
VOLUME 114
ISSUE 3
15 MARCH 2010
ELSEVIER
Remote Sensing of Environment
http://www.elsevier.com/locate/rse

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

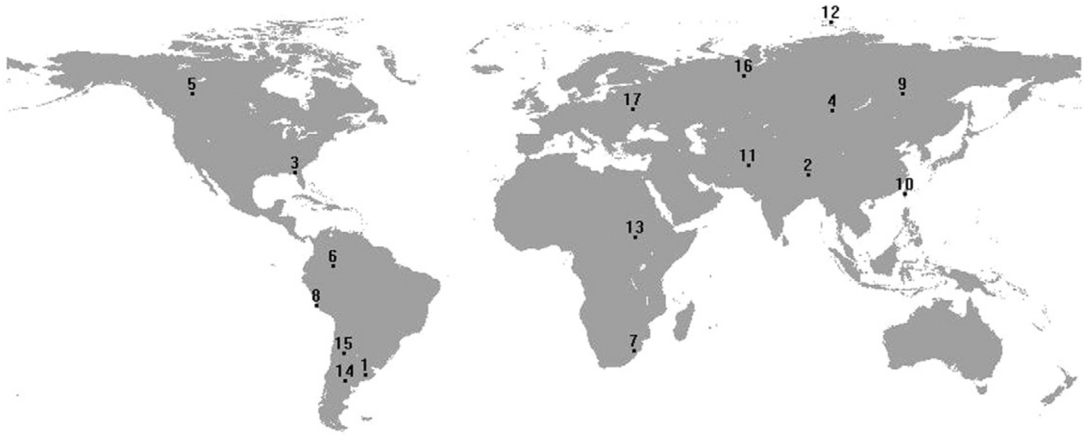


Fig. 1. Geographical distribution of the control points used for comparison of the reconstruction methods (see text for details). These control points have been selected randomly within each IGBP land cover class (shown in Table 1).

3.2. Double logistic fitting method

This approach has been previously applied in Julien and Sobrino (2009) to global GIMMS data, as a generalization of the method presented by Beck et al. (2006) for Siberia.

NDVI yearly evolutions are fitted to the following double logistic function (Beck et al., 2006):

$$NDVI(t) = wNDVI + (mNDVI - wNDVI) \times \left(\frac{1}{1+e^{-mS\times(t-S)}} + \frac{1}{1+e^{mA\times(t-A)}} - 1\right) \quad (1)$$

where NDVI(t) is the remotely sensed NDVI evolution for a given year (t = 0 to 364, in day of year), wNDVI is the winter NDVI value, mNDVI is the maximum NDVI value, S is the increasing inflection point (spring date), A is the decreasing inflection point (autumn date), mS is related to the rate of increase at S inflection point, and mA is related to the rate of decrease at A inflection point. All these parameters are retrieved iteratively on a pixel-by-pixel basis for the year 2006, by using the Levenberg–Marquardt technique (More, 1977). Since GIMMS NDVI images are bi-weekly, acquisition dates for each composite have been set to the day corresponding to the first day of the compositing period. A preliminary fit is conducted in order to estimate the dormancy period as the period before spring date and after autumn date. During this period, all eventual negative NDVI values are set to the highest positive value over the whole dormancy period. For those biomes with low NDVI amplitude variation throughout the year (lower than 0.1 NDVI unit), no fitting procedure is carried out, mNDVI and wNDVI being fixed to the mean value of NDVI over the considered year. These biomes correspond to arid or frozen areas, as well as cloud-free evergreen vegetation.

3.3. Iterative interpolation for cloud-free data reconstruction

This new method, named iterative Interpolation for Data Reconstruction (IDR), also exploits the tendency of cloud and atmospheric influence to lower NDVI values. Additionally, since NDVI is a proxy for vegetation greenness, its temporal variation should be smooth and continuous. Therefore, the following method was developed.

