585. Retrieved from: http://www.cje.net.cn/CN/abstract/abstract2272.shtml

Heumann, B. W., Seaquist, J. W., Eklundh, L., & Jönsson, P. (2007). AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982–2005. *Remote Sensing of Environment, 108*, 385–392. doi:10.1016/j.rse.2006.11.025.

Hird, J. N. & McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment, 113*, 248–258. doi:10.1016/j.rse.2008.09.003.

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing, 7*, 1417–1434. doi:10.1080/01431168608948945.

Jönsson, P. & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing, 40*, 1824–1832. doi:10.1109/TGRS.2002.802519.

Jönsson, P. & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences, 30*, 833–845. doi:10.1016/j.cageo.2004.05.006.

Julien, Y. & Sobrino, J. A. (2010). Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sensing of Environment, 114*, 618–625. doi:10.1016/j.rse.2009.11.001.

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*, 1271–1318. doi:10.1080/01431168508948281.

Kariyeva, J. & Van Leeuwen, W. J. D. (2011). Environmental drivers of ndvi-based vegetation phenology in Central Asia. *Remote Sensing, 3*, 203–246. doi:10.3390/rs3020203.

Kogan, F. & Sullivan, J. (1993). Development of global drought-watch system using NOAA/AVHRR data. *Advances in Space Research, 13*, 219–222. doi:10.1016/0273-1177(93)90548-P.

Landmann, T., Schramm, M., Huettich, C., & Dech, S. (2013). MODIS-based change vector analysis for assessing wetland dynamics in Southern Africa. *Remote Sensing Letters, 4*, 104–113. doi:10.1080/2150704X.2012.699201.

Li, J., Zhao, H., & Yang, X. (2011). Study on the seasonal dynamics of zonal vegetation of NDVI/EVI of costal zonal vegetation based on MODIS data: A case study of Spartina alterniflora salt marsh on Jiangsu Coast, China. *African Journal of Agricultural Research, 6*, 4019–4024. doi:10.5897/AJAR11.769.

Li, Z., Tang, H., Yang, P., Wu, W., Chen, Z., Zhou, Q., & Zou, J. (2012). Spatio-temporal responses of cropland phenophases to climate change in Northeast China. *Journal of Geographical Sciences, 22*, 29–45. doi:10.1007/s11442-012-0909-2.

Li, Z., Yang, P., Zhou, Q., Wang, Y., Wu, W., Zhang, L., & Zhang, X. (2009). Research on spatio-temporal pattern of crop phenological characteristics and cropping system in North China based on NDVI time series data. *Acta Ecologica Sinica, 29*, 6216–6226. Retrieved from: http://www.ecologica.cn/stxb/ch/reader/view_abstract.aspx?file_no=stxb200812180589&flag=1

Lovell, J. L. & Graetz, R. D. (2001). Filtering pathfinder AVHRR Land NDVI data for Australia. *International Journal of Remote Sensing, 22*, 2649–2654. doi:10.1080/01431160116874.

Lu, X., Liu, R., & Liang, S. (2007). Removal of noise by wavelet method to generate high quality temporal data of terrestrial MODIS Products. *Photogrammetric Engineering & Remote Sensing, 73*, 1129–1139. doi:10.14358/PERS.73.10.1129.

Markon, C. J., Fleming, M. D., & Binnian, E. F. (1995). Characteristics of vegetation phenology over the Alaskan landscape using AVHRR time-series data. *Polar Record, 31*, 179–190. doi:10.1017/S0032247400013681.

Michishita, R., Jin, Z., Chen, J., & Xu, B. (2014). Empirical comparison of noise reduction techniques for NDVI time-series based on a new measure. *ISPRS Journal of Photogrammetry and Remote Sensing, 91*, 17–28. doi:10.1016/j.isprsjprs.2014.01.003.

Motohka, T., Nasahara, K. N., Murakami, K., & Nagai, S. (2011). Evaluation of sub-pixel cloud noises on MODIS daily spectral indices based on in situ measurements. *Remote Sensing, 3*, 1644–1662. doi:10.3390/rs3081644.

Park, J., & Tateishi, R. (1998). *Correction of Time Series NDVI by the Method of Temporal Window Operation (TWO)*. Paper presented at the 19th Asian Conference on Remote Sensing, Manila

Park, S. (2013). Cloud and cloud shadow effects on the MODIS vegetation index composites of the Korean Peninsula. *International Journal of Remote Sensing, 34*, 1234–1247. doi:10.1080/01431161.2012.720043.

Ramon, S., Kamel, D., Andree, J., & Huete, A. (2010). *MODIS Vegetation Index User's Guide (MOD13 Series)*. Tucson, AZ: Vegetation Index and Phenology Lab, University of Arizona.

Reed, B. C., Schwartz, M. D., & Xiao, X. M. (2009). *Phenology of ecosystem processes*. New York, NY: Springer.

Roerink, G. J., Menenti, M., & Verhoef, W. (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing, 21*, 1911–1917. doi:10.1080/014311600209814.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., & Bounoua, L. (1996). A revised land surface parameterization (sib2) for atmospheric gcms. part i: Model formulation. *Journal of Climate, 9*, 676–705. doi:10.1175/1520-0442(1996) 009<0676:ARLSPF>2.0.CO;2.

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° by 1° NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing, 15*, 3519–3545. doi:10.1080/01431169408954343.

Shan, Z. & Xu, Q. (2013). Remote sensing of the seasonal naked croplands using series of ndvi images and phenological feature. *Intelligent Automation & Soft Computing, 19*, 207–216. doi:10.1080/10798587.2013.787188.

Taddei, R. (1997). Maximum Value Interpolated (MVI): A maximum value composite method improvement in vegetation index profiles analysis. *International Journal of Remote Sensing, 18*, 2365–2370. doi:10.1080/014311697217657.

Tuanmu, M., Viña, A., Bearer, S., Xu, W., Ouyang, Z., Zhang, H., & Liu, J. (2010). Mapping understory vegetation using phenological characteristics derived from remotely sensed data. *Remote Sensing of Environment, 114*, 1833–1844. doi:10.1016/j.rse.2010.03.008.

Velleman, P. (1980). Definition and comparison of Robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association, 75*, 609–615. doi:10.1080/01621459.1980.10477521.

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment, 114*, 2970–2980. doi:10.1016/j.rse.2010.08.003.

Viovy, N., Arino, O., & Belward, A. S. (1992). The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing, 13*, 1585–1590. doi:10.1080/01431169208904212.

Wu, W., Yang, P., Tang, H., Zhou, Q., Chen, Z., & Shibasaki, R. (2010). Characterizing spatial patterns of phenology in cropland of china based on remotely sensed data. *Agricultural Sciences in China, 9*, 101–112. doi:10.1016/S1671-2927(09)60073-0.

Xu, Q., Yang, G., Long, H., & Wang, C. (2013). Crop discrimination in shandong province based on phenology analysis of multi-year time series. *Intelligent Automation & Soft Computing, 19*, 513–523. doi:10.1080/10798587.2013.869109.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment, 84*, 471–475. doi:10.1016/S0034-4257(02)00135-9.