

Streamlining multitemporal vegetation indices for dependable crop growth monitoring in Himalayan foothill region

SANDEEP KUMAR SINGLA*, RAHUL DEV GARG and OM PRAKASH DUBEY

Geomatics Engineering Group, Department of Civil Engineering, Indian Institute of Technology (IIT) Roorkee, Roorkee, India
e-mail: sandy.dce2014@iitr.ac.in

MS received 25 October 2017; revised 15 December 2018; accepted 8 March 2019; published online 9 May 2019

Abstract. Satellite data in conjunction with geoinformatics are used to study the land cover change dynamics, extraction of crop information and the monitoring of crop growth. The information derived from the satellite may contain contaminated values due to the atmospheric effects, geometric errors, snow and clouds. These contaminated values can be identified and eliminated using time series analysis to further streamline for agricultural monitoring and prognostic applications. The inherent advantages and disadvantages of existing streamlining methods limit their usability in particular situation. The method proposed in this study synergizes the use of interpolation, running median and moving average. This has clearly shown the proposal's capability in preserving the trend in the series in addition to streamlining the temporal profile of satellite data in the Himalayan foothills. This will make the road map for satisfactory crop growth monitoring and crop yield estimation. Analysis based on the root mean square error and *F*-Test has been presented to deduce the results and interpretations.

Keywords. Remote sensing; NDVI; streamlining; crop yield estimation; crop growth monitoring; temporal data.

1. Introduction

1.1 Motivation

The utilization and application of remote sensing data for agricultural development was investigated since 1971 under Corn Blight Watch Experiment (CBWE) [1]. The capability to effectively extract the crop variety and crop growth information for prognostic analysis from remote sensing data is proven [2]. Multitemporal profiles of satellite data provide representative and spatially exhaustive information on the development of the model and for classification and pattern recognition [3]. Another experiment carried out by [4] using LANDSAT data was known as Crop Identification Technology Assessment for Remote Sensing (CITARS). It aimed at identification of two major crops, corn and soybean, and testing the concept of signature extraction. Large area crop inventory experiment (LACIE) [5] carried out during 1974–78 was a major international study carried out in the agricultural domain based on the satellite data.

In order to improve agricultural statistics, remotely sensed data have been investigated over the last decades [6, 7]. The research paper [8] presented a study on the use of vegetation indices based on satellite data for the crop

growth monitoring. Nevertheless, remotely sensed information may be contaminated due to a variety of factors such as clouds, snow, geometric errors and other atmospheric effects [9]. These errors always reduce the consistency, reliability and applicability of remote sensing data, especially for the agricultural applications [10, 11].

A variety of numerical solutions are available to surmount these effects on Normalized Difference Vegetation Index (NDVI) values due to contamination. One of the widespread approach is to carry out the time series analysis to streamline the time series of irregularly spaced observations [12]. These methods are based on some assumptions about NDVI that minor variations (upwards or downwards) in NDVI values are due to the phenological trend, whereas high-frequency variations amount to the sensory defects or atmospheric errors [13]. Therefore, it is often assumed that higher NDVI observations are generally the most accurate, and noise reduction techniques thus aim to preserve the upper envelope of NDVI values in a times series. The fitting methods, commonly known as regression-based methods, can be described using a function that may pass through some, but often few, of the original data points. On the other hand, the methods based on rules to replace each observed point by a linear combination of a set of nearby points in the profile are collectively known as filtering methods [14]. Regression-based methods may

*For correspondence

generate spurious oscillations that distort the time series [15]. Regression-based models generally bypass the too many high- and low-data points. Regression models are effective only in the case of very few contaminated points.

Among the widely used methods, Fourier series method has percolated in the data analysis [16]. The Fourier series method may be employed to represent noisy data by the amplitudes of cosine and sine functions. A process to remove the noise signal from the overall data depends upon the statistical information of the signal. The Fourier series is expanded to generate a spectrum related to the signal and noise. Fourier method for streamlining is very effective in terms of compression and data storage [17]. However, the Fourier series method may not be consistently applicable when there are frequent changes in the profile. It will significantly influence the crop growth monitoring, particularly during the growing period [13].

A comparative study of various methods to remove noise in the NDVI time series revealed that the performance of Savitzky–Golay (SG) method is the best in most of the situations [18]. Although a number of models have been investigated in the past for the streamlining of NDVI, yet there is a need of enhancement as per the study area and specific crop pattern.

Medians can be used to estimate the smooth component for the different points in the observed data as well as some extrapolated points. Performance of streamlining the series is influenced by the span of the data considered, which is the number of points chosen for taking the median at a time [19]. The outliers or the spikes in the data can be easily streamlined using the robustness of the median operator. It is to be mentioned that the median operator is nonlinear; consequently, the streamliners based on the medians are also nonlinear, making it robust. Other linear operators and the filters can be used once the outliers have been removed by resistant nonlinear running median streamlining approach.

A close look of the running median streamliner reveals that though it is efficient to streamline or reconstruct the time series, at the same time it also removes some key values that are essential for further prognostic analysis that requires NDVI values as basic input [20]. Therefore, for an efficient and dependable NDVI time series analysis, especially for the agricultural applications, it is imperative that the streamliner or the reconstruction method ensures that it modifies only contaminated values that do not follow the general crop growth pattern. This will help in preserving the uncontaminated points or peaks in the NDVI series. Under these practical situations, the moving average method followed by judicious selection of peaks is the best choice to reconstruct the time series purposefully efficiently.

It is worth mentioning that each of the methods has its own merits and demerits. This limits their usability in a particular situation [14, 21]. Therefore, appropriate blending of remotely sensed data and some

suitable techniques may be useful to reconstruct the time series. Keeping this in consideration, the objective of the proposed work is to design a robust and simple approach for streamlining of the temporal profile of NDVI values considering the crop growth pattern in the Himalayan foothill region so as to make a road map for dependable crop yield estimation.

