

Article

Automating the Derivation of Sugarcane Growth Stages from Earth Observation Time Series

Neha Joshi , Daniel M. Simms *  and Paul J. Burgess 

Faculty of Engineering and Applied Science, Cranfield University, Cranfield MK43 0AL, UK; n.joshi@cranfield.ac.uk (N.J.); p.burgess@cranfield.ac.uk (P.J.B.)

* Correspondence: d.m.simms@cranfield.ac.uk

Abstract: Sugarcane is a high-impact crop used in the majority of global sugar production, with India being the second largest global producer. Understanding the timing and length of sugarcane growth stages is critical to improving the sustainability of sugarcane management. Earth observation (EO) data have been shown to be sensitive to the variation in sugarcane growth, but questions remain as to how to reliably extract sugarcane phenology over wide areas so that this information can be used for effective management. This study develops an automated approach to derive sugarcane growth stages using EO data from Landsat-8 and Sentinel-2 satellite data in the Indian state of Andhra Pradesh. The developed method is then evaluated in the State of Telangana. Normalised difference vegetation index (NDVI) EO data from Landsat-8 and Sentinel-2 were pre-processed to filter out clouds and to harmonise sensor response. Pixel-based cloud filtering was selected over filtering by scene in order to increase the temporal frequency of observations. Harmonising data from two different sensors further increased temporal resolution to 3–6 days (70% of sampled fields). To automate seasonal decomposition, harmonised signals were resampled at 14 days, and low-frequency components, related to seasonal growth, were extracted using a fast Fourier transform. The start and end of each season were extracted from the time series using difference of Gaussian and were compared to assessments based on visual observation for both Unit 1 ($R^2 = 0.72$ – 0.84) and Unit 2 ($R^2 = 0.78$ – 0.82). A trapezoidal growth model was then used to derive crop growth stages from satellite-measured phenology for better crop management information. Automated assessments of the start and the end of mid-season growth stages were compared to visual observations in Unit 1 ($R^2 = 0.56$ – 0.72) and Unit 2 ($R^2 = 0.36$ – 0.79). Outliers were found to result from cloud cover that was not removed by the initial screening as well as multiple crops or harvesting dates within a single field. These results demonstrate that EO time series can be used to automatically determine the growth stages of sugarcane in India over large areas, without the need for prior knowledge of planting and harvest dates, as a tool for improving sustainable production.



Citation: Joshi, N.; Simms, D.M.; Burgess, P.J. Automating the Derivation of Sugarcane Growth Stages from Earth Observation Time Series. *Remote Sens.* **2024**, *16*, 4244. <https://doi.org/10.3390/rs16224244>

Academic Editors: Annarita D’Addabbo and Raffaella Matarrese

Received: 12 September 2024

Revised: 8 November 2024

Accepted: 11 November 2024

Published: 14 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sugarcane; phenology; growth stages; Earth Observation; sustainability; time series; sensor harmonisation; cloud filtering; seasonal decomposition; knee analysis; trapezoid growth model

1. Introduction

Sugarcane (*Saccharum officinarum*) is used in the production of 68% of global sugar and is also used for bioethanol [1]. It is a high-impact crop because of a long growing period and the large land, labour, water, and nutrient requirements needed for successful cultivation [2]. Global and national demands for sugarcane as food and bioenergy continue to increase, but the capacity to increase the area of sugarcane plantations is limited. Increased sugarcane yields are therefore dependent on increased production per unit area, with sustainable production being key to achieving this.

India is the second largest producer of sugarcane in the world [3], and the cultivation of sugarcane forms an important part of the Indian economy. National requirements for sugarcane are predicted to reach 600 million Mt by 2030, and global exports have risen

from 2.6% in 2010 to 6.3% in 2020 [4]. Sugarcane production also employs millions of farmers alongside transport and tertiary roles associated with the industry [5,6]. Sugarcane production in India is heterogeneous, with over 550 operational sugar mills across the country, with different ownership, management styles, and cane varieties that are exposed to a range of different environmental conditions. This is problematic, as the many independent and commercial farming communities do not have access to knowledge and physical tools like machinery, irrigation systems, and improved cultivars to enable sustainable sugarcane production [7]. As a result, the production of vast quantities of sugarcane may be unsustainable through widespread use of monocultures and inappropriate pesticide and fertilizer use [8], which negatively impact productivity and degrade soil and water resources to the point where short-term restoration becomes difficult [9]. Many businesses seek to be economically, environmentally, and socially sustainable, where profitable sugar production is combined with desirable employment and minimal negative environmental effects [3,6]. Hence, sugar mill managers are interested in improving their understanding of the source and management of their sugar.

Optimising agricultural practices such as fertilisation, irrigation, and pest and disease management requires information about the seasonal development of sugarcane [10], as resource requirements vary with sugarcane growth stage [5] and spatially between climatic zones. Accurate monitoring can identify biotic and abiotic stress in order to guide improved management aimed at reducing post-harvest losses and the cost of cultivation and increasing sucrose content at harvest [8]. However, sugarcane monitoring from field surveys at sample sites is costly, labour intensive, and time-consuming, especially as multiple surveys are required during the growing season [11]. These sample data will often not fully represent the heterogeneous nature of the sugarcane industry, and the resulting information quickly becomes outdated [12].

An alternative approach to field surveys is to combine Earth observation (EO) data from satellites with biophysical crop models [13]. Satellite sensors allow for large areas to be covered quickly and changes in biomass to be studied in near real-time [14,15]. They also capture information in non-visible wavelengths at a temporal frequency that would be too time consuming to collect through field visits, and they can image areas that are difficult to access on the ground [14,16].

Previous studies have used time series of readily available Landsat and Sentinel-2 optical imagery to derive information on sugarcane growth stages in Ethiopia, China, and India [12,17–19]. A major challenge when deriving sugarcane information from optical EO data is the low frequency of observations and gaps in the time series caused by clouds [19,20]. Using appropriate cloud filtering methods and parameters for temporal smoothing are important for optimising data retention, but parameterisation is subjective and left to the experience of the data analyst [21]. This is problematic for the development of algorithms, as these parameters vary spatially and temporally, affecting the interpretation of EO time series [22].

A second challenge is that existing methodologies require field-scale knowledge of the planting and harvest dates to isolate individual growing seasons from within a time series dataset before growth stages can be derived. Shape-fitting models have been used to monitor the growth and phenology of maize, soybean, wheat, barley, sorghum, rice, and cotton crops in the USA and China using MODIS vegetation indices time series data [23–26]. Shape-fitting model approaches work by averaging multiple-year growth cycles from EO time series data to estimate crop phenology. However, these methods require planting and harvesting dates derived from either field data or crop calendar information [23,25], or, where this information is not available, state-level [24] and national-level statistics [26] to isolate individual sugarcane growing seasons from long time series data and provide the field-scale growth information. These methods are not practical for Indian sugarcane, as sugarcane management is heterogeneous with phenology varying between different fields, which makes it difficult to collect ground-based phenology estimates as state-level and national-level statistics, and crop calendars cannot be used [12,18]. The averaging

of multiple-year growth cycles to generate shape models is also inappropriate for monitoring the growth of sugarcane crops, as the growth cycles and length of phenological stages for virgin and ratoon sugarcane differs. Differences in management can also affect the shape of the growth profiles. Studying these differences is important for the sustainability of sugarcane production and averaging multiple growth profiles would remove this information.

A third challenge is a lack of transferability between EO measurements of land surface phenology, based on temporal variation in plant vigour, and field survey-derived growth stages where field-scale phenological information is absent [27]. There are advantages for crop management if information derived from EO monitoring can be compared with growth stages familiar to producers using standard management practices in the field without needing to use survey-derived information.

This study seeks to overcome these challenges by developing an automated method to derive sugarcane growth stages from EO data by first splitting the temporal signal into individual seasons, followed by identifying features with the annual profiles that are representative of sugarcane phenology. The advantage of this approach is continuous field-level monitoring of sugarcane growth that can be used alongside bio-physical models and in combination with additional datasets, such as climate and yield measurements, as a tool for improving sustainable production. For example, mills could optimise the organisation of cane crushing operations at the field scale based on crop development before harvesting takes place.

2. Materials and Methods

The methodology starts with a description of the study areas and field data, the selection of satellite datasets, and pre-processing steps to increase the frequency of observations and filter clouds. Visual interpretation of sugarcane development stages was then used to validate an automated process for seasonal decomposition and deriving crop growth stages alongside a model of sugarcane crop growth. All image processing was coded in Python 3.7 using the Scipy scientific software stack [28] (Signal 1.5.2, Pandas 1.1.1 and Numpy 1.19.1 for image manipulation, functions for Gaussian filtering and fast Fourier transform). Map figures were created in ArcGIS Pro 2.7.0; other figures were made using Matplotlib 3.3.1 [29].

2.1. Study Areas and Field Data

Indian sugarcane factory locations in Chittoor District, Andhra Pradesh (Unit 1) and Zaheerabad in the Sangareddy district of Telangana (Unit 2) were used in this study (Figure 1). Both sites have a tropical climate with wet and dry seasons. Unit 1 data consisted of 337 sugarcane field polygons for the 2018 to 2019 season surveyed in 2018 with ground photography and observations of crop height and irrigation method (drip or flood irrigated). Unit 2 data consisted of 400 field polygons from the 2019 to 2020 season, surveyed in 2019 using the same protocol as Unit 1. Field polygons for both sites were digitized by walking the field perimeter with a GPS device. Field sizes ranged from 0.14 to 11.5 ha, with a median field size of 1.05 ha in Unit 1 and 0.63 ha in Unit 2. Sugarcane crops were a mixture of virgin cane (fields planted from cuttings) from the season surveyed and ratoon cane, which regrown after harvesting the previous season, planted between 2012 and 2018. Unit 1 fields were 29% virgin cane and 71% ratoon cane compared to 79% virgin cane and 21% ratoon cane in Unit 2. All sugarcane fields in Unit 1 and Unit 2 were irrigated, with 10% of the fields being drip irrigated in Unit 1 compared to 87% in Unit 2 and with the remainder being flood irrigated.

