Influence of using date-specific values when extracting phenological metrics from 8-day composite NDVI data

Asheer Bachoo*, Sally Archibald†
*abachoo@csir.co.za, Meraka Institute, CSIR, Pretoria, South Africa
†sarchibald@csir.co.za, National Resources and Environment, CSIR, Pretoria, South Africa

***Abstract*—Information of vegetation dynamics derived from remotely sensed data is essential for regional natural resource management. A common approach involves using an *n*-day composite time-series of normalized difference vegetation index (NDVI) satellite data for estimating vegetation phenology. The composite method was introduced to deal with problems associated with cloudy and noisy data but could possibly obscure the fine-scale timing of vegetation processes like leaf out and leaf drop, which can occur suddenly, often within a few days. Moreover, different methods for smoothing the NDVI curves also influence the accuracy of the vegetation parameters extracted from them. We investigate the difference between 8-day equal-interval (standard composite data) and 8-day date-specific NDVI data in their ability to extract phenological metrics of interest for ecologists and land managers. We also compare two different filtering algorithms - the Savitsky-Golay and the Gaussian filtering algorithms. Results from a typical savanna system in South Africa show that the Savitsky-Golay technique with date-specific NDVI values is the best method.**

I. Introduction

Time-series of coarse resolution satellite data contain indispensable information on the function and phenology of regional vegetation cover. Understanding the seasonal patterns of vegetation activity is essential to (i) the characterisation of vegetation, (ii) studying the impact of climate change and inter-annual variability, (iii) monitoring land degradation and (iv) detecting changes in land use/ land cover. A number of methods have been developed to extract seasonality parameters from long-term satellite data. The TIMESAT [1] software program is the most advanced, widely-used and user-friendly. However, the standard TIMESAT product might not be appropriate in systems that show great inter-annual and intra-annual variability, such as those found in southern Africa. Various alternative methods are assessed in this paper to investigate whether they provide any increased functionality over the TIMESAT method for semi-arid and tropical systems.

Daily satellite data are routinely combined into *n*-daily maximum NDVI value composites (MVC) to reduce the effect of clouds and the atmosphere, which both typically reduce NDVI values. A significant amount of noise nevertheless remains in the MVC data. The MVC process inevitably introduces an undetermined error of 0 to *n* days when estimating date-sensitive phenological parameters such as the start, peak and end of the growing season. This makes it very difficult to correlate these dates with in-situ vegetation or climate measurements. However, the MODIS 8-day composite surface reflectance product (MOD09_L3) [2] retains the specific date from which a pixel's value was taken (hereafter referred to as date stamp) and presents an opportunity to circumvent the aforementioned introduced errors. The current version of the TIMESAT program is unable to make use of this date stamp since its algorithms assume uniform temporal steps between NDVI data points (e.g. 10 days). The objectives of this paper are

1) to create four phenological curves using standard MVC and date-stamp adjusted MVC, and two different smoothing algorithms,
2) to extract key phenological metrics of interest for ecologist and land managers from these curves, and
3) to compare these metrics with those derived from field-data to derive a method that is both computationally tractable, and sufficiently accurate for use in a range of systems.

This research was done using data from a savanna system in southern Africa, which is prone to inter-annual variability in rainfall and sudden changes in vegetation cover.

II. Methodology

There are two steps to extracting date-specific phenology metrics from an NDVI profile. Firstly the time-series must be filtered to create a smoothed phenology curve and then various algorithms are applied to the curve to compute metrics for phenology description e.g. green-up date, maximum NDVI, length of growing season.

A. Filtering of noisy NDVI profiles

In the TIMESAT program, Jonsson and Eklundh [1], [3] implement three different methods based on least-squares fitting: Savitzky-Golay filtering, a combined polynomial and harmonic basis function fit and a least-squares fit to asymmetric Gaussian functions. These methods use years to determine the shape of the current year's curve, which could be problematic in environments with large variation in the magnitude and timing of vegetation growth. We implement the Savitsky-Golay filtering as well as Gaussian filtering, which we hypothesised might perform better in variable environments and with non-equal date stamps.