For each pixel, a time series of NDVI values are extracted from the GIMMS data. For each date, an alternative NDVI value is computed as the mean between the immediately preceding and following observations. An alternative NDVI time series is therefore obtained, and compared to the original time series. The date corresponding to the maximum difference between the alternative and original time series is identified, and the corresponding NDVI value in the original time series is replaced with the corresponding NDVI value in the alternative time series. This replacement is carried out only when the maximum difference between both time series is higher than 0.02 NDVI units. Then a new alternative time series is computed from the modified time series, and the process is iterated until convergence is reached. This process allows to progressively increase one by one the low and discontinuous NDVI values (corresponding to atmospherically contaminated values) until the upper envelope of the NDVI time series is reached. The methodology is somewhat similar to the one presented in Ma and Veroustraete (2006), with the difference that the IDR method is carried out from the data itself, and not from a comparison to an average of different years, which can be problematic for areas with high interannual variability, for areas suffering a land cover change, or when the acquired time series length is short.

3.4. Best reconstruction assessment

To determine the best reconstruction method among the 3 approaches presented above, some criteria have to be determined. As stated in the Introduction, an adequate reconstruction method complies two conditions on the fidelity of the reconstructed time series to the uncontaminated observations and on the correct identification of the contaminated observations. Since ground truth is unavailable, we decided to rely on the following criteria: distance to the raw observations and upper envelope proximity. Distance to the raw observations aims at quantifying the fidelity to the uncontaminated data, and is estimated as the average of the absolute differences between raw and reconstructed time series. Therefore, lower values for this distance criterion correspond to a better fidelity to original data. As for the upper envelope approximation criterion, it takes advantage of the fact that atmospheric contamination tends to decrease NDVI values. This criterion is thus estimated for the whole time series as the frequency of reconstructed observations with lower NDVI than the raw observations, ranging between 0 and 1. As for the preceding criterion, a lower value corresponds to a better proximity to the upper envelope of the raw time series. The analysis of both criteria therefore allows assessing the best reconstruction method for all classes.

4. Results

Examples of the original GIMMS NDVI values and the IDR NDVI values are shown in Fig. 2 for two dates (January and July 2006) in South America. This figure shows that high NDVI values can be reconstructed from contaminated data in January over the Amazonia region for example, where reconstructed NDVI values are clearly more homogeneous than in the raw data image.

