

# TIMESAT—a program for analyzing time-series of satellite sensor data<sup>☆</sup>

Per Jönsson<sup>a,b,\*</sup>, Lars Eklundh<sup>c,1</sup>

<sup>a</sup> *Division of Mathematics, Natural Sciences and Language, Malmö University, Malmö, Sweden*

<sup>b</sup> *Department of Physics, Lund University, Lund, Sweden*

<sup>c</sup> *Department of Physical Geography and Ecosystems Analysis, Lund University, Lund, Sweden*

Received 3 March 2003; received in revised form 14 April 2004; accepted 1 May 2004

## Abstract

Three different least-squares methods for processing time-series of satellite sensor data are presented. The first method uses local polynomial functions and can be classified as an adaptive Savitzky–Golay filter. The other two methods are more clear cut least-squares methods, where data are fit to a basis of harmonic functions and asymmetric Gaussian functions, respectively. The methods incorporate qualitative information on cloud contamination from ancillary datasets. The resulting smooth curves are used for extracting seasonal parameters related to the growing seasons. The methods are implemented in a computer program, TIMESAT, and applied to NASA/NOAA Pathfinder AVHRR Land Normalized Difference Vegetation Index data over Africa, giving spatially coherent images of seasonal parameters such as beginnings and ends of growing seasons, seasonally integrated NDVI and seasonal amplitudes. Based on general principles, the TIMESAT program can be used also for other types of satellite-derived time-series data.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Function fitting; Data smoothing; Seasonality; Phenology; TIMESAT; NOAA AVHRR; NDVI; CLAVR

## 1. Introduction

Time-series of Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974; Tucker, 1979; Holben et al., 1980), derived from e.g. NOAA/AVHRR, SPOT/VEGETATION or TERRA/MODIS spectral measurements, can be used to gain information on seasonal vegetation development. This information aids analyzes of the functional and structural characteristics of the global and regional land cover (Justice et al., 1985;

Malingreau, 1986; Townshend and Justice, 1986; Tucker et al., 1986), and adds to our current knowledge of global cycles of energy and matter (Running and Nemani, 1988; Rosenqvist et al., 2000; Sellers et al., 1997). Long time-series of NDVI data can also provide information on shifts in the spatial distribution of bioclimatic zones, indicating variations in large-scale circulation patterns or land-use changes.

Although the value of remotely sensed time-series data for monitoring vegetation seasons has been firmly established (Malingreau, 1986; Tucker et al., 1986), only a limited number of methods for exploring and extracting seasonality parameters from such data series have been developed. In this paper a FORTRAN90 program, TIMESAT, for extracting seasonal parameters is presented. The program uses an adaptive Savitzky–Golay filtering method and, optionally, newly developed methods based on upper envelope weighted fits to

<sup>☆</sup>Data on server at <http://www.iamg.org/CGEditor/index.htm>

\*Corresponding author. Tel.: +46-40-6657689.

E-mail addresses: per.jonsson@ts.mah.se (P. Jönsson), lars.eklundh@nateko.lu.se (L. Eklundh).

<sup>1</sup>P.Jönsson and L. Eklundh are recipients of a grant from the Crafoord Foundation.

harmonic and asymmetric Gaussian model functions (Jönsson and Eklundh, 2002). The program is tested with the  $8 \text{ km} \times 8 \text{ km}$  pixel resolution Pathfinder AVHRR Land (PAL) data set generated by NASA/NOAA (Townshend, 1994; James and Kalluri, 1994).

## 2. Data

NDVI data (Rouse et al., 1974) are organized in images (two-dimensional spatial arrays). Each image gives the NDVI values,  $I$ , in the array at a specified time  $t$ . By extracting NDVI values at a position  $(j, k)$  in the array for consecutive times, a time-series  $(t_i, I_i)$ ,  $i = 1, 2, \dots, N$  is obtained for this position (see Fig. 1). The NDVI data in this study are 10-day maximum-value composites (MVC). For each 10-day period the highest NDVI is selected to represent the period (Holben, 1986). The method reduces negatively biased noise due to interference of clouds and atmospheric constituents. However, negatively biased residual atmospherically related noise, as well as some noise due to other factors, such as surface anisotropy and sensor problems (Prince and Goward, 1996; Gutman, 1991), will remain in the data.

Cloudiness products are routinely generated from different sensor data (e.g. AVHRR, VEGETATION, MODIS). The products range from detailed measurements of cloud properties to simple binary classifications indicating presence or absence of clouds. The products may be useful for enhancing the quality of the time-series. PAL NDVI is accompanied by CLAVR (clouds from AVHRR). CLAVR is a cloud indicator based on thresholds of the AVHRR reflectance and thermal channels (Stowe et al., 1991). The original CLAVR data values lie between 1 and 30, and represent the three

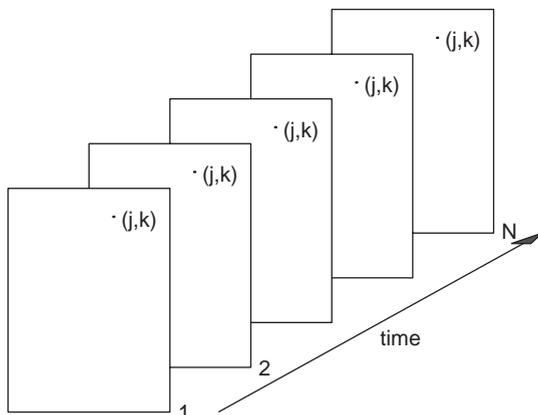


Fig. 1. NDVI data are organized in images. Each image gives NDVI values for time  $t$ . By extracting NDVI values at a spatial position  $(j, k)$  for consecutive times, a time-series  $(t_i, I_i)$ ,  $i = 1, 2, \dots, N$  of NDVI data is obtained for this position.

broad groups; cloudy (1–11), mixed (12–21), and clear (22–30). The CLAVR classification scheme is generalized to be globally valid, and CLAVR data have been shown to underestimate clear pixels (Vemury et al., 2001). Nevertheless the information may be used to improve NDVI estimates (Gutman and Ignatov, 1996).

## 3. Methodology

TIMESAT implements three processing methods based on least-squares fits to the upper envelope of the NDVI data. The first method uses local polynomial functions in the fitting, and the method can be classified as an adaptive Savitzky–Golay filter. The other two methods are ordinary least-squares methods, where data are fit to model functions of different complexity. All three processing methods use a preliminary definition of the seasonality (uni-modal or bi-modal) along with approximate timings of the growing seasons.

We start by a general description of least-squares fits to an upper envelope. This is followed by an account on how to determine the number of annual growing seasons and their approximate timing. The details of the three processing methods are given, and finally the extraction of seasonality information is described.

### 3.1. Least-squares fitting

Assume that we have a time-series  $(t_i, I_i)$ ,  $i = 1, 2, \dots, N$  and a model function  $f(t)$  of the form

$$f(t) = c_1 \varphi_1(t) + c_2 \varphi_2(t) + \dots + c_M \varphi_M(t), \quad (1)$$

where  $\varphi_1(t), \varphi_2(t), \dots, \varphi_M(t)$  are arbitrary basis functions. Then the best values, in the least-squares sense, of the parameters  $c_1, c_2, \dots, c_M$  are obtained by solving the system of normal equations

$$A^T A c = A^T b, \quad (2)$$

where

$$A_{ij} = \frac{\varphi_j(t_i)}{\sigma_i}, \quad b_i = \frac{I_i}{\sigma_i}. \quad (3)$$

Here  $\sigma_i$  is the measurement uncertainty of the  $i$ th data point, presumed to be known. If the measurement uncertainties are not known they may all be set to the constant value  $\sigma = 1$ .