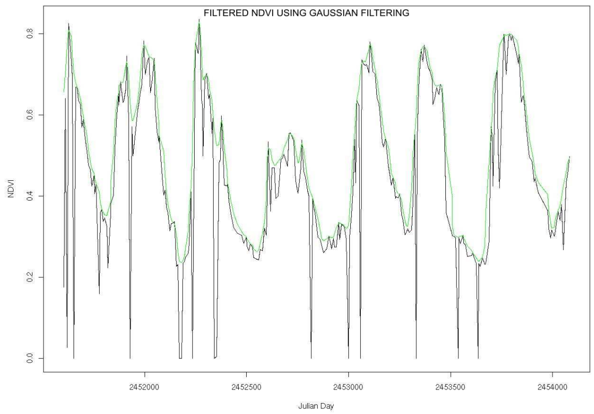

Fig. 1. *Gaussian filtering*

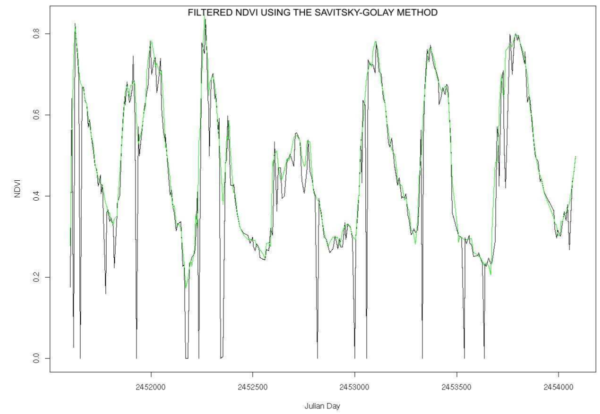

Fig. 2. *Savitsky-Golay filtering*

A maximum filter is first used to perform pre-filtering of noisy data. This step is necessary in order to remove sharp downward spikes that we can attribute to noise. Missing data is interpolated linearly. Our dataset contained very little missing data and they occurred within short intervals of time (usually 1 timestep). Thus, complex data interpolation was not implemented at this point in time.

1) Gaussian filtering: The Gaussian filter is a spatial filter that is used to smooth a signal [4]. It uses a ”bell-shaped” kernel that represents the shape of the Gaussian. The 1D Gaussian has the form

$$g(x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (1)$$

where σ is the standard deviation of the ”bell-shape”. The 1D signal is convolved with this kernel (in normalized form). The Gaussian filtering method works independently of the timesteps i.e. the spacing between time-stamps does not affect the output of the filter. Figure 1 shows a noisy NDVI profile with the Gaussian filtering output overlayed.

2) Savitsky-Golay filtering: This method smoothes and approximates data by replacing each data value x_i ($i = 1, \ldots, N$ where N is the number of data points) with the value of an approximated function at that point. The function is a quadratic polynomial fitted to the set of points X in a moving window centered at x_i. The width of the window controls the degree of smoothing. The quadratic polynomial is

$$f(t) = c_1 + c_2 t + c_3 t^2 \quad (2)$$

where $\{c_1, c_2, c_3\}$ are coefficients defining the polynomial. We fit the abovementioned function using a weighted least squares estimate (LSE) algorithm. Weights are computed so that the LSE algorithm is driven towards being asymmetrically biased so as to fit the upper envelope of NDVI values. The weight (w_i) for a data point x_i is derived from [5] and computed as follows:

$$w_i = \frac{1}{\sqrt{2}\sigma}\exp\left(s\left(\frac{x_i - \mu}{\sigma}\right)^2\right) \quad (3)$$

where

$$s = \begin{cases} -1 & \text{if } x_i - \mu < 0 \\ +1 & \text{otherwise} \end{cases} \quad (4)$$

The parameters σ and μ are the standard deviation and mean, respectively, of the data points in X.

The Savitsky-Golay filtering method is dependent on the timestep between data points when interpolating a polynomial function. Hence, these timesteps will affect the output of the filtering procedure and the parameters estimated will model the data in terms of the timesteps. Figure 2 shows a noisy NDVI profile with the Savitsky-Golay output overlayed.

