Software for analysis of vegetation indices dynamics

To cite this article: B V Sodnomov et al 2018 IOP Conf. Ser.: Earth Environ. Sci. 211 012083

View the article online for updates and enhancements.
Software for analysis of vegetation indices dynamics

B V Sodnomov, A A Ayurzhanaev, B Z Tsydypov, E Zh Garmaev, and A K Tulokhonov

Baikal Institute of Nature Management SB RAS, Ulan-Ude, 670047, Russia

Abstract. The article describes a software package designed for processing and analyzing of vegetation indices based on satellite data. A program is developed in the free-distributed environment Octave. The input data of the program are vegetation indices (MODIS NDVI and EVI). The data are pre-processed by removing the peak values using the Savitsky-Golay filter and a seasonal decomposition by a moving average. To exclude the influence of snow cover and the period without vegetation, data of the growing season are used in the calculation. The vegetation period is determined by phenological parameters of the beginning and end of the growth period. The program determines the phenological parameters by a double logistic function, an asymmetric Gaussian function, and a fourth-degree polynomial to approximate the original time series. After the preliminary processing, a linear trend of the series, its statistical significance, the minimum and maximum values, coefficients of variation, etc., are calculated. The correlation function between the vegetation indices and the time series set by the user is also implemented. The output data are exported in raster and text formats. In conclusion, an example of using the software for calculating the NDVI trends in the Uda River basin in Transbaikalia is given.

1. Introduction

Analysis of long-term series of remote sensing data is one of the methods for assessing the state of the earth’s surfaces, including the vegetation cover, and forecasting its change. Increased requirements on the analysis of time series and increased computer technology productivity demand creating new algorithms of satellite data processing, modifying the existing ones, and implementing them in software packages for the end user.

The software package developed by the authors is designed for processing and analyzing high temporal resolution satellite data of the MODIS spectroradiometer. The use of MODIS data is optimal for continuous monitoring of vegetation cover on a regional scale. It is noted that one of the key advantages of the MODIS spectroradiometer is the possibility of forming and using continuous time series of satellite data [1].

MODIS data are stored in the HDF format. The HDF file contains several layers with different information: NDVI, albedo, quality indicators, etc. The users need to perform several steps of preprocessing the original HDF files in order to extract specific data useful for time series analysis. The stages of preliminary processing are as follows:
1) downloading products from web archives for specific dates and scenes;
2) generation of mosaics, resizing and re-projection of images;
3) extraction of data from the layers;
4) converting data into an easier-to-use format.

Recently, several solutions have been developed to automate one or more of the above steps: NASA REVERB, MRT, MODIS R, MODISStp, etc. [2]. These products have a number of advantages, but they do not allow further analysis of time series. One of the most widely used tools for analyzing the time series of MODIS is the TIMESAT program [3]. The TIMESAT has extensive tools to remove peaks, data smoothing, seasonal decomposition, etc. Approximation of time series for the purpose of determining phenological parameters is carried out using an asymmetric Gaussian function and a double logistic function. The disadvantages of the TIMESAT are the need to convert data into a binary format using third-party software, lack of the possibility of polynomial approximation of time series and analysis of the contribution of different phases of the vegetation period to the overall trend.
2. Materials
Data of a MODIS spectroradiometer installed on the Terra and Aqua satellites were used. The 16-day composites of the normalized vegetation indices NDVI and EVI (product MOD13Q1) from 2000 to 2016 with a spatial resolution of 250 m in the sinusoidal projection were used to construct the time series [4]. The MVC (Maximum Value Compositing) method was used to create a composite, in which the maximum value for 16 days is selected as the vegetation index value. Using the MVC minimizes the impact of atmospheric effects, including clouds and haze [5].

3. Methods
3.1. Data filtration
Time series can be represented as a sum of three components: trend, noise, and the seasonal one. The trend component is of greatest interest from the point of view of analysis of long-term series.