### 3.2. On the use of ancillary cloud data for uncertainty estimation

In TIMESAT simple cloud classifications may be used as indicators of the uncertainty of the NDVI values. The cloudiness classifications are translated into weights that in turn determine the uncertainty of the NDVI values. With CLAVR data used in our study, the three groups

clear, mixed and cloudy are assigned weights  $w = 1, 0.5,$  and  $0,$  respectively. These weights are then transformed into measurement uncertainties  $\sigma$  using the relation

$$\sigma = \frac{1}{w + 0.0001}. \tag{4}$$

There are no general rules as for how the classification scheme or the weights should be chosen, and judicious settings are up to the user.

### 3.3. Adaption to the upper envelope

To take into account that most noise, even for data classified as clear by the cloud mask, is negatively biased, the determination of the parameters  $c_1, c_2, \dots, c_N$  of the model function is done in two or more steps. In the first step the parameters are obtained by solving the system of normal equations with  $\sigma_i$  obtained from the ancillary cloud data. Data points above the model function of the first fit are thought of as being more important, and in the second step the system is solved with the  $\sigma_i$  of the high data points decreased by some factor. If deemed important, the second step can be redone to obtain some sort of self-consistency. This multi-step procedure leads to a model function that is adapted to the upper envelope of the data (Fig. 2).

### 3.4. Determination of the number of seasons

The high level of noise often makes it difficult to determine the number of annual seasons based on data for only one year. Including data from surrounding years reduces the risk for erroneous determinations. In TIMESAT, data values  $(t_i, I_i), i = 1, 2, \dots, N$  for three

years are fit to a model function

$$f(t) = c_1 + c_2t + c_3t^2 + c_4 \sin(\omega t) + c_5 \cos(\omega t) + c_6 \sin(2\omega t) + c_7 \cos(2\omega t) + c_8 \sin(3\omega t) + c_9 \cos(3\omega t), \tag{5}$$

where  $\omega = 6\pi/N$ . The first three basis functions determine the base level and the inter annual trend whereas the three pairs of sine and cosine functions correspond to, respectively, one, two and three annual vegetational seasons.

The fitting procedure always gives three primary maxima. In addition, secondary and tertiary maxima may be found. If the amplitude of the secondary maxima exceeds a certain fraction of the amplitude of the primary maxima we have two annual seasons. For cases where the amplitude of the secondary maxima is low, the number of annual seasons is set to one. In Fig. 3(a) the amplitude of the secondary maxima is small, and the number of seasons is set to one. In Fig. 3(b) the amplitude of the secondary maxima is comparatively large, and the number of annual seasons is set to two.

### 3.5. Savitzky–Golay filtering

One way of smoothing data and suppressing disturbances is to use a filter, and replace each data value  $I_i, i = 1, \dots, N$  by a linear combination of nearby values in a window

$$\sum_{j=-n}^n c_j I_{i+j}. \tag{6}$$

In the simplest case, referred to as a moving average, the weights are  $c_j = 1/(2n + 1)$ , and the data value  $I_i$  is replaced by the average of the values in the window. The moving average method preserves the area and mean position of a seasonal peak, but alters both the width and height. The latter properties can be preserved by approximating the underlying data value, not by the average in the window, but with the value obtained from a least-squares fit to a polynomial. For each data value  $I_i$ , we fit a quadratic polynomial (higher order polynomials are not considered in this work)

$$f(t) = c_1 + c_2t + c_3t^2 \tag{7}$$

to all  $2n + 1$  points in the moving window and replace the value  $I_i$  with the value of the polynomial at position  $t_i$ . The procedure above is commonly referred to as an Savitzky–Golay filter (Press et al., 1992). To account for the negatively biased noise, the fitting is done in multiple steps as described in the previous section. The result is a smoothed curve adapted to the upper envelope of the NDVI values.

The width,  $n$ , of the moving window determines the degree of smoothing, but it also affects the ability to follow a rapid change (Press et al., 1992). In TIMESAT

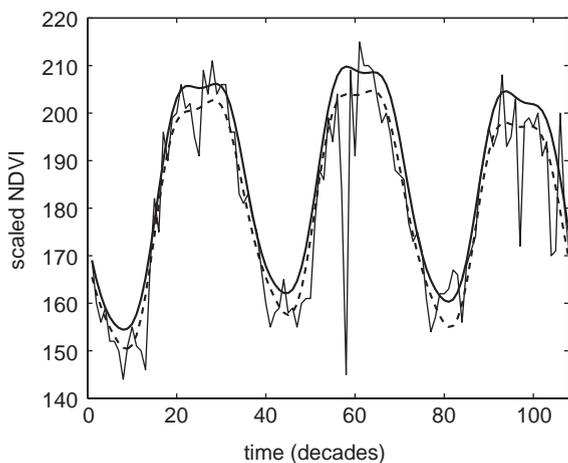


Fig. 2. Fitted polynomial and harmonic functions (see Section 3.6) from two-step procedure. Thin solid line represent original NDVI data. Thick dashed line shows fitted function from first step, whereas thick solid line displays fit from second step.

two values of  $n$  can be set by the user. The first value is used for data representing one annual season and the second is for data representing two, where the number of annual seasons is determined by the method outlined in the previous section. Even if the global settings of the moving window work fairly well, it is sometimes necessary to locally tighten the window. A typical situation is in semi-arid areas, where the vegetation sometimes responds almost instantaneously to precipi-

tation. To capture the corresponding sudden rise in data values, only a small window can be used. In the program the Savitzky–Golay filtering is performed using the global value  $n$  of the window. The filtered data are then scanned. If there is a large increase or decrease in an interval around a data point  $i$ , this data point will be associated with a smaller window. The filtering is then redone with the new locally adapted size of the window. The adaptive procedure is illustrated in Fig. 4.

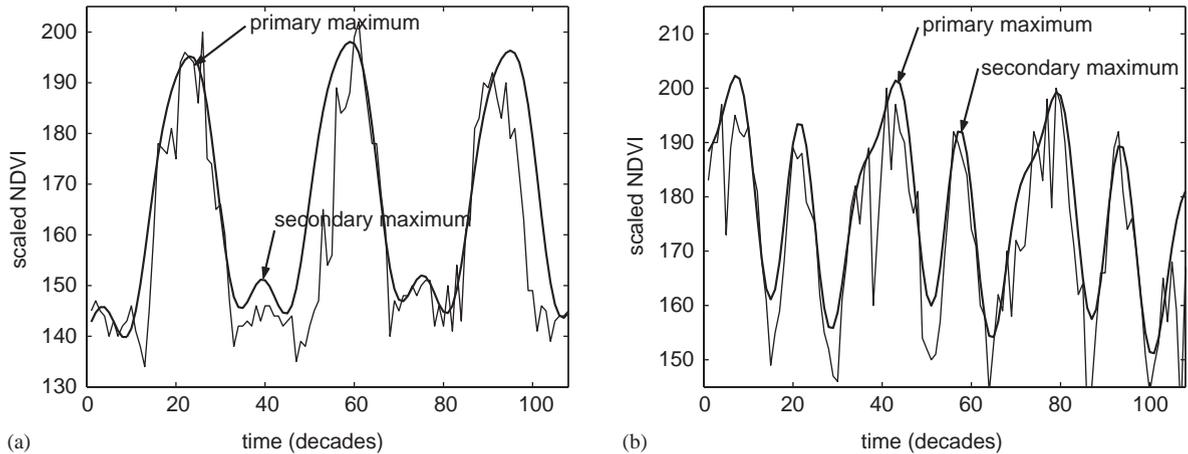


Fig. 3. Fits of NDVI to a basis consisting of a second order polynomial together with harmonic functions representing one two and three annual vegetation seasons. Thin solid line represents original NDVI data. Thick solid line shows fitted function. In (a) amplitude of secondary maxima is small, and number of annual seasons is set to one. In (b) amplitude of secondary maxima is comparatively large, and number of annual seasons is set to two.