The performance evaluation of streamlining time series analysis is essential. In this study it moves around bridging the gap between observed NDVI time series and streamlined or reconstructed time series. In [22], quality measures for assessing the performance of the streamlining are proposed. On the contrary, the authors of [21] simulated the time series and used the simulated series as a reference series. The series is again rebuilt by the addition of noise effects at particular levels. To evaluate the performance of the different streamlining approaches, the absolute difference between the noise-affected time series and reference time series has been taken into account; statistical measures and quality indicators such as *F*-value and root mean square error (RMSE) have been used.

1.2 Related work

A comprehensive review of the available literature revealed that the remotely sensed data are the nucleus of the land use and land cover information. Spatial and temporal variation in the satellite data may be effectively used to extract the information related to crop growth monitoring and for the study of the land cover change dynamics [23–26]. Multitemporal analysis of LANDSAT data may be effectively used to extract the information related to the sugarcane [27, 28]. LANDSAT 8 OLI bands are also suitable for the crop classification and crop growth monitoring [29] and the performance of LANDSAT OLI bands is much better than that of LANDSAT (Enhanced Thematic Mapper Plus) ETM+ bands [30].

Available data streamlining approaches may be broadly categorized into four major groups [18]. The categories have been represented diagrammatically in figure 1. The Best Index Slope Extraction (BISE) method [31] and the Modified BISE [32] method bank on the fixation of the threshold value of the tolerable increase in the curve. However, this strategy is generally subjective in addition to

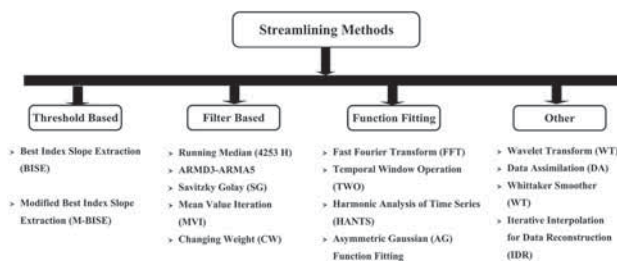


Figure 1. Categorization of streamlining methods.

the dependability on the area and the user and may lead to unreliable results. Data streamlining approaches based on Fourier series transformations [16, 33] are computationally inefficient for asymmetric and irregular NDVI data. The reason of inefficiency may be due to the dependence on symmetric cosine and sine functions. An approach based on AG function fitting [14] appears to be a more qualified candidate to streamline the NDVI time series. However, while implementing, it may not be easy to spot a rational and reliable set of minima and maxima for accommodating local functional variations. Consequently the approach becomes complex and time consuming. Under these circumstances an efficient method known as SG filter has been proposed in [15], which is an enhancement of SG method [34]. For successful application of this method, the availability of ancillary data related to the cloud flag band that is free from ambiguity is the basic requirement. In spite of this, it may experience an edge effect adversely affecting the eight endpoints.

Some other streamlining methods such as mean-value iteration (MVI) filter [35], double logistic function fitting [36], wavelet transform method [37, 38], the iterative Interpolation for Data Reconstruction (IDR) method [39], Changing-Weight Filter Method [40] and recently a method based on Moving Weighted Harmonic Analysis (MWhA) that is enhancement of Harmonic Analysis of Time Series (HANTS) algorithm have been proposed for reconstructing high-quality SPOT VEGETATION NDVI Time-Series Data [41]. The literature review goes some inconsistent conclusions on the streamlining performance of existing methods. To mention one of them, in [15] it was demonstrated that SG filter is superior to harmonic-based methods, whereas [42] explored on harmonic analysis method being much superior. These types of diverging conclusions can be ascribed to the dependence of the streamlining methods on specific initial conditions, including study areas, contamination factors, crop phenology and crop growth pattern. Thus, it is imperative to propose a new and robust method suitable for Himalayan foothills incorporating local contaminating agroclimatic factors.

1.3 Main contributions

In this study, an attempt has been made to analyse the effect of noise due to cloud cover and other associated climatic factors on the temporal NDVI profile. This situation is pronounced during the monitoring of crop growth in the Himalayan foothills region. Streamlining experiment was conducted by simulating the NDVI time series data of one-year memory period. The NDVI time series was generated from the data collected by LANDSAT 8 OLI. The generated series was streamlined followed by performance assessment.

This paper is organized in the following manner. In section 1, the introduction has been presented, which

includes motivation, literature review and main contributions of the proposed study. Brief review of the data required and the study area has been discussed in section 2. The proposed methodology has been discussed in section 3. Results and discussion are explained in section 4. The study has been concluded in section 5.

2. Description of the area and data

This research is focused on streamlining the temporal profile of NDVI to eliminate the contaminated values due to atmospheric effects, cloud cover or some kind of geometric errors. For the study, data and information from multiple sources have been fused to synthesize the analysis process. Details pertaining to the study area and the data set used in the study are described here.

2.1 Study area

The study has been carried out in a small experimental area in the Himalayan foothills (figure 2). The area is dominated by agricultural practices in Bhadarabad region and Haridwar district, Uttarakhand State, India. The area of the proposed work is enveloped by 29°55'35"N–29°56'10"N latitude and 77°57'40"E–77°58'5"E longitude.

2.2 Historical data

Historical sugarcane growth and yield data have been collected from the State Agriculture Department, sugarcane

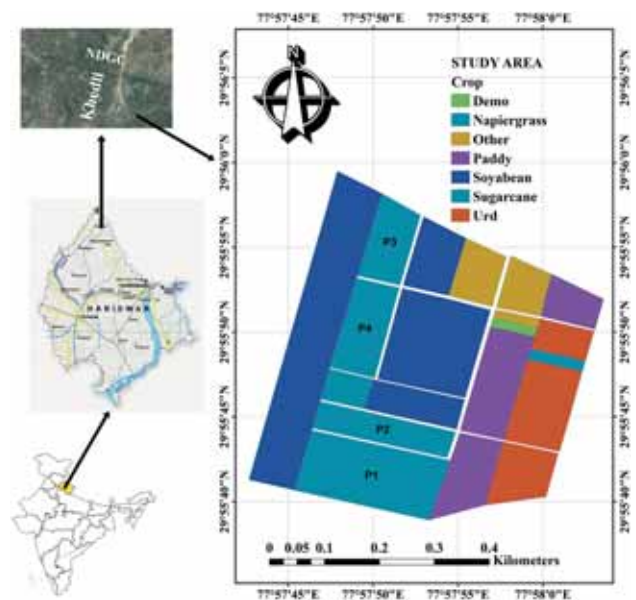


Figure 2. Study area.

industries and interaction with farmers. In addition, the sowing and harvesting periods of the different crops grown in the study area have been collected during field traversing and interaction with the farmers. The synthesized crop calendar of the study area is shown in figure 3.

