



Neha Joshi

EARTH OBSERVATION FOR
SUSTAINABLE SUGARCANE PRODUCTION

Faculty of Engineering and Applied Sciences

PhD Thesis

Academic Year: 2020-2024

Supervisors: Dr Daniel Simms and Professor Paul J. Burgess

November 2024



Faculty of Engineering and Applied Sciences
PhD Thesis

Academic Year 2020-2024

Neha Joshi

EARTH OBSERVATION FOR
SUSTAINABLE SUGARCANE PRODUCTION

Supervisors: Dr Daniel Simms and Professor Paul J. Burgess
November 2024

This thesis is submitted in fulfilment of the requirements for the
degree of PhD

© Cranfield University 2024. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

Academic integrity declaration

I declare that: the thesis submitted has been written by me alone.

- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

Sugarcane is a high impact crop and the source of most of the world's sugar, with India being the second largest producer and consumer of sugarcane in the world. The large resource requirements of sugarcane plantations, coupled with the demand for increased sugarcane production calls for the implementation of sustainable practices in sugarcane production over large areas quickly. Earth observation (EO) datasets and machine learning tools have been used in the past to monitor crop cultivation over large areas through spectral changes in surface reflectance that occur over time. Existing approaches that use EO for monitoring sugarcane cultivation over large areas are limited in scale by their reliance on crop calendars and phenological data derived from field surveys. Hence, this PhD aims to determine how new forms of EO imagery and machine learning tools can help to increase resource efficiency (land, labour, nutrients, water), in sugarcane production. The current state of the art for monitoring sugarcane production for improved sustainability was reviewed. Research gaps formed around using EO time-series data for monitoring sugarcane growth and management over time and the classification of sugarcane over large areas to overcome the limitations of field surveys. To overcome these limitations, new approaches were investigated to derive sugarcane growth stages using EO data from Landsat 8 and Sentinel-2 satellite data for sugarcane fields located in two

study sites, Unit 1 and Unit 2, situated in the India States of Andhra Pradesh and Telangana. Automated methods for decomposing time-series data into individual growing seasons were compared to visual observations which resulted in R^2 values of 0.72 and 0.84 for Unit 1 and R^2 values of 0.78 and 0.82 for Unit 2, when regressing the automated start and end of the season against the visually interpreted start and end of the season. Also, land surface phenology was related to agronomic practices using the FAO growth model resulting in R^2 values of 0.56 and 0.72 for Unit 1 and R^2 values of 0.36 and 0.79 for Unit 2, for the relationship between the automated and visually interpreted start and end of the mid-season growth stage. These results demonstrate that the EO time-series can 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. Synthetic aperture radar images were investigated to overcome challenges relating to unfiltered clouds and the frequency of observations. Growth information around tillering was present in the SAR data which was not present in the optical data and harvest could be more accurately determined compared to optical imagery. The results highlight the benefits of combining optical and SAR time-series information for monitoring sugarcane growth and management over time. Combining Sentinel-1 and NDVI time-series data allows managers of sugarcane factories to derive the timing of key sugarcane growing stages, harvests, and possible irrigation events. Seasonal decomposition was dependent on information regarding the year of the survey. To further overcome the dependency on survey data, an approach for automatic classification of sugarcane from 6-year long time-series data using Convolution Neural Networks was evaluated. The results include optimising parameters for the model, by analysing the effect of balancing classes in the training data, kernel size, and number of convolutional layers. The two best models for classifying sugarcane when trained using data from Unit 1, could generalise well and classify sugarcane in Unit 2. However, the performance of the two models improved when both were trained and tested using all data. These results demonstrated that a classification model, designed for the medical sector, could be successfully modified for agricultural applications. Individually the

algorithms derived in these studies provide decision support over large areas and automate information retrieval for sustainable resource management, yield, and biodiversity. Together the developed algorithms can be used as a strong tool to be implemented into existing sustainability frameworks for improving the sustainability of sugarcane production.

Keywords: Earth Observation; sustainability; sugarcane; sustainable development goals; resource optimisation; phenology; growth stages; time-series; sensor harmonisation; cloud filtering; seasonal decomposition; knee analysis; trapezoid growth model; classification; artificial intelligence; machine learning; convolutional neural networks; class imbalance; kernel size; generalisation;

Acknowledgements

Firstly, I would like to acknowledge and give a big thank you to my supervisors Dr Daniel Simms and Professor Paul Burgess for their continuous support and valuable guidance throughout my PhD Journey. Thank you both for supporting and guiding me with your experiences and constructive feedback, allowing me to improve and develop my academic research skills and do my best. Thank you both for your patience and making time for me even on weekends. Whenever I had a question or just needed to focus and organise my thoughts you were always there especially towards the end when everything felt like it was just too much. I also want to thank Dr Monica Rivas-Casado and Professor Peter Jarvis for their constructive feedback during my annual reviews which have been valuable.

I'd like to acknowledge to Natems for providing data used in this project and the Natural Environment Research Council (NERC) for funding this project [NE/S007350/1]. Thank you to the CENTA 2 team for organising valuable training events, even during the lockdowns, where many transferable skills were learnt, friendships and memories were made.

A special thank you to my parents, Dinesh and Neelam Joshi, without your unconditional love, support, prayers and blessings none of this would have been possible. Thank you for always believing in me even when I didn't believe in myself and for encouraging me to achieve highly and teaching me that that hard work always pays off. You are both my strength and the reason I am the person I am today. Thank you, Dad, for taking me to visit sugarcane plantations and factories in India after support from Natems was no longer available. Seeing your proud face as I explained my research to the factory owners in India is a memory I will forever hold on to. You both always tell me that you are proud of me no matter what, but I hope I've been able to make you both even prouder. Thank you, Mum and Dad, for never giving up on me.

A big thank you to both my sisters (Radhika and Anika Joshi), your impromptu karaoke sessions and dance performances kept me sane and provided me relief during the long days I spent sat at my computer. You always did your best to

cheer me up when I was sad even when I didn't ask. You are not just my sisters but my best friends as well, I am very lucky to have you both in my life.

I would like to express my gratitude to my mama (uncle), Dr Raj Sangar, if it weren't for his guidance and encouragement, I would have not had the confidence to apply for this PhD in the first place.

To the students in the downstairs PhD office of building 53, thank you for supporting me when I was on a rollercoaster of emotions (especially when trying to understand the bugs in my python code). Thank you for all the goofy laughs, chai, hot chocolate and coffee breaks over the years. I wish you all well and the very best for the future.

Finally, I want to thank God for providing me with this opportunity and giving me the strength and ability to undertake this research project to the best of my ability.

I would like to dedicate my thesis project to my Dadi Ma (grandmother), Raksha Sharma. I know that she is looking down at me from above with her beautiful smile and I hope that she is proud. She would have been extremely happy to see her granddaughters achieving highly.

Table of Contents

Academic integrity declaration.....	i
Abstract.....	i
Acknowledgements	iv
List of Figures.....	viii
List of Tables.....	xi
List of Abbreviations	xv
1 INTRODUCTION	16
1.1 Rationale	16
1.2 Aim and objectives	3
1.3 Structure.....	4
1.4 Overall methodological approach	6
2 LITERATURE REVIEW	16
2.1 Introduction	16
2.2 Survey data	16
2.3 Earth observation data	19
2.4 Monitoring and predicting yield at different growing stages.....	22
2.5 Deriving different growing stages for resource management	24
2.6 EO data for classifying sugarcane.....	27
2.7 Opportunities and Challenges	39
3 AUTOMATING THE DERIVATION OF SUGARCANE GROWTH STAGES FROM OPTICAL EARTH OBSERVATION TIME-SERIES.....	41
3.1 Introduction	41
3.2 Materials and Methods	44
3.3 Results	62
3.4 Discussion.....	71
3.5 Conclusions.....	78
4 DERIVING SUGARCANE GROWTH STAGES FROM SYNTHETIC APERTURE RADAR.....	79
4.1 Introduction	79
4.2 Materials and Methods	84
4.3 Results	93
4.4 Discussion.....	114
4.5 Conclusion	127
5 FINDING THE SWEET SPOT: AUTOMATING SUGARCANE DETECTION FROM LONG-TIME-SERIES DATA.....	128

5.1	Introduction	128
5.2	Materials and methods	130
5.3	Results	145
5.4	Discussion	156
5.5	Conclusions	163
6	SYNTHESIS	165
6.1	Introduction	165
6.2	Monitoring sugarcane production for improved sustainability	167
6.3	Deriving growth stages from Earth Observation time-series	170
6.4	Using machine learning to classify sugarcane	174
6.5	Practical Applications of monitoring sugarcane development from space: Challenges, Opportunities, and System Design	175
7	CONCLUSIONS AND FURTHER WORK	187
7.1	Conclusions	187
7.2	Recommendations for further work	189
8	REFERENCES	192
9	APPENDICES	203
9.1	CURES form	203
9.2	Sugarcane NDVI and VH-backscatter time-series profiles, with available photographs corresponding to the Start of the mid-season	204
9.3	Sugarcane NDVI and VH-backscatter time-series profiles, with photographs corresponding to the VH-backscatter derived initial knee	206
9.4	Data Sources	213
9.5	Thesis code description	214