B. Extraction of phenological metrics

The exact metric extracted from the phenology curves depends greatly on the algorithms used. TIMESAT uses a threshold approach to green-up detection (the date when the NDVI value increases above a certain distance from the minimum or maximum values [3]), which works well in systems with very predictable minimum and maximum NDVI values but could give odd behaviour in a drought year, for example. Similarly, the total annual production (area under the curve) is very sensitive to the definition of the minimum value. Moreover, there are some parameters, such as the ’peakiness’ of the phenology curve (amount of intra-annual fluctuation) that are not considered by the standard TIMESAT program. We have developed a range of algorithms for extracting 9 metrics of interest for researchers in semi-arid ecosystems:

The algorithms were developed with the assumption that the seasonal signals are unimodal (i.e. only one green-up and leaf fall per year). This assumption might not hold true in all semi-arid areas, but the expectation is that other metrics, such as the ’peakiness’ metric discussed earlier, would provide enough information to interpret the phenology of these ecosystems.

First the NDVI profile is split into year objects defined by the growing season (in southern Africa, from July 1 to June 30 of the following year). Thereafter, for each year, the maximum NDVI value and its date of occurrence are determined. In the next step, we compute a ”left” minimum value and its date

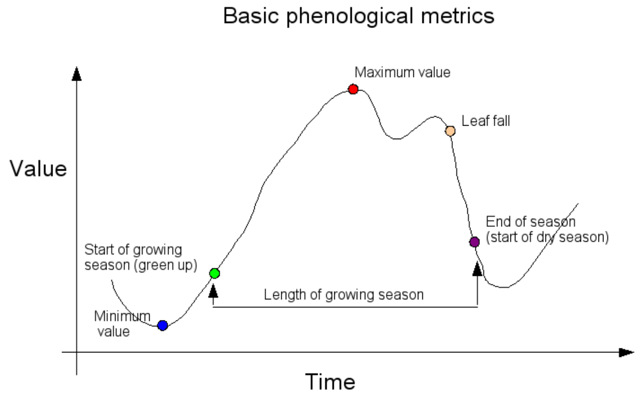

Fig. 3. *Some basic phenological metrics*

by scanning the data between the maximum of the previous year and the current year. A ”right” minimum is also computed using the current year and following year. Thus, we now have, for each year, a maximum value and left and right bounds for computing seasonal metrics. These values are then used to compute the various phenological metrics for each year.

The 9 metrics (some shown in Figure 3) extracted are:

1) Minimum value: minimum value on the smoothed NDVI curve for that year.
2) Date of minimum value.
3) Maximum value: maximum value on the smoothed NDVI curve for that year.
4) Date of maximum value.
5) Green up date (start of the growing season): the threshold method is used. When the smoothened curve crosses a value 10% greater than the difference between the maximum and the ”left” minimum value for that year, it is assumed to be green up day.
6) Leaf fall date (when trees start losing their leaves): the date when the maximum NDVI value has decreased by 10% of the distance between itself and the ”right” minimum value.
7) Date of start of dry season: defined similarly to greenup using the maximum NDVI value and the ”right” minimum value.
8) Length of growing season: this is the difference between green up date and start of the dry season.
9) Net Primary Productivity (NPP): this is the integral under the NDVI curve. Due to the contribution of bare soil to NDVI, a value of 0.1 instead of 0 was used as the minimum NDVI value (zero production).

C. Algorithm parameters

The parameters specified for the different algorithms were established from empirical results. The pre-filtering algorithm works on a radius of 1 pixel. In the case of the Savitsky-Golay filter, we use 7 data points to interpolate the function and a polynomial of order 2. The Gaussian filter has sigma set to 1.0. These parameters ensure a close approximation of the original signal and also fit the upper envelope of data points. The *n*-timestep moving average was computed using $n = 4$.

III. Experimental results

Experiments were conducted on time-series NDVI data from the Skukuza area in the Kruger National Park, South Africa (-25.02 S, 31.49 E). The data spanned the 2005/2006 growing year. The four smoothed curves produced were compared with phenology data collected in the field for the growth year 2005/2006. Weekly measures of tree greenness were collected by classifying the amount of leaves on ten randomly selected individuals of the four major tree species on a scale of 0 (no leaves) to 5 (fully green). This was then normalised to give a value between 0 and 1. Monthly measures of grass greenness were collected by clipping forty nine 1m$\times$1m quadrants. The samples were weighed before (wet biomass) and after drying (dry biomass), and the greenness of the grass calculated as

$$greeness = \frac{(wetBiomass - dryBiomass)}{wetBiomass} \quad (5)$$

which also gives a value between 0 and 1. The cover of each tree species was measured using a densiometer and the landscape greenness was then calculated by summing the greenness value of each component, modified by the fraction of the landscape that it covered.