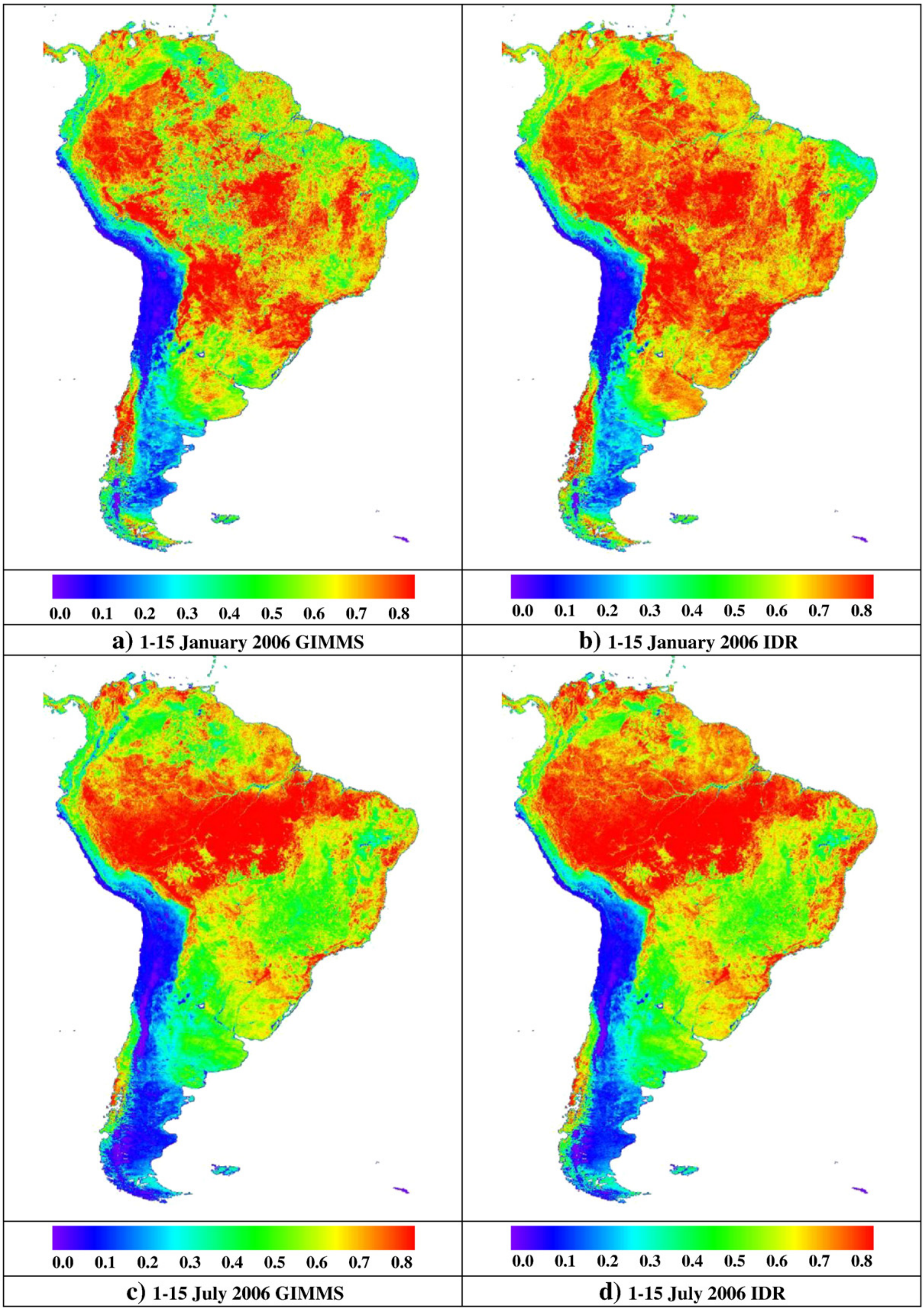