Ground truth information using GPS for the experimental area has been collected for the year 2015. This has been done twice a month, more or less synchronizing with the LANDSAT 8 overpass on the area. Preliminary analysis of the meteorological data shows that the average temperature varies between 2.40 and 42.60 °C, the relative humidity varies between 11% and 100%, average vapour pressure varies between 5.6 and 27.7 mm, monthly rainfall varies between 0.0 and 134.8 mm, and evaporation varies between 0.2 and 10 mm/day, whereas the wind velocity has been found to vary between 0.4 and 5.9 km/h. These meteorological parameters reconfirm that the study area is suitable for the sugarcane from meteorological consideration.

2.3 Satellite data

Satellite data and metadata from the LANDSAT 8 OLI (Operational Land Imager) for the entire year 2015 have been obtained for the proposed work. The different bands used to create temporal profile of NDVI values are presented in table 1.

3. Methodology

Reconstruction or streamlining of NDVI time series for removal of contaminated points such as cloud cover and noise is based on the following assumptions. (1) Vegetation indices are associated with the changes in vegetation and temporal profile of NDVI generally follows the crop

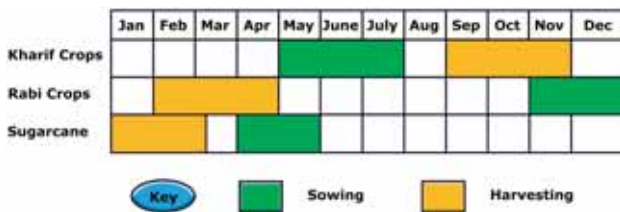


Figure 3. Sowing and harvesting periods of the study area.

Table 1. Wavelength and resolution of OLI bands used in the proposed study.

Band	Wavelength (μm)	Resolution (m)
Red (R)	0.64–0.67	30
Near infra-red (NIR)	0.85–0.88	30

calendar, crop condition and growth status. (2) Contaminated values of NDVI due to cloud and other atmospheric effects generally drag down the NDVI values, which consequently affects the growth monitoring of the vegetation and further these contaminated values are not attuned to the actual growth process of vegetation and crops. Taking these assumptions into consideration this study proposed a new methodology based on optimal selection of treated NDVI values to streamline the observed NDVI time series so as to preserve the general crop growth trend. A flow diagram of the proposed methodology has been given in figure 4.

3.1 Preprocessing

First of all, in this work a normalization formula has been used to convert the Digital Number (DN) values to reflectance values (physical measurements of the part of the solar energy reflected by earth features) based on the LANDSAT 8 radiance re-scaling factors available in the metadata (MTL) file [43]. Mathematically, this transformation without incorporating adjustments for sun angle is expressed as

$$\rho\lambda' = M_{\rho}Q_{cal} + A_{\rho} \tag{1}$$

where

$\rho\lambda'$ is Top of Atmosphere (TOA) planetary reflectance, M_{ρ} band-specific multiplicative re-scaling factor, A_{ρ} is band-specific additive re-scaling factor and Q_{cal} is quantized and calibrated pixel values (DN).

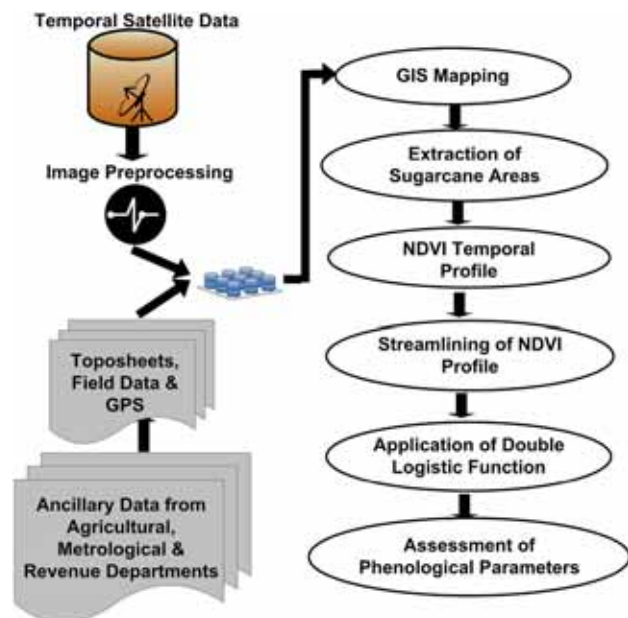


Figure 4. Flow diagram of the proposed methodology.

The corresponding reflectance value has been calculated using Eq. (1), which has been further corrected for the sun angle using

$$\rho\lambda = \frac{\rho\lambda'}{\cos(\theta_{SZ})} = \frac{\rho\lambda'}{\sin(\theta_{SE})} \quad (2)$$

where

$\rho\lambda$ is sun-angle-corrected TOA planetary reflectance,
 θ_{SE} is local sun elevation angle and
 θ_{SZ} is local solar zenith angle.

3.2 Vegetation indices

Vegetation indices may be used to relate the crop parameters with remotely sensed data. These indices are the amalgamation of the different bands of satellite data. These indices are modulated by the interaction of solar radiation with crop photosynthesis. A number of vegetation indices have been proposed by various researchers in the past. A few commonly used indices are NDVI, Soil Adjusted Vegetation Index (SAVI) and Ratio Vegetation Index (RVI). A comprehensive list of various vegetation indices has been given by Basso *et al* [44].

