Spatio-Temporal Reconstruction of MODIS NDVI Data Sets Based on Data Assimilation Methods

Juan Gu *, Xin Li and Chunlin Huang

Cold and Arid Region Environmental and Engineering Research Institute, CAS, Lanzhou, 730000, China

Abstract. Consistent Normalized Difference of Vegetation Index (NDVI) time series, as paramount and powerful tool, can be used to monitor ecological resources that are being altered by climate and human impacts, since its temporal evolution is strongly linked to changes in the state of land surface. However, the noise caused mainly by cloud contamination, heavy aerosol, atmospheric variability and signal of background soil and bi-directional effects impedes NDVI data from being further applied. In this work, data assimilation method for NDVI was proposed to reconstruct high-quality spatially and temporally continuous MODIS NDVI data. The historical MODIS NDVI data are used to generate the background field of NDVI based on a simple three-point smoothing technique, which can generally capture the annual feature of vegetation change. At every time step, the quality assurance (QA) flags in MODIS VI products were adopted to determine empirically the weight between background field and observation of NDVI. Additionally, the gradient inverse weighted (GIW) filter algorithm is adopted further to remove spatial discontinuity. Finally, the more reliable NDVI data can be generated. This method is implemented by the 16-Day L3 Global 1km SIN Grid NDVI data sets covered west China during 2003-2006. Results indicate that the newly developed method is easy and effective in reconstructing high-quality MODIS NDVI time series.

Keywords: data assimilation, gradient inverse weighted filter, MODIS NDVI, spatio-temporal reconstruction

1. Introduction

Normalized difference vegetation index (NDVI) is the most widely used vegetation index due to their simplicity, ease of application, and wide-spread familiarity. Time-series MODIS NDVI products provide consistent, spatial and temporal comparisons of global vegetation conditions which are used to monitor the Earth’s terrestrial photosynthetic vegetation activity in support of phenologic, change detection, and biophysical interpretations (Van Leeuwen et al., 1999). The existing MODIS VI compositing algorithm includes three separate components: maximum value composite (MVC), constraint view angle - maximum value composite (CV-MVC), and bidirectional reflectance distribution function composite (BRDF-C) (Huete et al., 1994). However, the current NDVI product is still spatio-temporal discontinuous due to cloud cover, seasonal snow, atmospheric variability, bi-directional effects and instrument problems (Xiao et al., 2003; Moody et al., 2005). These biases limit the application of NDVI in vegetation dynamics monitoring and global change research.

Most often-used NDVI data sets utilized the Maximum Value Composite (MVC) algorithm to obtain a higher percentage of clear-sky data, but significant residual effects remain (Holben, 1986). Moreover, a number of mathematical filters have been applied to reduce noise and reconstructing high-quality time-series NDVI data sets in recent years, which are generally grouped into two types: (1) noise-reducing in the frequency domain such as fourier-based fitting methods (Roerink et al., 2000); (2) noise-reducing in the temporal domain such as the best index slope extraction algorithm (BISE) (Viovy et al., 1992), the weighed least-squares linear regression approach (Swets et al., 1999), the modified BISE filtering (Lovell and Graetz,
2001), the asymmetric Gaussian function fitting approach (Jonsson and Eklundh, 2002), the polynomial least squares operation (PoLeS) approach (Jose et al. 2002), Savitzky-Golay filtering (Chen et al., 2004) and mean-value iteration filter (Ma et al., 2006). Although the current methods have been widely used to reconstruct the NDVI profile, these methods exist several drawbacks and do not make full use of information available (Jonsson and Eklundh, 2002; Chen et al. 2004). If we can merge more auxiliary information available during reconstructing time-series NDVI data, the quality of NDVI data will be improved significantly. This idea is data assimilation originated from meteorology and oceanography (Daley, 1991), which has been widely used for initializing numerical weather prediction models (Daley, 1991) and improving estimation of soil moisture profile by assimilating remote sensing data or observations in situ in recent years (Huang et al., 2008a, 2008b). During the late two years, data assimilation methods have also been used to reconstruct remote sensing data dynamically. Gu et al. (2006) adopted optimal interpolation method to generate appropriate MODIS LAI data. He (2007) performed an experiment based on the gradient inverse weighted filter and objective analysis for improving estimation of time-series MODIS LAI data products. However, limited research focused on NDVI time series.

In this paper, we develop a simplified data assimilation method based on NDVI quality assurance (QA) data sets for reconstructing high quality, spatially and temporally continuous MODIS NDVI products. The scheme was tested and evaluated by MODIS 16-Day L3 Global 1km SIN Grid VI data sets (MOD13A2).

2. Data Analysis Scheme

In the following, we first briefly introduce the simplified data assimilation scheme, and then describe the main steps for implementing the new method according to the flowchart shown in Fig. 1.