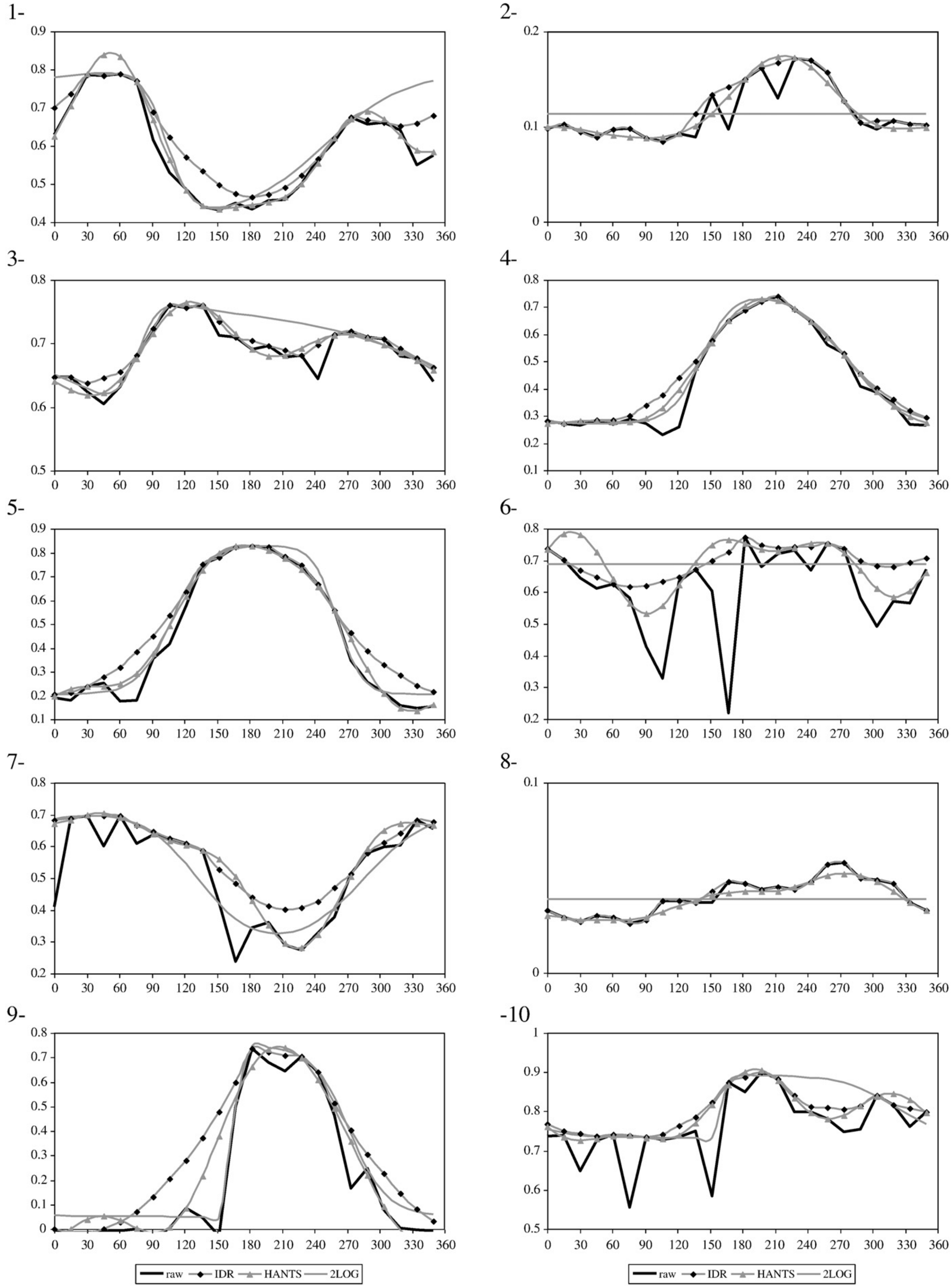