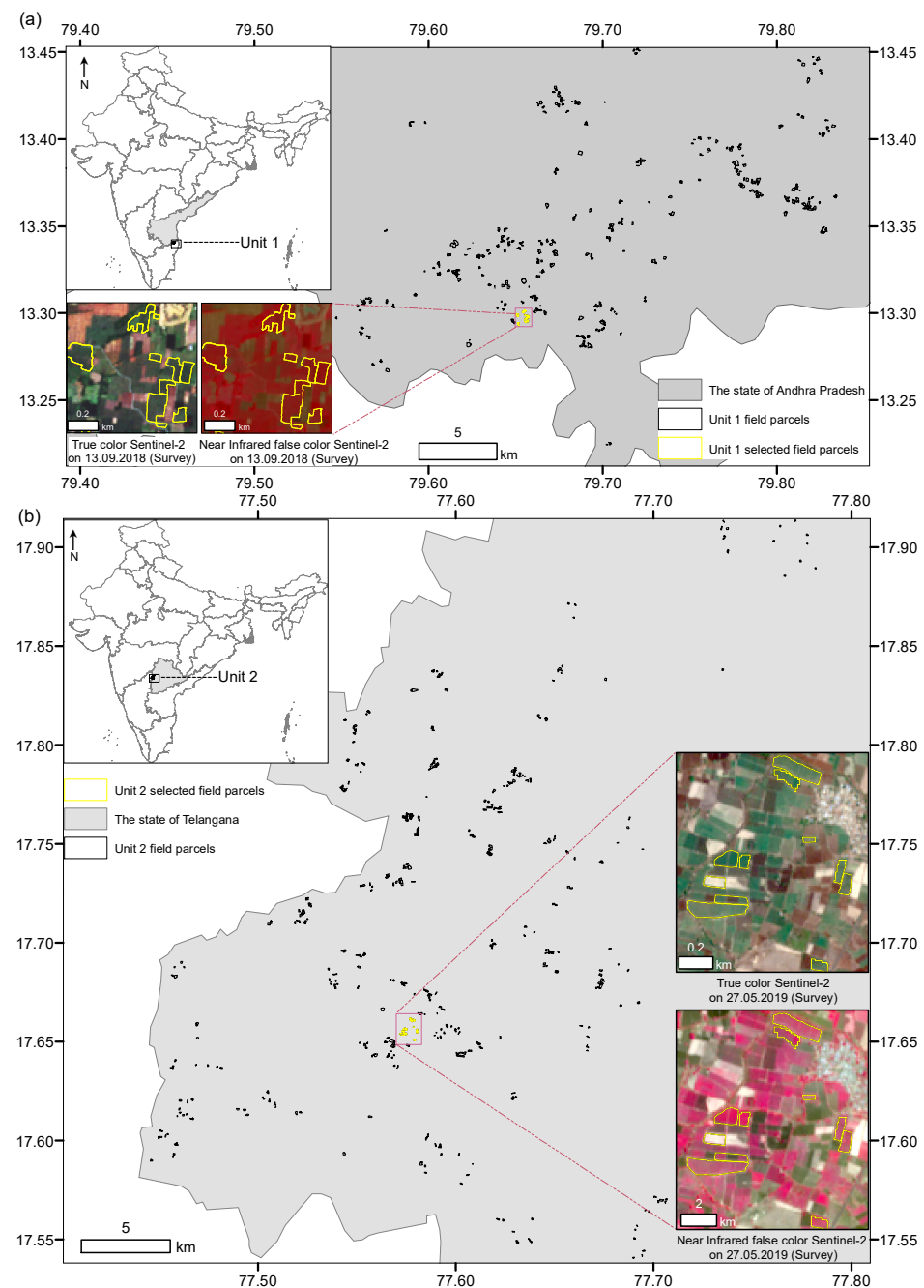


Figure 1. Sugarcane fields associated with processing Units 1 and 2 located in southern India. Sugarcane area Unit 1 is situated in (a) Andhra Pradesh, and (b) Unit 2 is in Telangana.

2.2. EO Datasets

A major consideration of this study was access to temporal EO datasets that can be used in the context of sustainable development. The study limited itself to data available on the Google Earth Engine (GEE) platform, which allows for the integration and manipulation of datasets, regardless of format and datatype, to carry out different tasks rapidly without the need for image data download, pre-processing, and computing resources [30]. Prior to cloud platforms like GEE, producing a time series dataset involved a time-intensive process of individually downloading and pre-processing images before building a time series of observations. The type of dataset available on the GEE platform varied in terms of the temporal availability and spatial and spectral resolution. The MODIS surface reflectance (SR), Landsat-8 Top of Atmosphere (L8 TOA), and Sentinel-2a and Sentinel-2b

Top of Atmosphere (S2 TOA) were identified as three EO datasets that provide reflectance information in the red (R) and near-infrared (NIR) wavelength bands for India, suitable for vegetation monitoring using the Normalised Difference Vegetation Index (Equation (1)) (Table 1). Despite the high revisit time of MODIS, the 250 m (6.25 ha) pixel resolution was deemed to be too coarse in comparison to the median field size across the two units (0.63–1.05 ha). L8 TOA with a resolution of 30 m (0.09 ha) and an S2 TOA resolution of 10 m (0.01 ha) are both smaller than the median field size, so all but the smallest fields will be made up of multiple pixels at these resolutions (Table 1), and they were considered as potential EO datasets for deriving sugarcane phenology.

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R}) \quad (1)$$

Table 1. Characteristics of three Earth observation datasets of red and near-infrared radiation (NIR) in South India. The information for MODIS is based on the Terra (EOS AM) and Aqua (EOS PM) satellites combined. Sentinel-2 data are comprised of data from two satellites: Sentinel 2a, which was launched in 2015, and Sentinel 2b, which was launched in 2017.

Sensor	Launch Year	Revisit Time	Resolution, m
MODIS	2002	1 to 2 days	250
Sentinel-2	2017	5 days	10
Landsat-8	2013	16 days	30

Sentinel and Landsat TOA reflectance was used to increase the number of cloud-free observations and span the full time series of field parcel data, as SR data was only available at the GEE platform for the 2019 season onward. This study recognises that the sensitivity of TOA data to changes in the atmosphere can limit their use, but previous studies have found that they can still successfully exploit the spectral differences for monitoring crop growth over time [12].

2.3. Selecting a Sugarcane Growth and Development Model

The development of sugarcane can be described as four stages [5,31,32] using a trapezoidal model to describe biomass change over time through changes in NDVI or crop evapotranspiration (Figure 2). From initial planting, the establishment stage of sugarcane typically takes 45 to 50 days until approximately 10% of the ground is covered with vegetation. This is followed by an extension stage during which branching from the primary shoot occurs, increasing the number of stalks and overall crop cover to 70–80% [5], typically lasting about 70–75 days [5,31]. The extension stage is followed by the mid-season stage, which can occur from 120 to 250 days after planting, typically lasting 130 to 220 days [5,31]. This is the stage in which the crop achieves its maximum canopy cover and accumulates the most biomass before the crop starts to senesce. The senescence stage can occur from 250 to 360 days after planting and typically lasts between 110 and 140 days. During this stage, fructose and glucose are converted to sucrose [5,31].

After harvest, a second *ratoon* crop can grow from the base of the harvested cane [33]. Sugarcane can grow up to eight ratoon crops from initial planting, but with every ratoon, there is a tendency for reduction in yield [34]. The timing and length of growth stages for ratoon crops also differ from the initially planted cane, with the overall period of growth being shorter for ratoon cane as the establishment of new roots is not needed [33]. In the case of ratoon cane grown in tropical climates like India, the durations of the establishment, extension, mid-season, and senescence periods are typically 30, 50, 180, and 60 days, respectively, with the total length of the season being less than one year at around 320 days [31].

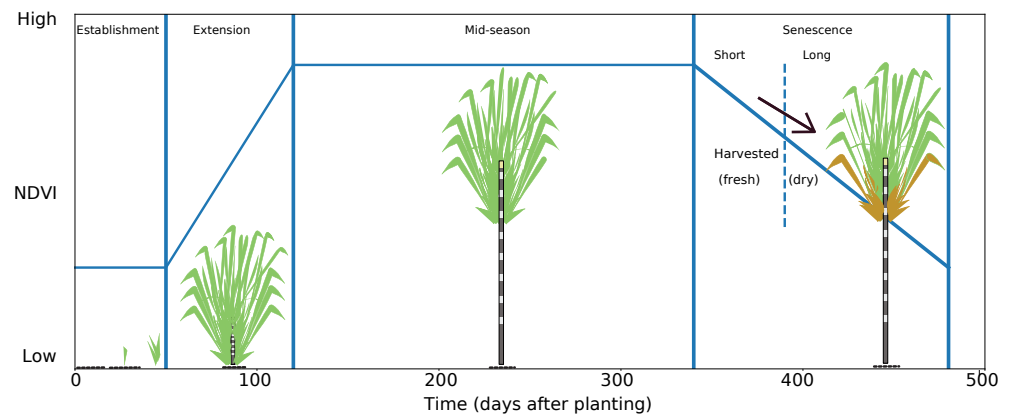


Figure 2. A schematic diagram depicting the four stages of sugarcane growth in India based on the FAO model for sugarcane evapotranspiration [31].

2.4. Establishment of Data-Processing Method

A method for image preprocessing and the extraction of individual sugarcane seasons was developed to automate the determination of sugarcane development stages from EO data (Figure 3), which is detailed in the following sections. A small number of fields with an area less than an L8 image pixel (900 m^2) were removed from the dataset, which reduced the number of fields studied from 337 to 326 for Unit 1 (the calibration dataset) and 400 to 356 for Unit 2 (the evaluation dataset). The mean NDVI L8 TOA and S2 TOA time series for each field was downloaded and processed using the GEE python API.

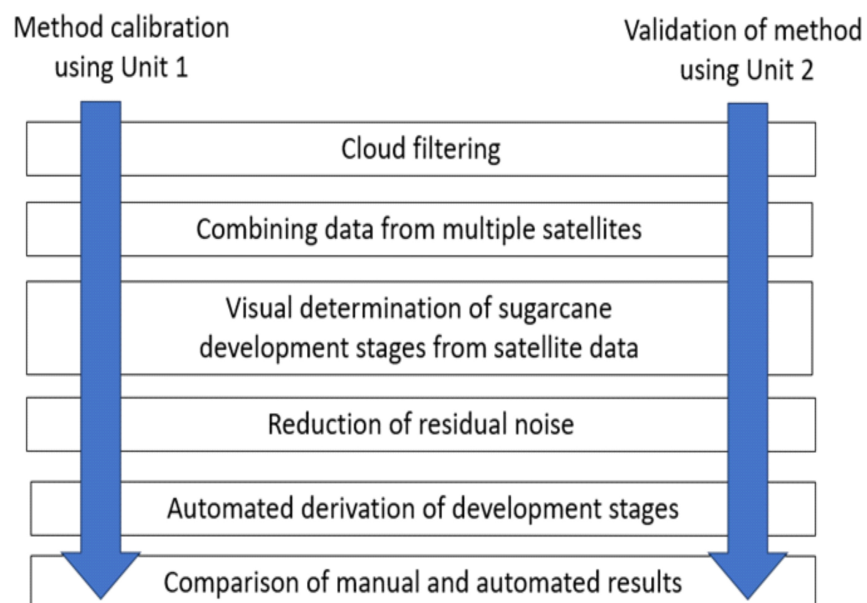


Figure 3. The methodology was established using data from Unit 1 and then evaluated using data from Unit 2.