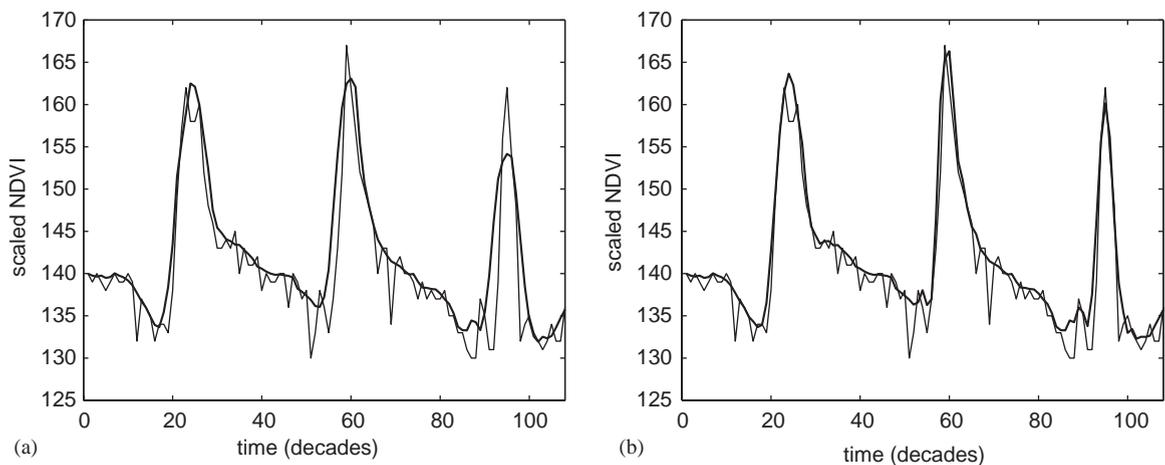


Fig. 4. Upper envelope Savitzky–Golay filtered data. Time is in ten-day steps (decades). In (a) filtering is done with  $n = 5$ , which is too large for filtered data to follow sudden increase and decrease of underlying data values. A scan of filtered data identifies data points for which there are large increases or decreases in surrounding intervals. Setting  $n = 3$  for these points and redoing filtering gives curve in (b). Note improved fit at rising edges and at narrow seasonal peaks.

### 3.6. Least-squares fits to a polynomial and harmonic basis

In this least-squares method a combined polynomial and harmonic basis

$$f(t) = c_1 + c_2t + c_3t^2 + c_4 \sin(\omega t) + c_5 \cos(\omega t) + c_6 \sin(2\omega t) + c_7 \cos(2\omega t) + c_8 \sin(3\omega t) + c_9 \cos(3\omega t), \quad (8)$$

where  $\omega = 6\pi/N$ , is fit to a set of data points  $(t_i, I_i)$ ,  $i = n_1, \dots, n_2$  in an interval around the central peak (cf Menenti et al., 1993; Sellers et al., 1994). The difference compared to the method for identifying the number of annual seasons is that we are now fitting in a smaller interval representing one year. The interval is between the minima to the left and to the right of the central peak. In order to improve the fit at the beginning and end of this interval it has shown useful to enlarge the fitting interval so that it comprises a little more than one year. As in the previous cases the fitting is done in steps to account for the negatively biased noise. In Fig. 5 a fit to the combined polynomial and harmonic basis is displayed along with the original data.

### 3.7. Least-squares fits to asymmetric Gaussian functions

In the Gaussian method (Jönsson and Eklundh, 2002) local model functions are fit to data in intervals around maxima and minima. The local model functions have the general form

$$f(t) \equiv f(t; c_1, c_2, a_1, \dots, a_5) = c_1 + c_2g(t; a_1, \dots, a_5), \quad (9)$$

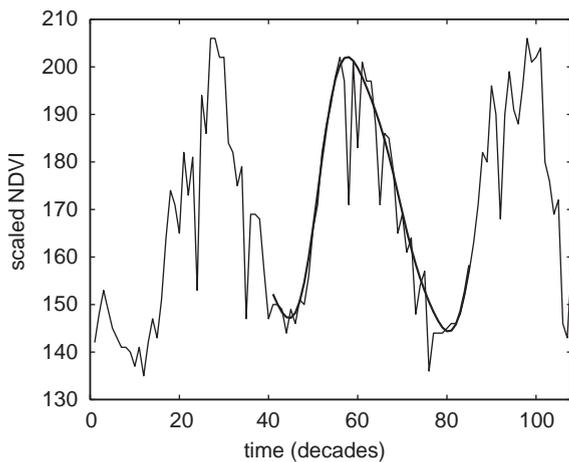


Fig. 5. Fits of NDVI to a basis consisting of a second order polynomial together with harmonic functions representing one, two and three annual vegetation seasons. Thin solid line represents original NDVI data. Thick solid line shows fitted function. Fitting interval is between minima to left and to right of central peak.

where

$$g(t; a_1, \dots, a_5) = \begin{cases} \exp\left[-\left(\frac{t-a_1}{a_2}\right)^{a_3}\right] & \text{if } t > a_1, \\ \exp\left[-\left(\frac{a_1-t}{a_4}\right)^{a_5}\right] & \text{if } t < a_1 \end{cases} \quad (10)$$

is a Gaussian type of function. The linear parameters  $c_1$  and  $c_2$  determine the basis level and the amplitude. For the Gaussian function,  $a_1$  determines the position of the maximum or minimum with respect to the independent time variable  $t$ , while  $a_2$  and  $a_3$  determine the width and flatness (kurtosis) of the right function half. Similarly,  $a_4$  and  $a_5$  determine the width and flatness of the left half. The effects of varying the parameters  $a_2, \dots, a_5$  are shown in Fig. 6.

In order to ensure smooth shapes of the model functions, consistent with what is observed for NDVI data, the parameters  $a_2, \dots, a_5$  are restricted in range. For example,  $a_3$  and  $a_5$  are assumed to be larger than two in order to avoid a cusp at the matching point  $t = a_1$  of the Gaussian function.