Fig. 2. NDVI values for South America before and after the application of the IDR (iterative Interpolation for Data Reconstruction) method for two compositing periods. A striking improvement can be observed over the Amazonia for the January image, although more subtle differences can be observed over the whole continent for both January and July images.

The reconstructed NDVI time series with the HANTS, double logistic (2LOG) and IDR approaches for each control point are shown in Fig. 3. This figure shows that for all control points, the GIMMS data present some discontinuities, even in the case of deserts and semi-deserts, although discontinuities for these land covers have a smaller amplitude. In a general manner, the HANTS approach tends to


1-
2-
3-
4-
5-
6-
7-
8-
9-
-10
raw IDR HANTS 2LOG
raw IDR HANTS 2LOG

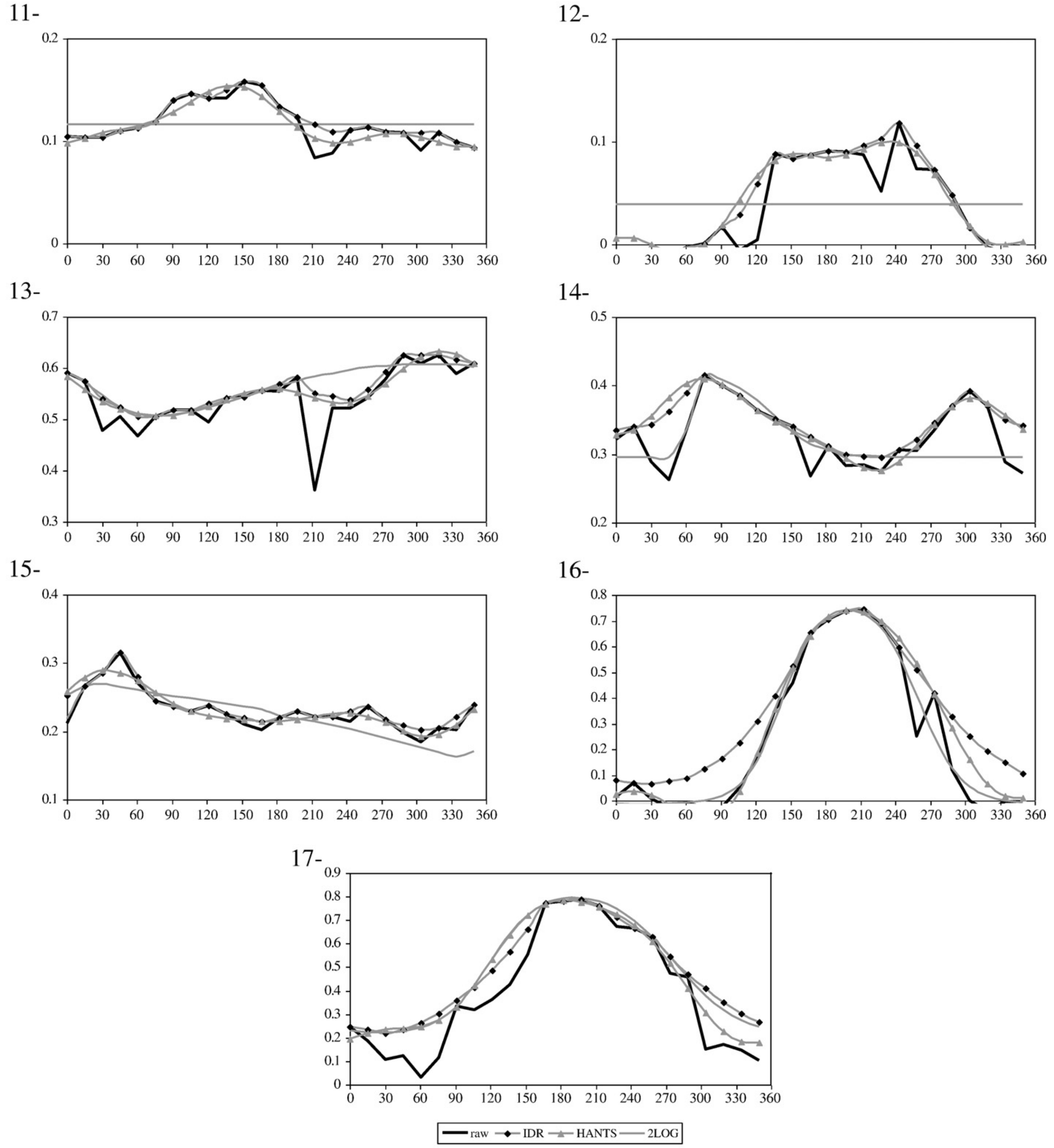


Fig. 3. Comparison of the HANTS, double logistic (2LOG), and iterative interpolation of data reconstruction (IDR) approaches with the original GIMMS data for the different IGBP control points described in Table 1 and Fig. 1. Note that vertical scale is not uniform (for better visualization), and therefore the contamination in non-vegetated profiles (2, 8, 11 and 12) is low.

approach the upper envelope of the NDVI time series from below, except when a plateau is present, for which maximum NDVI is overestimated (see for example control point 1). Moreover, in the presence of several successive atmospherically contaminated values (control points 6, 13, 14, and 17), the HANTS reconstructed time series tend to underestimate NDVI values for the dates surrounding persistent contamination.

The double logistic approach implemented in this work does not carry out a double logistic fit when NDVI annual amplitude (once contaminated values are discarded) is lower than 0.1 NDVI units. This explains the constant reconstructed profile for control points 2, 8, 11 and 12, which all correspond to low vegetated land covers, and for control point 6, corresponding to evergreen forest. Due to its definition, the double logistic function describes correctly vegetation

Table 2
Determination of the best fitting approach as regards distance to the raw data and proximity to its upper envelope for each control points (IDR: iterative Interpolation for Data Reconstruction; HTS: HANTS algorithm; 2LOG: double logistic fitting procedure). See text for details.