List of Figures

Figure 1.1. Thesis overview: describing the role of each chapter for fulfilling each objective.....	5
Figure 1.2. Sugarcane fields surveyed and managed by Natems.	9
Figure 1.3. Example field polygon visualised using high resolution satellite imagery and field photography.	10
Figure 1.4 The set of guidelines used to verify the GPS data provided by Natems.	11
Figure 2.1. Stages of the Indian sugarcane growth cycle and their optimal resource requirements.....	19
Figure 2.2. Damage to a sugarcane stem, resulting from a shoot Boorer infestation	26
Figure 2.3. A sugarcane field affected by red rot disease	29
Figure 3.1. Sugarcane fields associated with processing Units 1 and 2 located in southern India.	46
Figure 3.2. A schematic diagram depicting the four stages of sugarcane growth in India based on the FAO model for sugarcane evapotranspiration	49
Figure 3.3. The methodology was established using data from Unit 1 and then evaluated using data from Unit 2.	51
Figure 3.4. The results of Methods 1 and 2 for cloud filtering	52
Figure 3.5. A frequency diagram presenting the average number of cloud-free observations received by each field, from three datasets, in one year.	53
Figure 3.6. Relationship between the Landsat 8 TOA NDVI observations and the Sentinel-2 TOA observations for Unit 1.	54
Figure 3.7. An illustration of how field reference data of sugarcane growth stages were manually interpreted.....	55
Figure 3.8. The effect of changing the best index slope extraction (BISE) sliding period on the calculated start and end of the sugarcane growth cycle.	58
Figure 3.9. Time-series NDVI decomposed into individual seasons using the low-frequency components of a fast Fourier transform for a sugarcane field with four ratoons.	60
Figure 3.10. A schematic diagram illustrating how the start of the mid-season and the end of the mid-season were derived.....	61
Figure 3.11. Unit 1 sugarcane fields used for algorithm development showing the relationship between automated and manually derived start and end of season, using 14-day resampled time-series data.	63
Figure 3.12. Unit 2 sugarcane fields used for validation showing the relationship between the automated and manually derived start and end of season using 14-day resampled time-series.....	64

Figure 3.13. Sugarcane fields from Unit 1 used for calibration, showing the relationship between automated and manually derived start and end of mid-season, using 14-day resampled time-series.	66
Figure 3.14. Sugarcane fields from Unit 2 used for validation, showing the relationship between automated and manually derived start and end of mid-season, using 14-day resampled time-series.	67
Figure 4.1 A schematic diagram of the sugarcane growth stages that can be automatically derived from processed NDVI time-series using methods developed in Chapter 3.	80
Figure 4.2. The location of sample sugarcane fields in the Telangana and Andhra Pradesh states of India.	85
Figure 4.3. An example of the respective mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 for a field in Unit 1 for five seasons (a,b,c), and mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 of the same field for the 2018 2019 sugarcane growing season only (d,e,f).	87
Figure 4.4. An example of the respective mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 for a field in Unit 2 for five seasons (a,b,c), and mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 of the same field for the 2019-2020 sugarcane growing season only (d,e,f).	88
Figure 4.5. The methodology was established using data from Unit 1 and then evaluated using data from Unit 2.	89
Figure 4.6. Relationship between the automated (p) and manually derived (m) start of the season for fields in Unit 1 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults	94
Figure 4.7. Relationship between the automated (p) and manually derived (m) end of the season (EOS) for fields in Unit 1 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.	95
Figure 4.8. The relationship between the Sentinel-1 back scatter derived start of the season (SOS) and SOS manually interpreted from raw NDVI time-series profiles for fields located in Unit1 (a-c). The relationship between the SOS derived by Sentinel-1 back scatter and the SOS automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (d-f). ...	98
Figure 4.9. The relationship between the Sentinel-1 back scatter derived end of the season (EOS) and EOS manually interpreted from raw NDVI time-series profiles (a-c) for fields located in Unit 1. The relationship between the EOS derived by Sentinel-1 back scatter and the EOS automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (d-f). .	100
Figure 4.10. The relationship between the Sentinel-1 back scatter derived second knees and the start of mid-season manually interpreted from raw NDVI time-series profiles (a-c) for fields located in Unit 1. The relationship between the start of mid-season derived by Sentinel-1 back scatter and the	

initial knee automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (d-f).	102
Figure 4.11. The relationship between the Sentinel-1 back scatter derived second knees and the end of mid-season manually interpreted from raw NDVI time-series profiles (a-c) for fields located in Unit 1. The relationship between the end of mid-season derived by Sentinel-1 back scatter and the initial knee automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (d-f).	104
Figure 4.12. Relationship between the automated (p) and manually derived (m) start of the season for fields in Unit 2 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.	105
Figure 4.13. Relationship between the automated (p) and manually derived (m) end of the season for fields in Unit 2 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.	106
Figure 4.14. The relationship between the Sentinel-1 back scatter derived start of the season (SOS) and SOS manually interpreted from raw NDVI time-series profiles for fields located in Unit2 (a-c). The relationship between the SOS derived by Sentinel-1 back scatter and the SOS automatically derived by processed NDVI time-series profiles for fields located in Unit 2 (d-f). .	109
Figure 4.15. The relationship between the Sentinel-1 back scatter derived end of the season (EOS) and EOS manually interpreted from raw NDVI time-series profiles for fields located in Unit2 (a-c). The relationship between the EOS derived by Sentinel-1 back scatter and the EOS automatically derived by processed NDVI time-series profiles for fields located in Unit 2(d-f). ...	111
Figure 4.16. The relationship between the Sentinel-1 back scatter derived initial knees and the start of mid-season manually interpreted from raw NDVI time-series profiles (a-c) for fields located in Unit 2. The relationship between the start of mid-season derived by Sentinel-1 back scatter and the initial knee automatically derived by processed NDVI time-series profiles for fields located in Unit 2 (d-f).	112
Figure 4.17. The relationship between the Sentinel-1 back scatter derived second knees and the end of mid-season manually interpreted from raw NDVI time-series profiles (a-c) for fields located in Unit 2. The relationship between the end of mid-season derived by Sentinel-1 back scatter and the initial knee automatically derived by processed NDVI time-series profiles for fields located in Unit 2 (d-f).	114
Figure 4.18. Mean VH from processed Sentinel-1 Ground Range Detected time-series profiles of an example sugarcane field during the 2018 season, together with pre-processed mean NDVI time-series profiles.....	118
Figure 4.19 Mean VH from processed Sentinel-1 Ground Range Detected time-series profiles of an example sugarcane field during the 2018 season, together with pre-processed mean NDVI time-series profiles.....	120

Figure 4.20. A schematic diagram demonstrating the different aspects of sugarcane growth and development derived from the processed NDVI and VH dB time-series data.....	121
Figure 5.1. A schematic diagram demonstrating the process of convolution in 1-dimensional time-series data and red, green, and blue (RGB) images. ..	131
Figure 5.2. Sugarcane and other crop fields associated with processing Unit 1 located in the southern Indian state of Andhra Pradesh.	135
Figure 5.3. Sugarcane and other crop fields associated with processing Unit 2 located in the southern Indian state of Telangana.	136
Figure 5.4. The architecture of the initial convolutional neural network model for sugarcane detection in long-time-series data.	139
Figure 5.5. The set of guidelines used to manually label the training data as either belonging to the no-data, sugarcane or other crop class.	140
Figure 5.6. A schematic diagram demonstrating how recall, precision and F1 scores were calculated for sugarcane and other crops using the confusion matrixes for each crop.	142
Figure 5.7. The architecture of Model 1 for sugarcane detection in long-time-series data.	152
Figure 5.8. The architecture of Model 2 for sugarcane growth detection in long-time-series data.	153
Figure 6.1 Thesis outputs and their direct and indirect links to eleven components of the Target 2.4 of the UN Sustainable Development Goals related to society, economy, and the environment.	170
Figure 6.2. The sugarcane supply chain, variables, stakeholders and goals at each stage.	177
Figure 6.3. Applications of remote sensing at different stages of the sugarcane supply chain and it's benefits.....	180

List of Tables

Table 1.1. Number of sugarcane fields before and after verification	12
Table 2.1. List of themes and sub-indicators associated with the SDG indicator 2.4.1 (adapted from FAO,2019).....	17
Table 2.2. Four indicators in the Bonsucro Production standard that could be related to sugarcane growth stages and/or Earth observation data.....	21
Table 2.3. Four environmental and two social sustainability indicators in the Bonsucro Production standard that could be related to Earth observation data (Bonsucro, 2023)	28
Table 3.1. Characteristics of three Earth observation datasets of red and near-infrared radiation in South India.	48