The local model functions are well suited for describing the shape of the scaled NDVI time-series in overlapping intervals around maxima and minima. Given a set of data points  $(t_i, I_i)$ ,  $i = n_1, \dots, n_2$  in an interval around a maximum or a minimum, the parameters  $c_1, c_2$  and  $a_1, \dots, a_5$  are obtained by minimizing the merit function

$$\chi^2 = \sum_{i=n_1}^{n_2} \left[ \frac{f(t_i; c_1, c_2, a_1, \dots, a_5) - I_i}{\sigma_i} \right]^2. \quad (11)$$

The function depends non-linearly on the parameters  $a_1, \dots, a_5$  and in TIMESAT the minimization is done using a quasi-Newton method (Dennis et al., 1981a,b). Also in this case the fitting is done in steps to account for the negatively biased noise.

The local model functions describe NDVI data very well in broad intervals around maxima and minima. At the limbs, however, the fits are less good. Denoting the local functions describing the NDVI variation in intervals around the left minima, the central maxima and the right minima by  $f_L(t)$ ,  $f_C(t)$ , and  $f_R(t)$  (see Fig. 7 (a)–(c)), the global function  $F(t)$ , that correctly models the NDVI variation in full interval  $[t_L, t_R]$ , is

$$F(t) = \begin{cases} \alpha(t)f_L(t) + [1 - \alpha(t)]f_C(t) & t_L < t < t_C, \\ \beta(t)f_C(t) + [1 - \beta(t)]f_R(t) & t_C < t < t_R. \end{cases} \quad (12)$$

Here  $\alpha(t)$  and  $\beta(t)$  are cut-off functions that in small intervals around  $(t_L + t_C)/2$  and  $(t_C + t_R)/2$ , respectively, smoothly drop from 1 to 0. Loosely speaking, the global function  $F(t)$ , shown in Fig. 7(d), assumes the character of  $f_L(t)$ ,  $f_C(t)$  and  $f_R(t)$  in, respectively, the left, central and right part of the interval  $[t_L, t_R]$ . The merging of local functions to a global function is a key feature of the method. It increases the flexibility and allows the

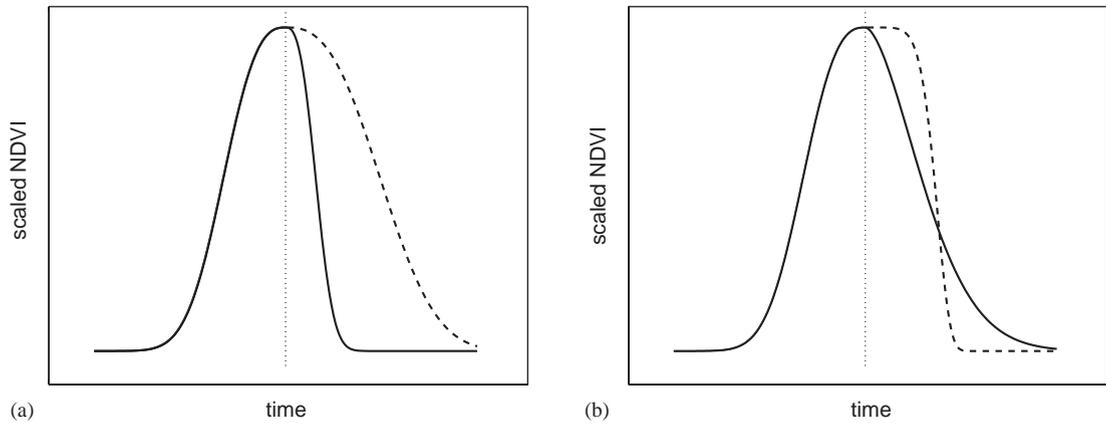


Fig. 6. Effect of parameter changes on local functions. In (a) parameter  $a_2$ , which determines width of right function half, has been decreased (solid line) and increased (dashed line) compared to value of left half. In (b) parameter  $a_3$ , which determines flatness of right function half, has been decreased (solid line) and increased (dashed line) compared to value of left half.

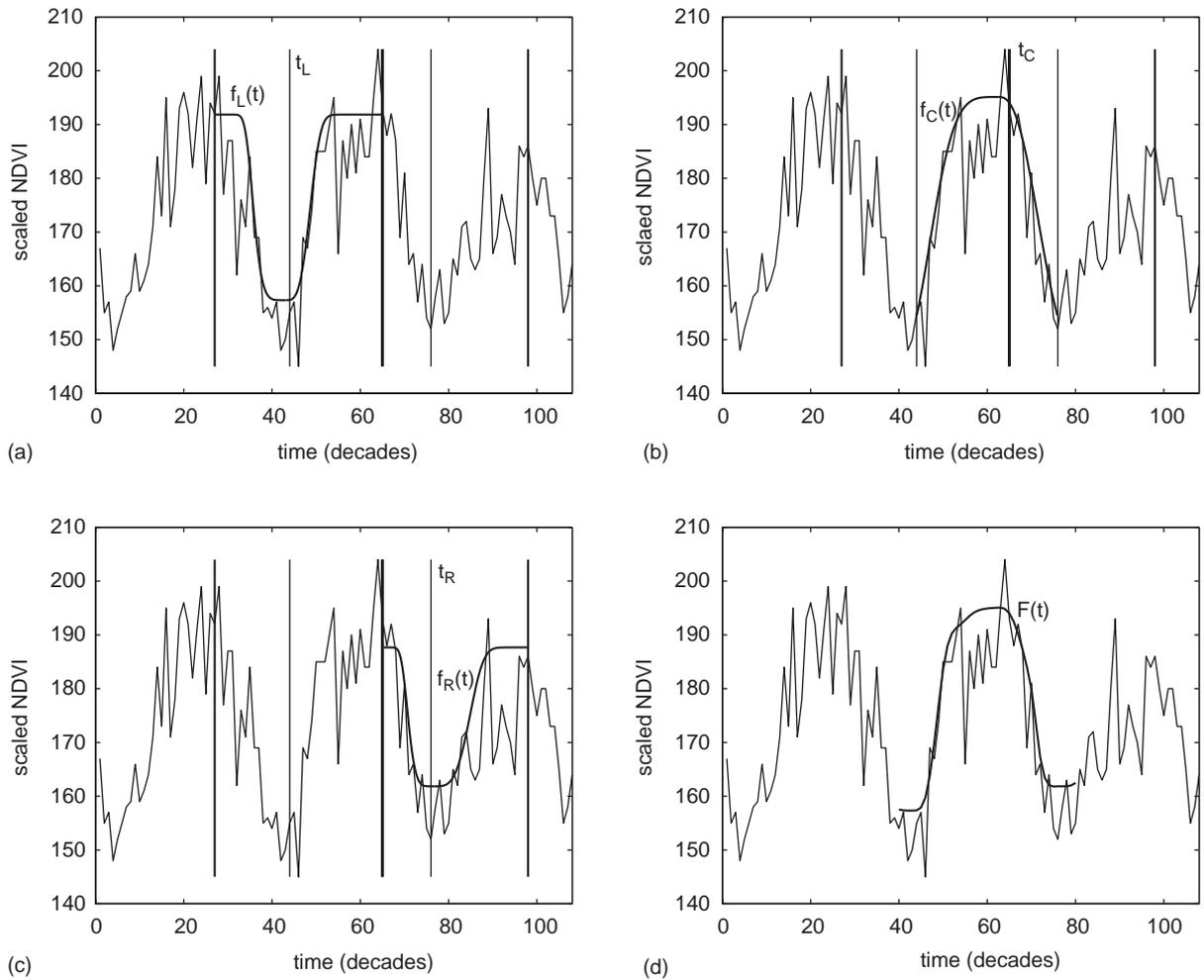


Fig. 7. (a)–(c) Display local model functions fitted to, respectively, left minimum, central maximum, and right minimum. (d) shows global model function that is obtained by merging three local functions.

fitted function to follow a complex behavior of the time-series (Jönsson and Eklundh, 2002).