Class	Distance			Upper envelope		
	IDR	HTS	2LOG	IDR	HTS	2LOG
1	36.79	**15.29**	50.67	**0.00**	0.42	0.21
2	**4.83**	8.13	X	**0.00**	0.38	X
3	**8.79**	11.08	20.00	**0.00**	0.50	0.29
4	27.42	20.71	**20.50**	**0.00**	0.33	0.25
5	56.33	**27.79**	32.83	**0.00**	0.33	0.29
6	81.33	**78.50**	X	**0.00**	0.29	X
7	60.08	**48.21**	50.63	**0.00**	0.33	0.46
8	**0.25**	2.33	X	**0.00**	0.54	X
9	103.29	62.63	**57.71**	**0.00**	0.33	0.17
10	39.00	**35.92**	46.38	**0.00**	0.38	0.33
11	**3.25**	6.92	X	**0.00**	0.58	X
12	**7.08**	13.54	X	**0.00**	0.33	X
13	**19.67**	22.25	29.71	**0.00**	0.33	0.42
14	**20.17**	22.67	24.96	**0.00**	0.42	0.46
15	**5.71**	10.08	20.33	**0.00**	0.42	0.63
16	95.63	47.71	**36.04**	**0.00**	0.33	0.46
17	85.75	**77.96**	89.67	**0.00**	0.29	0.17

with one growing cycle, which helps describe NDVI time series correctly for control points 3, 10 and 13 in spite of persistent cloud contamination, although this same characteristic hinders the reconstruction for control point 14, which presents an apparent double growing season. Moreover, this double logistic approach does not identify the upper envelope of the original time series for control points 7, 13, 14 and 15.

Finally, the IDR approach identifies correctly the upper envelope of the NDVI time series, although it tends to overestimate the upper envelope for the dormancy period of the vegetation when the transition between vegetation activity and dormancy is abrupt (control points 1, 7 and 9). Moreover, this approach conserves the data values for the pixels not identified as contaminated during the iterative reconstruction, which is not always the case with the other two approaches (see for example control points 2, 6, 8, 13, 14, and 15).

Using the two criteria presented in the Methodology section, we can assess quantitatively which approach performs best. Values for these criteria for all control points and all 3 approaches are presented in Table 2. As regards distance to original data, the IDR method provides the highest fidelity to raw data in 8 cases (control points 2, 3, 8, 11, 12, 13, 14 and 15), while the HANTS approach provides highest fidelity in 6 cases (control points 1, 5, 6, 7, 10 and 17). The double logistic approach provides the highest fidelity to raw data in only 3 cases (control points 4, 9 and 16). As for the proximity to the upper envelope of the raw data, the IDR method performs best for all control points. The consideration of both criteria therefore identifies the IDR approach as the best of the three for the removal of atmospheric contamination.

5. Discussion

Direct comparison of Fig. 2 with the IGBP classification (not shown, see Loveland et al., 2000) mentioned in the Methodology section shows that intra-class homogeneity is higher for the reconstructed NDVI than for the raw data. For example, it is clearly possible to make out the limits of Amazonia (IGBP-BATS class 6) from the July IDR image (red and orange colors), whereas Amazonia's northern border is more difficult to identify in the July GIMMS composite. This is also true for January composite, where Central Amazonia appears with NDVI values ranging from 0.4 to 0.8 in GIMMS composite, which is surprising given the expected evergreen character of the vegetation. On the other hand, in the IDR composite, this range is decreased to 0.6–0.8, which seems more acceptable. Specific geographic features such as river beds can also be distinguished more clearly on the IDR composites for both January and July.

As stated in the Data section, no indication is provided as whether the GIMMS data is atmospherically contaminated. However, if the number of pixels corresponding to interpolated data is compared to the number of pixels considered as contaminated by the three approaches presented above, their orders of magnitude (around 600 000 pixels out of almost 3 000 000 land pixels) are similar during northern hemisphere winter, due mainly to the "possibly snow" GIMMS flag. In boreal summer, the number of pixels flagged out in the GIMMS data decreases (less than 100 000 pixels), although the number of pixels identified as contaminated by all three approaches stays more or less constant through year 2006. Please note that the HANTS approach tends to identify two to three times more contaminated pixels than the other two approaches, due to its tendency to overestimate maximum NDVI values (see for example control points 1 and 6 in Fig. 3).

