A SIMPLIFIED DATA ASSIMILATION METHOD FOR RECONSTRUCTING TIME-SERIES MODIS NDVI DATA

Juan Gu, Xin Li, Chunlin Huang

Cold and Arid Regions Environmental and Engineering Research Institute,
Chinese Academy of Sciences, China
E-mail:azalea_gu@163.com

ABSTRACT

Normalized difference vegetation index (NDVI) is the most widely used vegetation index due to its simplicity, ease of application, and wide-spread familiarity. Time-series NDVI products have been proven to be a powerful tool to learn from past events, monitor current natural-resource conditions, extract canopy biophysical parameters and forecast terrestrial ecosystems on different scales. However, the current NDVI product is still spatiotemporally discontinuous mainly due to cloud cover, seasonal snow and atmospheric variability. In this work, a simplified data assimilation method is proposed to reconstruct high-quality time-series MODIS NDVI data. Results indicate that the newly developed method is easy and effective in reconstructing high-quality MODIS NDVI time series.

Index Terms— time series, NDVI, reconstruct, data assimilation

1. INTRODUCTION

Consistent Normalized Difference Vegetation Index (NDVI) time series has proven to be a tool for detecting and quantifying large-scale changes in plant and ecosystem processes associated with global change [1, 2]. However, NDVI data are often degraded by atmospheric conditions such as cloud cover, dust and aerosol, which impedes NDVI data from being further applied in monitoring the impact of climate change.

Several methods exist to minimize these biases of NDVI data sets in recent years, which are generally grouped into two types including noise-reducing in the frequency domain [3] and in the temporal domain [4-10]. However, these methods still have several drawbacks and do not make full use of information available [7,8]. In recent years, data assimilation method originated from meteorology and oceanography [11] has been paid more and more attentions to land surface science and hydrology [12], as well as high-quality time-series data reconstruction [13,14].

In this work, a simplified data assimilation method by the use of the ancillary MODIS NDVI QA data is proposed to reconstruct high-quality time-series MODIS NDVI data. Using MODIS NDVI data from 2003 to 2006, we illustrate the performance of the method for Heihe River Basin (96°-102°E, 37°-43°N) in west China and compare it with three existing methods. Results indicate that the newly developed method is easy and self-correcting to reconstruct high-quality MODIS NDVI time series.

2. METHOD DESCRIPTION

Similarly to other strategies for reconstructing high-quality time-series NDVI data [3-10], our method is also based on two general assumptions: (1) the NDVI data from a satellite sensor is primarily related to vegetation changes; and (2) clouds and poor atmospheric conditions usually depress NDVI values. In the following, we will briefly introduce the main steps of the method according to the flowchart shown in Fig.1.

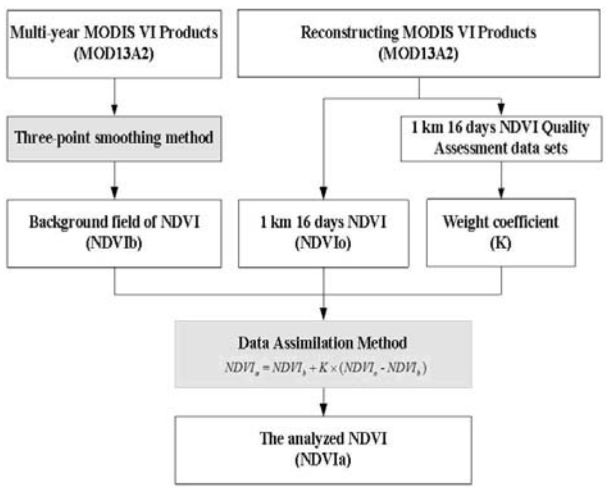

Fig.1. The flowchart of reconstructing MODIS NDVI data

2.1. Data Assimilation Scheme

In a general data assimilation scheme, the model operator, the observation operator, the error estimators, and the minimization algorithm need be determined.

A cost function is adopted to integrate the above components by fitting both the background field and time-dependent observations. It can be expressed as

$$J(x_a) = (H(x_a) - y_o)^T R^{-1} (H(x_a) - y_o) + (x_a - x_b)^T B^{-1} (x_a - x_b) \quad (1)$$

where *J(.)* is the 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 and *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.