3.3 Generation of NDVI temporal profile

The NDVI [45] has been computed using

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3)$$

where R represents the reflectance in the red band while NIR is the near-infrared band reflectance value of LANDSAT 8 OLI (table 1). The value of NDVI varies between -1 and $+1$. Generally, the negative values and values close to zero are specific to soil devoid of vegetation or of sparse vegetation, whereas the surfaces covered by dense and healthy vegetation have the values $0.7-1.0$. The generated temporal NDVI profile of four agricultural fields (P1–P4) containing different varieties of sugarcane in the study area for the year 2015 is shown in figure 5.

Visual analysis of figure 5 in conjunction with general crop growth information pertaining to the study area confirms the belief that some of the values are contaminated. Field traversing and ground truth information revealed that the contamination of NDVI values is due to the presence of integrated noise popping up as the cumulative effect of cloud cover and other climatic reasons. The objective of the proposed study is to update or streamline these errors and contaminated values with nearly true information. For this, these contaminated NDVI values will be transformed using some robust smoothing filter to streamline the NDVI temporal profile suitable for the Himalayan foothills area.

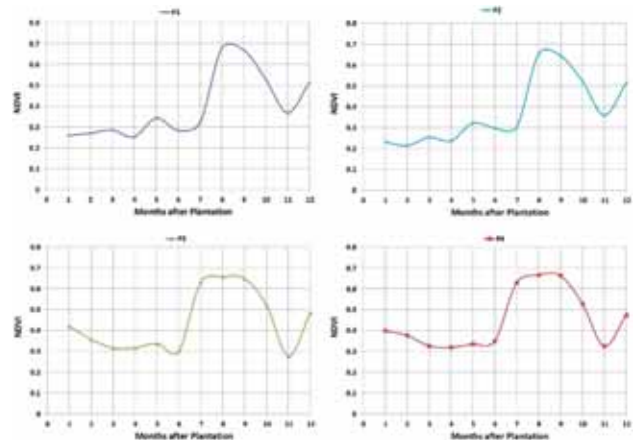


Figure 5. Temporal profile of NDVI for sugarcane fields (P1–P4).

3.4 Fourier series transformation

As mentioned earlier, Fourier series transformation is a widely used technique for data fitting. Fourier series, which is a series of sums of sine and cosine functions, is used to represent a function with a constant period of oscillation. The wavelet components of sine and cosine at a particular frequency are known as the harmonics of the Fourier series. The number of harmonics influences the performance of streamlining the series. Crop growth profile satisfies all the conditions of the simple harmonic motion. Figure 6 shows the concept of using Fourier series to estimate the underlying curve in a periodic time series. The Fourier series or harmonic analysis considering the NDVI temporal profile for the streamlining or reconstruction of the profile or time series data consisting of unwanted errors or gaps has been mathematically represented as

$$U_t = a_0 + a_1 \cos\left(\frac{2\pi}{\lambda}t\right) + a_2 \cos\left(\frac{2\pi}{\lambda}2t\right) + \dots + b_1 \sin\left(\frac{2\pi}{\lambda}t\right) + b_2 \sin\left(\frac{2\pi}{\lambda}2t\right) + \dots \quad (4)$$

where $b_i = \frac{2}{n} \sum_{t=1}^n U_t \sin\left(\frac{2\pi}{t}it\right), (i = 1, 2, \dots)$

$$a_i = \frac{2}{n} \sum_{t=1}^n U_t \cos\left(\frac{2\pi}{t}it\right), (i = 1, 2, \dots)$$

$$a_0 = \frac{1}{n} \sum_{t=1}^n U_t$$



Figure 6. Fourier series to estimate the underlying curve.

and n denotes the number of constant components or terms in the time series. In general form, the streamlining may be represented as

$$U'(t) = U(t) + \varepsilon(t) \quad (5)$$

where $U(t)$ is constructed series, $U'(t)$ is the original series, $\varepsilon(t)$ is the error series and t is the time at which the value is obtained (observed); a_i and b_i are the cosine and sine wavelets, respectively, with frequency f . The computed value of a_0 is the mean of the given time series.

3.5 SG filtering

The authors in [34] proposed a polynomial function fitting method based on a weighted least squares regression method. This filtering technique preserved the features of the data set such as relative maxima, minima and widths. The filter may be treated as a weighted moving average filter with weights given as a polynomial of a certain degree. This filter may be implemented on any consecutive data when the data points are located in fixed and uniform intervals along the chosen abscissa. The curve evolved by plotting the points must be continuous and more or less smooth. In this study, the data set belongs to NDVI temporal values observed at regular intervals. NDVI values are distributed in 16-day composites, which do not mean that the data for any given pixel occurred on a regular frequency of 16 days. However, in this study, SG filter is used to find out its impact on the NDVI temporal profile. The filter algorithm is as follows:

$$NDVI'_j = \frac{\sum_{i=-n}^n C_i}{N} NDVI_{j+i} \quad (6)$$

where $NDVI_j$ is the original NDVI value, $NDVI'_j$ is the filtered NDVI value, C_i is the coefficient for the i^{th} NDVI value of the filter (smoothing window) and N is the number of convoluting integers equal to the smoothing window size ($2n + 1$). The index j is a running index of the original ordinate data table. The smoothing array (filter size) consists of $2n + 1$ points, where n is the half-width of the smoothing window. A larger value of n produces a smoother result at the expense of flattening sharp peaks.