- **Step 1:** The multi-year NDVI data for 2003 to 2005 were extracted from MODIS VI products to generate background field of NDVI based on three-point smoothing method. The smoothing results can describe the general feature of mean NDVI at every pixel that occurred during past years.
- **Step 2:** The grid data of 1km 16-day composite NDVI in 2006 and the corresponding QA data were extracted from MODIS VI products to be fed into the data analysis scheme. Then, the optimal NDVI value of in 2006 was calculated.
- **Step 3:** Post-processing based on gradient inverse weighted (GIW) filter was conducted.

![Fig. 1: The flowchart of reconstructing MODIS NDVI data.](image)

In general, a data assimilation scheme requires four elements of information: the model operator, the observation operator, the error estimators, and the minimization algorithm (Huang et al, 2008a, 2008b). A cost function is adopted to integrate the above components by fitting both the background field and time-dependent observations, and can be expressed as

\[
J(x_o) = (H(x_o) - y_o)^T R^{-1} (H(x_o) - y_o) + (x_o - x_b)^T B^{-1} (x_o - x_b)
\]  

(1)
where $J(.)$ is cost function, $H$ is observation operator, which is used to convert state variables to observations, $y_o$ is observation and $R$ is observation error covariance, $x_a$ is the analyzed state variables, $x_b$ is the background field, $B$ is background field error covariance. The main purpose of data assimilation is to minimize the cost function and obtain the optimal estimation of state variables. Considering the case of a single observation NDVI, the cost function can be simplified as

$$J(NDVI) = \frac{(NDVI_a - NDVI_b)^2}{B_{NDVI}} + \frac{(NDVI_a - NDVI_o)^2}{R_{NDVI}}$$

(2)

where $NDVI_a$, $NDVI_b$, and $NDVI_o$ are the analyzed value, background field and observation of NDVI, respectively. $B_{NDVI}$ and $R_{NDVI}$ are background and observation error variance of NDVI. The analyzed value of NDVI is simply found by setting the gradient of this cost function to zero:

$$\nabla J(NDVI) = 2\left(\frac{(NDVI_a - NDVI_b)}{B_{NDVI}} + \frac{(NDVI_a - NDVI_o)}{R_{NDVI}}\right) = 0$$

(3)

From the above formula, the analyzed value of NDVI is given as

$$NDVI_a = NDVI_b + K \times (NDVI_a - NDVI_b)$$

(4)

$$K = \frac{B_{NDVI}}{B_{NDVI} + R_{NDVI}}$$

(5)

where $K$ is weight coefficient. If the difference (error) associated with the background field are large compared with the observations, then $K$ will be large and a large correction will be made to the background to get the optimal value, and vice versa. However, it is very difficult to directly estimate the error due to the variable cloud and viewing geometry at every time step. Thus we try to apply the QA flag to determine empirically the $K$ of the corresponding pixel locations without calculating the background and observation errors in this study. When the QA flag is larger than 5, we consider that there exists large observation error and the corresponding $K$ is set to zero. If $K$ is equal to 1, the analyzed value is equal to the observation of NDVI. If $K$ is equal to zero, the optimal value is given as background field of NDVI.

### 2.1. Generation of Background Field

The method also assumes that noise signal is always negatively biased which is mainly caused by clouds and poor atmospheric conditions (Chen et al., 2004; He, 2007). Following this, a simple three-point smoothing technique (Gu et al., 2006) is applied to generate a background field for NDVI ($NDVI_b$) in combination with a selection of maximum NDVI values at every time step.

$$NDVI_b(t) = \text{MAX}[NDVI_o(t), (0.5*NDVI_o(t) + 0.25*(NDVI_o(t-1) + NDVI_o(t+1)))]$$

(6)

where $NDVI_o(t-1)$ and $NDVI_o(t+1)$ represent the observation data at the previous and the following time-step, respectively. In order to remove the noise more efficiently, the smoothing formula is repeated three times in this study. Results show that the smoothing method based on three-point “upper envelope” works well for single perturbations (errors) and do not remove the noise effectively for continuous contaminated data.

The final step for temporal smoothing background field of NDVI is the averaging for the 3 years (N=3) available in the context of this study (i.e., 2003–2005):

$$NDVI_{\text{mean}}(t) = \frac{1}{N} \sum_{\text{year}}^{2003,2005} NDVI_b(t, \text{year})$$

(7)

$$NDVI_b(t) = 0.5*NDVI_{\text{mean}}(t) + 0.25*(NDVI_{\text{mean}}(t-1) + NDVI_{\text{mean}}(t+1))$$

(8)

This method ensures that the smoothing profile of NDVI background field can capture the general feature that occurred during the past years. The first three years of NDVI data were used to generate background field of NDVI and the last year of NDVI data were applied as test data to evaluate the performance of our method.