In this study, both the state variable and observation are NDVI value, so the cost function can be simplified as

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

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 can be obtained by setting the gradient of this cost function to zero:

$$\nabla J(NDVI_a) = 2\frac{(NDVI_a - NDVI_b)}{B_{NDVI}} + 2\frac{(NDVI_a - NDVI_o)}{R_{NDVI}} = 0 \quad (3)$$

From the formula (3), the analyzed value of NDVI is given as

$$NDVI_a = NDVI_b + K \times (NDVI_o - NDVI_b) \quad (4)$$

$$K = \frac{B_{NDVI}}{B_{NDVI} + R_{NDVI}} \quad (5)$$

where *K* is weight coefficient. The general idea of data assimilation method used in this study is to determine the optimum values for the elements of *K* so that the total analysis error is minimized. If the errors associated with the background field are larger compared with the observations, then *K* will be high and most corrections will be made to the background to get the optimal value. Conversely, if the background errors are small compared with the observation errors, *K* will be small, and the background value will dominate the final analyzed value.

2.2. Generation of Background Field

We extracted 16-Day L3 Global 1km SIN Grid NDVI data sets of the Heihe River Basin in west China from MODIS vegetation index (VI) products (MOD13A2), spanning the period from 2003 to 2006. It is produced from surface reflectance data corrected for molecular scattering, ozone absorption, and aerosols [15-17]. Due to the basic hypothesis, the “upper NDVI envelope” smoothing strategy is usually adopted to reconstructing NDVI data [3-10].

In this paper, a simple three-point smoothing technique is combined with a selection of maximum NDVI values in order to provide the background field for NDVI at every time step [13].

$$\begin{aligned} NDVI_b'(t) = MAX[&NDVI_o(t), (0.5 * NDVI_o(t) \\ &+0.25 * (NDVI_o(t-1) + NDVI_o(t+1)))] \end{aligned} \quad (6)$$

where *t*-1 represents the data at the previous time-step and *t*+1 represents the data at the following time-step. In order to remove the noise more efficiently, the smoothing formula is repeated three times in this study. However, the NDVI evolution obtained by using three-point “upper envelope” smoothing method shows that this smoothing method works well when the perturbations (errors) are only for one data point 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) :

$$NDVI_{mean}(t) = \frac{1}{N} \sum_{year}^{2003,2005} NDVI_b'(t, year) \quad (7)$$

$$\begin{aligned} NDVI_b(t) = 0.5 * NDVI_{mean}(t) + \\ 0.25 * (NDVI_{mean}(t-1) + NDVI_{mean}(t+1)) \end{aligned} \quad (8)$$

This method ensures that the background field of NDVI is smooth and of a higher quality, which represent the gradual process of annual vegetation cycle and can capture the general feature that occurred during the past years.

2.3. Determination of Weight

Since cloud and viewing geometry are variable at every time step [15], it is very difficult to directly estimate background and observation errors for NDVI data. In this assimilating scheme, NDVI QA data sets are used to determine the weight required by data assimilation method in every time step. QA data sets provided by MODIS VI products can really reflect the quality of NDVI data [15,16]. So we try to apply the QA data sets to determine empirically the *K* of the corresponding pixel locations without calculating the background and observation errors in this study.

The QA data sets are combined with 16 bits. The bits 2-5 are called the VI usefulness index. Its value for a pixel is determined from multiple conditions, including aerosol quantity, atmospheric correction conditions, cloud cover, shadow, and sun-target-viewing geometry, so it is a high resolution quality indicator for us. Specific score is assigned to each condition and a sum of all the scores gives a usefulness index value for each pixel. According to VI usefulness index, we assigned empirically a value to *K* for each QA flag range from 0 to 1 and *K* are all zero when QA flag is greater than 5.

3. RESULTS

The simplified assimilation of MODIS NDVI data is done for 2006 following the strategy above. For the detailed assessment of the new method, comparisons will be carried out in the following from spatial pattern and temporal evolution, respectively.

3.1. Comparison at the Image Level

To analyze the quality of the reprocessed data, we compared the MODIS NDVI maps obtained from the observed and the corresponding reconstructing results. Figure 2 is an example of NDVI images of Heihe River Basin before and after reconstrucing on Julian days 49 and 321 of 2006.