3.6 Streamlining by running medians

The running medians approach for streamlining time series analysis aims to retain the general characteristics of the series. This method of streamlining is based on the assumption that the entire series can be obtained as the sum of smooth as well as non-smooth or rough components [46]. Based on the literature review and the consideration of the

robustness as well as computational efficiency and the complexity, a running median streamliner-5 [47] has been used, which is defined by Eq. (7):

$$Z_t = \text{median}(NDVI_{t-2}, NDVI_{t-1}, NDVI_t, NDVI_{t+1}, NDVI_{t+2}) \quad (7)$$

where $NDVI_t$ is the original time series data. Values at the endpoints are streamlined using different versions of this method. A formula based on median of 3 values is given by Eq. (8):

$$Z_t = \text{median}(NDVI_{t-1}, NDVI_t, NDVI_{t+1}) \quad (8)$$

whereas streamliner based on 1 median may be defined by $\text{median}(NDVI_t)$; however, practically it does nothing.

3.7 Streamlining by moving average

The time series analysis revolves on the assumption that for the perfectly streamlined NDVI time series, each NDVI value should come within the reach of the average of its preceding and succeeding points. One very commonly used streamliner to reconstruct the time series is MVI filter and its approach is purely based on this assumption. Moving average streamlining method consists of measurement of trend by smoothing out the fluctuations in the data by means of a moving average. Moving average of extent m is a series of successive averages or the arithmetic means of m points at a time, starting with 1st, 2nd, 3rd point and so on. The formula for the 3-point moving average method [47] is given by

$$M_t = \frac{NDVI_{t-1} + NDVI_t + NDVI_{t+1}}{3} \quad (9)$$

Practical experience in the study area indicates that none of the afore-mentioned methods in isolation are sufficiently efficient to carry out the effective reconstruction of temporal profile of NDVI. Therefore, appropriate blending of some suitable techniques may be useful to reconstruct the time series data. For this, a new approach for streamlining the NDVI profile is being proposed in this study.

3.8 Streamlining by optimal selection of treated NDVI values

The NDVI temporal profile will be streamlined in such a way that resulted NDVI temporal profile delivers the near-real-time growth curve to be effectively used for the agricultural applications. For achieving this, the proposed streamlining process has been decomposed into following modules.

- Module M1: running medians streamlining.
- Module M2: moving average streamlining.

- Module M3: cubic interpolation and selection of the optimal values.

Modules M1 and M2 have already been described in section 3.6 and 3.7. For implementing module M3, this study makes use of the cubic interpolation and maximum value among the values obtained in the first two modules and the original data values. This step is proposed to be applied just to preserve the key points and peaks of the NDVI profile.

Cubic spline approach may be effectively employed to produce smooth curves passing through the given set of points. This method is based on the creation of smooth cubic spline curves over consecutive intervals of the given values and these curves are very smooth at the knot points also. Spline fitting is of recent origin and has been employed by various researchers to interpolate the temporal data.

Cubic spline to interpolate $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ where x_i are the values of NDVI and y_i are the values of the time of observations, may be given as

$$f''(x) = \frac{1}{h} [(x_{i+1} - x)f''(x_i) + (x - x_i)f''(x_{i+1})] \quad (10)$$

where h is the time interval between two observations. After the integration of Eq. (10), we get

$$f(x) = \frac{1}{h} \left[\frac{(x_{i+1} - x)^3}{3!} f''(x_i) + \frac{(x - x_i)^3}{3!} f''(x_{i+1}) \right] + a_i(x_{i+1} - x) + b_i(x - x_i) \quad (11)$$

where the constants a_i, b_i are obtained by substituting the values of $y = f(x)$ at x_i and x_{i+1} .

After streamlining the NDVI temporal profile with preliminary modules, the double logistic equation [48] has been used to simulate the growth of sugarcane in the study area. The growth of sugarcane consists of two stages, leaf expansion and senescence, and each of the stages may be described by a singular sigmoid function.

The double logistic equation has been used to simulate the variation of NDVI over time to further analyse the phenological behaviour of sugarcane in the study area. The equation is represented as

$$NDVI(t) = NDVI_{max}(NDVI_{max} - NDVI_{min}) \left[\frac{1}{1 + e^{-a(t-b)}} + \frac{1}{1 + e^{c(t-d)}} - 1 \right] \quad (12)$$

where t is the time and $NDVI(t)$ is the numerical value of NDVI at the given time t for a specific cropping period; $NDVI_{max}$ and $NDVI_{min}$ variables are, respectively, the maximum and minimum numerical values of NDVI time series. The constants a, b, c and d are the fitting parameters, where b and d are associated with the rates of increasing inflection point a and decreasing inflection point c , respectively.

All the modules and steps of the methodology have been implemented with the help of the following algorithm:

Algorithm 1: Streamlining algorithm.

Step 1: Inputs and initialization

Compilation of input data: ground truth, meteorological, ancillary information and LANDSAT OLI data

Step 2: Preprocessing

Analysis based on metadata (MTL) files: Reading different parameters from metadata files, data sufficiency checking, selection of the imagery (cloud cover less than 30%) and field traversing

-layer stack of temporal images

-Digital Number (DN) to reflectance [Eqs. (1) and (2)]

-NDVI temporal profile from reflectance [Eq. (3)]

Step 3: Streamlining of temporal NDVI

-Calculation of mean NDVI (for pixels obtained in Step 2)

-Generation of scatter plot

-Smoothing of the obtained curve based on Fourier series

-SG filter streamlining

-Running medians streamlining [Eqs. (7) and (8)]

-Enhancement based on moving point average [Eq. (9)]

-Optimal value selection using cubic spline method [Eqs. (10) and (11)]

-Double logistic method [Eq. (12)]

4. Results and discussion

This study has been carried out in an experimental area, a part of the Himalayan foothill region of Haridwar District, Uttarakhand, India. Four sugarcane agricultural fields (P1–P4) have been identified and selected for the experimental work. The scatter plots of the original temporal NDVI values of LANDSAT 8 OLI sensor have been diagrammatically presented in figure 5. Preliminary analysis of figure 5 revealed the significance and presence of the contaminated values, specifically in the initial stage of the plant growth. Probable contaminated errors have been indicated in figure 5. The contaminated errors have been detected due to the presence of abnormal decrease in NDVI, and the numerical value of NDVI at these points does not fit into the general crop growth model. Also, the ground truth information and field traversing brought out the fact that during this period the general crop growth was normal. Consequently, there is the requirement of streamlining the temporal profile of NDVI values so as to monitor and predict the crop growth in an appropriate manner. As mentioned earlier the literature review indicated that use of Fourier series and SG filter may be suitable for streamlining the NDVI profile in this area. Therefore first of all Fourier Series has been used to streamline the observed NDVI profile. Figure 7 shows the scatter plot of observed and streamlined NDVI time series using Fourier series having 3

harmonics, whereas figure 8 depicts the streamlining of the series using SG filter and the original value. Figure 9 shows the observed and smoothed NDVI time series using an optimal selection of treated NDVI values, the proposed method in this study. Figures 7, 8 and 9 present the data for the original series of temporal NDVI values, measured

through satellite images of LANDSAT 8 OLI, whereas the predicted series of temporal NDVI values is obtained through the different methods of streamlining.