2.5. Cloud Filtering

The presence of clouds in the EO images obscures surface reflectance and causes erroneous low estimates of the NDVI where the cloud is thin or occupies a fraction of an image pixel (Figure 4). To maximise data retrieval and minimise the effect of cloud and haze, we compared a percentage of cloud acceptance approach (20% or less cloud cover in each image tile) (Method 1) with a method that uses the cloud mask bands BQA and QA60 for removing pixels containing cloud from the L8 TOA and S2 TOA datasets, respectively

(Method 2). With the S2 TOA data, Method 1 was able to remove many of the low NDVI values (grey lines in Figure 4a), but some erroneous low values remained (drop in NDVI during the mid-season of the red time series in Figure 4a). In contrast, Method 2 was able to remove a greater number of erroneous low values, and hence, this method was used for the remaining analysis. Following the removal of the low NDVI values caused by cloud cover, the NDVI results generally showed a trapezoidal growth profile (see the blue time series profiles in Figure 4), following the trapezoidal model by [31] depicted in Figure 2.

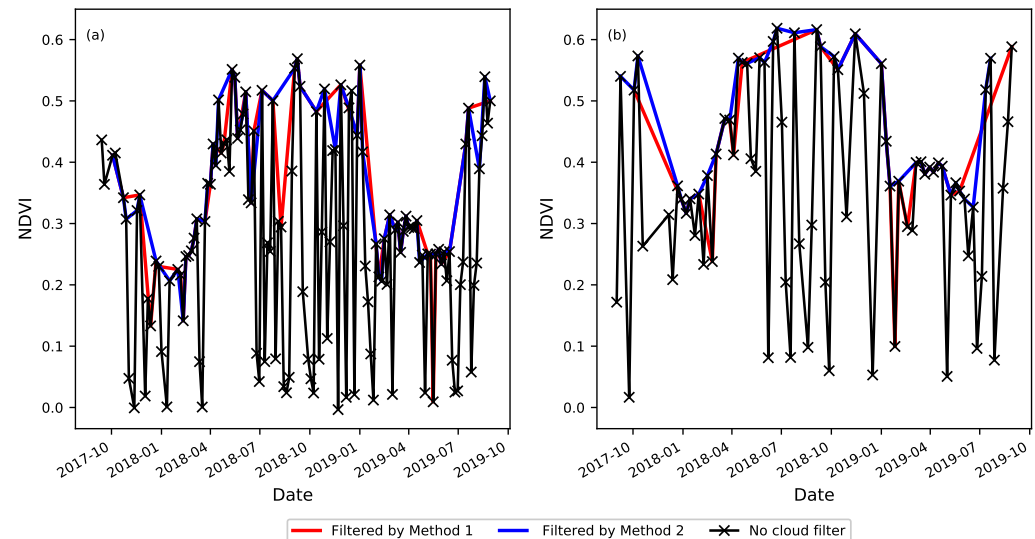


Figure 4. Before cloud filtering, cloud cover caused many low values of the Normalised Difference Vegetation Index (NDVI) (grey lines) from the Sentinel-2 (left in (a)) and Landsat-8 (right in (b)) data for sugarcane fields from September 2017 to October 2019. Using Method 1 to filter the clouds removed many of the low NDVI values (red lines), but more low values were removed using Method 2 (blue lines).

2.6. Combining and Harmonising Data from Multiple Sensors

A negative effect of cloud filtering is the reduced frequency of the remaining valid measurements. To address this and to minimise gaps in data at key growth stages, we developed an approach to combine observations from L8 TOA and S2 TOA data. By using both the S2 TOA and L8 TOA time series, over 70% of the 337 fields in Unit 1 returned 60–95 observations for the year 2018 (one observation every 3–6 days), compared to only 40–80 observations (one observation every 4–9 days) if only using S2 TOA or only 10–20 observations (one observation every 18–36 days) if only using L8 TOA (Figure 5).

One known issue with combining NDVI data from two satellites is the systemic measurement differences between sensors [35]. A correlation of paired NDVI measurements which were collected at the same location and on the same day from L8 TOA and S2 TOA showed a positive bias in the L8 data compared with S2 of 0.16 at low values of NDVI, decreasing as NDVI increases (Figure 6). These data were used in a linear line calibration to correct the L8 values to the S2 (Equation (2)), with a with a root-mean-square error (RMSE) of 0.09. This approach was used instead of an existing product, such as Harmonised Landsat and Sentinel-2 (HLS), as the GEE platform currently does not provide the Sentinel-2 component of HLS (HLSS30). We also tested the effect of applying local regression parameters between the study areas and found that calibration was site specific despite their close proximity (neighbouring states).

$$\text{NDVI}_{\text{L8}} = 0.78\text{NDVI}_{\text{S2}} + 0.16, \quad (2)$$

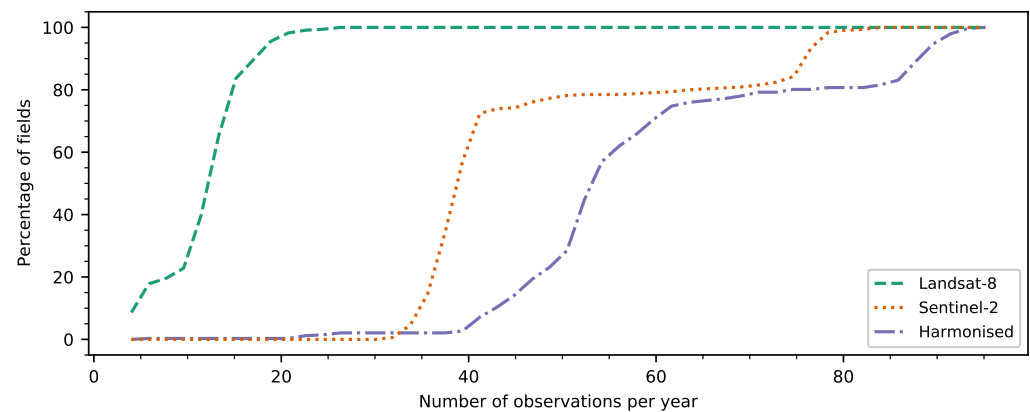


Figure 5. A frequency diagram presenting the average number of cloud-free observations received by each field, from three datasets, in one year (1 January 2018–1 January 2019). The number of observations received by the Landsat-8 (L8 TOA) dataset is presented by the green line, and the number of observations received by the Sentinel-2 dataset (S2 TOA) is presented by the orange line. The purple line represents the harmonised (L8/S2 TOA) dataset, which is the Landsat-8 and Sentinel-2 dataset combined.

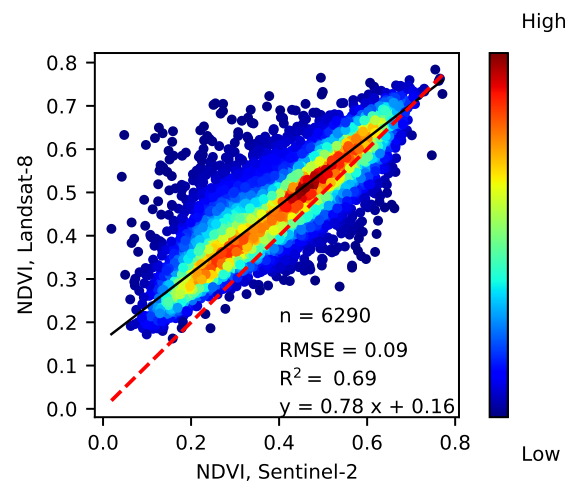


Figure 6. Correlation between the Landsat-8 (L8 TOA) NDVI observations and the Sentinel-2 (S2 TOA) observations for Unit 1. The solid black line shows ordinary least-squares (OLS) regression of the Landsat-8 data against the Sentinel-2 data. Only NDVI values in the range of 0.0 to 1.0 are illustrated. The hashed red line shows a reference 1:1 line, and the colour bar represents the density of observations.

2.7. Visual Determination of Sugarcane Phenology

Using the harmonised L8 and S2 datasets, the NDVI data of each individual field of the 337 fields in Unit 1 and the 456 fields in Unit 2 were visually inspected in order to determine a date for the start and end of the season (SOS and EOS, respectively) and the start and end of the mid-season stage of growth (SMS and EMS, respectively) (Figure 7) using the criteria described in (Table 2). It is important to acknowledge that there will be differences between physical planting dates and the beginning of the crop growth cycle derived from the satellite data because growth prior to the green-up (where NDVI starts to increase) will not be detected. Cover crops and irrigation events can also affect the interpretation of the profile early on [36]. Therefore, the SOS in this study refers to the first point of the sugarcane growing season detected from EO imagery (Figure 7).

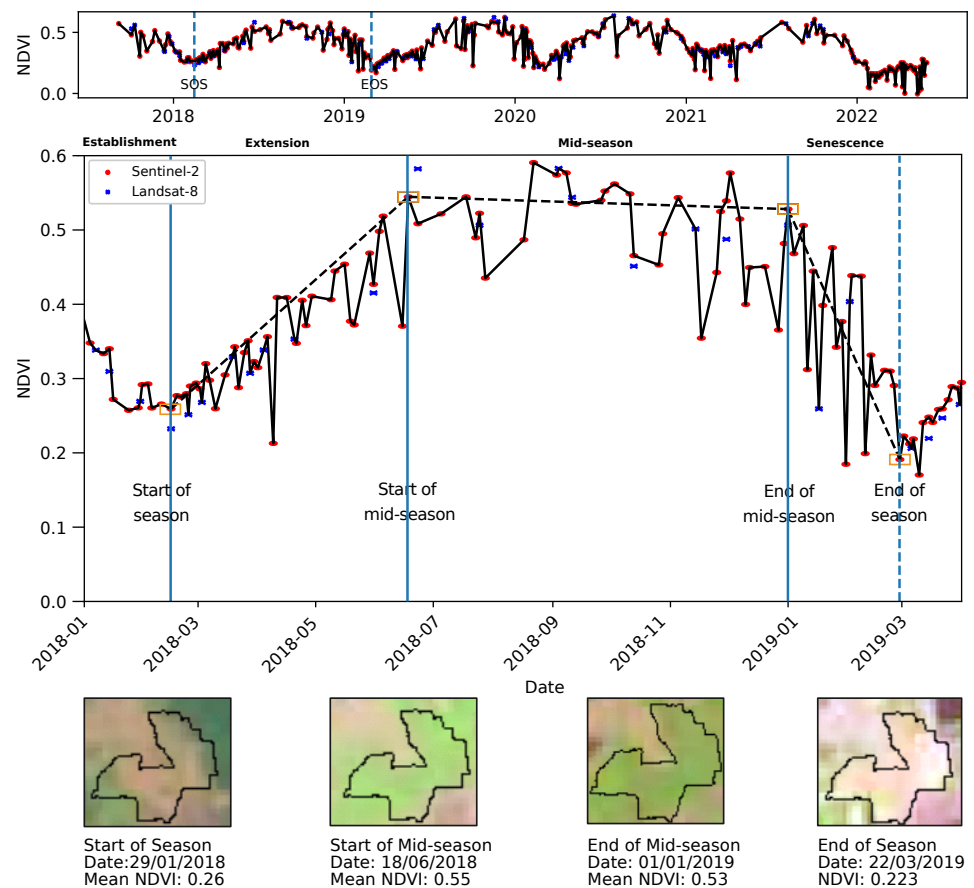


Figure 7. An illustration of how field reference data of sugarcane growth stages were manually interpreted using an example sugarcane field in Unit 1 for the 2018 season. The cloud-filtered NDVI time series data for the example field are displayed in grey, with Sentinel-2 observations marked in red and Landsat-8 observations marked in blue. An idealized trapezoid profile for the field marked with hashed black lines superimposed on top of the NDVI time series profile helped with the manual interpretations of growth stages. The start of the season was manually interpreted as the first point of the sugarcane growing season detected from EO imagery, and the end of the season was interpreted as the last low observation after the end of the mid-season before a rapid increase in NDVI (marking the start of a new growth cycle).