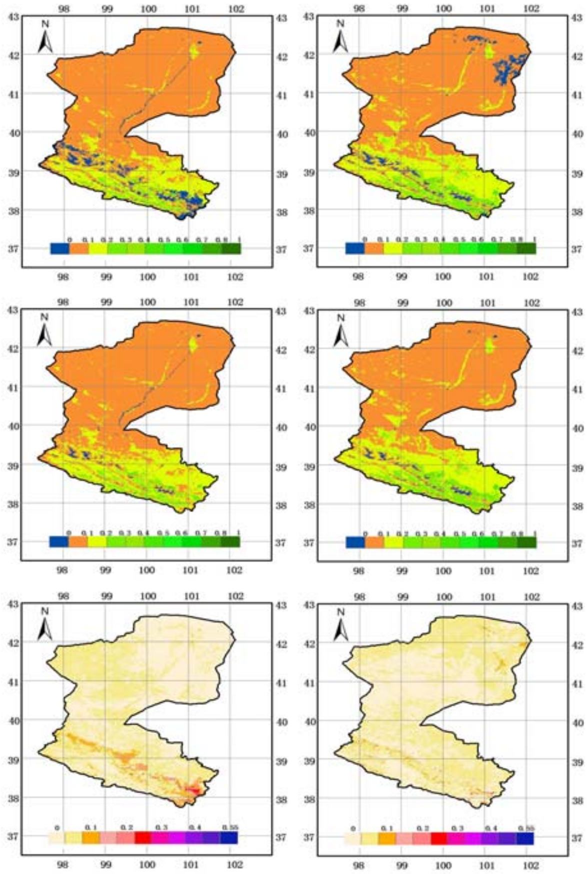

Fig.2. Comparison of the observed (up), analyzed (middle) and their difference (down) on days 49 (left panel) and 321 (right panel), 2006, respectively.

Generally, the reconstructed NDVI are in better agreement in the spatial pattern with the observed data. The significant improvements are observed in the southeast part on Julian day 49 and northeast part on Julian day 321 with an increase of 0.1-0.3. On the basis of QA flag, most of problematic points in the original image caused by cloud contamination and snow influence have been successfully identified and corrected to give higher values.

By contrast with land use map, it is advantageous to represent the real landscape of Heihe River Basin after reconstructing at the image level and improve data spatial uniformity while preserving seasonal changes for different land cover types.

3.2. Comparison at the Point Level

To further investigate the capacity of the scheme, the comparison is done with three often-used reconstructing methods.

In general, the annual variation of NDVI mainly depends on the type of land cover. Three regular annual cycles of vegetation growth are observable, which are selected from test points in Fig. 3 for representing deciduous broadleaf forest and shrubland. From the figures, we can see that the highest yearly variation of NDVI is always in the leaf growing and falling seasons of the study region.

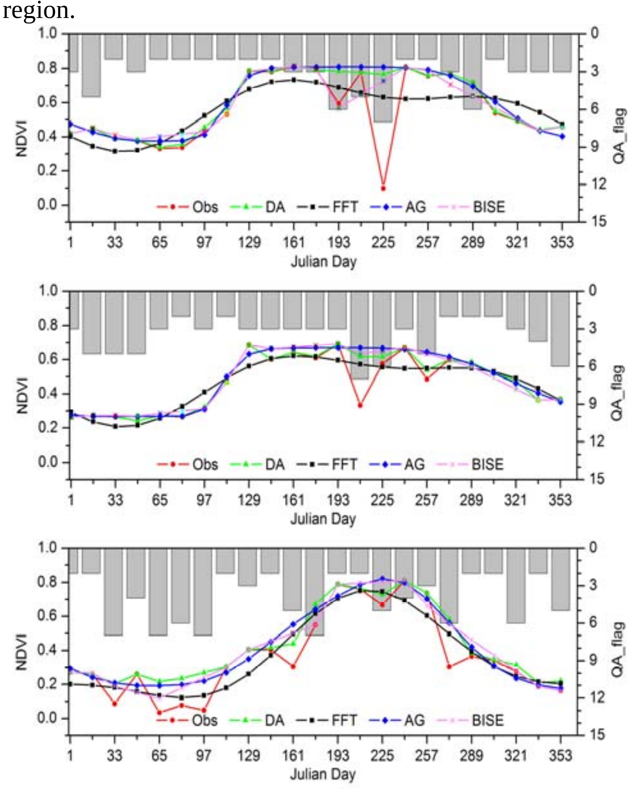

Fig.3. The final NDVI time series of three test points generated by the new method, the FFT method, the asymmetric gauss fitting and the BISE algorithm.

Comparison with the background value, we conclude that a large percent of NDVI observations are lower than that when QA flag > 5, and most of observations are very close to that when QA flag < 5. So the background field of NDVI can supply valuable information about multi-year average profile of vegetation. Though the background value has limitation in capturing the detail variation, it can also be used to identify and correct the noisy points to near the real condition as soon as possible. During the data analysis process, the ancillary QA flags are effective assistant to determine the weight of NDVI background in order to avoid the loss of useful information.

Comparison analysis also show that our method can describe the real NDVI temporal profile better than the Fourier series as well as the BISE algorithm, and slightly better than the asymmetric gauss fitting can. As Fig. 3 demonstrates, the BISE algorithm with the sliding window of three in this work can remove a large amount of noise in the time series, but it is not suitable for eliminating the continuous noise. The Fourier series fitting method can make the smoothest profile of NDVI time series. However, it is unable to describe the abrupt state and end of the growing before the spring increase and makes NDVI profile lose many local features. Under the condition of consecutive noises, BISE and FFT can not work well for the significant perturbation and generate the false exaggerated NDVI value (see in Fig. 3). The reconstructing scheme proposed in this work can remove noise effectively and the correct parts of the original data are also retained perfectly. Considering the determination of weights, our method make full use of the ancillary QA information and the multi-year average condition to adjust the problematic points. So based on the above both restrictions, the simplified data assimilation method are capable of supplying a reliable NDVI data. In the four methods, AG method has a similar result as ours, but it is more complex to conduct.

