Remote Sensing of Environment 114 (2010) 618–625

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Comparison of cloud-reconstruction methods for time series of composite NDVI data

Yves Julien *, José A. Sobrino

Global Change Unit, Image Processing Laboratory, University of Valencia, Spain

ARTICLE INFO

Article history:
Received 10 July 2009
Received in revised form 17 September 2009
Accepted 8 November 2009

Keywords:
GIMMS
NDVI
IDR
Atmospheric contamination
Data reconstruction

ABSTRACT

Land cover change can be assessed from ground measurements or remotely sensed data. As regards remotely sensed data, such as NDVI (Normalized Difference Vegetation Index) parameter, the presence of atmospherically contaminated data in the time series introduces some noise that may blur the change analysis. Several methods have already been developed to reconstruct NDVI time series, although most methods have been dedicated to reconstruction of acquired time series, while publicly available databases are usually composited over time. This paper presents the IDR (iterative Interpolation for Data Reconstruction) method, a new method designed to approximate the upper envelope of the NDVI time series while conserving as much as possible of the original data. This method is compared quantitatively to two previously applied methods to NDVI time series over different land cover classes. The IDR method provides the best profile reconstruction in most cases. Nevertheless, the IDR method tends to overestimate low NDVI values when high rates of change are present, although this effect can be lowered with shorter compositing periods. This method could also be applied to data before compositing, as well as to reconstruct time series for other biophysical parameters such as land surface temperature, as long as atmospheric contamination affects these parameters negatively.

1. Introduction

Remotely sensed data are the most widely used means of studying global vegetation change, especially in light of climate change concerns. However, cloud presence sometimes contaminates these data, and therefore obscures the observations in the visible to thermal infrared wavelengths. Other atmospheric contamination, such as dust, ozone or aerosols also adds noise in these data, as well as do bidirectional effects (Gutman, 1991; Holben & Fraser, 1984; Li & Strahler, 1992). Several approaches have been developed to identify clouds in the data, which can be divided in two groups. The first group, which relies on a spectral approach, uses all available spectral information to determine if a specific pixel includes cloud information. This approach is based on the spectral signature of the different clouds to identify them in the data, and therefore differs between sensors. An example of this approach is the method of Saunders and Kriebel (1988), developed for the AVHRR (Advanced Very High Resolution Radiometer) sensor onboard NOAA (National Oceanic & Atmospheric Administration) satellite series. The method presented by Ackerman et al. (1998) for the MODIS (MODerate resolution Imaging Spectroradiometer) sensor onboard the AQUA and TERRA platforms also relies on this approach. However, these spectral approaches do not provide any means to estimate the missing data due to the presence of atmospheric contamination.

On the other hand, the second group of methods, based on a temporal approach, does provide an estimation of the missing values through temporal interpolation. This group of methods exploits the fact that the retrieved data are linked to biological processes, and therefore should present continuity through time. Several methods have been presented to identify and interpolate contaminated values in time series data (Beck et al., 2006; Chen et al., 2004; Jönsson & Eklundh, 2002, 2004; Ma & Veroustraete, 2006; Roerink et al., 2000; van Dijk et al., 1987; Viovy et al., 1992), the latest methods usually performing better than the previous ones (Hird & McDermid, 2009). The criteria usually followed to assess the best reconstruction are its fidelity to the original cloud-free data and its ability to identify cloud contaminated values. Validation of the reconstructed time series is usually qualitative, since spatially extensive measurements (usually of the order of one square kilometer) under clear-sky conditions would be needed for a quantitative validation. Note that these methods do not distinguish between clouds and other atmospheric contamination of the data.

However, several of these methods (van Dijk et al., 1987; Viovy et al., 1992) were designed with their application to daily time series in mind, and therefore are difficult to apply on composited time series. Indeed, most of the publicly available databases of remotely sensed data for Earth observation, such as Pathfinder AVHRR Land (Smith et al., 1997) or GIMMS (Global Inventory Modeling and Mapping Studies; Tucker et al., 2005) are composited. This compositing aims at lowering atmospheric and cloud influence, as shown in Holben (1986), with different compositing periods ranging usually from 8 to 15 days. Even though composite data presents lower atmospheric contamination than

* Corresponding author. Tel.: +34 96 354 3781.
E-mail address: yves.julien@uv.es (Y. Julien).

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.11.001

raw time series, this composition process does not eliminate atmospheric contamination. For example, cloud cover can persist longer than the compositing period for some time periods (rainy season) or over some specific areas (tropical rainforests).

The work presented here introduces a new method for time series reconstruction, and compares this method to two already validated methods (Julien & Sobrino, 2009; Roerink et al., 2000) that can be implemented for composite data. These methods are tested on global GIMMS data for 2006.

2. Data

The GIMMS dataset (Pinzon, 2002; Pinzon et al., 2005; Tucker et al., 2005) compiles NDVI (Normalized Difference Vegetation Index; Tucker, 1979) images acquired by AVHRR sensor aboard NOAA satellites. The database ranges from July 1981 to December 2006. The data are composited over approximately 15 day periods (13 to 16 days) with the MVC (Maximum Value Compositing) technique (Holben, 1986), which minimizes the influences of atmospheric aerosols and clouds. The more than 25 years of data have been covered by 6 different satellites: NOAA-7, 9, 11, 14, 16 and 17. NDVI images are obtained from AVHRR channels 1 and 2 data, which correspond respectively to red (0.58 to 0.68 μm) and infrared wavelengths (0.73 to 1.1 μm).