Cavernous study of figures 7, 8 and 9 brings out the fact that both Fourier series wavelet components are not able to preserve the general trend of the NDVI time series for the study area. Fourier series combination of both cosine and sine wavelets leads to representation of the NDVI time series data to some extent only. However, the SG filter has markedly improved the trend preservation of the series. Nevertheless, one peculiar drawback that has been noticed is that the SG filter has shifted the peak towards higher days. This feature of SG filter may tend to be the biggest limitation while using the NDVI time series data for any successful crop modelling and prognostic study. The method proposed in this study, which synergies the integrated use of running median, moving average and cubic interpolation, has clearly shown its capability in preserving the trend in the series in addition to streamlining the NDVI time series in the Himalayan foothills.

Further, a closer look of these figures clearly brings out another fact that there are two distinguishable patterns of NDVI variations: one for P1, P2 and another for P3, P4. Peculiarly, it has been observed that towards the end of the growing season, NDVI values are almost same for both the patterns, which is around 260 days from the Julian day. At this juncture the field traversing confirmed the fact that plants P1 and P2 belong to a specific variety of sugarcane, in contrast with notable biological peculiarity from P3 and P4. This is inferred from the fact that that P3 and P4 show higher NDVI during the initial stage of the crop growth. At the critical stage, however, both patterns attain more or less similar numerical value of NDVI.

The phenological attributes of the sugarcane have been parameterized and classified based on the phase, variance, mean and the variation in peaks with reference to the time period in the cropping season. The extracted characteristics of the field type have been given in table 2.

In order to assess the streamlining performance of investigated methods quantitatively, the statistical indicator RMSE has been used in this study. The RMSE gives the weighted variations in errors (residual) between the original and streamlined series. This is numerically calculated as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (s_{y_i} - o_{y_i})^2} \quad (13)$$

where n denotes number of measurements, s_y is the streamlined NDVI while o_y is the original time series. The values of RMSE obtained for the Fourier series, SG filter and the proposed method have been tabulated in table 3. Values of this table bring out the fact that the RMSE values range from 0.2094 to 0.2214 for Fourier series approach, for SG filter it is 0.0477–0.0623 and for the proposed approach the range is from 0.0427 to 0.0616. This is

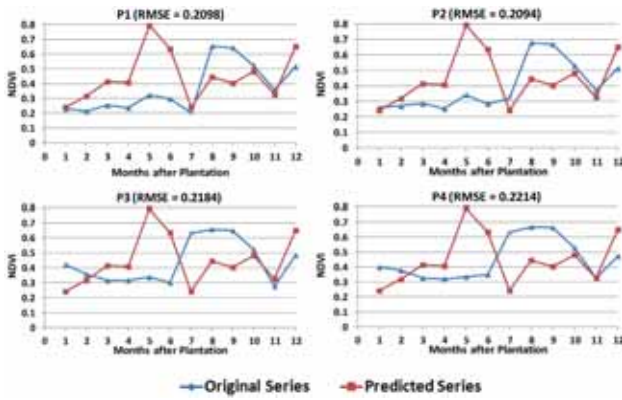


Figure 7. Streamlining temporal profile of NDVI (Fourier series method).

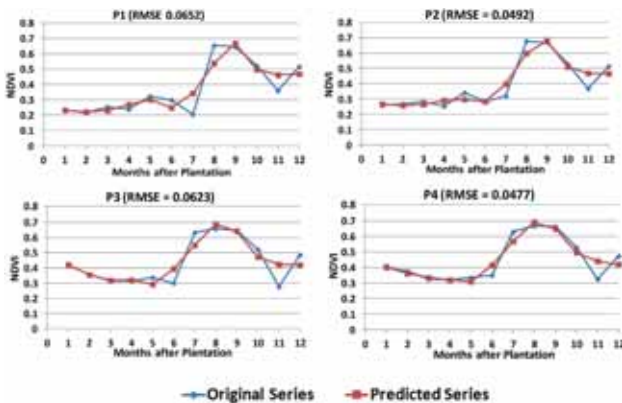


Figure 8. Streamlining temporal profile of NDVI (SG filter).

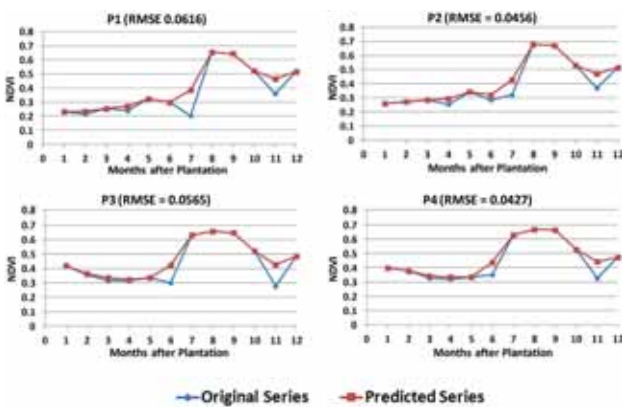


Figure 9. Streamlining temporal profile of NDVI (proposed method).

Table 2. Phenological characteristics of each pattern.