4. CONCLUSIONS

In this work, we use time-series data smoothing method to generate the NDVI background field, and takes advantage of ancillary QA data sets of NDVI product to determine the weight. Then the analyzed NDVI is calculated by the MODIS NDVI observation, background field, and empirical weight.

The final NDVI analyses based on this work are of higher quality, have smoother temporal evolution, and are more consistent with land cover databases. Additionally, this method can reduce continuous contamination in NDVI evolution without the effects of the adjacent points in temporal NDVI profile, because it uses the synchronous QA flag to estimate NDVI and make relevant correction.

The simplified assimilation method proposed in this work is very simple in theory and easy to implement for reconstructing time-series MODIS NDVI data at different intervals. The analyzed NDVI are more realistic and appropriate for global vegetation dynamics research.

5. ACKNOWLEDGEMENTS

This research is supported by the CAS (Chinese Academy of Sciences) Action Plan for West Development Project (grant number: KZCX2-XB2-09), the NSFC (National Science Foundation of China) project (grant number: 40771036) and the Chinese COPES project (GYHY200706005).

6. REFERENCES

[1] J. Fang, S. Piao, Z. Tang, et al.“Interannual variability in net primary production and precipitation,” *Science*, 293, pp. 1723a, 2001.

[2] J. Penuelas, L. Filella, “Responses to a warming world,” *Science*, 294, pp.793-794, 2001.

[3] G. J. Roerink, M. Menenti, W. Verhoef, “Reconstructing cloud free NDVI composites using Fourier analysis of time series,” *Int. J. Rem. Sens.*, 21, pp.1911-1917, 2000.

[4] N. Viovy, O. Arino, A. S. Belward, “The best index slope extraction (BISE): a method for reducing noise in NDVI time series,” *Int. J. Rem. Sens.*, 13, pp.1585-1590, 1992.

[5] P. J. Sellers, C. J. Tucker, G. J. Collatz, et al. “A global 1 by 1 NDVI data set for climate studies: 2. The generation of global fields of terrestrial biophysical parameters from the NDVI, *Int. J. Rem. Sens.*, 151, pp.3519-3545, 1994.

[6] J. L. Lovell, R. D. Graetz, “Filtering pathfinder AVHRR Land NDVI data for Australia,” *Int. J. Rem. Sens.*, 22, pp. 2649-2654, 2001.

[7] P. Jonsson, L. Eklundh, “Seasonality extraction by function fitting to time-series of satellite sensor data,” *IEEE Trans. Geosci. Rem. Sens.*, 40, pp.1824-1832, 2002.

[8] J. Chen, P. Jönsson, M. Tamura, et al., “A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter,” *Rem. Sens. Environ.*, 91, pp.332-344, 2004.

[9] E. G. Moody, M. D. King, S. Platnick, et al., “Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products, ” *IEEE Trans. Geosci. Rem. Sens.*, 43, pp.144-158, 2005.

[10] M. Ma, F. Veroustraete, “Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China, ” *Adv. Space Res.*, 37, pp.835-840, 2006.

[11] Daley, R. *Atmospheric Data Analysis* ,Cambridge University Press, New York, pp. 457, 1991.

[12] X. Li, C. L. Huang, T. Che, et al., “Development of a Chinese land data assimilation system: Its progress and prospects,” *Prog. Nat. Sci.*, 17, pp. 881-892, 2007.

[13] Y. Gu, S. Belair, J. F. Mahfouf, et al. “Optimal interpolation analysis of leaf area index using MODIS data,” *Rem. Sens. Environ.*, 104, pp.283-296, 2006.

[14] B. He, “A simple data assimilation method for improving the MODIS LAI time-series data products based on the object analysis and gradient inverse weighted filter,” *Chin. Opt. Lett.*, 5, pp.367-369, 2007.

[15] A. Huete, H. Liu, “An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS,” *IEEE Trans. Geosci. Rem. Sens.*, 32, pp.897-905, 1994.

[16] A. Huete, K. Didan, T. Miura, et al., “Overview of the radiometric and biophysical performance of the MODIS vegetation indices,” *Rem. Sens. Environ.*, 83, pp.195-213, 2002.

[17] V. Leeuwen, A. R. Huete, T. W. Laing, “MODIS vegetation index compositing approach: A prototype with AVHRR data,” *Rem. Sens. Environ.*, 69, pp.264-280, 1999.