In order to remove the outliers and to reduce the noise level, the Savitzky-Golay filter is used [6]. This filter is optimal from the point of view of squared difference minimizing. The Savitzky-Golay filter is calculated as follows:

$$Y_i = \frac{\sum_{j=-m}^{i+m} C_i NDVI_{j+i}}{N},$$

where $C_i$ is the convolution coefficient for the $i$-th NDVI value, $NDVI_{j+i}$ are the initial NDVI values, $N$ is the window width ($2m + 1$), $j$ is the convolution coefficient number, and $m$ is half the width of the window. The width of the window is chosen to be a multiple of the vegetation period.

The moving average filter is used to remove the seasonal component and smooth the time series. The width of the window is also chosen in a multiple vegetative period. The moving average filter is a special case of the Savitsky-Golay filter, but instead of the convolution coefficients, the weight coefficients are used. In the case of a simple moving average, all weight coefficients are equal to 1.

3.2. Determination of phenological parameters
To reduce the influence of periods with open soil (early spring, late autumn) and snow (winter) on the trend, it is required to limit the time series to the values of the growing season [3, 7]. There are different approaches to determine the beginning and end of the growing season. For example, 1) when the vegetation index (VI) reaches an experimentally identified threshold value; 2) through search for points with maximum acceleration of growth or decrease in the zeros of the second derivative function [8].

When using the threshold method for the start (SOS) and end (EOS) of the vegetation period, a value when the NDVI becomes correspondingly higher and lower than the threshold level is taken annually for each pixel. Further, the constant length of the vegetation period (LOS) is determined for each pixel. For the beginning of the vegetation period, the most recent value is chosen, and for its end, the earliest one, i.e. the duration of the period is chosen to be minimal. Then the data of each year are limited to the period of vegetation. Also, there is an option in the algorithm which allows the user to change the threshold value.

To determine the phenological parameters, a preliminary approximation of the seasonal dynamics of the VI is required. To approximate the seasonal dynamics of the NDVI, a double logistic function (DLF), an asymmetric Gaussian function (AGF), and a fourth-degree polynomial are used. The DLF was first proposed in [9], and it is often used in analyzing the intra-annual dynamics of the VI [10, 11, 12].

The function has 6 coefficients, which denote the key phenological parameters determined by means of the VI: the minimum (winter) and maximum (summer) annual values (wNDVI and mNDVI, respectively), the date (S) and maximum growth rate (mS) (nominal start of the growing season), the date (A) and maximum fall speed (mA) (nominal end of the growing season):

$$NDVI(t) = wNDVI + (mNDVI - wNDVI) \times$$
The asymmetric Gaussian function combines the properties of both the power and exponential functions and has the following form:

\[
NDVI(t) = \begin{cases} 
  w_{NDVI} + (m_{NDVI} - w_{NDVI}) \times \exp \left( -\frac{t - t_{\text{max}}}{w_r} \right), & \text{if } t > t_{\text{max}} \\
  w_{NDVI} + (m_{NDVI} - w_{NDVI}) \times \exp \left( -\frac{t_{\text{max}} - t}{w_l} \right), & \text{if } t < t_{\text{max}}
\end{cases}
\]

where \( w_{NDVI} \) is the minimum annual value, \( m_{NDVI} \) is the maximum annual value, \( t_{\text{max}} \) is the maximum value index, \( w_r \) and \( w_l \) are the width of the right and left parts of the function, respectively, \( f_r \) and \( f_l \) are the steepness of the right and left parts of the function, respectively.

The first and second derivatives of the polynomial of the fourth degree have three significant extremum points, each of which is confined to changes in the phenological state of vegetation: the attainment of the maximum value, as well as the dates of the beginning of the NDVI growth in spring and autumn fall. The function is represented by the following formula:

\[
NDVI(t) = a(1) \cdot t^4 + a(2) \cdot t^3 + a(3) \cdot t^2 + a(4) \cdot t + a(5).
\]

4. Implementation of the algorithm

The developed software for processing time series has a data preprocessing block that allows excluding the use of additional software. The processing units allow not only to perform time series analysis, but also to involve additional data to obtain spatial correlation.

The software package is written in the Octave language, compatible with the MATLAB high-level language intended for numerical calculations. The source code for the Octave is freely distributed under the terms of the GNU General Public License.