Table 3.2. Development stages and manual method for determination from the time-series	56
Table 3.3. Linear relationship between the automated (p) and manually derived (m) start of the sugarcane season and the end of the season for fields in Unit 1 before and (*) after the removal of outliers.	63
Table 3.4. Linear relationship between the automated (p) and manually derived (m) start of the sugarcane season and the end of the season for fields in Unit 2 before and (*) after the removal of outliers.	65
Table 3.5. Relationship between the automated (p) and manually derived (m) start of the mid-season and end of the mid-season for fields in Unit 1 before and (*) after the removal of outliers.	66
Table 3.6. Relationship between the predicted (SMSp) and manually derived (SMSm) date of the start of the mid-season and the predicted (EMSp) and manually derived (SMSm) date of the end of the mid-season for fields in Unit 2, including the effect of removing outliers.	67
Table 3.7. 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. (2016) and Doorenbos (1977).....	70
Table 4.1. Calculated duration (days) of five sugarcane growth stages for virgin (v) and ratoon (R) crops Unit 1 and Unit 2 in days, together with standard deviations. Values from Molijin et al (2019), Mall et al. (2016) and Doorenbos (1977) are also presented.	122
Table 5.1. The number of profiles in the training and test profiles for sugarcane and other crops with an imbalanced and balanced weight.	144
Table 5.2. A confusion matrix depicting the effect of balancing the training data on the classification of sugarcane and other crops.....	146
Table 5.3. Recall, precision, and F1 scores of the convolutional neural network (CNN) model for classifying sugarcane and other crops for imbalanced and balanced training data	147
Table 5.4. The impact of changing the kernel size for the first and second convolutional layer, on the recall, precision, and F1 scores of the convolutional neural network model for classifying sugarcane and other crops with imbalanced data	148
Table 5.5. The impact of changing the kernel size, for the first and second convolutional layer, on the recall, precision, and F1 scores of the convolutional neural network model for classifying sugarcane and other crops with balanced data	149
Table 5.6. A confusion matrix depicting the results of the initial model ran using; kernel sizes 8,8 (in the first and second convolutional layer) and imbalanced training data; balanced training data and kernel sizes 6,8 and 8,8 (in the first and second convolutional layer).....	150

Table 5.7. The impact of the number of convolutional layers on the recall, precision and F1 scores of the convolutional neural network model for classifying sugarcane and other crops using both balanced and imbalanced training data and kernel sizes 8,8.	151
Table 5.8. A confusion matrix depicting the performance of Model 1 and Model 2 trained using Unit 1 data and evaluated using Unit 2 data.....	154
Table 5.9. Model 1 and Model 2 performance for identifying sugarcane fields located in Unit 2. Both models were trained using Unit 1 data and evaluated using Unit 2 data.....	155
Table 5.10. A confusion matrix depicting the performance of Model 1 and Model 2 trained and evaluated using Unit 2 data (70%, 30% split)	156
Table 5.11. Model 1 and Model 2 performance for identifying sugarcane fields located in Unit 2. Both models were trained and evaluated using Unit 2 data (70%, 30% split).	156
Table 9.1. The profile and photograph of 1 st year ratoon sugarcane fields, first planted in 2018, located in Unit 2.	205
Table 9.2. The profile and photographs of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 4ft and wet when surveyed on 27th April 2019	205
Table 9.3. The profiles and photographs of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 2ft and wet when surveyed on 24 th April 2019.	206
Table 9.4. The profile and photographs of virgin a sugarcane field, first planted in 2019, located in Unit 2. This field was first drip irrigated, and the conditions of the field was reported as being 2ft and dry when surveyed on 24 th April 2019.	207
Table 9.5. The profiles and photographs of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 3ft and wet when surveyed on 24th April 2019.	208
Table 9.6. The profile and photograph of a virgin sugarcane field, first planted in 2019, located in Unit 2. This field was first drip irrigated, and the conditions of the field was reported as being 3ft and dry when surveyed on 24 th April 2019.....	210
Table 9.7. The profiles and photographs of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 4ft and wet when surveyed on 24th April 2019.	211
Table 9.8. The profiles and photographs of 1 st year ratoon sugarcane fields, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and	

the conditions of the fields were reported as being 2ft and wet when surveyed on 24th April 2019.	211
Table 9.9. The profile and photograph of a 1 st year ratoon sugarcane field, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 3ft and wet when surveyed on 24th April 2019.	212
Table 9.10. The profile and photograph of a 1 st year ratoon sugarcane field, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 5ft and wet when surveyed on 24th April 2019.	212

List of Abbreviations

1D	One dimensional
2D	Two dimensional
AI	Artificial Intelligence
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
dLUC	direct Land use change
DoG	Difference of the Gaussians
DOS	Day of Season
EO	Earth Observation
EOS	End of season
EMS	End of Mid-season
EVI	Enhanced Vegetation Index
FFT	Fast Fourier Transform
GEE	Google Earth Engine
GHG	Greenhouse gas
GIS	Geographical Information Science
GPS	Global positioning system
GUD	Green-up date
HCVs	High Conservation Values
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LSWI	Land Surface Water Index
LUCC	Land use and land cover change
ML	Machine Learning
mNDWI	Modified Normalized Difference Water Index
MODIS	Moderate Resolution Imaging Spectrometer
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
RADAR	Radio detection and ranging
RF	Random Forest
RGB	Red, Green, Blue
SAR	Synthetic Aperture Radar
SDGs	Sustainable Development Goals
SD	Senescent date
SDPS	Start date of peak season
SMV	Support Vector Machine
SOS	Start of Season
SMS	Start of Mid-season

SVM-RBF	Support Vector Machine with Radial Basis Function
VH	Sentinel-1 cross-polarized backscatter
VV	Sentinel-1 co-polarized backscatter
UN	United Nations

1 INTRODUCTION

1.1 Rationale

Sugarcane (*Saccharum officinarum*) is a tall perennial grass species native to South East Asia and Melanesia (Singh and Tiwari, 2018). The crop can produce high quantities of sucrose and hence it is widely grown for sugar and ethanol, and products such as bagasse, cane wax, molasses, jaggery (Gur), and fibre cake (Abdel-Rahman and Ahmed, 2008; Singh and Tiwari, 2018).

In 2000, more than 50% of the world's sugar was sourced from sugarcane (Abdel-Rahman and Ahmed, 2008), but this has since increased to over 80% with sugarcane being grown for profit in over 100 countries worldwide (Singh and Tiwari, 2018). Brazil, India, and China are the three largest producers of sugarcane today (Lin et al., 2009; Singh, 2018). The demand for sugarcane is increasing with global population growth and the increased use of biomass energy to mitigate climate change (Singh and Tiwari, 2018). But as the demand for sugarcane production increases, and the socio-economic and environmental states of sugarcane growing countries change, it would be quite possible to soon see changes in the dominant sugarcane growing countries (Singh and Tiwari, 2018).

India is the second largest producer of sugarcane in the world, and sugarcane provides employment for about 6 million farmers, and millions more employees with transport and tertiary roles associated with the industry (Mall et al., 2016). The main sugarcane producing states are Punjab, Bihar, Uttar Pradesh and Maharashtra (Singh and Tiwari, 2018). To enable sustainable long-term production, sugarcane production must be economically, environmentally and socially sustainable (Singh and Tiwari, 2018).

Economic sustainability relates to the profitability of production for workers and farmers which provides an incentive for continued production (Blair and Crocker, 2000; Meyer et al., 1996; Singh and Tiwari, 2018). Environmental sustainability concerns the minimal degradation of natural resources to enable continued production over many generations (Singh and Tiwari, 2018). Social sustainability relates to the social acceptance of continued sugarcane production by a range of rural and urban stakeholders (Conway, 1985; Edwards, 1989; Singh and Tiwari, 2018).

Changes in temperatures and rainfall related to climate change have affected sugarcane yields (Mall et al., 2016). Since 1904, the mean temperature in India has increased by 0.6°C, the number of heavy rainfall events in India has risen, and the number of low and medium rainfall events have fallen (Goswami et al., 2006; Mall et al., 2016). Mall et al. (2016) reports that between now and 2050, the number of heavy rainfall days in western and central India will increase by 15 days and mean temperatures by 2-4°C. Further predicted increases in temperature by 2080 are predicted to result in a 10-25% reduction in agricultural productivity (Mahato, 2014; Mall et al., 2016). It is therefore important to understand how climate change affects productivity and how management can be modified to maintain levels of sugarcane production in a sustainable way (Mall et al., 2016).

To increase the sustainability of sugarcane production and therefore sugarcane productivity, it is important to understand the spatial and temporal distribution of sugarcane growing conditions within and across different sugarcane growing regions. Knowing the spatial and temporal distribution of growing conditions allow for better decisions, such as the best time to plant/harvest, or when and where to apply fertilisers and/or irrigation, to be made (Nandhini and Padmavathy, 2017; Singh and Tiwari, 2018). In the past, researchers have attempted to understand the spatial and temporal distribution of sugarcane typically using national statistics and/or field datasets or the use of Earth Observation (EO) (occasionally combined with field data).

Many farming communities in developing countries do not have access to the right knowledge and physical tools (e.g. machinery, irrigation systems, improved cultivars) to produce sugarcane sustainably (Mnisi, 2012; Cliff, 2012). Hence much sugarcane production is carried out through unsustainable practices, such as continuous monocropping and inappropriate pesticide and fertiliser use, which are beginning to degrade soil and water resources (Singh and Tiwari, 2018).

EO can be used as a tool to tackle spatially and temporally complex problems, such as the identification and monitoring of sugarcane production over large areas. Data is collected using satellite sensors that orbit the earth, allowing for large areas to be covered quickly and changes to be studied in near real time (Fu et al., 2020). EO imagery can capture information beyond the visible spectrum such as near infra-red (NIR) radiation, and which is too time consuming to collect through field visits or in areas inaccessible to humans (Stubbings et al., 2019).