Sr. no.	Type of plant	NDVI characteristics	Plant abb.
1	Sugarcane type 1	Increases in steady and slow state	P1, P2
2	Sugarcane type 2	Starts with initial decrease and then increases instantly in the middle of growing season	P3, P4

Table 3. RMSE values for different methods in the study.

Method	P1	P2	P3	P4
Fourier	0.2098	0.2094	0.2184	0.2214
SG	0.0652	0.0492	0.0623	0.0477
Proposed	0.0616	0.0456	0.0565	0.04279

indicative of the fact that numerically the proposed method is capable of yielding better results as compared with both the Fourier series and SG filter. This is in line with the visual analysis of figure 9. Consequently, this demonstrates the usability of the proposed approach for streamlining the observed NDVI values for crop growth monitoring in the Himalayan foothills area. Besides giving better values of RMSE, the proposed method has been found to preserve the peak values for the NDVI, which is very essential for the prognostic analysis of NDVI series for the crop yield estimation, which banks on the vegetation indices obtained from satellite data. The analysis of the results also revealed that the selection and performance of streamlining method is site specific, crop specific and depends on the temporal variations in the NDVI time series.

Statistical analysis based on *F*-Test has been performed to further investigate the results obtained with the proposed method. The *F*-Test has been performed at 5% level of significance for checking the equivalence of the variances of the obtained results. The obtained statistical values related to the *F*-Test have been presented in table 4. The obtained values indicate that the null hypothesis of equal variances is valid in all the cases. Further, since the RMSE values obtained for the proposed method are the lowest, it has been concluded that the performance of the proposed method is significantly better than those of other methods when applied in the study area.

Table 4. Statistical analysis based on *F*-Test.

Method	P1	P2	P3	P4
Fourier	1.027346	1.029164	1.057071	1.052035
SG	1.278015	1.195713	1.314613	1.190095
Proposed	1.157439	1.100273	1.354964	1.211276

F-critical = 2.81793.

Based on the analysis after the application of double logistic equation, it has been found that plants P1 and P2 have lower numerical values in the initial stages of the crop growth and instant high values near the maturity period while P3 and P4 have medium high values in the starting and very steady progress in the middle of the season and high values towards the end of the season. The appearance of outliers in the NDVI temporal profile may be attributed to a variety of factors ranging from the atmospheric effects to varying degree of cloud cover. It is important to point out that the field traversing confirms that atmospheric effects and cloud cover are causative contamination factors in the memory period of the data.

5. Conclusions

NDVI images were extracted from LANDSAT OLI sensor having temporal resolution of 16 days. The sugarcane growth cycle is spread over the end of January and start of February to the end of following year January. The collected LANDSAT data were processed from January 2015 to December 2015 to include the complete growth cycle. A brief preliminarily analysis has been conducted to extract the potential sugarcane pixels. The streamlining algorithms based on running median as well as moving average and spline interpolation were implemented to streamline the NDVI profile. Quantitative and qualitative analysis based on the RMSE and *F*-Test was presented to deduce the suitability and interpretations. The study indicates that the quantitative performance of the Fourier series approach is the lowest and performances of SG filter and proposed approach are nearly same. However the optimal selection of treated NDVI values approach yields better results for the study area. This is attributed to the fact that the SG filter shifts the peaks. The filtered or streamlined LANDSAT time series NDVI then can be used to develop the classification rules of sugarcane type and crop growth monitoring in the region. The results revealed that the performance of smoothing or reconstruction method depends highly upon the variations in the numerical values of temporal NDVI. In spite of the fact that the present study is based on the data for one single year only, it covers the entire crop growth season and hence it is reasonable to deduce that the observed accuracy of the proposed method is operational in achieving the global normal accuracy in Himalayan foothill areas having diverse agricultural practices.