### 3.8. Extraction of seasonal parameters

Seasonal data are extracted for each of the growing seasons of the central year (see Fig. 8). The beginning of a season, marked by (a) in the figure, is defined from the filtered or fitted functions as the point in time for which the value has increased by a certain number, currently set to 10% of the distance between the left minimum level and the maximum. The end of the season (b) is defined in a similar way. The mid of a season is difficult to define but a reasonable estimate is obtained as the position (e) between the positions (c) and (d) for which the value of the fitted function has increased to 90% of distance between, respectively, the left and right minimum levels and the maximum. The amplitude (f) of the season is obtained as the difference between the peak value and the average of the left and right minimum values.

The annual integrated NDVI is frequently used as a measure of net primary production (Running and Nemani, 1988; Goward and Dye, 1987; Ruimy et al., 1994). To give a good estimate of the production of the seasonally dominant vegetation type it is also of interest to compute the integrated NDVI over the growing season, i.e. between the start and end of the season. In TIMESAT there are two integrals over the growing season. The first integral (h), given by the area of the region between the fitted function and the average level

of the left and right minima, represents the seasonally active vegetation, which may be fairly small for evergreen areas. The second integral (i), given by the area of the region between the fitted function and the zero level, represents the total vegetation production. In evergreen areas the first integral may be small even if the total vegetation production is large.

The rate of increase in NDVI during the beginning of the season is theoretically related to the physiognomy of the vegetation and can be estimated by looking at the ratio between the amplitude and the time difference between the season start and the mid of the season. Another interesting quantity is the asymmetry, which can be defined as the ratio of the time differences between the mid of the season and the start and end of the season. A value of the asymmetry that is smaller than one (positive skewness) indicate a rapid rise and a slow fall. Asymmetries larger than one (negative skewness), on the other hand, are indicative of a slow rise and rapid fall.

## 4. Implementation of the methods

The methods are implemented in a general FORTRAN90 computer program, TIMESAT, that processes NDVI data for pixels in a selected geographical area. By default the Savitzky–Golay filtering is always performed. Fits to the combined polynomial and harmonic basis and to the Gaussian functions are optional. Below are some comments on the organization of the program.

### 4.1. Overview of the TIMESAT program

The subroutine `timesat` is the main driver. After some initializations input data controlling the calculation is read from file. The files containing NDVI and ancillary cloud data are checked and opened. The subroutine `gensincos` computes the nine sine and cosine functions of harmonic basis on the grid  $t_i$ ,  $i = 1, \dots, n$ . If fits to asymmetric Gaussian functions are requested, indicated by `gauss = true`, then the subroutine `gengauss` computes Gaussian functions  $g(t_i; a_1, a_2, \dots, a_5)$  on the grid  $t_i$ ,  $i = 1, \dots, n$  for a number of combinations of the non-linear parameters  $a_i$ . The harmonic basis functions together with the Gaussian functions are used repeatedly and are kept in core. After these initial steps the loop over pixels starts. For each pixel the time-series and corresponding cloud data are extracted and sent to the subroutine `ts` that controls the processing of single time-series. From `ts` the extracted seasonal parameters are returned. These parameters are then printed to various output files.

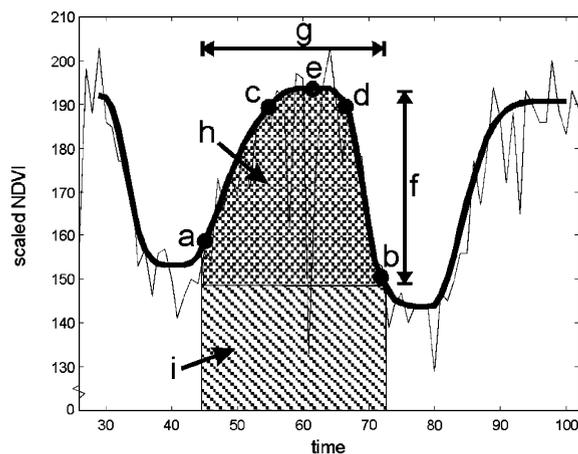


Fig. 8. Some of seasonality parameters computed in TIMESAT: (a) beginning of season, (b) end of season, (c) left 90% level, (d) right 90% level, (e) peak, (f) amplitude, (g) length of season, (h) integral over growing season giving area between fitted function and average of left and right minimum values, (i) integral over growing season giving area between fitted function and zero level.

```

program timesat
initializations
read input data from file
allocate needed memory
open in- and output files
call gensincos
if gauss = true call genfunc
loop over pixels
  for each pixel extract the time-series
  and the corresponding cloud data
  call ts
  write output parameters to files
end loop over pixels
end program timesat

```

#### 4.2. Processing of single time-series

The next level in the program structure is to process single time-series. In the subroutine `datachecks` different checks on data are performed. Checks are done to ensure that the pixel is not over water, single spikes in the data are removed, or more accurately, the corresponding weight is set to zero, checks on missing data are performed and finally pixels over arid areas, with signal variation less than a defined threshold, are detected and removed. After these data checks the subroutine `season` is called. Here the number of annual seasons together with approximate positions for maxima and minima are determined according to the prescription in Section 3.1. Depending on the input data `ts` calls the subroutines that performs the Savitzky–Golay filtering (`savgol`), least-squares fits to the polynomial and harmonic basis (`polharm`) or least-squares fits to local asymmetric Gaussian functions that are subsequently merged to a global function (`globalgauss`). Calls to these routines are followed by calls to `parameters` that extracts seasonal parameters from the filtered or fitted functions.

```

subroutine ts
call datachecks
call season
call savgol
call parameter
if harm = true
  call polharm
  call parameters
end if
if gauss = true
  call globalgauss
  call parameters
end if
end subroutine ts

```

#### 4.3. Fits to asymmetric Gaussians and merging to global functions

Of the subroutines in `ts` only `globalgauss` has a structure complicated enough to motivate a more detailed description. The routine starts by a loop over intervals. For time-series with one annual season these intervals are: the left minimum interval, the central maximum interval, and the right minimum interval (see Fig. 7). For each interval there is a call to `fit` that controls the non-linear fit of the asymmetric Gaussian. The subroutine `fit` returns the locally fitted asymmetric Gaussian functions that are merged to a global function describing the full vegetational season.

```

subroutine globalgauss
loop over fitting intervals
  call fit
end loop over fitting intervals
merge local functions to a global one
end subroutine globalgauss

```

When fitting the local functions we loop over uncertainty settings to adapt to the upper envelope. After the sigma are set the loop over parameter combinations  $a_i$  starts. For each combination a linear-least squares fit to  $c_1 + c_2g(t; a_1, a_2, \dots, a_5)$  is performed, and the merit function  $\chi^2$  is computed. The combination of  $a_i$  giving the smallest  $\chi^2$  is used as input values for `nonlinlsq` that determines the final parameter values. After the non-linear fits are done a number of consistency checks are made to weed out erroneous solutions.

```

subroutine fit
loop over uncertainty settings to
  set sigma
  loop over non-linear parameters a(i)
  perform liner least-squares fit to function (9)
  compute chisquare value
  end loop over non-linear parameters
  determine the parameter combination a(i) that
  gives
  the smallest chisquare
  call nonlinlsq
  perform consistency checks
end loop over uncertainty settings
end subroutine fit