This dataset, in spite of its limitation to NDVI data (no other channel information is available), presents several improvements regarding its predecessor, the PAL (Pathfinder AVHRR Land) dataset (Smith et al., 1997). The first improvement consists of a better data process, including navigation, sensor calibration and atmospheric correction for stratospheric aerosols. Another main improvement regards the correction of NOAA's orbital drift (Price, 1991), through the empirical mode decomposition (EMD) technique (Pinzon et al., 2005). The work presented here has been carried out using GIMMS NDVI data for the year 2006 only. The GIMMS data are provided along with flags, which indicate whether the data were obtained directly from satellite data, or if the data have been obtained from spline interpolation or from seasonal profiles, and if the data may correspond to snow. Therefore, no indication is provided as whether the data is atmospherically contaminated, and the GIMMS online documentation (available from the GIMMS website at http://glcf.umiacs.umd.edu/library/guide/GIMMSdocumentation_NDVIg_GLCF.pdf) does not provide any information on the reason why spline interpolation or seasonal profiles have been used.

Validity of the GIMMS dataset has been discussed in previous studies (Tucker et al., 2005; Zhou et al., 2001), so it is not assessed here. However, the GIMMS group itself points out two problems with the data: the volcanic eruption of Mount Pinatubo in June 1991, which decreased NDVI values, affecting particularly tropical regions; and the corrections made for extremely high solar zenith angles during winter for areas north of 65° N. Additionally, the GIMMS group advises not to draw local conclusions from the data since its NDVI present generalized patterns.

3. Methodology

We present here briefly the two methodologies to which we compare a new method thereafter described in detail. The comparison was carried out on GIMMS NDVI data for year 2006, identifying as contaminated the pixels for which the difference between original GIMMS data and reconstructed time series is higher than 0.05 NDVI units. This was done over different land covers identified from the IGBP (International Geosphere–Biosphere Programme) classification (Loveland et al., 2000), which was resampled to the GIMMS spatial resolution by agglomeration of the IGBP pixels. To this end, for each GIMMS resolution pixel, the ensemble of geographically overlapped original IGBP pixels was considered, and an IGBP class was assigned to the GIMMS resolution pixel only when 90% of the original IGBP pixels were from the same land cover class, which ensured a good homogeneity of intra pixel class description. The used IGBP classes (BATS — Biosphere–Atmosphere Transfer Scheme) are presented in Table 1, which correspond to all classes except inland and ocean water. For each one of these classes, one control point was chosen randomly to compare the three mentioned methods. The geographical location of these control points is shown in Fig. 1.

3.1. HANTS algorithm

The HANTS (Harmonic Analysis of NDVI Time Series) algorithm (Menenti et al., 1993; Roerink et al., 2000; Verhoef et al., 1996) was developed with the application to time series of NDVI images in mind. These images are usually composited by means of the so-called Maximum Value Compositing (MVC) algorithm in order to suppress atmospheric effects. These always have a negative influence on the NDVI and therefore taking the maximum value of the NDVI over a limited period tends to remove most contaminated observations. The HANTS algorithm also exploits this negative effect of atmospheric contamination on the NDVI, but in a different way. In HANTS a curve fitting is applied iteratively, i.e. first a least squares curve is computed based on all data points, and next the observations are compared to the curve. Observations that are clearly below the curve are candidates for rejection due to atmospheric contamination, and the points that have the greatest negative deviation from the curve therefore are removed first. Next a new curve is computed based on the remaining points and the process is repeated. Pronounced negative outliers are removed by assigning a weight of zero to them, and a new curve is computed. This iteration eventually leads to a smooth curve that approaches the upper envelope over the data points. In this way atmospheric contaminated observations have been removed and the amplitudes and phases computed are much more reliable than those based on a straightforward FFT (Fast Fourier Transform).

For our analysis of GIMMS NDVI of year 2006, the HANTS parameters were set as follows (for more detail on these parameters, see Roerink et al., 2000):

– number of frequency: 3 (24 = yearly; 12 = half-yearly; 8 = tri-yearly)
– suppression flag: low,
– invalid data rejection threshold: low threshold: 0; high threshold: 1,
– fit error tolerance: 0.02,
– degree of overdeterminedness: 5.

Another implementation of the HANTS algorithm for multi-temporal vegetation analysis can be found in Julien et al. (2006).

Table 1
Correspondence between control points and associated land covers in the IGBP classification scheme. See Fig. 1 for their geographical distribution.

Control point	Associated land cover
1	Crops, mixed farming
2	Short grass
3	Evergreen needleleaf trees
4	Deciduous needleleaf trees
5	Deciduous broadleaf trees
6	Evergreen broadleaf trees
7	Tall grass
8	Desert
9	Tundra
10	Irrigated crops
11	Semi-desert
12	Ice caps and glaciers
13	Bogs and marshes
14	Evergreen shrubs
15	Deciduous shrubs
16	Mixed forest
17	Forest/field mosaic