Table 2. Development stages and manual method for determination from the time series.

Development Stage	Manual Method of Determination
Start of season (SOS)	First value before rapid increase in NDVI
Start of mid-season (SMS)	Point at which the rate of rapid increase in NDVI is reduced, where the values of NDVI start to plateau
End of mid-season (EMS)	Last point of the plateau before a consistent decrease in NDVI
End of senescence (EOS)	Last low NDVI observation before an increase in NDVI (marking the start of a new growth cycle)

2.8. Splitting Time Series Data into Seasons

After cloud filtering and harmonisation of the sensor data, three additional steps were required before automating the derivation of the four development stages outlined above. Step one was to remove residual noise not removed during cloud filtering, step 2 was decomposing the time series into individual seasons, and step three was deriving the development stages for each sugarcane season.

Remaining noise in the profiles could be identified and excluded during manual determination of development stages. However, further processing was required to minimise residual noise within the profiles, as automated analysis is sensitive to outliers. We used a best index slope extraction (BISE) algorithm [37] to complete this. It had three assumptions attached to the NDVI values recorded within each sugarcane field: First, the presence of clouds will only decrease values; second, data transmission errors will cause abnormally high values; and third, sudden reductions in NDVI caused by cutting are likely to persist because vegetation re-growth is typically slow. In this analysis, observations were removed if the NDVI value was greater than 20% of the difference between the first low value and the previous high value in a pre-defined *sliding period*. If the selected sliding period, or window, is too short, then there is a reduced capacity to removed noise, whereas a sliding period that is too long can remove too much data. It was therefore important to select an appropriate sliding period for automation of sugarcane growth stages [38].

A sliding window for the BISE was chosen by examining which period would enable the removal of erroneous troughs, defined as a low point in the profile, within the sugarcane season whilst also minimising the RMSE between the predicted and the manually extracted start and end dates of the season for the calibration dataset (Unit 1). Values of sliding period ranging from 5 to 60 days were examined at increments of 5 days. The upper value of 60 days was based on the maximum interval between observations for the 337 fields from 1 January 2018 to 1 January 2019 (see Figure 5). Using the manually interpreted growth cycles for each field in Unit 1 for the 2018/19 (see Table 2 and Figure 7) growing season, the value for noise (N) was derived from the trough count (TC) minus 1 ((Equation (3)). This is because the trapezoid model [31] presents one trough at the end of the season, so any additional troughs would be the result of noise (See Table 2).

$$N = TC - 1 \quad (3)$$

The amount of noise removed (NR) was then calculated as the difference between the trough count in the sliding period (TCsp) and the trough count in the raw data (TCrd) ((Equation (4)).

$$NR = TCsp - TCrd \quad (4)$$

The RMSE between the SOS and EOS for the 2018 season was then calculated from a comparison of the manually derived SOS and EOS dates from the harmonised L8 and S2 TOA cloud-filtered dataset (from now on referred to as the raw data) and the BISE-filtered data at each sliding period. Increasing the sliding period from 5 to 40 days led to the removal of erroneous troughs (black line in Figure 8); there was minimal effect on increasing the window length to more than 40 days (Figure 8). However, a negative effect of increasing the sliding period was to increase the RMSE between the predicted and the manually extracted start and end dates of the season calibration dataset (Figure 8). A BISE sliding window period of 30 days was selected.

The SOS and EOS for each growing season, defined as a distinct seasonal cycle in the time series, were automatically extracted using the differences of Gaussians method (DoG) [39], which uses the difference between two smoothed versions of the time series to identify significant changes in magnitude corresponding to troughs. First, a fast Fourier transform (FFT) was used to simplify the time series into seasonal (low) frequency components. The FFT [40] transforms a time-varying signal $f(t)$ into its magnitude and phase components $F(u)$ in the frequency domain (Equation (5)),

$$F(u) = \sum_{t=0}^{N-1} f(t) e^{-i \frac{2\pi}{N} ut} / N \quad u = 0, \dots, n-1 \quad (5)$$

where u is the frequency component of the time-varying signal t , and N is the length of the time series. Removing higher-frequency components is straight forward, and the signal can be transformed back into the time domain to produce a smoothed signal. The frequencies

to retain were determined from the maximum number of cycles expected within a year (365 days) over the total timespan of the NDVI profile.

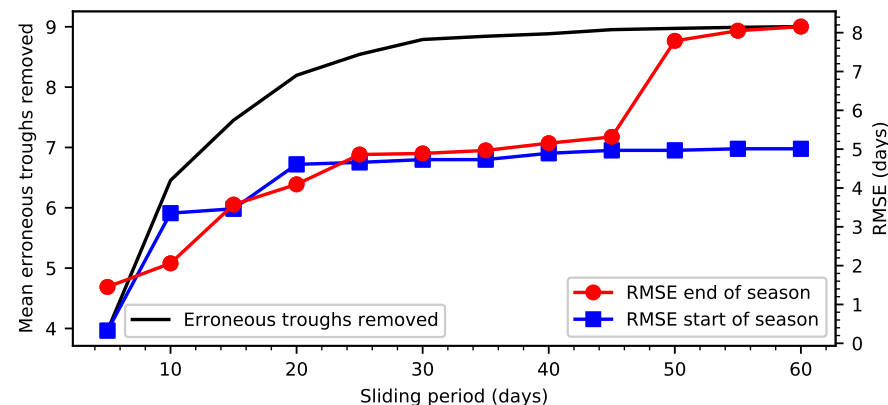


Figure 8. The effect of changing the best index slope extraction (BISE) sliding period on the number of additional erroneous troughs removed per field per growth cycle (left-hand scale) and the root mean square error (RMSE) (right-hand scale) in the calculation of the start and end of the sugarcane growth cycle.

The input signal was resampled, as the FFT requires an input signal at a known frequency with equally spaced intervals and no missing data. Where there were multiple good-quality (cloud-free with minimal noise) observations within the resampling period, the mean value of all good-quality observations was used to represent the observation of that period. Where there were no good-quality observations in the interval period, linear interpolation was used to fill the data gap, using adjacent observations. The impact of the resampling intervals was assessed against manual assessments of the start and end dates of the season. A 14-day interval was chosen as a compromise between excluding periods with missing data (requiring interpolation) and over smoothing of the signal. Peaks were then extracted from the DoG and used to identify the edges of seasonal growth cycles, corresponding to periods of low vegetation cover, by taking the lowest value from the original 14-day resampled signal within a 5-point buffer (Figure 9). The FFT (over) smoothed profile was retained for further processing to extract the start and end of the mid-season (see Section 2.9), then discarded.

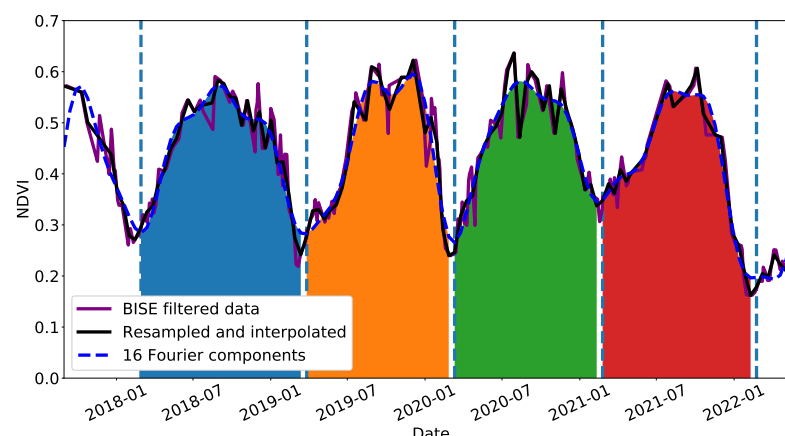


Figure 9. Time series NDVI decomposed into individual seasons using the low-frequency components of a fast Fourier transform for a sugarcane field with four ratoons.

To minimise the effect of the monsoon season, the algorithm for splitting the long time series data into growth cycles was programmed with two additional conditions. The first condition was that the latest date that the SOS could be derived was one month before

the monsoon season of the specific study site. The second condition was that the earliest date that the EOS could be derived was one month after the monsoon season of the specific study area. The derived SOS and EOS were compared with manually determined start and end dates for both units using regression analysis and calculation of the RMSE.

2.9. Fitting the Trapezoid Model of Crop Development

The trapezoid model of crop development was fitted to the decomposed seasonal profiles by identifying two points defining the plateau of the mid-season. The FFT-smoothed profiles for each field (λ) (see processing step in Section 2.8) were used to select the values of SMS and EMS from each seasonally decomposed 14-day resampled profile. First, the profiles for each field were split into two parts: λ_{p1} ranging from λ_1 to $\lambda_{max}+2$, where λ_{max} is the maximum NDVI, and λ_{p2} ranging from λ_{mid} to λ_n .

To calculate the SMS, λ_{p1} was first normalised (x and y values ranging from 0 to 1), λ_{p1norm} , and a knee algorithm, based on the work of Satopaa et al. [41], was used to locate the index of the value on λ_{p1norm} with the largest perpendicular Euclidean distance from the line extended from λ_{p1norm_1} to λ_{p1norm_n} (Figure 10a,b), which is the index of the point SMS on the 14-day resampled profiles (Figure 10e).