As commented in the Results section, the IDR approach tends to overestimate NDVI values for vegetation dormancy period in the case of rapid transitions. This is due to both the difficulty for the averaging technique used in this method to handle rapid transitions (Jönsson & Eklundh, 2006; van Dijk et al., 1987), and to the temporal characteristics of the compositing of the GIMMS data over periods of approximately 15 days (13 to 16 days), since vegetation is usually assumed to be stable over 10 days. Shorter compositing period would lower this overestimation since the transitions would appear less abrupt, since the sampling interval of the vegetation signal would be shorter. Moreover, such abrupt transitions can be observed in a limited number of cases, corresponding to crops with high vegetation density and rapid growth, such as maize or sunflower, or in cases of snowfalls over evergreen vegetation. However, due to the spatial scale of the publicly available databases (several square kilometers), these effects are usually averaged over different species, which tend to decrease the rates of change of the observed vegetation.

Previous studies confirm that the length of the applied temporal window is of highest importance when considering noise reduction. For example, Viovy et al. (1992) advised to gather information on vegetation phenology before choosing analysis time windows for their noise reduction method. As for the TIMESAT software (Jönsson & Eklundh, 2002, 2004, 2006), one of its input parameters is the temporal window size needed for the Savitsky–Golay filter. This choice of external parameterization emphasizes the importance of the temporal window analysis for noise reduction technique in NDVI time series. The method proposed by Chen et al. (2004) does not require any parameterization of time window length for signal reconstruction, although it needs cloud flags to increase its efficiency.

The IDR approach could be applied to shorter compositing periods with improved results since the key aspect of this approach is the continuity of the time series. Thus, the shorter is the compositing period, the higher the continuity of NDVI data for all land covers is. As a consequence, the IDR method could also be applied to data before compositing, though this aspect is beyond the scope of this work, as is its application to other parameters than NDVI, as long as the atmospheric contamination affects these parameters negatively. For example, Julien et al. (2006) showed that the HANTS approach could be successfully applied to land surface temperature parameter for vegetation studies.

However, several aspects of the approach need further analysis and research. First, a comparison of the IDR approach with the method of Ma and Veroustraete (2006) still has to be conducted. This comparison has not been carried out in this paper since its aim was to develop an approach which could be applicable on a yearly basis, with the application to land cover change detection in mind. This comparison could also be extended to other previous methods, and to a higher number of control points, since the presented quantitative approach for best fit assessment allows for easier comparison between methods.

The main difference of the IDR approach with the previously developed methods (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) resides in the fact that the upper envelope of the NDVI time series is always conserved with the IDR approach, while other approaches may decrease some high values (as visible in Fig. 3). This particular feature of the IDR method therefore entrusts the removal of positive outliers (due for example to data transmission errors or to twilight terminator effect) to the pre-processing of the data. However, publicly available NDVI multi-temporal data are provided after this pre-processing step has been carried out, which makes the IDR method well suited for noise removal in these cases.

6. Conclusion

This work has presented a new approach for NDVI composite time series reconstruction from atmospheric contamination, the IDR method, which has been compared quantitatively to two previously developed approaches, the HANTS and the double logistic methods. The IDR method has been shown to outperform the two other approaches as regards conservation of uncontaminated NDVI values and approximation to NDVI time series upper envelope. This method tends to overestimate NDVI values for the vegetation dormancy period in a few cases, especially land covers with high rates of change, although these cases tend to be seldom throughout the globe due to the spatial scale of the publicly available data. However, shorter periods of compositing would lower this effect. The IDR method could also be applied to NDVI time series before compositing, and to other biophysical parameters such as land surface temperature for cloud-free time series reconstruction, as long as the atmospheric contamination affects these biophysical parameters negatively.

Acknowledgement

The authors wish to thank the CEOP-AEGIS project (European Union, project FP7-ENV-2007-1 Grant nr. 212921), the EODIX project (Ministerio de Ciencia e Inovación, project AYA2008-05965-C04-01/ESP) and the WATCH project (European Union, Contract No. 036946) for their financial support.

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., & Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research*, *103*(32), 141−157.