There are two types of remote sensing sensors. Active sensors emit their own energy and the backscatter from this is used to image surface features, while passive sensors take advantage of the absorption and reflectance differences of external electromagnetic energy, (i.e., from the sun) by different earth surface features (Toth and Jóźków, 2016). The interaction between sunlight and different land use and land cover types, produce varying spectral signatures relating to structural differences such as differences in canopy architecture and geometric structure, foliar chemistry (pigments in leaves like *chlorophyll a* and *b*, nitrogen content and leaf water content). Leaf area index (LAI), which describes the ratio of vegetative canopy leaf area (m^2) to the unit area of ground they occupy (m^2) (Watson, 1947), can also be derived from optical satellite data (Abdel-Rahman and Ahmed, 2008). Hence there is the potential opportunity to identify sugarcane straight from the spectral response curves of remotely-sensed images (Townshend et al., 1991).

1.2 Aim and objectives

This research investigates how sugarcane can be monitored from long term satellite programmes and new satellite constellations in near real time for the

sustainable production of sugarcane. The aim of this research is to determine how new forms of EO imagery and machine learning tools can help to increase resource efficiency (land, labour, nutrients, water), in sugarcane production. It is proposed that this aim can be achieved by addressing five objectives.

1. To review the current state of the art for monitoring sugarcane production for improved sustainability.
2. To develop and critically evaluate a methodology for extracting sugarcane growth information from EO time-series data.
3. To develop and critically evaluate a machine learning method, for accurate detection and classification of sugarcane using available sugarcane datasets.
4. To evaluate how information derived from EO datasets can be used within existing frameworks for improving the sustainability of sugarcane production.
5. To inform the debate on how information from EO can improve the sustainability of Indian sugarcane cultivation.

1.3 Structure

The thesis is made up of six chapters formatted as journal papers followed by a summary of the findings and conclusions. The experimental work was designed, coordinated, and conducted by Neha Joshi at Cranfield University while each chapter was authored by Neha Joshi and edited by Dr Daniel Simms and Professor Paul Burgess.

This research was supported by funding from NETA Analytics and the Natural Environment Research Council (NERC), through the Central England NERC Training Alliance (CENTA2) Doctoral Training Partnership [grant number NE/S007350/1]. Photographs taken by Neha Joshi through field visits were used to support the narrative of the literature review. GPS data showing the location of each sugarcane field were used alongside field photographs and additional information relating to irrigation type, year of planting, height at survey. These datasets were provided by NETA, to support the findings in Chapters 3-5. Ethical

approval was granted by Cranfield University before research activities took place (Appendix 9.1).

This introduction contains the background to sustainable sugarcane production and introduces the concept of remote sensing and its applications for sustainable sugarcane production. The remaining chapters reflect the above objectives and follow the structure outlined in Figure 1.1

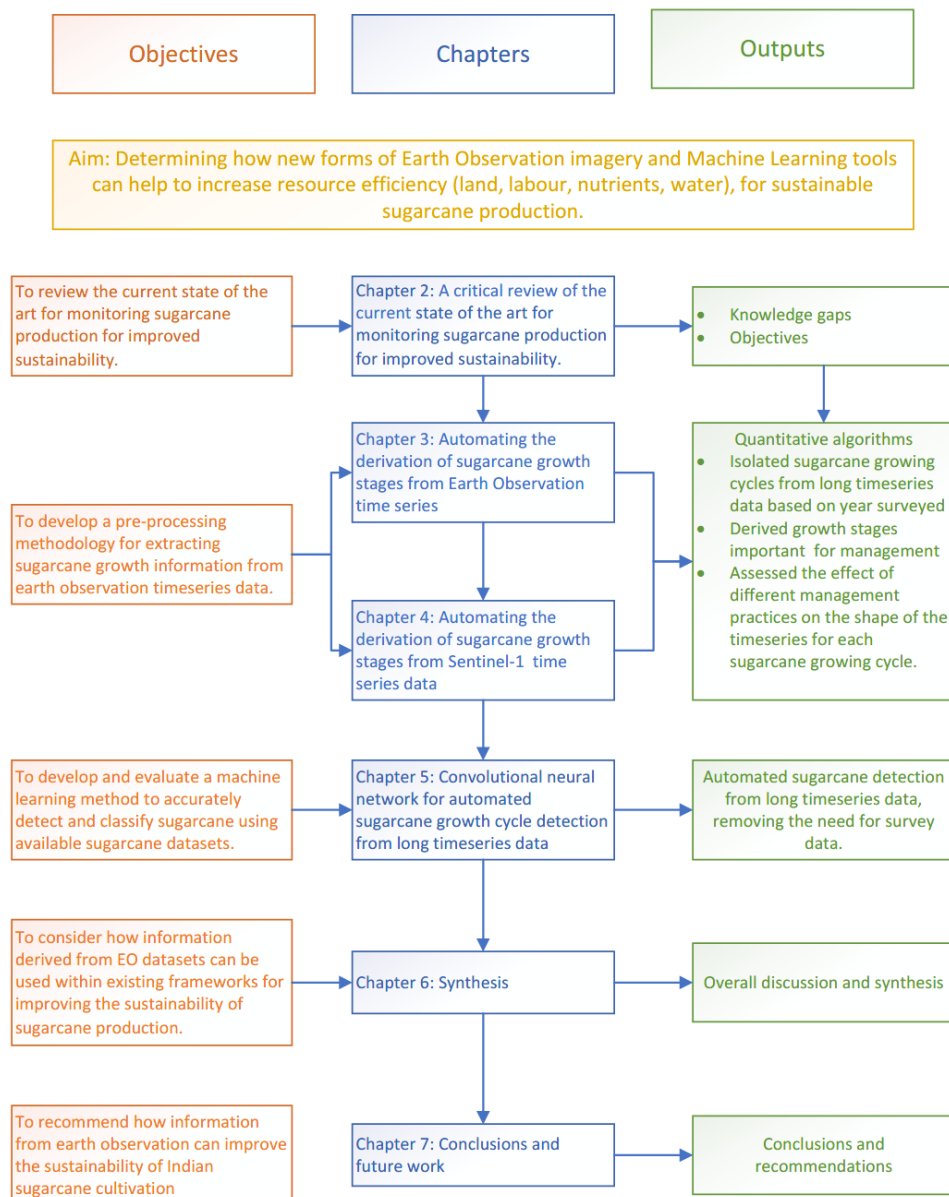


Figure 1.1 Thesis overview: describing the role of each chapter for fulfilling each objective

The first step to improving the sustainability of sugarcane production, is having effective strategies in place for monitoring sugarcane production to establish baselines for management. Chapter 2 addresses Objective 1 and reviews the current state of the art for monitoring sugarcane production for improved sustainability, and possible gaps in the literature.

Chapter 3 addresses Objective 2 and presents a pre-processing methodology for extracting sugarcane growth information from EO time-series data. This novel methodology was developed to tailor the pre-processing and harmonisation process to the Landsat 8 and Sentinel-2 time-series data to improve the removal of noise and agreement between the data derived from the two sensors, optimising the number of cloud-free observations.

Chapter 4 further develops the method outlined in Chapter 3 using synthetic aperture radar (SAR) data derived from Sentinel-1 for an improved seasonal splitting by overcoming the limitations of clouds. The identification of the tillering stage and harvesting new information about changes in the structure of sugarcane over time.

Chapter 5 presents a binary sugarcane/ not sugarcane classifier that was developed by combining Unit 1 and Unit 2 time-series datasets and splitting them as training (70%) and validation (20%) data.

Chapter 6 presents a synthesis of the results from each chapter, their contribution to knowledge and limitations, and discusses their applications in existing frameworks for sustainable sugarcane production. A short conclusion is then provided followed by recommendations for improving the sustainability of Indian sugarcane cultivation.

1.4 Overall methodological approach

This section provides an overview of the methodological framework employed in this thesis, which comprises six key steps.

Step 1 involved conducting a comprehensive literature review to establish the theoretical foundation for modelling plant growth and senescence.

Step 2 focused on the selection of two case study sites, referred to as Unit 1 and Unit 2. This step also included identifying specific fields within these units for monitoring sugarcane growth over time, as well as collecting corresponding field reference data.

Step 3 involved the acquisition of EO data corresponding to the same fields identified in Step 2.

Step 4 focused on the development of a framework for monitoring sugarcane growth using satellite-based remote sensing.

Step 5 entailed the classification of sugarcane within the EO time series datasets to capture spatial and temporal growth patterns.

1.4.1 Literature review

The role of the literature review was to construct a theoretical background of sugarcane growth and senescence, particularly how climate affects productivity and how management can be modified to maintain levels of productivity as the climate changes. Findings from the literature review showed that, in addition to space, optimal conditions for high yielding sugarcane changes temporally, particularly across four key stages of sugarcane growth (Lin et al.,2009; Mall et al.,2016). Optimising sugarcane management practices therefore requires information about seasonal methods, but current methods used to acquire this information is reliant on field data, which is both labour intensive and limited in spatial extent.