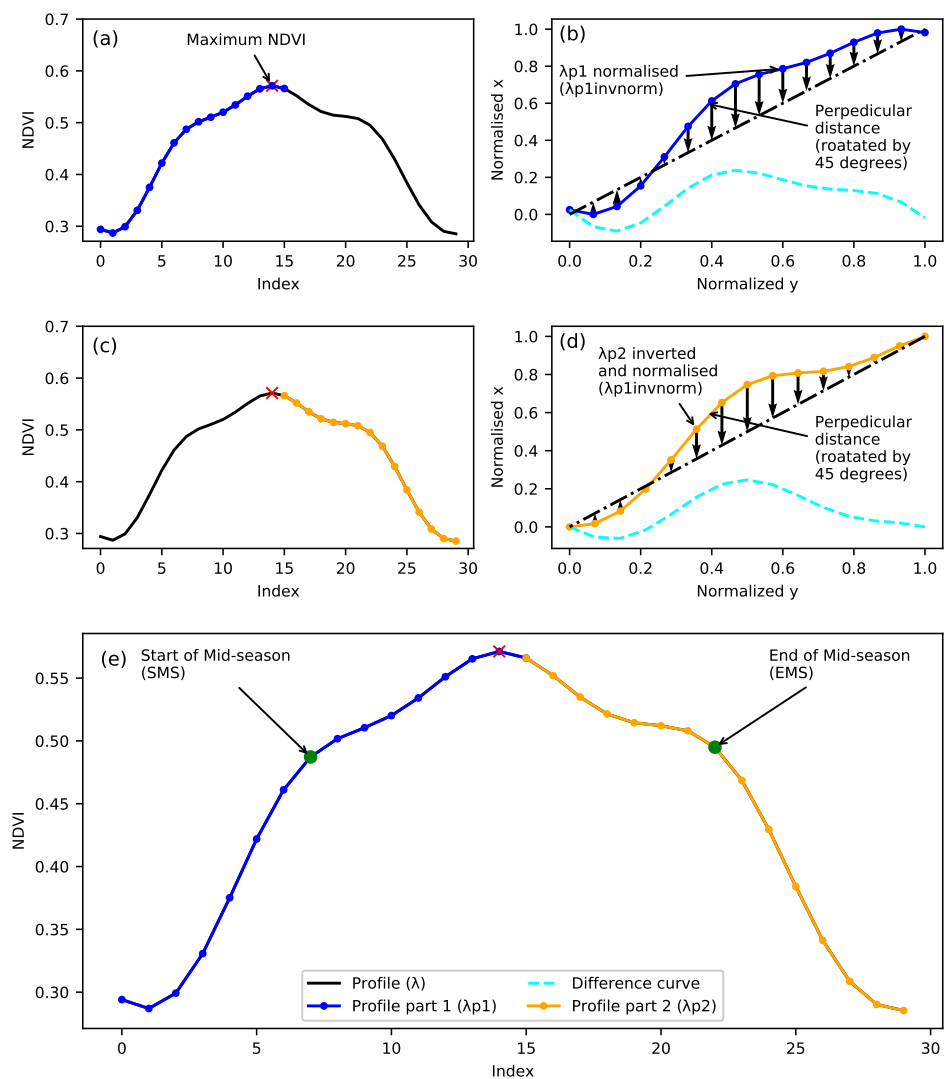


Figure 10. A schematic diagram illustrating how the start of the mid-season (SMS) and the end of the mid-season (EMS) were derived.

To calculate the EMS, the order of the values of λp_2 was reversed and then normalised $\lambda p_2 \text{invnorm}$. The same approach was then used to locate the index of the value of largest perpendicular Euclidean distance from the line $\lambda p_2 \text{invnorm}_1$ to $\lambda p_2 \text{invnorm}_n$ (Figure 10c,d), which is the reverse index of the second knee on the 14-day resampled profiles (Figure 10d).

3. Results

This section presents manual and algorithm-derived development stages for the calibration dataset from Unit 1 and then the validation dataset from Unit 2. To aid the comparison, 1 September was selected as the nominal start (day 1) of the year, before the start of planting for both growing areas.

3.1. Start and End of Season

Manual and automatically derived SOS ranged from 65 to 347 days (Figure 11). An initial comparison of the two methods showed an R^2 value of 66% and a RMSE of 26 days (Table 3). The relatively low R^2 and high range was related to three outliers with predicted SOS 150 days greater than the manual labelling (Figure 11). Subsequent analysis of those pixels showed that the overestimate was related to a period of cloud cover during the mid-season which was not removed during filtering. This resulted in a split of the sugarcane profile into two separate seasons, with the much later SOS for the second half of the split detected as the 2017 season. With the removal of these three points, the R^2 increased to 72%, and the RMSE was 27 days (Table 3).

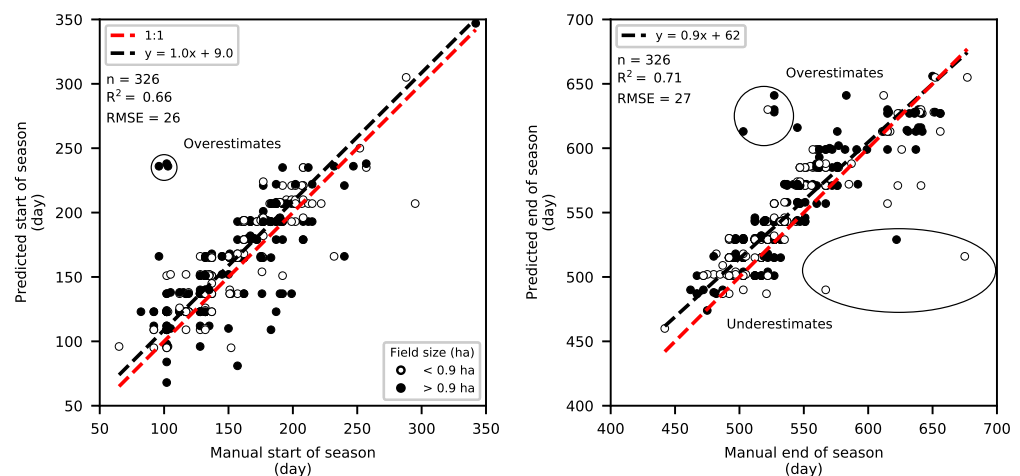


Figure 11. Unit 1 sugarcane fields used for algorithm development showing the relationship between automated and manually derived (left) start of the season and (right) end of season, using 14-day resampled time series, where day is the number of days after 1 September 2017. Equations show the linear relationship before the removal of outliers.

Table 3. Linear relationship between the automated (p) and manually derived (m) start of the sugarcane season (SOS) and the end of the season (EOS) for fields in Unit 1 before and (*) after the removal of outliers.

Relationship	Field Count	Equation	R^2	RMSE, Days
Start of season	326	$SOS_p = 1.00SOS_m + 9$	0.66	26
	323 *	$SOS_p = 0.99SOS_m + 9$	0.72	27
End of season	326	$EOS_p = 0.90EOS_m + 62$	0.71	27
	305 *	$EOS_p = 0.90EOS_m + 64$	0.84	21

Estimates of the EOS for Unit 1 ranged from 422 to 656 days with an R^2 value of 71% and an RMSE of 27 days (Table 3). Analysis of Figure 11 shows two groups of outliers. The underestimation of the EOS for three sites was again related to cloud cover over the field during the mid-season, which caused a premature split in the season. In contrast, substantial overestimates for five fields were related to mixed planting within the field parcel with multiple harvest dates. These were identified as errors in the digitisation of field boundaries, with multiple sugarcane fields grouped into the same polygon despite their differing growth stages and different harvest dates. With the removal of these outliers, the R^2 increased to 84%, and the RMSE decreased to 21 days (Table 3).

For the validation dataset from Unit 2, the SOS for the 359 sugarcane fields ranged from 94 to 332 days, with an R^2 value of 32% and an RMSE of 37 days (Table 4). There was a cluster of outliers showing overestimation in SOS Figure 12 caused by unfiltered cloud coverage, the effect also found in Unit 1. Other anomalies were related to the mixing of crops within field parcels, which resulted in an early prediction of the start of the season. After the removal of these data anomalies, the R^2 value increased to 78%, and the RMSE reduced to 23 days (Figure 12). The EOS estimates ranged from 422 to 722 days (Figure 12) with an R^2 value of 81% and an RMSE of 17 days, with no outliers.

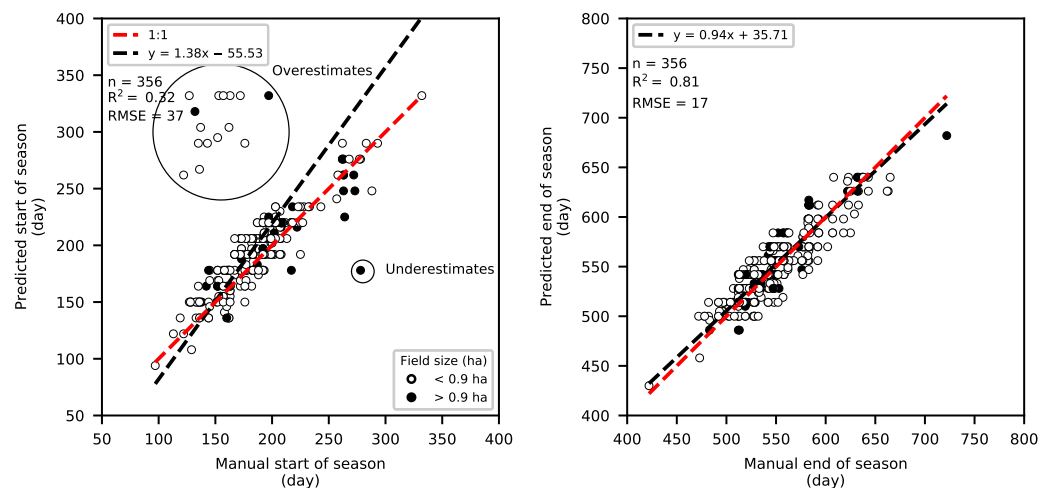


Figure 12. Unit 2 sugarcane fields used for validation showing the relationship between the automated and manually derived (left) start of the season and (right) end of season using 14-day resampled time series, where day is the number of days after 1 September 2017. Results are not affected by field size. Equations show the linear relationship before the removal of outliers.

Table 4. Linear relationship between the automated (p) and manually derived (m) start of the sugarcane season (SOS) and the end of the season (EOS) for fields in Unit 2 before and (*) after the removal of outliers.

Relationship	Number of Data		R^2	RMSE, Days
Start of season	356	$SOS_p = 1.38SOS_m - 56$	0.32	37
	344 *	$SOS_p = 1.08SOS_m + 18$	0.78	23
End of season	356	$EOS_p = 0.94EOS_m + 36$	0.81	17

3.2. Start and End of the Mid-Season

For the calibration dataset of Unit 1, the derived dates for the SMS ranged from 132 to 431 days after the nominal date of 1 September 2017 (Figure 13). The corresponding days for the EMS ranged from 334 to 599 days (Figure 13). Prior to the removal of outliers, the R^2 was 51% for the SMS and 27% for the EMS, and the RMSE for the SMS and EMS was 40 and 39 days, respectively.

Again, it was possible to identify outliers in the comparison of SMS because of the error in the seasonal decomposition caused by unfiltered cloud coverage (Figure 13). After the removal of these outliers, the automated analysis could explain 72% of the variation in the manually determined SMS dates. A similar knock-on effect can be seen with the prediction of EMS, with cloud cover during the mid-season resulting in an error in the EOS and, hence, an early end to the mid-season Figure 13. In the case of the three outlier fields where EMS was overestimated, incorrect field boundaries containing multiple crops masked the shape of the sugarcane curve, and no knee was detected. After the removal of these outliers, the automated analysis could explain 56% of the variation in EMS compared to manual determination, with an RMSE of 31 days (Table 5).