Once the four sets of phenological metrics were produced they were compared with metrics manually extracted from the time series by three ecologists familiar with the ecosystem. This was considered ’reality’ and the accuracy of each smoothing algorithm and data type were tested by determining the difference between ’reality’ and each method for each metric. Some methods worked better for certain metrics than for others. To determine the overall best method we normalised all differences to a percentage of the total possible difference for each metric: 365 days for the date metrics, 0.8 for the NDVI metrics and 24 for the NPP metric i.e. a difference in greenup dates of 30 days was considered to be 30/365*100 percent accurate and a difference in maximum values of 0.005 was considered to be 0.005/0.8*100 percent accurate. The overall accuracy of each method was then calculated as the average percentage accuracy over all metrics.

Table III presents the error rates and average error rates for each smoothed NDVI profile when computing the metrics. Differences from the ”true” metrics are reported as a percentage of maximum possible error to make the accuracy of the date metrics comparable with the accuracy of the NDVI metrics. The date stamp data, smoothened using the Savitsky-Golay method, produces the least inaccuracy (19.8%). If the same Savitsky-Golay filter is applied to the composited data, it generates the worst accuracy (25.4%). This implies that date stamp time data is more accurate than composited time data for computing seasonal metrics (when using the Savitsky-Golay method). In the case of 8-day composite time intervals, the most accurate extraction of metrics (22.8%) is almost the same as the worst case scenario whens using date stamps (23.1%).

Metric	8-day composite (G)	8-day composite (SVG)	Date stamp (G)	Date stamp (SVG)	Average
Max value	0.0	0.0	0.0	0.0	0.0
Max value date	14.8	23.2	16.7	14.7	17.3
Minvalue	0.0	0.0	0.0	0.0	0.0
Minvalue date	55.0	47.3	56.3	12.2	42.7
Greenup date	26.8	27.8	27.7	32.3	28.7
Leaf fall date	40.2	50.5	40.7	56.3	46.9
Start of dry season	23.3	26.0	26.2	23.8	24.8
Length of growing season	42.2	51.5	37.5	36.8	42.0
NPP	2.8	2.4	2.8	2.4	2.6
Average error (over metrics)	22.8	25.4	23.1	19.8	22.8

TABLE I
ERROR RATES (%) FOR THE GAUSSIAN (G) AND SAVITSKY-GOLAY (SVG) OUTPUTS

IV. Future work

The work presented in this paper shows a promising approach to improving the accuracy of automatically extracted phenological metrics by using date stamped date. The first results presented here are based on a single growing season and it is envisioned that the data set will be expanded so that longer time-series are introduced and processed. However, a constant problem is the lack of sufficient long term field measurements that can be used to verify the outputs of the algorithms for longer time-series data. It is hoped that this problem will be adequately addressed in the near future. A more difficult problem evident in time-series analysis is the issue of missing data. This will be addressed in future algorithm implementations.

V. Conclusion

The results presented in this paper demonstrate that the use of actual date stamped NDVI data can improve the accuracy of algorithms that compute phenological metrics by as much as 5.6%. On average, the Savitsky-Golay filter will improve this accuracy by 3%. In addition, we have also shown that the Savitsky-Golay filtering method works better than the Gaussian filter.

References

[1] P. Jonsson and L. Eklundh, "TIMESAT - a program for analyzing time-series of satellite sensor data," *Computers and Geoscience*, vol. 30, pp. 833–845, 2004.

[2] C. J. E.F. Vermote, N.Z. El Saleous, "Atmospheric correction of modis data in the visible to middle infrared- First results," *Remote Sensing of Enviroment*, vol. 83, pp. 97–111, 2002.

[3] P. Jonsson and L. Eklundh, "Seasonality extraction by function fitting to time-series of satellite sensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 8, pp. 1824–1832, 2002.

[4] R. Gonzalez and R. Woods, *Digital image processing*. Addison-Wesley Publishing Company, 2002.

[5] J.F. Hermance, R.W. Jacob, B.A. Bradley and J.F. Mustard, "Extracting phenological signals from multi-year AVHRR NDVI time series: Framework for applying high-order annual splines with roughness dampening," in press.