References

- [1] MacDonald R B, Bauer M E, Allen R D, Clifton J W, Erickson J D and Landgrebe D A 1972 *Results of the 1971 Corn Blight Watch Experiment*. LARS Tech. Rep. 107-107
- [2] Murthy R S, Venkataratnam L and Saxena R K 1983 Application of remote sensing techniques for land evaluation and classification for agriculture. *Proc. Indian Acad. Sci. (Eng. Sci.)* 6: 177–188
- [3] Laxman S and Sastry P S 2006 A survey of temporal data mining. *Sadhana* 31: 173–198
- [4] Bauer M E, Cary T K, Davis B J and Swain P H 1975 *Crop identification technology assessment for remote sensing (CITARS)*. NASA-CR-147389, LARS-INFORM-NOTE-072175, pp. 1–59
- [5] Bauer M E, McEwen M C, Malila W A and Harlan J C 1979 *Design, implementation and results of LACIE field research*. LARS Tech. Rep. 102579, pp. 1037–1066
- [6] Dempewolf J, Adusei B, Inbal B R, Hansen M, Potapov P, Khan A and Barker B 2104 Wheat yield forecasting for Punjab province from vegetation index time series and historic crop statistics. *Remote Sens.* 6: 9653–9675
- [7] Dadhwal V K, Singh R P, Dutta S and Parihar J S 2002 Remote sensing based crop inventory: a review of Indian experience. *Trop. Ecol.* 43: 107–122
- [8] Silleos N G, Alexandridis T K, Gitas I Z and Perakis K 2006 Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto Int.* 21: 21–28
- [9] Deekshatulu B L and Krishnan R 1983 Basic research problem in remote sensing. *Proc. Indian Acad. Sci. (Eng. Sci.)* 6: 337–354
- [10] Mulianga B, Bégué A, Clouvel P and Todoroff P 2015 Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing. *Remote Sens.* 7: 14428–14444
- [11] Wei W, Wu W, Li Z, Yang P and Zhou Q 2016 Selecting the optimal NDVI time-series reconstruction technique for crop phenology detection. *Intell. Autom. Soft. Comput.* 22: 237–247
- [12] Holben B N 1986 Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* 7: 1417–1434
- [13] van Dijk A, Callis S L, Sakamoto C M and Decker W L 1987 Smoothing vegetation index profiles: an alternative method for reducing radiometric disturbance in NOAA/AVHRR data. *Photogramm. Eng. Remote Sens.* 53: 1059–1067
- [14] Jonsson P and Eklundh L 2002 Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans. Geosci. Remote Sens.* 40: 1824–1832
- [15] Chen J, Jonsson P, Tamura M, Gu Z, Matsushita B and Eklundh L 2004 A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* 91: 332–344
- [16] Menenti M, Azzali S, Verhoef W and Swol R 1993 Mapping agroecological zones and time lag in vegetation growth by means of fourier analysis of time series of NDVI images. *Adv. Space Res.* 13: 233–237
- [17] Kosarev E L and Pantos E 1983 Optimal smoothing of 'noisy' data by fast Fourier transform. *J. Phys. E: Sci. Instrum.* 16: 537–543
- [18] Geng L, Ma M, Wang X, Yu W, Jia S and Wang H 2014 Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe river basin, China. *Remote Sens.* 6: 2024–2049
- [19] Velleman P F and Hoanglin D C 1981 *Applications, basics and computing of exploratory data analysis*. Boston, MA: Duxbury Press
- [20] Reed B C, Brown J F, VanderZee D, Loveland T R, Merchant J W and Ohlen D O 1994 Measuring phenological variability from satellite imagery. *J. Veg. Sci.* 5: 703–714
- [21] Hird J N and McDermid G J 2009 Noise reduction of NDVI time series: an empirical comparison of selected techniques. *Remote Sens. Environ.* 113: 248–258
- [22] Atzberger C and Eilers P H C 2011 Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *Int. J. Remote Sens.* 32: 3689–3709
- [23] Tucker C J, Townshend J R and Goff T E 1985 African land-cover classification using satellite data. *Science* 227: 369–375
- [24] Ahamed T, Tian L, Zhang Y and Ting K C 2011 A review of remote sensing methods for biomass feedstock production. *Biomass Bioenergy* 35: 2455–2469
- [25] Morel J, Todoroff P, Bégué A, Bury A, Martiné J F and Petit M 2014 Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: a case study on Reunion Island. *Remote Sens.* 6: 6620–6635
- [26] Dangwal N, Patel N R, Kumari M and Saha S K 2016 Monitoring of water stress in wheat using multispectral indices derived from Landsat-TM. *Geocarto Int.* 31: 682–693
- [27] Gers C J 2003 Relating remotely sensed multi-temporal Landsat 7 ETM+ imagery to sugarcane characteristics. In: *Proceedings of the South African Sugar Technology Association*, pp. 1–7
- [28] Rao P V K, Rao V V and Venkataratnam L 2002 Remote sensing: a technology for assessment of sugarcane crop acreage and yield. *Sugar Tech.* 4: 97–101
- [29] Ke Y, Im J, Lee J, Gong H and Ryu Y 2015 Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. *Remote Sens. Environ.* 164: 298–313
- [30] Jia K, Wei X, Gu X, Yao Y, Xie X and Li B 2014 Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto Int.* 29: 941–951
- [31] Viovy N, Arino O and Belward A S 1992 The best index slope extraction BISE: a method for reducing noise in NDVI time-series. *Int. J. Remote Sens.* 13: 1585–1590
- [32] Lovell J L and Graetz R D 2001 Filtering pathfinder AVHRR land NDVI data for Australia. *Int. J. Remote Sens.* 22: 2649–2654
- [33] Roerink G J, Menenti M and Verhoef W 2000 Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *Int. J. Remote Sens.* 21: 1911–1917
- [34] Savitzky A and Golay M J E 1964 Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36: 1627–1639
- [35] Ma M and Veroustraete F 2006 Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Adv. Space Res.* 37: 835–840

- [36] Beck P S A, Atzberger C, Hogda K A, Johansen B and Skidmore A K 2006 Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Remote Sens. Environ.* 100: 321–334
- [37] Galford G L, Mustard J F, Melillo J, Gendrin A, Cerri C C and Cerri C E P 2008 Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens. Environ.* 112: 576–587
- [38] Zhang S, Lei Y, Wang L, Li H and Zhao H 2011 Crop classification using MODIS NDVI data denoised by wavelet: a case study in Hebei plain, China. *Chin. Geogr. Sci.* 21: 322–333
- [39] Julien Y and Sobrino J A 2010 Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sens. Environ.* 114: 618–625
- [40] Zhu W, Pan Y, He H, Wang L, Mou M and Liu J 2012 A changing-weight filter method for reconstructing a high-quality NDVI time series to preserve the integrity of vegetation phenology. *IEEE Trans. Geosci. Remote Sens.* 50: 1085–1094
- [41] Yang G, Shen H, Zhang L, He Z and Li X 2015 A moving weighted harmonic analysis method for reconstructing high-quality spot vegetation NDVI time-series data. *IEEE Trans. Geosci. Remote Sens.* 53: 6008–6021
- [42] Atkinson P M, Jeganathan C, Dash J and Atzberger C 2012 Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sens. Environ.* 123: 400–417
- [43] Chander G, Markham B L and Helder D L 2009 Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* 113: 893–903
- [44] Basso B, Cammarano D and Carfagna E 2013 Review of crop yield forecasting methods and early warnings. In: *Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics*, FAO Headquarters, Rome, Italy, pp. 1–56
- [45] Rouse J W, Haas R H, Schell J A and Deering D W 1974 Monitoring vegetation systems in the great plains with ERTS. In: *Technical Presentations, Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, vol. I, p. 309
- [46] Tukey J W 1977 *Exploratory data analysis*. Reading: Addison-Wesley
- [47] Jin Z and Xu B 2013 A novel compound smoother—RMMEH to reconstruct MODIS NDVI time series. *IEEE Geosci. Remote Sens. Lett.* 10: 942–946
- [48] Shabani A, Sepaskhah A R, Kamgar-Haghighi A A and Honar T 2018 Using double logistic equation to describe the growth of winter rapeseed. *J. Agric. Sci.* 156: 37–45