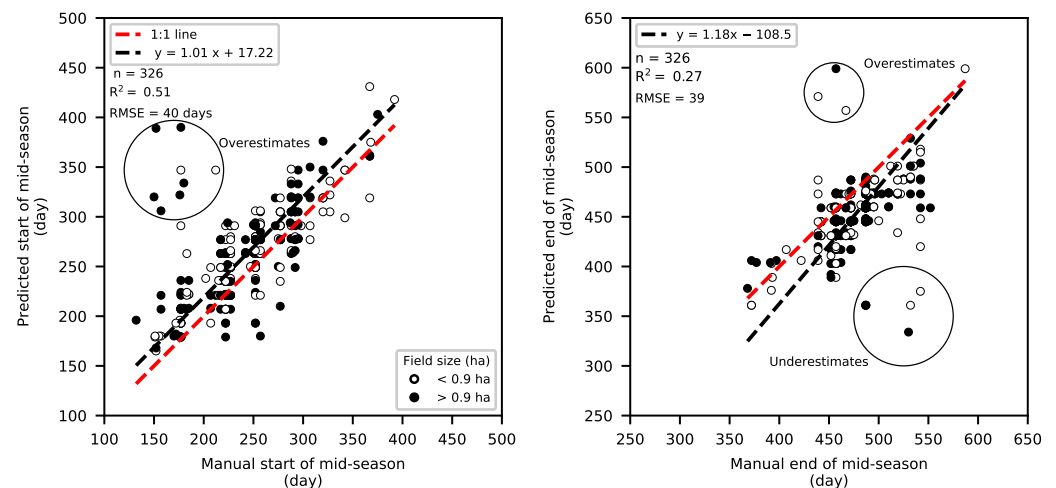


Figure 13. Sugarcane fields from Unit 1 used for calibration, showing the relationship between automated and manually derived (left) start of mid-season and (right) end of mid-season, using 14-day resampled time series, where day is the number of days after 1 September 2017. The equations show the linear relationship after the removal of outliers.

Table 5. Relationship between the automated (p) and manually derived (m) start of the mid-season (SMS) and end of the mid-season (EMS) for fields in Unit 1 before and (*) after the removal of outliers.

Relationship	Number of Data		R^2	RMSE, Days
Start of mid-season	326	$SMS_p = 1.01SMS_m + 17$	0.51	40
	318 *	$SMS_p = 0.99SMS_m + 17$	0.72	29
End of mid-season	326	$EMS_p = 1.18EMS_m - 109$	0.27	39
	316 *	$EMS_p = 0.90EMS_m + 23$	0.56	31

For Unit 2, the derived dates for the SMS ranged from 192 to 416 days after the nominal date of 1 September 2018 (Figure 14). The corresponding days for the end of the mid-season ranged from 361 to 612 days (Figure 14), with the initial analysis showing R^2 values for SMS and EMS of 65% and 28%, respectively, and an RMSE of 30 days (Table 6). Similar outliers to those observed on Unit 1 were identified in the Unit 2 validation set, which were caused by unfiltered cloud cover flattening out the end of the season profile or causing an error in the season decomposition. After the removal of outliers, the R^2 for the start and end of the mid-season increased to 79% and 36%, respectively, and the corresponding RMSEs decreased to 23 and 27 days.

Table 6. Relationship between the predicted (SMS_p) and manually derived (SMS_m) date of the start of the mid-season season and the predicted (EMS_p) and manually derived (SMS_m) date of the end of the mid-season for fields in Unit 2, including the effect of removing outliers.

Relationship	Number of Data		R^2	RMSE, Days
Start of mid-season	356	$SMS_p = 1.10SMS_m - 23$	0.65	30
	344 *	$SMS_p = 1.08SMS_m - 17$	0.79	23
End of mid-season	356	$EMS_p = 1.15EMS_m - 79$	0.28	30
	352 *	$EMS_p = 0.93EMS_m + 20$	0.36	27

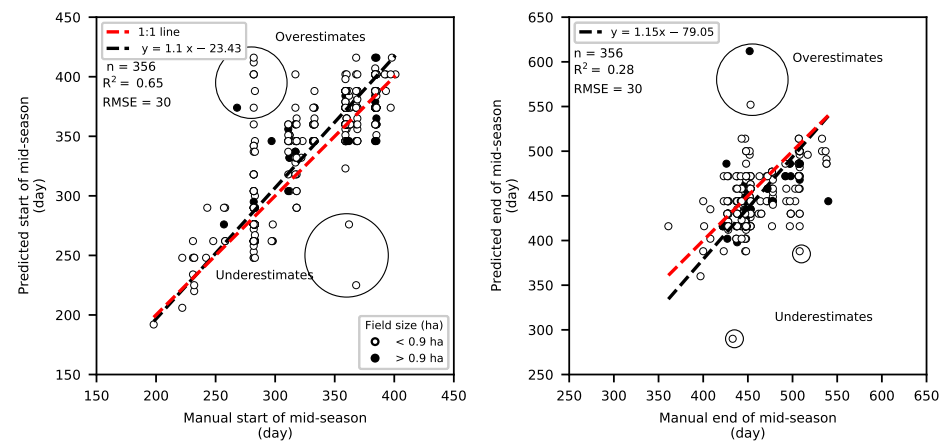


Figure 14. Sugarcane fields from Unit 2 used for validation, showing the relationship between automated and manually derived (left) start of mid-season and (right) end of mid-season, using 14-day resampled time series, where day is the number of days after 1 September 2018. The equations show the linear relationship after the removal of outliers.

3.3. Sugarcane Phenology

The timings of the mean automated growth stages, predicted by the algorithm, were compared to their respective mean manual estimates (through visual interpretation) and those reported by [5] for sugarcane grown in Uttar Pradesh, India and the FAO trapezoid model for sugarcane grown in tropical climates [31] (Table 7).

Table 7. Calculated duration of sugarcane growth stages establishment, extension, mid-season, and senescence for Unit 1 and Unit 2 in days derived from visual observation and automated calculation together with standard deviations, compared to estimates from Mall et al. [5] and Doorenbos [31]. Establishment was not separated from extension in the results of this study, as the start of the season was the point of green-up.

	Establishment	Extension	Mid-Season	Senescence	Total Growth Cycle
Unit 1 man _(V)	-	98 ± 33	216 ± 40	62 ± 46	376 ± 50
Unit 1 auto _(V)	-	114 ± 39	178 ± 45	93 ± 39	385 ± 46
Unit 1 man _(R)	-	86 ± 37	239 ± 50	62 ± 44	387 ± 49
Unit 1 auto _(R)	-	95 ± 33	195 ± 46	97 ± 39	386 ± 47
Unit 2 man _(V)	-	146 ± 47	124 ± 48	88 ± 41	357 ± 44
Unit 2 auto _(V)	-	142 ± 51	100 ± 53	104 ± 37	345 ± 46
Unit 2 man _(R)	-	156 ± 57	119 ± 53	91 ± 42	366 ± 57
Unit 2 auto _(R)	-	151 ± 66	102 ± 55	102 ± 32	355 ± 56
Mall et al. [5]	45	75	130	110	360
FAO _(V)	50	70	220	140	480
FAO _(R)	30	50	180	60	320

For Unit 1 sugarcane, the manual and automated mean length of the extension stage for the manually derived dataset were 89 and 99 days, respectively (Table 7). The mean lengths of the growing season from the two methods of 385–387 days were also similar. However, the automated method tended to predict a shorter mid-season and longer senescence stage than the manual observations (Table 7). The mean length of the extension period in Unit 2 of 144–148 days was greater than that observed in Unit 1. However, the length of the mid-season stage of only 100–123 days was substantially less than the value of 190–235 days in Unit 1. The overall length of the growth cycle in Unit 1 of 347–359 days was less than the value of 385–387 days in Unit 1. As a whole, there was less variation between the derived growth stages in Unit 2, with the senescence stage presenting the greatest difference between the manual and automated growth stages, 88 and 103 days, respectively, compared to variation across all growth stages for Unit 1.

We compared the mean manual estimates and automated growth stages to those reported by previous studies (Table 7). The length of the mid-season and senescence stages were similar to the values reported by Mall et al. [5] for Uttar Pradesh, India and Doorenbos [31] for the Tropics. The length of the reported manual and automated estimates of the extension stage for both Unit 1 and Unit 1 were much longer than those reported by Mall et al. [5] and Doorenbos [31]. This suggests that the point in the FAO model where establishment becomes extension is different in the time series data, with the SOS occurring at some point between the establishment and extension growth stages.

4. Discussion

Growth stage information is important for sustainable sugarcane management as it allows for the optimisation of resource use at different stages of the sugarcane growth cycle. The advantage of using EO data to explore biomass change over time across large areas is well documented, and there has been a notable increase in the number of studies using EO to explore sugarcane growth. However, the applications of previous studies to assess sustainable sugarcane management at the field scale are limited by three main factors: low frequency of observations after cloud removal, reliance on prior knowledge of planting and harvest times before the intermediate growth stages can be defined [12,17–19], and ambiguity and lack of transferability between EO-derived growth stages and field-derived growth stages important for management where field survey data are unavailable.

To overcome the above challenges, in this study, we investigated methods and optimised pre-processing parameters for cloud and noise removal. We also automated the splitting of long time series data into growing seasons to then automate the identification of sugarcane growth stages within two Indian study sites. The FAO model of sugarcane growth was chosen as a simplified model to relate the physical principles of remote sensing to field-level development stages important for management.

4.1. Increasing Observation Frequency in Cloudy Locations

In contrast to previous studies, this study investigated methods for the removal of cloud coverage and optimised pre-processing parameters specifically for the time series analysis of sugarcane fields in India. This was important to effectively remove cloud coverage whilst maximising the number of cloud-free observations per field per growing season.

We found that pixel-based cloud filtering appeared to be more effective at removing erroneously low values of NDVI recorded by S2 TOA and L8 TOA than the use of scene-wide filtering through a >20% cloud mask. The pixel-based cloud filtering method was able to remove more noise and resulted in fewer data gaps within the time series. Finding this sweet spot between the amount of noise removed and number of observations retained was important for the extraction of useful information by computer algorithm for the automated identification of growth stages, and after pixel-based cloud filtering, the trapezoid shape of sugarcane growth could be visualised from changes in NDVI over time (Figure 4).

The monsoon season in our study areas in India typically occurs during the three months from June to September. During this period, there is a high probability of cloud

coverage, and hence, the frequency of good-quality observations to monitor changes in NDVI is low. To address this, we combined L8 TOA and S2 TOA NDVI values by developing a local regression calibration, which was similar to the regression reported by others [35]. The harmonisation worked well, increasing the frequency of cloud-free observations from every 18–36 and 4–9 days, respectively, for L8 TOA and S2 TOA time series datasets, to every 3–6. However, gaps in the data still remained, with cloud cover persisting for long periods of time during the monsoon.