### 2.2. Post-processing by gradient inverse weighted method

Since the above steps only take into account values in the temporal neighborhood, we still can not obtain high-quality spatially continuous NDVI data by the above steps. Then, we conduct post-processing additionally by gradient inverse weighted (GIW) filter to remove spatial discontinuity and enhance surface
features. The smoothing scheme is based on the observation that variations of gray levels inside a region are smaller than those between regions (Wang et al., 1981). A neighborhood about the interpolated point is identified and a weighted average is taken of the observation values within this neighborhood. And the weighting coefficients are the normalized gradient inverses between the center point and its neighbors. The general description is shown the following:

Let \( NDVI_d(i,j) \) be the gray level of analyzed NDVI at coordinate \((i,j)\). Define

\[
d(i; j;k,l) = |NDVI_d(i+k, j+l) - NDVI_d(i, j)|
\]

(9)

where \( k, l = -1,0,1 \), but \( k \) and \( l \) are not equal to zero at the same time. We denote this vicinity as \( V(i,j) \).

Then the output of the GIW filter \( (NDVIs) \) has the form

\[
NDVI_s(i, j) = 0.5 NDVI_d(i, j) + 0.5 y(i, j)
\]

(11)

where

\[
y(i, j) = \sum_{k,l \neq (0,0)} W(k,l) NDVI_d(i+k, j+l)
\]

(12)

and

\[
W(k,l) = \delta(i+j,k+l) \left[ \sum_{k,l \neq (0,0)} \delta(i+j,k+l) \right].
\]

(13)

3. Results and Analysis

MOD13A2 16-day composite products, covered west China (73-112° E, 26 - 50° N), were collected from 2003 to 2006 for evaluating the performance of our method proposed in this work. To analyze the quality of the reprocessed data, we compared the final reconstructing results with the original MODIS NDVI data for a particular summer day (on day 177) since the maximum total cloud amount in west china mainly occurred during from April to July (Chen et al., 2005).

MODIS NDVI maps obtained from the observed and the corresponding reconstructing results on day 177 are provided in Fig. 2. \( NDVIs \), \( NDVI_s \), and \( NDVI_d \) represent the original MODIS NDVI data and reconstructing results before and after the spatial filter, respectively. The observed and reconstructing NDVI are in better agreement in the spatial pattern as shown in the upper panel of Fig. 2. The main improvements of NDVI data are observed in the south part of west China covered by dense vegetation with an increase of
0.1-0.3 (Fig. 2(c)). As indicated in Fig. 2(d), it is advantageous in removing noise caused by cloud contamination and representing landscapes after the post-processing.

Indeed, the annual variation of NDVI mainly depends on the type of land cover. For shrub, crop and meadow, the highest yearly variation of NDVI is always in the leaf growing and falling seasons in west China. As for the Gobi, the annual NDVI value is very low and almost no significant variation, the differences is slight between the observed and optimal MODIS NDVI.

To further investigate the de-noising performance of this method, we analyzed the feature of NDVI time evolution (in Fig. 3) during 2006 for the observation (NDVI_o), background field (NDVI_b), reconstructing value after spatial filter (NDVI_s) combined with the corresponding QA flag of the selected test pixels. A very regular annual cycle of vegetation growth is observable in Fig. 3.

In the data analysis scheme, we make full use of QA information of MODIS VI data. QA has always been considered as high resolution quality indicator (Huete et al., 1999), which is combined with 16 bits and the bits 2-5 are called the VI usefulness index. On the basis of QA flag, most of problematic points caused by cloud contamination has been successfully identified and corrected. As Fig. 3 indicated, a large percent of NDVI observations are lower than the NDVI background value when QA flag > 5, but most of them are very close to that when QA flag < 5. During the data analysis process, the ancillary QA flags are effective means to determine the weight of NDVI background in order to avoid the loss of useful information.

4. Conclusions

The primary results in this work show that the simplified data assimilation method is able to reconstruct high-quality spatio-temporal continuous MODIS NDVI datasets efficiently. In this scheme, the multi-year average has been taken into account for generating background field to describe the general propagation of NDVI time-series. QA information of NDVI product was for the determination of the weight and the gradient inverse weighted method is used to spatial smooth processing. Finally, the optimal NDVI reconstructing results can be obtained. Analysis revealed that most of problematic points caused by cloud contamination can be successfully identified and corrected, and the reconstructing NDVI data are more realistic and more consistent with landscape. Besides, the algorithm is able to be executed automatically for NDVI data at different intervals. In future study, we will pay more attention to finding a more effective approach for the determination of weight.

5. Acknowledgements

This research is supported by the CAS (Chinese Academy of Sciences) Action Plan for West Development Project (grant number: KZCX2-XB2-09) and the NSFC (National Science Foundation of China) project (grant number: 40771036). Authors thank NASA Distributed Active Archive Centre for processing and providing MODIS VI data.
6. References