```

#### 4.4. MATLAB programs to view fits

TIMESAT is accompanied by a number of MATLAB script files that are used to view the original data

together with the fitted functions. The files are listed in the table below.

file	viewed properties
data.m	views the NDVI data together with the weights obtained from cloud data. Data points classified as cloudy are marked by crosses. Data points classified as mixed are marked by circles. The script allows the user to test various parameter settings including the window size for the Savitzky–Golay method and the parameter that determines the degree of adaptation to the upper envelope.
season.m	views the fit to the polynomial and harmonic basis that is used to determine the number of annual seasons. Minima and maxima that defines the positions between which the different local Gaussian functions are fitted (see Section 3.6) are marked by vertical lines.
savgol.m	views the Savitzky–Golay filtered data. The beginning and end of the season are marked by circles. The position of the maximum is indicated by a vertical line.
harm.m	views the fitted polynomial and harmonic function. The beginning and end of the season are marked by circles. The position of the maximum is indicated by a vertical line.
gauss1.m	gives three figures with the local Gaussian functions fitted to the left minimum, the central maximum and the right minimum.
gauss2.m	views the global model function obtained by merging the local ones. The beginning and end of the season are marked by circles. The position of the maximum is indicated by a vertical line.

## 5. Input and output data

The TIMESAT program can be used in MATLAB or image mode. In MATLAB mode the user interactively steps through the pixels. For each pixel the program outputs a number of ASCII files that can be processed by the MATLAB files listed above. This is useful for debugging purposes and for testing parameter settings for the program. In the image mode the program

automatically loops through all pixels and outputs a number of image files of seasonal properties. The image files are in binary 32-bit real (single precision) raster format, sequentially organized by row and column ( $j, k$ ), and can be viewed using IDRISI, PCI, or any equivalent image processing software.

The input data for the program are entered interactively or can alternatively be supplied in an input file `input.txt`. Detailed instructions on how to enter input data and to how prepare the input file are given in the file `instructions.pdf`, that accompanies the TIMESAT program.

## 6. Test run

TIMESAT was tested on 10-day MVC AVHRR NDVI data for Africa for the time period 1998–2000. Fig. 9 presents results from a run in MATLAB mode. Here the MATLAB files listed above have been used to display original NDVI data together with fitted functions for a pixel (318,300), which is located in Mali.

Fig. 10 presents results from a run in image mode. The upper left panel displays the number of annual growing seasons. The upper right panel shows the start of the first growing season of 1999 obtained from the fitted asymmetric Gaussian functions. The colors represent the 10-day period of the start, and vary from 1 (1st to 10th of January) to 36 (21st to 31st of December). White pixels represent missing data or areas where the data could not be processed. Near the South Atlantic coast the season starts around decade 5 (end of February). The starting date then shifts toward later dates until the border of the Sahara, where it falls at about decade 25 (beginning of October). The overall patterns appear to be in general agreement with the climatic patterns, which in Africa are the results of the movements of the Inter Tropical Convergence Zone (ITCZ) (Landsberg, 1972). The lower left panel displays the amplitude obtained from the fitted asymmetric Gaussian functions, of the first growing season of 1999, i.e. the difference between the maximum function value and the basis level. The lowest amplitudes are found in and around desert areas, where the overall values of NDVI are very low. The highest amplitudes are found in semi-arid areas, where the vegetation development follows a marked annual cycle of growth and decline coupled to the rainfall variations. In moist areas, e.g. parts of West Africa and areas close to the Equator, the amplitude is fairly low. In these areas rainfall is more abundant and more evenly distributed over the year, resulting in a weaker seasonal variation in vegetation greenness. The lower right panel displays the asymmetry or skewness of the first growing season of 1999. The skewness was obtained from the Savitzky–Golay filtered data. One notes the strong negative skewness in a belt along the Saharan fringe.