To overcome the limitations of field data, the second part of the literature review evaluated the potential use of EO data for monitoring sugarcane growth over time and space. The review highlighted the importance of EO time-series data for monitoring sugarcane phenology over time and distinguishing sugarcane from spectrally similar land cover types. However, the phenology derived from EO timeseries data is often not comparable with the growths stages that producers are familiar with. There is also no standardised time-series approach for the mapping of sugarcane, making comparisons between regions difficult and there is no combined approach for classifying and monitoring sugarcane growth at the same time. The identification of these knowledge gaps therefore allowed for the

development of the trapezoidal model for both monitoring and classifying sugarcane growth over time, which was then corroborated with observed data, for Indian sugarcane fields, in the later chapters that followed.

In Chapter 3, for example, the start, end and mid-season stages were derived using the simplified FAO trapezoidal model (Doorenbos, 1977) and a combination of optical EO datasets. In Chapter 4, the same model was used to improve the splitting of the season and the identification of additional growth stages such as the tillering stage and harvesting using structural information derived from SAR data. The Trapezoid model also allowed for the development of a binary sugarcane/ not sugarcane classifier for EO time-series data in Chapter 5.

1.4.2 Study area and field reference data

Sugarcane development was observed for sugarcane fields located across two study sites, Unit 1 and Unit 2, which were respectively located in the Indian states of Andhra Pradesh and Telangana (Figure 1.2). These fields were monitored and managed by the sugar producing company Natems (the parent company of NETA) through field surveys and the sugarcane grown in these fields were directly supplied to the company's mills. Unit 1 was surveyed by Natems in 2018, and Unit 2 was surveyed in 2019.

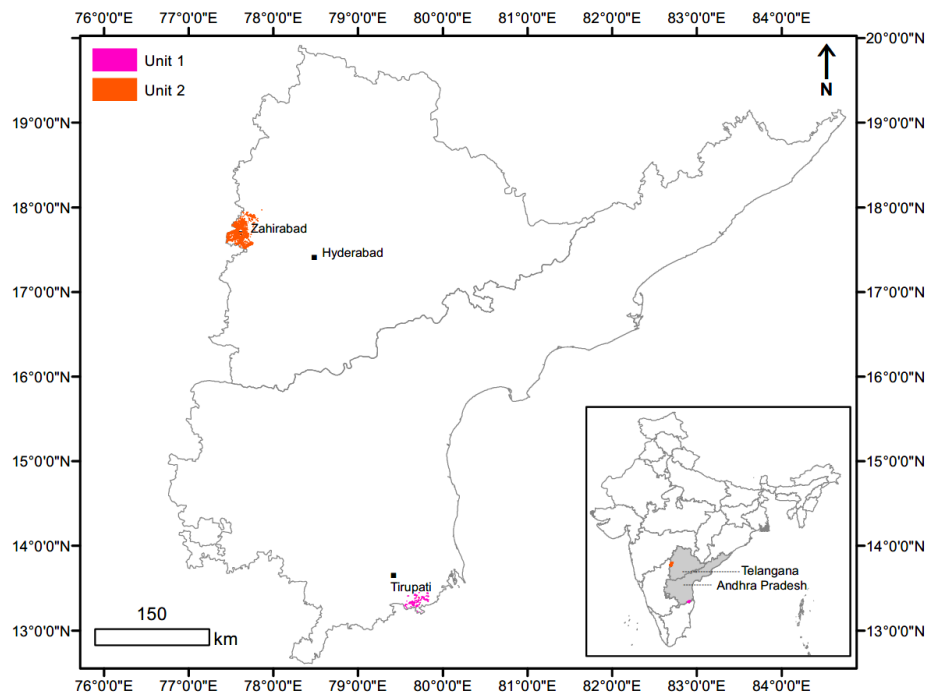


Figure 1.2 Sugarcane fields surveyed and managed by Natems.

The reason for selecting Unit 1 and Unit 2 as specific study sites for this thesis was twofold. Firstly, the management of fields located in Unit 1 and Unit 2 were different which made for a good comparison between the two sites. Unit 1 could be characterised as an area where flood irrigation and monocropping were common practices. Unit 2, on the other hand, could be characterised as an area where drip irrigation and crop-rotation were common practices. Secondly, the field data for these study areas were readily available, at a time when carrying out field visits was not possible due to the 2020 global pandemic. During the 2018 and 2019 surveys GPS data showing the location of each sugarcane field was collected, by walking the field perimeter with a handheld GPS device, alongside information relating to irrigation type, year of planting and the height of each field at the time of survey. The photographs of the surveyed fields and fields north of the surveyed fields (northern fields) were also provided by Natems and used in this research (Figure 1.3). This data was received in the form of individual folders with .kml files for the GPS coordinates and .jpg files for the photographs of each field. Hence the first step carried out was the re-organisation of the data for each

study area into two separate folders, one containing only the individual .kml files associated with each field and another containing only the field photographs. This was done using command prompt.

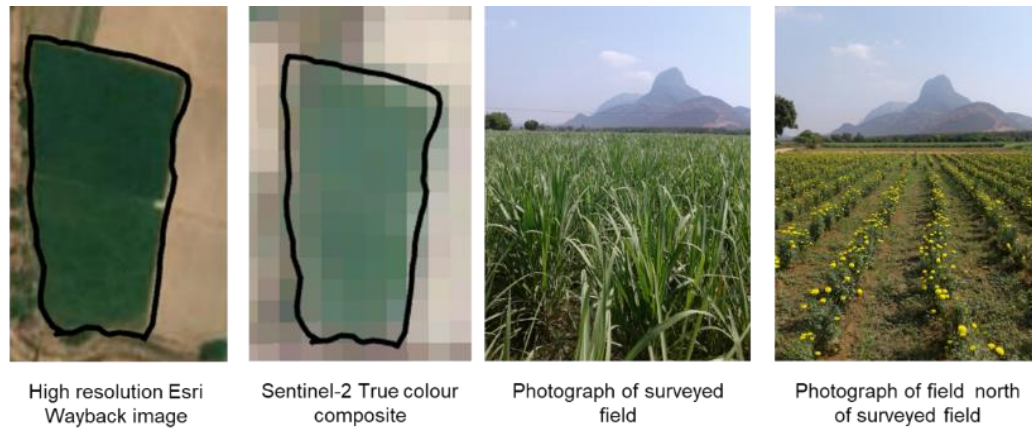


Figure 1.3 An example field polygon visualised using high resolution satellite imagery and field photography.

After organising the data, the individual .kml files, for each study area, were loaded into a GIS software to produce a geodatabase, for Unit 1 and Unit 2 respectively. The geodatabases were then converted into shapefiles and the GCS_WGS_1984 map projection was used to visualise and clean the dataset by verifying that the supplied polygons were in fact of sugarcane using the supplied field photographs and the guidelines outlined in Figure 1.3.

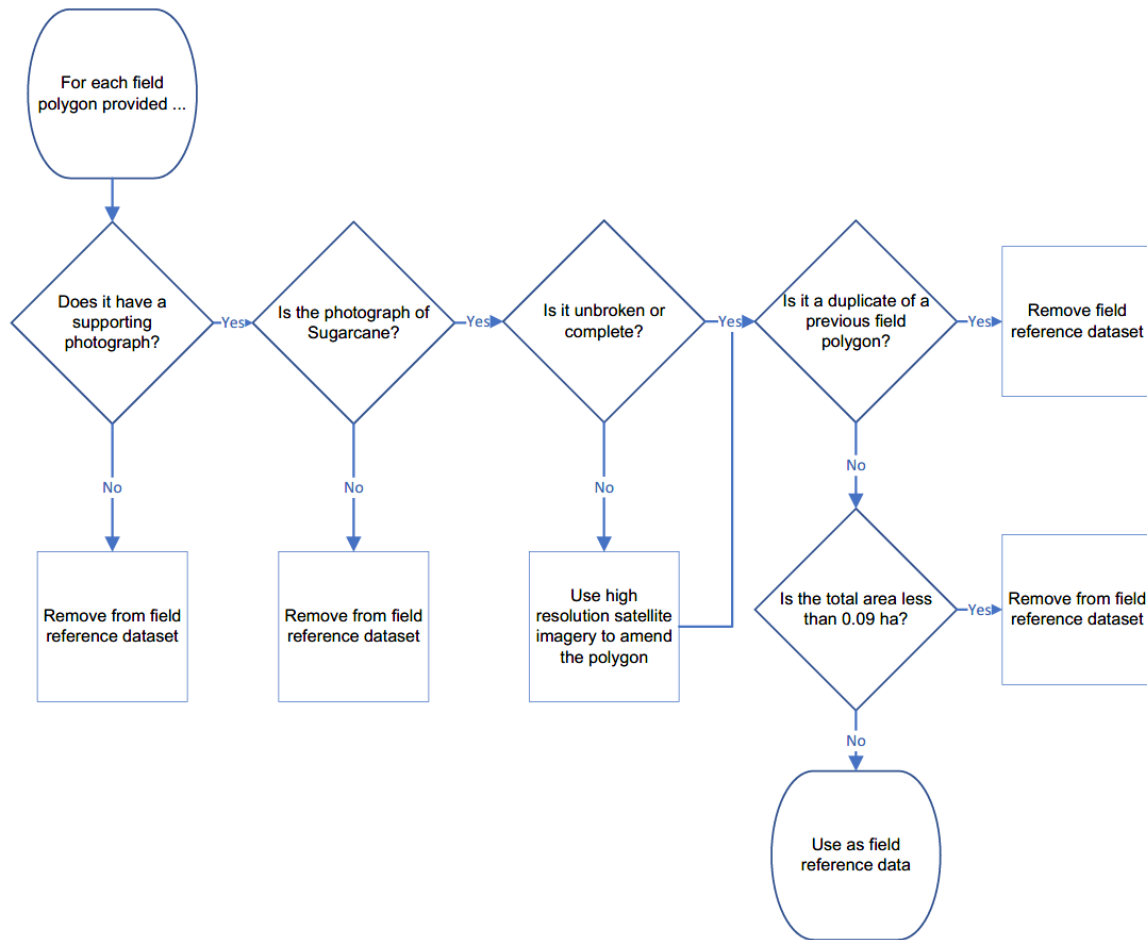


Figure 1.4 The set of guidelines used to verify the GPS data provided by Natems.