We therefore investigated the optimum BISE sliding window for the removal of additional noise during the monsoon season in Unit 1 and a resampling interval for the splitting of the long time series data into seasons. This was important, because the lack of definitive rules for pre-processing, i.e., determining an acceptable threshold for the BISE filter, makes methods difficult to replicate. The results showed that a sliding window period of 30 days and a resampling interval of 14 days gave the best balance of minimising erroneous troughs and the root mean standard error whilst retaining sufficient information. The BISE filter with a sliding period of 30 and a 14-day resampling interval was applied to the time series data for fields in both Unit 1 and Unit 2 for the removal of additional noise and, thereby, automation of growth stage identification. The above method worked well for removing cloud coverage in Unit 1, with clouds affecting the identification of growth stages in only a small number of fields. In contrast to the Unit 1 dataset, the major cause for outliers in the evaluation dataset (Unit 2) was cloud coverage. Resampling and smoothing where there was the continued presence of some NDVI data affected by clouds during the mid-season resulted in significant over- and underestimation of the start and end of the mid-season for Unit 2 (Figure 14). This suggests that there is still room for further improvement of the removal NDVI data on cloudy days. With the removal of such outliers, it was possible to develop an algorithm that could explain 78% of the variation in a manually determined estimate of the SOS evaluation dataset of 338 fields. Like the calibration dataset, the errors from the prediction of the SOS and EOS cascaded and affected the prediction of the SMS and EMS of the mid-season. Compared to the calculation of the SOS and EOS, the algorithm was less able to predict the manually determined values of SMS and EMS. A reason for this could be the tailored nature of the determined BISE sliding period, as resampling was conducted on Unit 1 while the results were simply applied to Unit 2. To improve the identification of sugarcane phenology from Unit 2, it may be necessary to tailor pre-processing parameters to specific study areas and growth seasons in future studies. The 2018–2019 growth season in Unit 1 may have contained fewer cloudy observations than the 2019–2020 growth season in Unit 2. Further investigation is also required to assess the impact of the SR and HLS datasets on the algorithm sensitivity.

It is also important to note that the impact of cloud coverage was more severe in small fields (less than 0.9 ha); therefore, increasing the spatial resolution of the satellite data used or supplementing missing NDVI observations with synthetic aperture radar (SAR) data from Sentinel-1 back scatter may help to overcome this challenge. Sentinel-1 data come from active instruments that have been used in the past to monitor vegetation growth over time without being affected by clouds [11,42–44]. However, it is important to note that Sentinel-1 time series data reflect structural changes in vegetation over time, while NDVI time series data are related to biomass. Future studies should consider the differences in the data captured by passive and active sensors, which could be harmonised to improve optically derived sugarcane growing stages.

4.2. Time Series for Management

Previous studies on determining the duration of sugarcane growth stages have been based on the pre-specification of planting and harvesting dates [12,17,19,45]. By contrast, this study demonstrates the capacity to determine the growth stages of sugarcane without knowing the planting and harvesting dates, and this is especially important for India, where planting and harvest dates are variable and cannot simply be assumed. Due to the

lack of a requirement for specifying planting and harvesting dates, the approach described here can be used over substantially wider areas.

Another difference between this study and previous studies is the use of the simplified FAO trapezoid model for deriving field-level growth stages. The novel method developed in this study, based on the FAO model, is advantageous, as it was able to provide more information about growth stages important for management than previously used simplified models like double logistic curves [45], as key stages are not removed through over-smoothing. The use of the FAO trapezoidal model of sugarcane growth stages was advantageous, as information regarding differences in the timing and length of growth stages between virgin cane, ratoons, and different management practices were not masked by averaging multiple-year growth cycles. This paper presents a novel method for deriving the stages of the FAO trapezoid model for sugarcane growth by applying a knee function originally designed for computer programmers to a biological system using the physical principles of remote sensing. For computer programmers, the knee analysis is based on the relationship between cost to increase tunable parameters and the corresponding performance benefit, with the knee being defined at the point of levelling off. This paper found that leveling-off effects (knees) are also present in EO data that capture information about biological systems, with NDVI stabilising at the SMS as the number of new tillers (new leaves) stabilises and the focus of growth moves to stem-elongation (an effect which is not captured by NDVI). This is important for sustainable sugarcane production because the mid-season stage is particularly sensitive to drought stress, whereas drought stress is less critical during the senescence stage [31]. It is important to note that the methodology developed in this study used a complete growth cycle of observations to derive the sugarcane growth stages, as our focus was on splitting long time series data into multiple seasons. The use of historic time series data to monitor phenology over large areas and time scales is useful for establishing baselines for management where yields and intervention practices are variable. Within-season monitoring is useful for precision agriculture, allowing for intervention strategies to be put in place whilst the sugarcane crop is still growing, for example, to reduce post-harvest losses. Future studies should therefore investigate whether the methodology developed in this study could be used for within-season crop phenology detection, i.e., whether an entire season of data is needed for deriving sugarcane phenology.

Many sugarcane factory owners want more clarity on when sugarcane is grown and harvested; hence, the described trapezoidal approach for describing sugarcane growth phases using EO data could have benefits for field level crop management and regional predictions of sugarcane yields. For example, it is important for sugarcane to be transported to the mills for crushing as soon as it is harvested in order to minimise post-harvest losses [46]. In addition to the timing of growth stages, monitoring the length of growth stages is also important for management. For example, the longer length of the mid-season in Unit 1 sugarcane fields than the standards stated by [5,31] could indicate that the fields were harvested after the sugarcane had started to senescence. This information is important for sustainable sugarcane production because sugarcane that is harvested after senescence or before the mid-season has lower yields resulting from reduced sucrose concentrations in the stalks [47], and is therefore not a sustainable use of resources.

The Indian state of Telangana also faced extreme drought between 2016 and 2018. The shorter mid-season in Unit 2 could indicate earlier harvests due to the onset of drought. Prematurely harvesting the cane during the mid-season could have been a way for farmers to provide the mills with some cane before the full effects of the drought had completely caused the crop to fail. Such observations highlight the benefits of combining EO measurements with agrometeorological data. There is also a difference between the proportion of drip and flood irrigated fields in Unit 1 and Unit 2; therefore, future work should also investigate whether the irrigation type has an impact on the derived length of growing seasons.

4.3. Improving the Automation for Deriving Growth Stages

Although ground surveys can provide detailed data, their collection can be costly and labour intensive, and hence, this type of data collection is rare for large areas [11]. The validity of ground surveys is also difficult to assess. The method used in this paper takes the mean NDVI within the field boundary; therefore, correct GPS field boundaries were important, as averaging the NDVI values of two opposing growth cycles will cause them to cancel each other out.

Mixed cropping before the planting of the sugarcane crop resulted in the underestimation of the SMS. The early identification of the end of the season due to cloud coverage resulted in the early prediction of the SMS for two fields. The late identification of the end of season resulting from the point of harvest coinciding with a cloud event in turn resulted in the overestimation of the end of mid-season for two fields.

Changing the boundary after harvest resulted in erroneous overestimates of the EOS for five sugarcane fields located in Unit 1. With the removal of such outliers, it was possible to develop an algorithm that could explain 73% and 86% of the variation in a manually determined estimate of the start and the end of a sugarcane growth cycle, respectively, for a calibration dataset of 337 individual sugarcane fields. To address these problems, there is a need to develop an approach which can separate the mixed fields, as errors cascade and affect how the other phenological stages are derived. This can be seen in the prediction of the start and end of the mid-season for Unit 1, which was less able to predict the start and the end of the mid-season stage for fields with boundary faults. To overcome this problem, it will be necessary for future studies to determine the growth stages for individual pixels within a field or group similar pixels before extracting time series data.

Time series-based approaches for classifying sugarcane have been found to be more effective than single-date mapping approaches, as they allow for land cover to be assessed continually over many different spatial and temporal resolutions [48]. Sugarcane is a dynamic crop, and features used for sugarcane detection, e.g., spectral reflectance, texture, shape/size, and height change with development stage and management practices such as irrigation. The inclusion of temporal information should also enable the clearer discrimination of sugarcane from other land cover types such as paddy rice and maize, which may have a similar signal on a single date. Sugarcane, for example, typically has a single annual cropping cycle, whereas paddy rice and maize can be sown and harvested two to three times within a single year [12,49,50]. Mapping the spatial and temporal distribution of sugarcane fields over large areas in order to identify the source of sugarcane supplied to mills is therefore the first step to using EO for sustainable sugarcane production. Future work should conduct an investigation into whether the growth stages derived in this study could also be used to aid classification in addition to management.

5. Conclusions

Estimates of EO growth stages can be improved by pre-processing time series data to maximise information retrieval from cloudy images and effectively fuse data from multiple sensors. We found a relationship between land-surface phenology and sugarcane growth models that demonstrates how tools for making management decisions within the sugarcane growing season can be informed in near-real time using open-source data and cloud-based tools. The results also demonstrate a fully automatic approach to separating the time series data of NDVI into seasons with variable start and end dates. Such an approach can be useful for managers of sugarcane factories in terms of understanding the location and timing of sugarcane harvests. This information can be used in conjunction with external weather datasets to aid sustainable decision making about planting, harvesting, irrigation, and fertiliser application, to minimise post-harvest losses, increase sucrose content within the plant, and increase the sustainability of cultivation.

In addition, the utility of the automatic decomposition of time series data into seasonal profiles is not limited to sugarcane crops and has much wider applications, such as crop classification and information on crop rotations.

Author Contributions: Conceptualization and funding acquisition, D.M.S.; methodology, software, formal analysis, and writing—original draft, N.J.; supervision and writing—review and editing, D.M.S. and P.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Environment Research Council (NERC) sponsored by Central England NERC Training Alliance (CENTA2) Doctoral Training Partnership [grant number NE/S007350/1].