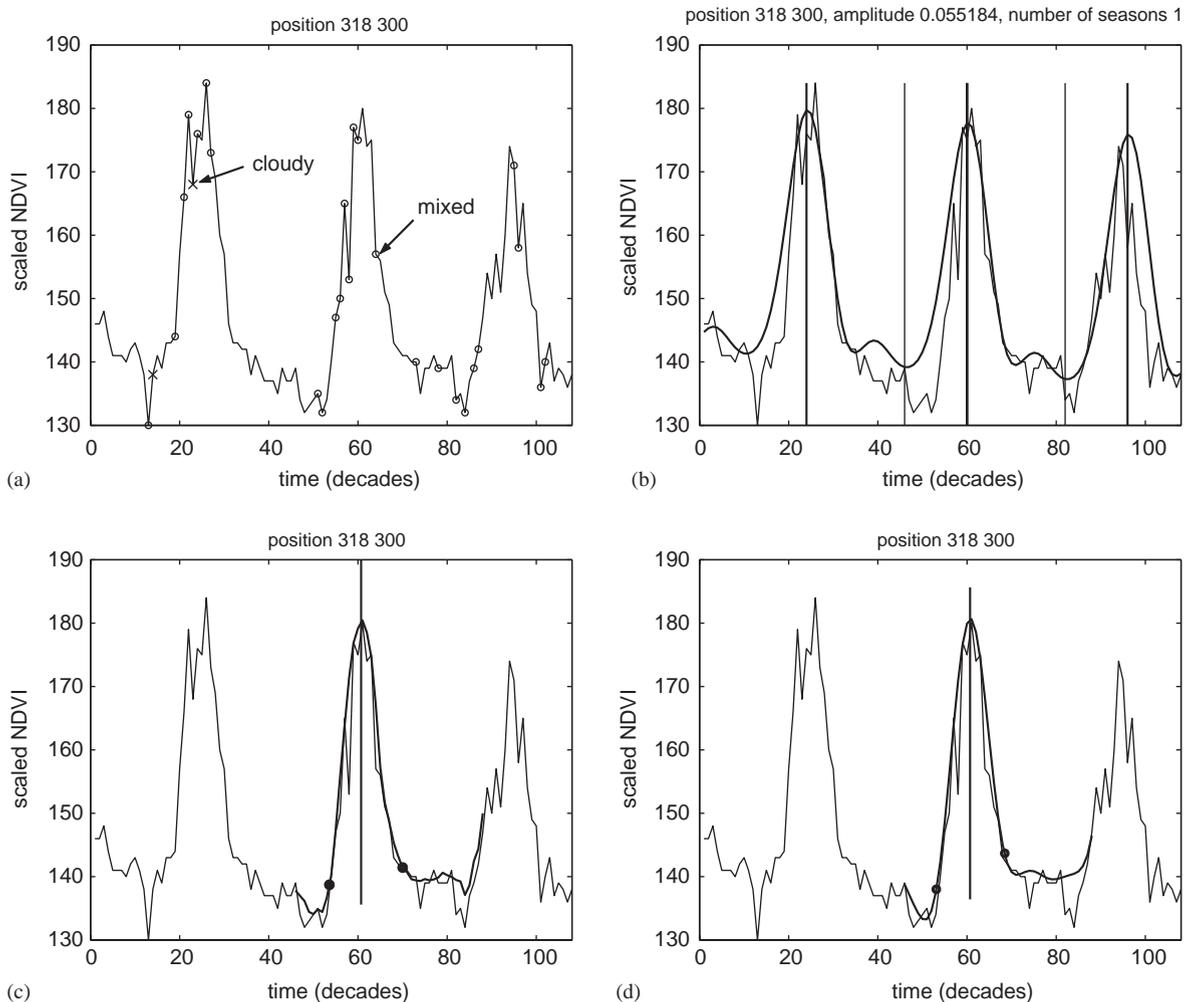


Fig. 9. Original NDVI data and fitted functions displayed using MATLAB m-files. (a) displays original NDVI data together with CLAVR information. Data points classified as cloudy (weight 0) are marked by crosses. Data points classified as mixed (weight 0.5) are marked by circles. (b) displays fit to polynomial and harmonic basis that is used to determine number of annual seasons. Minima and maxima that defines positions between which different local Gaussian functions are fitted (see Section 3.7) are marked by vertical lines. (c) views Savitzky–Golay filtered data. Beginning and end of season are marked by circles. Position of maximum is indicated by a vertical line. (d) displays fitted polynomial and harmonic functions. Beginning and end of season are marked by circles. Position of maximum is indicated by a vertical line.

Precipitation in this area falls during a very short period when the ITCZ is at its northernmost position (Nieuwolt, 1982). The shape of the NDVI curve suggests a very rapid response to rainfall, followed by a slower decay as the vegetation withers.

## 7. Discussion

### 7.1. On the use of the different methods

The Savitzky–Golay filtering and the fitting to asymmetric Gaussians can be classified as local and

semi-local methods, respectively. The fitting to the harmonic basis, on the other hand, is a global method since it uses data for a full year. The advantage of local and semi-local methods is that they work properly even for time-series that are quasi-periodic. The global method, that has a fixed period of 1 year, compare unfavorably in these cases. For NDVI data relatively unaffected by noise, the Savitzky–Golay filtering method works very well. Since the method is local it is able to follow also more complex behaviors such as a rapid increase followed by a decreasing plateau (see Fig. 11).

For noisy time-series the Savitzky–Golay filtered data are sometimes difficult to interpret. In these cases it

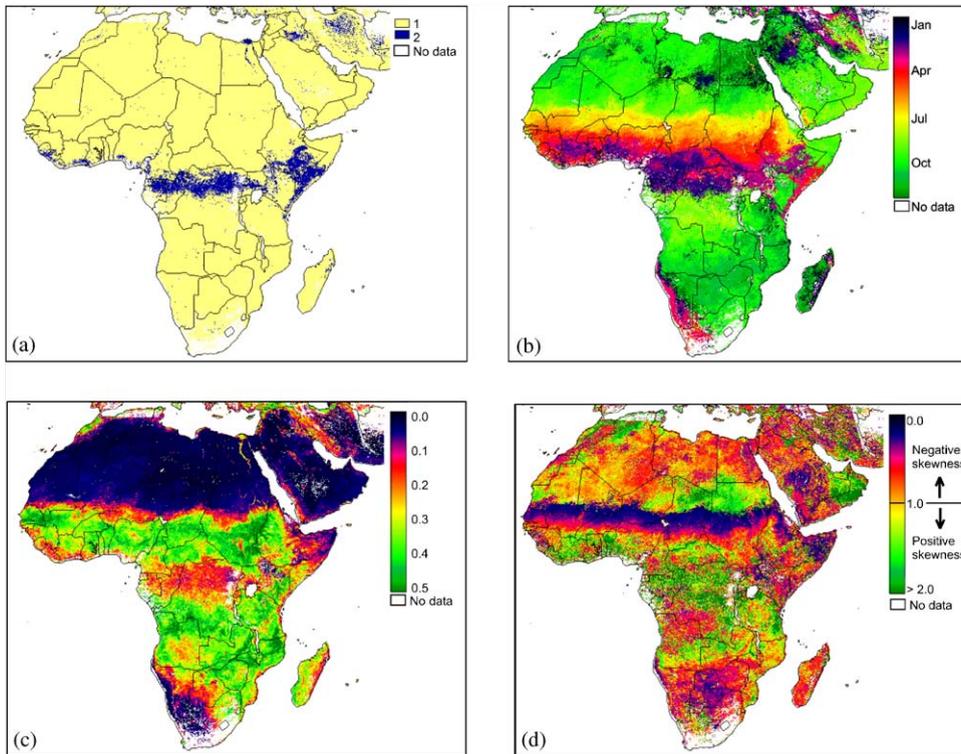


Fig. 10. Extracted seasonal parameters for Africa during 1999. (a) number of annual vegetation seasons. (b) start of first growing season obtained from fitted asymmetric Gaussian functions. (c) amplitude of first growing season obtained from fitted asymmetric Gaussian functions. (d) asymmetry of first growing season obtained from Savitzky–Golay filtered data.

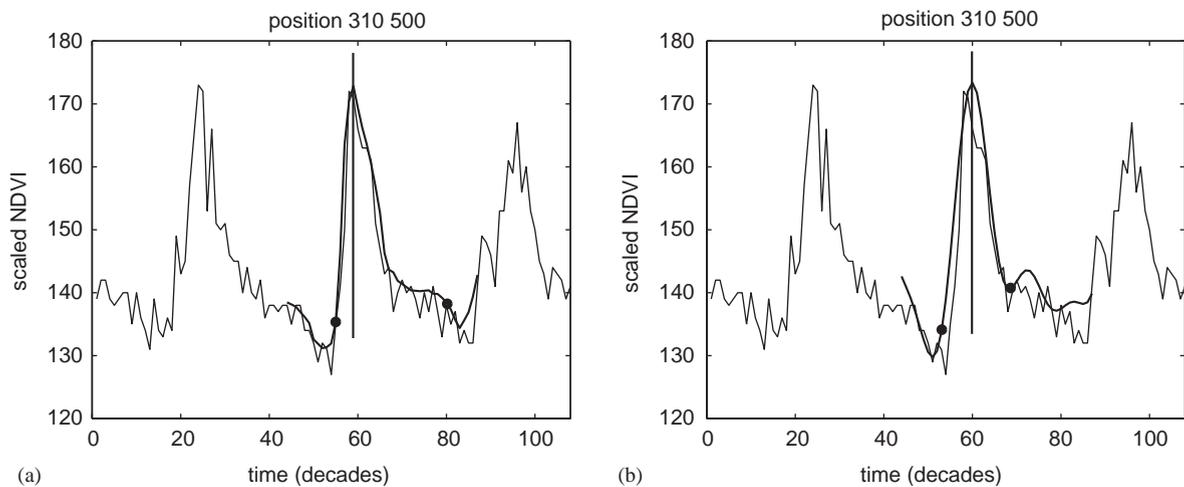


Fig. 11. Growing season is characterized by a narrow peak followed by a decreasing plateau. (a) filtered values from Savitzky–Golay method. (b) function from least-squares fits to combined harmonic and polynomial basis. In this case narrow peak gives rise to spurious oscillations. Note also that beginning of season is inaccurately determined.