After cross-checking and correcting any incorrectly digitised field polygons, a total of 337 sugarcane fields located in Unit 1 and 400 sugarcane fields located in Unit 2 were used to carry out this research (Table 1.1) with over 90% of the supplied data for Unit 2, not meeting the guidelines outlined in Figure 1.4. The sizes of the studied sugarcane fields ranged from 0.14 to 11.5 ha, where the median field size was 1.05 ha for Unit 1 and 0.63 ha for Unit 2. The studied sugarcane crops were a mixture of virgin cane (fields planted from cuttings) from the season surveyed and ratoon cane, which was regrown after harvesting the previous season, planted between 2012 and 2018. In Unit 1, the sugarcane fields consisted of 29% virgin cane and 71% ratoon cane, whereas Unit 2 comprised 79% virgin cane and 21% ratoon cane. All fields in both units were irrigated; however, only 10% of the

fields in Unit 1 utilised drip irrigation compared to 87% in Unit 2. The remaining fields in each unit were irrigated using flood irrigation.

Table 1.1 Number of sugarcane fields before and after verification

Study area	Field data before verification	Field data after verification
Unit 1	354	326
Unit 2	4660	356

For the classification portion of this research photographs of the fields north of each sugarcane field, now referred to as the northern fields, were used alongside high resolution EO data (ESRI wayback) to provide information about other crops. Where the photographs and ESRI wayback images showed evidence of other crops being grown in the northern fields, field boundaries were digitised using the digitisation tool in ArcGIS Pro. In total 78 and 103 northern field boundaries were digitised for Unit 1 and Unit 2 respectively, with field sizes ranging from 0.11 ha to 10.87 ha.

1.4.3 Earth Observation datasets

The general approach was to automate the monitoring of sugarcane fields using open access EO datasets available at the Google Earth Engine (GEE) platform, with the methods being developed using EO time-series data for fields located in Unit 1 and evaluated using EO time-series data for fields located in Unit 2.

The study limited itself to data available on the GEE platform as it allows for tasks such as the integration and manipulation of datasets, regardless of format and datatype, to be carried out rapidly without the need for image data download, pre-processing, and computing resources (Gorelick et al., 2016). The type of datasets available on the GEE platform varied in terms of their temporal availability and spatial and spectral resolution, therefore Landsat 8, Sentinel-2 and Sentinel-1 were selected as appropriate EO datasets for monitoring sugarcane development

across the two study areas as they were able to provide information at the appropriate spatial, spectral and temporal resolution for monitoring sugarcane development overtime. Once selected, the GEE API in python was used to obtain mean reflectance values for the red and Near-infrared (NIR) electromagnetic wavelength bands of Landsat 8 and Sentinel-2 satellite and mean VH and VV backscatter values from Sentinel-1, for each sugarcane and northern field located in Unit 1 and Unit 2. Field mean values of NDVI and the VH/VH cross ratio (CR) were then calculated using Equation 1.1 and Equation 1.2 respectively.

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R}) \quad \text{Equation 1.1}$$

$$\text{CR} = \text{VH} / \text{VV} \quad \text{Equation 1.2}$$

Before analysis the NDVI timeseries profiles were further processed. This included, maximising the number of cloud free observations by evaluating two methods for cloud filtering and harmonising the Landsat 8 and Sentinel-2 time-series datasets by correcting the Landsat 8 observations to the Sentinel-2 observations, using local regression parameters. Residual noise not removed during cloud filtering was also removed using the best index slope extraction (BISE) algorithm, where an appropriate sliding window size was chosen through experimentation.

1.4.4 Monitoring sugarcane growth from space

The simplified FAO trapezoidal model for sugarcane growth, outlined by Doorenbos (1977), was used in this study to depict key growth stages important for sugarcane management and provide an insight into effective management strategies under different climate scenarios. To automate the derivation of key growth stages related to sugarcane management using EO, the processed time-series datasets were resampled to have equal intervals, and a fast Fourier transform (FFT) algorithm was used to simplify each timeseries into low

frequency components. A difference of the Gaussian (DoG) algorithm was then used to decompose the long pre-processed timeseries data into individual agricultural seasons and a knee and inverse knee algorithm was developed and implemented based on the work of Satopaa et al. (2011) to derive the growth stages corresponding to the FAO model for sugarcane growth.

A date for the four key growth stages of the FAO model for sugarcane growth were manually derived by visually inspecting the processed timeseries. The manually derived growth stages were then compared to the automated growth stages using orthogonal regression and statistical values of RSME and R^2 which provided information about the agreement between the manual and automated growth stages used to for validation. Photographs taken within ± 3 days of the date of the derived automated growth stages were also used for further validation and where photographs were unavailable, Sentinel-2 band 4,3,2 True colour composite (TCC) and band 8,4,3 NIR False colour composite (FCC) images were used.

1.4.5 Classifying sugarcane within EO timeseries datasets

The methods for automatically classifying sugarcane from within the processed timeseries datasets containing six seasons of growth, was developed by modifying an approach developed for the medical sector. The developed model was based on the approach described by Kensert et al. (2021) for the automated detection of peaks in reversed-phase liquid chromatography using the Python package Tensorflow. This model was chosen to classify sugarcane because of the similarity in behaviour between the unique trapezoid shape of sugarcane within time-series datasets and peaks in chromatogram datasets. The developed approach allowed for the localisation and classification of sugarcane to be carried out in the same step to overcome the limitations of classification approaches that require crop calendar information to isolate patterns in time-series data before classification can take place.

The custom code developed for this thesis and the data used have been is hosted in a private GitHub repository. For more information about the code and data, including details on how to access the files, see Appendix 9.4 and 9.5.

2 LITERATURE REVIEW

2.1 Introduction

An improved understanding of the location and development of sugarcane crops should provide options for managers to improve technical, economic and environmental performance of sugarcane production. For the farmer, this could include the best time to plant, harvest, fertilise and irrigate (Nandhini and Padmavathy, 2017). For a processor or government official, this could include information on the area and potential yields of sugarcane in a given season or area. The traditional method for determining the location and temporal characteristics was to use local surveys. However, this can be enhanced through the use of EO imagery. The objective of this literature review is to review the current state of the art for monitoring sugarcane production for improved sustainability.

2.2 Survey data

Sugarcane production can affect a number of the seventeen goals, and sub-goals, set in the United Nations Sustainable Development Goals (SDGs) for 2030 (Kavvada et al., 2020). Target 2.4 under the “Zero Hunger” goals is to “ensure sustainable food production systems and implement resilient agricultural practices”. There are a wide number of attributes that can be associated with sustainability including soil health, water, pesticide and fertiliser use (Table 2.1). The UN (2024) recommend that the data to monitor these attributes can be based on “remote sensing, GIS, models, agricultural surveys, household surveys, administrative data or environmental monitoring systems”.

Table 2.1. List of themes and sub-indicators associated with the SDG indicator 2.4.1
(adapted from FAO,2019)

Dimension	Sub-indicator	Theme
Economic	Farm output value per hectare	Land productivity
	Net farm income	Profitability
	Risk mitigation mechanisms	Resilience
Environmental	Prevalence of soil degradation	Soil health
	Variation in water availability	Water use
	Management of fertilisers	Fertiliser pollution risk
	Management of pesticides	Pesticide risk
	Use of biodiversity-supportive practices	Biodiversity
Social	Wage rate in agriculture	Decent employment
	Food insecurity experience scale (FIES)	Food security
	Secure tenure rights to land	Land tenure

Narayanamoorthy (2004) used samples from 100 fields to assess the effectiveness of different irrigation methods for improved sugarcane productivity in the state of Maharashtra, India. The analysis showed that the drip method of irrigation (DMI) could increase sugarcane yield and reduce costs. Although the above research is useful in helping gain an understanding of the spatial and temporal distribution of optimal sugarcane growth conditions for a specific study area, this paper was only able to cover one state. India is a large country with varying climates (Mall et al., 2016), and sugarcane is grown in many states across the country. Optimal growing conditions for sugarcane vary from region to region, as well as the varieties of sugarcane species planted (Singh, 2018). The climate in states like Uttar Pradesh, Bihar, Punjab and Haryana is subtropical whilst the climate in Maharashtra, Gujarat, Tamil Nadu, Andhra Pradesh and Karnataka is

tropical with optimal growing conditions for sugarcane, varying between the two climate groups (Gujja et al., 2009; Nandhini and Padmavathy, 2017).