Data Availability Statement: Data supporting this study such as the location of sugarcane fields are restricted due to commercial restrictions imposed by the industry partner. However, the derived growth stage information is available from Cranfield Online Research Data at <https://doi.org/10.57996/cran.ceres-2631>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. FAO. *World Food and Agriculture—Statistical Yearbook 2021*; FAO: Rome, Italy, 2021.
2. Hess, T.; Aldaya, M.; Fawell, J.; Franceschini, H.; Ober, E.; Schaub, R.; Schulze-Aurich, J. Understanding the impact of crop and food production on the water environment—Using sugar as a model. *J. Sci. Food Agric.* **2013**, *94*, 2–8. [[CrossRef](#)] [[PubMed](#)]
3. Inman-Bamber, N.; Bonnett, G.; Smith, D.; Thorburn, P. Sugarcane physiology: Integrating from cell to crop to advance sugarcane production. *Field Crop. Res.* **2005**, *92*, 115–117. [[CrossRef](#)]
4. Solomon, S. Sugarcane production and development of sugar industry in India. *Sugar Tech* **2016**, *18*, 588–602. [[CrossRef](#)]
5. Mall, R.K.; Sonkar, G.; Bhatt, D.; Sharma, N.K.; Baxla, A.K.; Singh, K.K. Managing impact of extreme weather events in sugarcane in different agro-climatic zones of Uttar Pradesh. *Mausam* **2016**, *67*, 233–250. [[CrossRef](#)]
6. Nandhini, T.K.D.; Padmavathy, V. A study on sugarcane production in India. *Int. J. Adv. Res. Bot.* **2017**, *3*, 13–17. [[CrossRef](#)]
7. Mnisi, M.S. The concept of sustainable sugarcane production: Global, African and South African perceptions. *Afr. J. Agric. Res.* **2012**, *7*, 4337–4343. [[CrossRef](#)]
8. Singh, P.; Tiwari, A.K. *Sustainable Sugarcane Production*; Apple Academic Press: New York, NY, USA, 2018. [[CrossRef](#)]
9. Gujja, B.; Loganandhan, N.; Vinod Goud, V.; Agarwal, M.; Dalai, S. *Sustainable Sugarcane Initiative—Improving Sugarcane Cultivation in India—Training Manual Developed by WWF India and ICRISAT*; ICRISAT: Patancheru, India, 2010.
10. Li, Y.R. Growth and development of sugarcane (*Saccharum* spp. Hybrid) and its relationship with environmental factors. In *Agro-Industrial Perspectives on Sugarcane Production under Environmental Stress*; Springer Nature: Singapore, 2022; pp. 1–11. [[CrossRef](#)]
11. Olfindo, N.T., Jr.; de la Cruz, R.M.; Borlongan, N.J.B.; Marciano, J.J.S., Jr.; Olalia, L.C. Sugarcane plantation mapping using dynamic time warping from multi-temporal Sentinel-1A radar images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-3–2020*, 519–524. [[CrossRef](#)]
12. Wang, J.; Xiao, X.; Liu, L.; Wu, X.; Qin, Y.; Steiner, J.L.; Dong, J. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sens. Environ.* **2020**, *247*, 111951. [[CrossRef](#)]
13. Kussul, N.; Lavreniuk, M.; Kolotii, A.; Skakun, S.; Rakoid, O.; Shumilo, L. A workflow for Sustainable Development Goals indicators assessment based on high-resolution satellite data. *Int. J. Digit. Earth* **2019**, *13*, 309–321. [[CrossRef](#)]
14. Fu, W.; Ma, J.; Chen, P.; Chen, F. Remote sensing satellites for digital Earth. In *Manual of Digital Earth*; Springer: Singapore, 2019; pp. 55–123. [[CrossRef](#)]
15. Sudmanns, M.; Tiede, D.; Lang, S.; Bergstedt, H.; Trost, G.; Augustin, H.; Baraldi, A.; Blaschke, T. Big Earth data: disruptive changes in Earth observation data management and analysis? *Int. J. Digit. Earth* **2019**, *13*, 832–850. [[CrossRef](#)]
16. Stubbings, P.; Peskett, J.; Rowe, F.; Arribas-Bel, D. A hierarchical urban forest index using street-level imagery and deep learning. *Remote Sens.* **2019**, *11*, 1395. [[CrossRef](#)]
17. Dimov, D.; Uhl, J.H.; Löw, F.; Seboka, G.N. Sugarcane yield estimation through remote sensing time series and phenology metrics. *Smart Agric. Technol.* **2022**, *2*, 100046. [[CrossRef](#)]
18. Zheng, Y.; Li, Z.; Pan, B.; Lin, S.; Dong, J.; Li, X.; Yuan, W. Development of a phenology-based method for identifying sugarcane plantation areas in china using high-resolution satellite datasets. *Remote Sens.* **2022**, *14*, 1274. [[CrossRef](#)]
19. Yeasin, M.; Haldar, D.; Kumar, S.; Paul, R.K.; Ghosh, S. Machine learning techniques for phenology assessment of sugarcane using conjunctive SAR and optical data. *Remote Sens.* **2022**, *14*, 3249. [[CrossRef](#)]
20. Meraner, A.; Ebel, P.; Zhu, X.X.; Schmitt, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 333–346. [[CrossRef](#)]
21. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [[CrossRef](#)]

22. Wilson, A.M.; Parmentier, B.; Jetz, W. Systematic land cover bias in Collection 5 MODIS cloud mask and derived products—A global overview. *Remote Sens. Environ.* **2014**, *141*, 149–154. [\[CrossRef\]](#)
23. Sakamoto, T.; Wardlaw, B.D.; Gitelson, A.A.; Verma, S.B.; Suyker, A.E.; Arkebauer, T.J. A Two-Step Filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sens. Environ.* **2010**, *114*, 2146–2159. [\[CrossRef\]](#)
24. Sakamoto, T. Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops—ScienceDirect. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 176–192. [\[CrossRef\]](#)
25. Zhou, M.; Ma, X.; Wang, K.; Cheng, T.; Tian, Y.; Wang, J.; Zhu, Y.; Hu, Y.; Niu, Q.; Gui, L.; et al. Detection of phenology using an improved shape model on time-series vegetation index in wheat. *Comput. Electron. Agric.* **2020**, *173*, 105398. [\[CrossRef\]](#)
26. Liu, L.; Cao, R.; Chen, J.; Shen, M.; Wang, S.; Zhou, J.; He, B. Detecting crop phenology from vegetation index time-series data by improved shape model fitting in each phenological stage. *Remote Sens. Environ.* **2022**, *277*, 113060. [\[CrossRef\]](#)
27. Kooistra, L.; Berger, K.; Brede, B.; Graf, L.V.; Aasen, H.; Roujean, J.L.; Machwitz, M.; Schlerf, M.; Atzberger, C.; Prikaziuk, E.; et al. Reviews and syntheses: Remotely sensed optical time series for monitoring vegetation productivity. *Biogeosciences* **2024**, *21*, 473–511. [\[CrossRef\]](#)
28. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [\[CrossRef\]](#)
30. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2016**, *202*, 18–27. [\[CrossRef\]](#)
31. Doorenbos, J. Crop water requirements. *FAO Irrig. Drain. Pap.* **1977**, *24*, 1–144.
32. Alface, A.B.; Pereira, S.B.; Filgueiras, R.; Cunha, F.F. Sugarcane spatial-temporal monitoring and crop coefficient estimation through NDVI. *Rev. Bras. Eng. Agrícola Ambient.* **2019**, *23*, 330–335. [\[CrossRef\]](#)
33. Xu, F.; Wang, Z.; Lu, G.; Zeng, R.; Que, Y. Sugarcane ratooning ability: Research status, shortcomings, and prospects. *Biology* **2021**, *10*, 1052. [\[CrossRef\]](#)
34. Singh, P.; Rai, R.K.; Suman, A.; Srivastava, T.K.; Singh, K.P.; Arya, N.; Yadav, R.L. Soil-root interface changes in sugarcane plant and ratoon crops under subtropical conditions: Implications for dry-matter accumulation. *Commun. Soil Sci. Plant Anal.* **2014**, *46*, 454–475. [\[CrossRef\]](#)
35. Zhang, H.K.; Roy, D.P.; Yan, L.; Li, Z.; Huang, H.; Vermote, E.; Skakun, S.; Roger, J.C. Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sens. Environ.* **2018**, *215*, 482–494. [\[CrossRef\]](#)
36. Zhou, Q.; Guan, K.; Wang, S.; Hipple, J.; Chen, Z. From satellite-based phenological metrics to crop planting dates: Deriving field-level planting dates for corn and soybean in the U.S. Midwest. *ISPRS J. Photogramm. Remote Sens.* **2024**, *216*, 259–273. [\[CrossRef\]](#)
37. Viovy, N.; Arino, O.; Belward, A.S. The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. *Int. J. Remote Sens.* **1992**, *13*, 1585–1590. [\[CrossRef\]](#)
38. Xu, X.; Conrad, C.; Doktor, D. Optimising phenological metrics extraction for different crop types in Germany using the moderate resolution imaging spectrometer (MODIS). *Remote Sens.* **2017**, *9*, 254. [\[CrossRef\]](#)
39. Zhou, J.; Jia, L.; Menenti, M.; Gorte, B. On the performance of remote sensing time series reconstruction methods—A spatial comparison. *Remote Sens. Environ.* **2016**, *187*, 367–384. [\[CrossRef\]](#)
40. Li, S.; Xu, L.; Jing, Y.; Yin, H.; Li, X.; Guan, X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102640. [\[CrossRef\]](#)
41. Satopaa, V.; Albrecht, J.; Irwin, D.; Raghavan, B. Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MI, USA, 20–24 June 2011. [\[CrossRef\]](#)
42. Vavlas, N.C.; Waine, T.W.; Meersmans, J.; Burgess, P.J.; Fontanelli, G.; Richter, G.M. Deriving Wheat Crop Productivity Indicators Using Sentinel-1 Time Series. *Remote Sens.* **2020**, *12*, 2385. [\[CrossRef\]](#)
43. Bao, X.; Zhang, R.; Lv, J.; Wu, R.; Zhang, H.; Chen, J.; Zhang, B.; Ouyang, X.; Liu, G. Vegetation descriptors from Sentinel-1 SAR data for crop growth monitoring. *ISPRS J. Photogramm. Remote Sens.* **2023**, *203*, 86–114. [\[CrossRef\]](#)
44. den Besten, N.; Steele Dunne, S.; Mahmud, A.; Jackson, D.; Aouizerats, B.; de Jeu, R.; Burger, R.; Houborg, R.; McGlinchey, M.; van der Zaag, P. Understanding Sentinel-1 backscatter response to sugarcane yield variability and waterlogging. *Remote Sens. Environ.* **2023**, *290*, 113555. [\[CrossRef\]](#)
45. Zhu, W.; He, B.; Xie, Z.; Zhao, C.; Zhuang, H.; Li, P. Reconstruction of vegetation index time series based on self-weighting function fitting from curve features. *Remote Sens.* **2022**, *14*, 2247. [\[CrossRef\]](#)
46. Misra, V.; Mall, A.; Solomon, S.; Ansari, M.I. Post-harvest biology and recent advances of storage technologies in sugarcane. *Biotechnol. Rep.* **2022**, *33*, e00705. [\[CrossRef\]](#)
47. Marin, F.R.; Rattalino Edreira, J.I.; Andrade, J.F.; Grassini, P. Sugarcane yield and yield components as affected by harvest time. *Sugar Tech* **2021**, *23*, 819–826. [\[CrossRef\]](#)
48. Morel, J.; Todoroff, P.; Bégué, A.; Bury, A.; Martiné, J.F.; Petit, M. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. *Remote Sens.* **2014**, *6*, 6620–6635. [\[CrossRef\]](#)

49. Birdi, P.K.; Kale, K. Identification of growth stage of sugarcane crop using decision tree for Landsat-8 data. In Proceedings of the 38th Asian Conference on Remote Sensing, New Delhi, India, 23–27 October 2017.
50. Cruz-Sanabria, H.; Sanchez, M.G.; Rivera-Caicedo, J.P.; Avila-George, H. Identification of phenological stages of sugarcane cultivation using Sentinel-2 images. In Proceedings of the 2020 9th International Conference On Software Process Improvement (CIMPS), Mazatlan, Sinaloa, Mexico, 21–23 October 2020. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



DSPACE

<https://dspace.org/>

Automating the derivation of sugarcane growth stages from Earth observation time series

Joshi, Neha; Simms, Daniel M.; Burgess, Paul J.

2024-11-14

MDPI

<https://doi.org/10.3390/rs16224244>

Joshi N, Simms DM, Burgess PJ. (2024) Automating the derivation of sugarcane growth stages from Earth observation time series. Remote Sensing, Volume 16, Issue 22, November 2024, Article number 4244