The developed software package is based on a version of the algorithm presented in [13]. The package has a friendly user interface (Figure 1) and includes modules that provide the following tasks:

1. Reading of files of the original MODIS HDF-format, including selection of a layer and a fragment of data. The data is written into a cube two axes of which are spatial and the third is a temporal one.
2. Creating tiles. This module will allow one to split the data cube into parts. Subsequently the user can read not the whole data set for processing, but only the necessary tile, which greatly speeds up the processing process. The size of the tiles (and, correspondingly, their number) is determined by the user.
3. Export of time series.
4. Checking the data for errors and missing values, their replacement. The missing value is replaced by the average value of the composite in the preceding and subsequent years or by the average over the entire series.
5. Determination of the phenological parameters by logistic, Gaussian and polynomial functions.
6. Data filtering and smoothing. To remove outliers, the Savitsky-Golay filter is used. The order of the polynomial is 2, and the window width is equal to the number of composites in a year. To remove the seasonal component, the moving average is used.
7. Determination of the linear trend, its statistical significance. The obtained smoothed data are used to construct a linear regression model and determine the slope ratio (trend). The statistical significance of the trend is determined using the Student’s test. A significant trend is taken at a 95% level (\( p < 0.05 \)). The user has the ability to change the level of significance.
8. Calculation of the mean, minimum, and maximum values, the coefficient of variation, and other characteristics of the time series.
9. Calculation of spatial correlation. The possibility of calculating the Spearman correlation coefficient for each pixel is realized, the calculation can be made both with a time series of the processed data and with the user data. For example, time series of meteorological parameters, tree-ring
indices, indices of oscillations can serve as user data. For correct calculation of the correlation coefficient, the series of user data are reduced to the same dimension as the series of satellite data; otherwise, the ability to reformat the latter is provided.

10. Saving of data. The user can save the processing results in txt and tiff formats.

5. Example of software use
Figure 2 shows the result of processing of a NDVI time series from 2000 to 2016 with the use of the developed software for the territory of the Uda River basin. This territory is one of the most economically developed lands of Transbaikalia. It is revealed that the areas with a negative NDVI trend are confined to the steppe communities that are located in the intermountain basins. Areas with a positive NDVI trend refer to forest vegetation not affected by fires or logging.

![Figure 1. The software.](image)

![Figure 2. Map of NDVI trends of the Uda river basin.](image)
6. Conclusions
A software evaluation system to analyze the dynamics of vegetation indices has been created. The software package has a wide use and allows analyzing data without using any third-party software. According to the results of the study, a certificate of state registration of the computer program has been obtained. In the future it is planned to modify the complex to make it more convenient for the end user, as well as to expand the functionality of the algorithm in terms of accounting for landscape typing, features of the relief, etc.

Acknowledgements
This study was supported by the Russian Foundation for Basic Research (project no. 17-05-01059) and under a BINM SB RAS state order.

7. References
[1] Bartalev S et al 2003 Int. J. of Rem. Sens. 24 (9) 1977-82 (In Russian)
[2] Busetto L and Ranghetti L 2016 Computers & Geosciences 97 40-8
[3] Jonsson P and Eklundh L 2004 Computers & Geosciences 30 833–45
[4] Huete A et al 1999 Modis vegetation index (MOD13): Algorithm theoretical basis document, Version 3 (USGS Land Process Distributed Active Archive Center) p 129
[5] Van Leeuwen W J D et al 1999 Remote Sens. Environ. 69 264-80
[6] Savitzky A and Golay M J E 1964 Anal. Chem. 36 1627–39
[7] Pang G. et al 2016 Quaternary International 444 Part A 87-96
[8] Ayurzhanaev A A et al 2017 Proc. IV Int. Scient. Conf. Regional problems of the Earth remote sensing, September 12-15, 2017 Krasnoyarsk (Krasnoyarsk: Siberian Federal University Press) pp 187-90 (In Russian)
[9] Fischer A 1994 Int. J. Remote Sens. 15 1421–46
[10] Beck P S A et al 2006 Remote Sens. Environ. 100 321–34
[11] Busetto L et al 2010 Global Change Biology 16 2504–17
[12] Butt B et al 2011 Remote Sens. Environ. 115 3367–76
[13] Sodnomov B V et al 2018 J. Sib. Fed. Univ. Eng. Technol. 11 61-68 (In Russian)