In an assessment of Indian sugarcane production between 2000 and 2010, Nandhini and Padmavathy (2017) utilised government-published statistics and published literature on a state-wide level, for all Indian sugarcane-producing states. This research identified Uttar Pradesh as India's most productive sugarcane-producing state due to favourable rainfall, and Assam as India's least productive sugarcane-producing state. Sugarcane producers in Assam also tended to use less technology e.g., no irrigation mechanisms to offset the impact of less rainfall, as well as the fragmented fields (Nandhini and Padmavathy, 2017). Unfavourable weather events affected sugarcane production in all sugarcane-growing states during the 10-year study period especially changes to the timing of the monsoon season. Nandhini and Padmavathy (2017) therefore concluded that the implementation of irrigation mechanisms would be needed, to increase sugarcane productivity in India. In addition to spatial differences, optimal growing conditions for sugarcane crops also vary with time. A study by Mall et al. (2016) for sugarcane grown in the Indian State of Uttar Pradesh, found that each of the four stages of the sugarcane growth cycle had unique optimal temperature, water, soil nutrients and sunlight hour needs (See Figure 2.1). The grand growth stage, for example, was more dependent on moisture than any other stage, requiring a minimum of 1000mm of water and a humidity level of 80% for effective growth (Mall et al., 2016). This is because cane formation and elongation occur during the grand growth stage, with moisture affecting flowering and internodal length (Mall et al., 2016). The study by Nandhini and Padmavathy (2017) was therefore limited and did not provide information regarding the best times to irrigate the fields (at which growth stages) and how much water to use as this information was not available from government-published statistics and published literature.

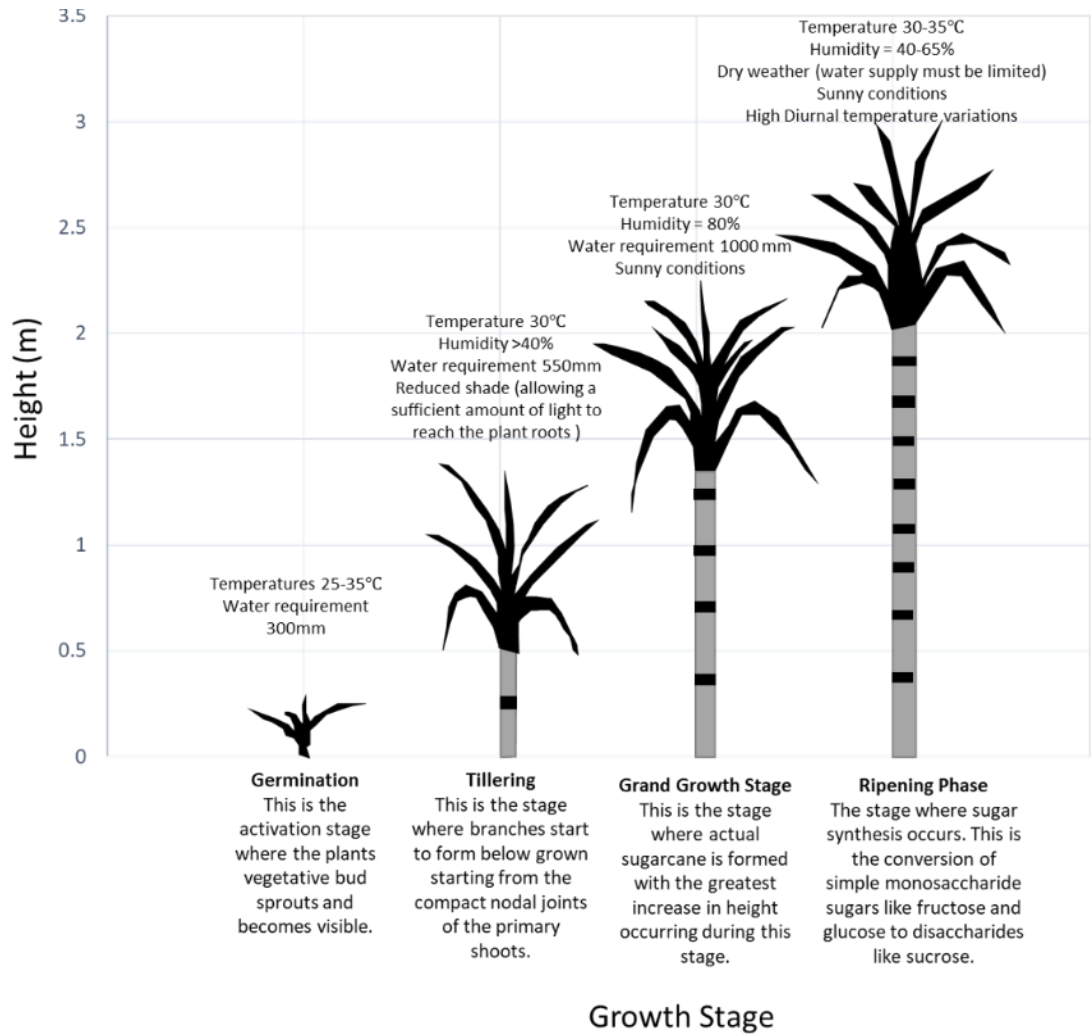


Figure 2.1. Stages of the Indian sugarcane growth cycle and their optimal resource requirements (Adapted from: Lin et al. (2009) and Mall et al. (2016)).

2.3 Earth observation data

Both the studies by Nandhini and Padmavathy (2017) and Narayanamoorthy (2004) used historic field data. However, sugarcane production can change with time (Fu et al., 2020) and the resulting information collected from historic field surveys quickly becomes outdated. Another limitation of sugarcane monitoring from field surveys is the labour-intensive, costly and time-consuming nature of data collection, especially as multiple surveys are required during the growing season (Olfindo et al., 2020; Wang et al., 2020) and the sampled data often does not fully represent the heterogeneous nature of the sugarcane industry (Fu

et al., 2020). EO imagery can be used as an alternate approach for monitoring sugarcane by providing near-real-time information about changes in biomass for large areas (Fu et al., 2020; Joshi et al., 2016; Sudmanns et al., 2020). EO information can also capture information in non-visible wavelengths at a temporal frequency that would be too time-consuming to collect through field visits and can image areas that are difficult to access on the ground (Fu et al., 2020; Stubbings et al., 2019).

EO could be used to improve the sustainability of sugarcane production by monitoring sugarcane for a single growing season at different growing stages, and across different growing seasons and years. Monitoring sugarcane production at different growing stages in one growing season and across growing seasons is important for improving the sustainability of sugarcane production, depending on whether the production of sugarcane in particular areas is deemed sustainable, the information derived from EO sensors should allow for intervention strategies to be put in place to improve the sustainability of sugarcane production.

It can be beneficial for sugarcane producers to use benchmarking frameworks as a tool for monitoring and improving the sustainability of sugarcane production and to compare different areas, years, and farms. Benchmarking frameworks can be defined as a shared set of indicators for data collection, implementing a mutual language and a criterion of data collection for sustainability. The Bonsucro Production Standard is an internationally recognised framework for sustainable sugarcane production and has been used for improving the sustainability of sugarcane production in countries such as Brazil, Australia and India (Smith et al., 2019; Bonsucro, 2023). In 2016, 4.4 % of global sugarcane was produced under the Bonsucro Production Standard guidelines and global companies such as Coca Cola and Pepsi co have made public pledges to sustainably source 100% of their sugarcane procurement, in line with the Bonsucro Production Standard guidelines (Smith et al., 2019). The Bonsucro Production Standard is a tool with five key principles that aim to help farmers and millers improve and verify their sugarcane production practices as sustainable, to provide assurance to buyers. The five principles in the Bonsucro Production Standard are to 1) assess

and manage environmental, social and human rights risks; 2) respect labour rights and occupational safety and health standards; 3) manage input, production, and processing efficiencies to enhance sustainability; 4) actively manage biodiversity and ecosystem services; and 5) continuously improve other key areas of the business (Bonsucro, 2023). Each principle comes with a set of indicators that address different economic, environmental and social components of sustainability, some of which can be monitored using EO data (Table 2.2).

Table 2.2. Four indicators in the Bonsucro Production standard that could be related to sugarcane growth stages and/or Earth observation data

Indicator	Definition
3.1.1	“The operator ensures that yields of production are above the threshold set by the climatic zone map
3.1.2	“The operator maximises sugar content in cane”
4.3.5	“The operator maximises irrigation productivity
4.4.3	“The operator maximises the efficiency of agrochemicals applied”.