Beck, P., Atzberger, C., Hogda, K. A., Johansen, B., & Skidmore, A. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, *100*(2006), 321−334.

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*, 332−334.

Gutman, G. (1991). Vegetation indices from AVHRR: An update and future prospects. *Remote Sensing of Environment*, *35*, 121−136.

Hird, J. N., & McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, *113*(2009), 248−258.

Holben, B. N. (1986). Characteristics of maximum-value composite image from temporal AVHRR data. *International Journal of Remote Sensing*, *7*, 1417−1434.

Holben, B. N., & Fraser, R. S. (1984). Red and near-infrared sensor response to off-nadir viewing. *International Journal of Remote Sensing*, *5*, 145−160.

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.

Jönsson, P., & Eklundh, L. (2004). TIMESAT — A program for analyzing time-series of satellite sensor data. *Computers and Geoscience*, *30*, 833−845.

Jönsson, P., & Eklundh, L. (2006). TIMESAT — A program for analyzing time-series of satellite sensor data, *Users Guide for TIMESAT 2.3*, available at (last access 27 August 2009): http://www.nateko.lu.se/personal/Lars.Eklundh/TIMESAT/timesat2_3_users_manual.pdf.

Julien, Y., & Sobrino, J. A. (2009). Global land surface phenology trends from GIMMS database. *International Journal of Remote Sensing*, *30*(13), 3495−3513.

Julien, Y., Sobrino, J. A., & Verhoef, W. (2006). Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment*, *103*(2006), 43−55.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., et al. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*. (No. 6 & 7), 1303−1330.

Li, X., & Strahler, A. (1992). Geometric-optical bidirectional reflectance modelling of the discrete crown vegetation canopy: Effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, *30*, 276−292.

Ma, M., & Veroustraete, F. (2006). Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Advances in Space Research*, *37*, 835−840.

Menenti, M., Azzali, S., Verhoef, W., & van Swol, R. (1993). Mapping agro-ecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images. *Advances in Space Research*, *13*, 233−237.

More, J. J. (1977). The Levenberg–Marquardt algorithm: Implementation and theory. In G. A. Watson (Ed.), *Numerical analysisLecture Notes in Mathematics*, *Vol. 630*. Springer-Verlag.

Pinzon, J. (2002). Using HHT to successfully uncouple seasonal and interannual components in remotely sensed data, *SCI 2002 Conference Proceedings*, July 14–18. Orlando, Florida.

Pinzon, J., Brown, M. E., & Tucker, C. J. (2005). Satellite Time Series Correction of Orbital Drift Artifacts Using Empirical Mode Decomposition. In N. E. Huang & S. S. P. Shen (Eds.), *EMD and Its Applications*, *Vol. 10*. (pp. 285−295). Singapore: World Scientific.

Price, J. C. (1991). Using spatial context in satellite data to infer regional scale evapotranspiration. *IEEE Transactions in Geoscience and Remote Sensing*, *28*, 940−948.

Roerink, G. J., Menenti, M., & Verhoef, W. (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, *21*(No. 9), 1911−1917.

Saunders, R. W., & Kriebel, K. T. (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, *9*. (No. 1), 123−150.

Smith, P. M., Kalluri, S. N. V., Prince, S. D., & Defries, R. (1997, Jan). The NOAA/NASA pathfinder AVHRR 8-km land data set. *Photogrammetric Engineering and Remote Sensing*, *63*(1), 12−31.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*, 127−150.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., et al. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, *26*(No. 20), 4485−4498.

van Dijk, A., Callis, S., Sakamoto, C., & Decker, W. (1987). Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data. *Photogrammetric Engineering and Remote Sensing*, *53*, 1059−1067.

Verhoef, W., Menenti, M., & Azzali, S. (1996). A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981–1992). *International Journal of Remote Sensing*, *17*(no. 2), 231−235.

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

Zhou, L., Tucker, C. J., Kaufmann, R. K., Slayback, D., Shabanov, N. V., & Myneni, R. B. (2001). Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *Journal of Geophysical Research*, *106*(D17), 20069−20083.