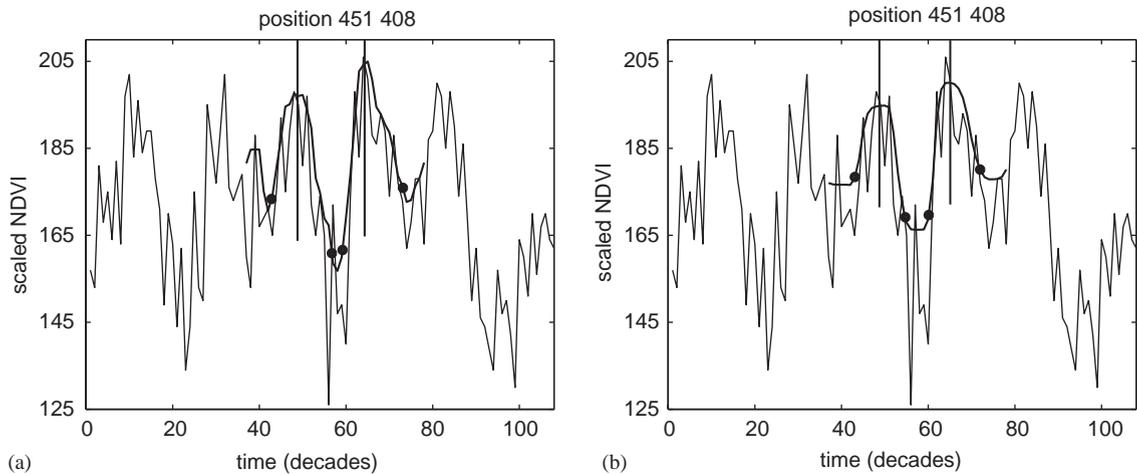


Fig. 12. Noisy time-series from a region with two annual vegetation seasons. (a) filtered values from Savitzky–Golay method. (b) function from fits to asymmetric Gaussians.

becomes necessary to apply restrictions, and force the data into the fixed functional form that can be spanned by the asymmetric Gaussians or the combined harmonic and polynomial basis. An example of this is given in Fig. 12. Here the noise in the Savitzky–Golay filtered data resulted in an erroneous determination of the time for the end of the first season. The asymmetric Gaussian method on the other hand is less sensitive to the noise and seems to give better predictions for the beginnings and ends of the seasons.

### Acknowledgements

Data used by the authors in this study include data produced through funding from the Earth Observing System Pathfinder Program of NASA's Mission to Planet Earth in cooperation with National Oceanic and Atmospheric Administration. The data were provided by Earth Observing System Data and Information System (EOSDIS), Distributed Active Archive Center at Goddard Space Flight Center which archives, manages, and distributes this data set.

### References

- Dennis, J.E., Gay, D.M., Welsch, R.E., 1981a. An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software* 7, 348–368.
- Dennis, J.E., Gay, D.M., Welsch, R.E., 1981b. Algorithm 573, NL2SOL—an adaptive nonlinear least-squares algorithm [E4]. *ACM Transactions on Mathematical Software* 7, 369–383.
- Goward, S.N., Dye, D.G., 1987. Evaluating North American net primary productivity with satellite observations. *Advances in Space Research* 7, 165–174.
- Gutman, G.G., 1991. Vegetation Indices from AVHRR: an update and future prospects. *Remote Sensing of Environment* 35, 121–136.
- Gutman, G., Ignatov, A., 1996. The relative merit of cloud/clear identification in the NOAA/NASA Pathfinder AVHRR Land 10-day composites. *International Journal of Remote Sensing* 17, 3295–3304.
- Holben, B.N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing* 7, 1417–1443.
- Holben, B.N., Tucker, C.J., Fan, C.J., 1980. Spectral assessment of soybean leaf area and leaf biomass. *Photogrammetric Engineering and Remote Sensing* 46, 651–656.
- James, M.E., Kalluri, S.N.V., 1994. The Pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing* 15, 3347–3363.
- Jönsson, P., Eklundh, L., 2002. Seasonality extraction and noise removal by function fitting to time-series of satellite sensor data. *IEEE Transactions of Geoscience and Remote Sensing* 40, 1824–1832.
- Justice, C.O., Townshend, J.R.G., Holben, B.N., Tucker, C.J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* 6, 1271–1318.
- Landsberg, H.E., 1972. *Climates of Africa*. World Survey of Climatology. Elsevier, Amsterdam, 604pp.
- Malingreau, J.-P., 1986. Global vegetation dynamics: satellite observations over Asia. *International Journal of Remote Sensing* 7, 1121–1146.
- Menenti, M., Azzali, S., Verhoef, W., van Swol, R., 1993. Mapping agroecological 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.

- Nieuwolt, S., 1982. Tropical Climatology. Wiley, Chichester, UK, 207pp.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in FORTRAN. The Art of Scientific Computing, 2nd Edition. Cambridge University Press, Cambridge 963pp.
- Prince, S.D., Goward, S.N., 1996. Evaluation of the NOAA/NASA Pathfinder AVHRR land data set for global primary production modelling. *International Journal of Remote Sensing* 17, 217–221.
- Rosenqvist, Å., Imhoff, M., Milne, A., Dobson, C., 2000. Remote sensing and the Kyoto protocol—a workshop summary. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B7, Amsterdam, pp. 1278–1285.
- Rouse, J.W.J., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. In: *Third ERTS Symposium*, NASA SP-351, Washington DC, pp. 309–317.
- Ruimy, A., Saugier, B., Dedieu, G., 1994. Methodology for the estimation of terrestrial net primary production from remotely sensed data. *Journal of Geophysical Research* 99, 5263–5283.
- Running, S.W., Nemani, R.R., 1988. Relating seasonal patterns of the AVHRR vegetation index to simulated photosynthesis and transpiration of forests in different climates. *Remote Sensing of Environment* 24, 347–367.
- Sellers, P.J., Tucker, C.J., Collatz, G.J., Los, S.O., Justice, C.O., Dazlich, D.A., Randall, D.A., 1994. A global 1° by 1° NDVI data set for climate studies. Part 2: the generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing* 15, 3519–3545.
- Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F.G., Berry, J.A., Collatz, G.J., Denning, A.S., Mooney, H.A., Nobre, C.A., Sato, N., Field, C.B., Henderson-Sellers, A., 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science* 275, 502–509.
- Stowe, L.L., McClain, E.P., Carey, R., Pellegrino, P., Gutman, G., Davis, P., Long, C., Hart, S., 1991. Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Advances in Space Research* 3, 51–54.
- Townshend, J.R.G., 1994. Global data sets for land applications from the Advanced Very High Resolution Radiometer: an introduction. *International Journal of Remote Sensing* 15, 3319–3332.
- Townshend, J.R.G., Justice, C.O., 1986. Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing* 7, 1435–1445.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8, 127–150.
- Tucker, C.J., Justice, C.O., Prince, S.D., 1986. Monitoring the grasslands of the Sahel 1984–1985. *International Journal of Remote Sensing* 7, 1571–1581.
- Vemury, S., Stowe, L., Anne, V., 2001. AVHRR pixel level clear-sky classification using dynamic thresholds (CLAVR-3). *Journal of Atmospheric and Oceanic Technology* 18, 169–186.