Remote sensing sensors can be active or passive. Active sensors release their own energy, and the backscatter is used to determine surface features. Passive sensors, on the other hand, take advantage of differences in the absorption and reflectance of external electromagnetic energy (i.e., from the sun), by different Earth surface features (Toth and Jóźków, 2016). The interaction of sunlight with different land use and land cover types produces varying spectral signatures relating to structural differences between plants and soils. These structural differences can include differences within canopy architecture and geometric structure, foliar chemistry (pigments in leaves like chlorophyll a and b, nitrogen exposure and leaf water content) and agronomic parameters (like Leaf Area index (LAI) (Abdel-Rahman and Ahmed, 2008). Townshend et al. (1991) suggested the possibility of distinguishing sugarcane from other land cover types using only the

spectral signatures taken from ground surveys of sugarcane and other known land cover types. However, factors such as sun-sensor geometry, mixed pixels for images with low to medium spatial resolution, atmospheric effects, cloud cover and the radiometric properties of different sensors, reduce the capability of identifying sugarcane straight from the spectral response curves of known land cover types (Townshend et al., 1991). Statistical classification models which group similar earth surface features together, are therefore used to offset the impact of the above factors (Vikesh et al., 2010).

Active and passive sensors have been used to monitor sugarcane production at different growing stages in one season and across growing seasons, using different models and combinations of sensors. Having a whole range of proposed methods for sugarcane monitoring is problematic because methodological differences between these studies make replication and comparison between different sugarcane growing regions difficult. Hence, the next section reviews the current EO methods for monitoring sugarcane production within and across growing seasons

2.4 Monitoring and predicting yield at different growing stages

Indicator 3.1.1 of the Bonsucro Production standard highlights the importance of ensuring that yields are above the threshold set by the climatic zone map. Accurate monitoring of yield is important for indicator 3.1.1 of the Bonsucro production standard, as it could enable biotic and abiotic stress detection based on management practices at different growing stages to guide improved management aimed at reducing the cost of cultivation, post-harvest losses, and increasing sucrose content at harvest (Singh and Tiwari, 2018). By observing the continuous relationship between sugarcane yield and socio-economic factors like, weather patterns, water availability, soil nutrient stress, management strategies, pests, and disease, the detection of any anomalies which might indicate stress becomes easier (Mulianga et al., 2013). The identification of early stress warning signs can allow strategies to be put in place to reduce damage to yield and sugarcane health (Bégué et al., 2010).

Mulianga et al. (2013) investigated the use of Moderate Resolution Imaging Spectrometer (MODIS) derived NDVI data for different sugarcane management zones in Kenya over an eight-year period (2002–2010) to forecast annual yield. The results of this study found that normalising NDVI values using a weighting method that considered the growth period of the sugarcane crop (wNDVI), produced yield estimates that were more consistent with historical yield and significant (P -value = 0.001) than using NDVI integrated over the calendar year. The results also highlighted the difficulty in forecasting annual yield using low-resolution MODIS data as the sizes of sugarcane fields were less than a 250 m pixel and had no seasonal cropping calendar. The recommendations from this research were for future research to monitor sugarcane yields using high image resolution satellite imagery allowing for crop phenology to be extracted from individual plots.

A study by Mello et al. (2014) also noted the advantages of yield monitoring at multiple occasions during sugarcane growth to account for the impact of changes to environmental conditions such as the heavy rainfall and frost on sucrose concentration thus, yield. This study used 7-day smoothed real-time MODIS NDVI sugarcane time-series profiles alongside official yearly yield from 2003 to 2011, reported by the Brazilian Institute of Geography and Statistics to produce weekly estimates of sugarcane yield for 5 municipalities located in São Paulo State, Brazil. The results of this study highlighted that the period of full development (when sugar accumulation ends) produced the best estimates for remote sensing-based yield. The harvest season on the other hand produced the worst EO based estimates of yield, with many overestimates.

As part of their attempt to forecast the future of sugarcane yield in the Bundaberg region of Australia, Rahman and Robson (2016) also produced a time-series to identify the optimum time for sugarcane yield estimation based on historical relationships using 98 archived Landsat images captured between 2001 and 2015. After masking the Landsat data to vector layers containing pre-defined location and extent of each sugarcane field, Rahman and Robson (2016) calculated the green normalised difference vegetation Index (GNDVI) for each field. By averaging GNDVI across all fields within the Bundaberg region of

Australia for each time point contained within the time-series, Rahman and Robson (2016) identified that yields could best be related to GNDVI data from April when sugarcane crops were at peak growth i.e. when GNDVI reached a maximum.

Direct yield estimates from vegetation index time-series data, however, can be limited by inter-annual variations in yield. A study carried out by Dimov et al. (2022) demonstrated the advantages of using dimensionality-independent phenology metrics derived from daily linear interpolated Sentinel-2 NDVI time-series data as predictor variables to a Random Forest (RF) regression model for estimating sugarcane yield in Ethiopia. In addition to the peak of the season (POS), which was also identified by Rahman and Robson (2016) as an important part of the sugarcane seasonal time-series for identifying yield, the study by Dimov et al. (2022) also identified season duration and green-up rate (gradient between the start of the season (SOS) and POS) as important metrics for estimating sugarcane yield. The results of the study by Dimov et al. (2022), also found that the characteristics of areas of sugarcane with high yields and healthy biomass were high growth rates (gradient between SOS and POS), average sum (sum of average NDVI values) and maximum sum (sum of maximum NDVI values), and low senescence rates (gradient between POS and the end of season (EOS)).

2.5 Deriving different growing stages for resource management

There are advantages for crop management if the information derived from EO monitoring can be compared with growth stages familiar to producers using standard management practices in the field. The integration of EO-derived information regarding the seasonal development of sugarcane with external datasets about climate and farming practices could aid the optimisation of agricultural practices such as fertilisation, irrigation, and pest and disease management (Li, 2022; Mishra et al., 2023) as management varies with different sugarcane growth stages and spatially between climatic zones (Mall et al., 2016).

However, the phenological stages derived from EO data in the studies by Rahman and Robson (2016) and Dimov et al. (2022) for estimating yield, lack transferability to field survey derived growth stages important to management.

Water is the main limiting factor during the mid-season (Carr and Knox, 2011; Mall et al., 2016). If weather and yield data is also available, the impact of irrigation type and the amount of water used at specific growth stages and under different weather conditions could be assessed to help inform future management practices for improved yields and irrigation productivity (Mall et al., 2016). Limiting the supply of water to sugarcane crops is also recommended during the late season, so deriving this stage is useful for informing farmers on when to limit water supply (Mall et al., 2016). Using historic EO time-series data to monitor stages over large areas and time scales is useful for establishing baselines for irrigation management where yields and intervention practices are very variable.

Near real-time EO time-series data is also important for improving the sustainability of sugarcane. To maximise the sugar content in the cane (Indicator 3.1.2 of the Bonsucro production standard), a recommendation given to sugarcane factories is to collect sugarcane from the fields within 24 hours after harvest to reduce post-harvest losses (Misra et al., 2022). The use of near-real-time EO data could therefore help identify the time delay between harvest and collection for use in the factory, and it could be used to help plan factory routes and allocate resources for crushing.

The use of agrochemicals such as fertilisers and pesticides is common in sugarcane management. The optimal use of nutrients and pesticides can be related to specific growth stages. Borers are pests that can damage sugarcane stems by entering the crop through small holes in the epidermal tissues of leaves and stalks, they then proceed to feed on the tissues, which can promote fungal growth and reduce sucrose content and yields by contaminating sugarcane juice (Mishra et al., 2023) (Figure 2.2). To manage early shoot borer, *Chilo infuscatellus* Snellen (Lepidoptera: Crambidae) pests, it is recommended for farmers to start “earthing up” (creating a mound of soil around the base of the

sugarcane stem) and apply fertilisers at the pre-monsoon season (Mishra et al., 2023).



Figure 2.2. Damage to a sugarcane stem, resulting from a shoot Borer infestation

A study by Yeasin et al. (2022) was able to model the germination, tillering, grand-growth and maturity stages of sugarcane growth in Haryana, India using EO data derived from Sentinel-1, Sentinel-2 and seven machine learning algorithms. The logistic regression, decision tree, random forest, artificial neural network, support vector machine, naïve Bayes, and fuzzy rule-based systems algorithms were trained and validated using field derived phenological stages from 40 sugarcane fields with an average size of 3–3.5 ha at 5 temporal points split using an 80:20 ratio. The results of this study found little difference in the performance of the machine learning models, although the random forest, neural network, and SVM models performed slightly better than the other four models. NDVI and WDWI indices were found to be the most responsive Sentinel-2 derived datasets to different sugarcane growth stages while the VV and VH polarization were most useful for discriminating growth stages using Sentinel-1 data. The best performing models were those that used both Sentinel-1 and Sentinel-2 datasets

rather than using Sentinel-1 and Sentinel-2 datasets separately. The results of this study are important as the derived phenology is familiar to producers using standard management practices in the field and they highlight the benefits of combining the use of optical and SAR datasets. However, the application of this approach over larger study areas is limited with the need to derive training data from field surveys for five different time periods. Future research should therefore consider methods for overcoming the reliance on survey data for a more robust approach for deriving sugarcane growth stages from EO imagery.

2.6 EO data for classifying sugarcane

Using EO datasets for classifying sugarcane is important for improving the sustainability of sugarcane production for two reasons. Mapping provides an understanding of where sugarcane cultivation is, overcoming the reliance on the GPS field datasets used in the studies above to study sugarcane phenology over time. If sugarcane plantations can be located across large areas and multiple years without GPS data, comparisons in productivity and yield can be made based on differences in environmental conditions and management practices between different years. Using EO datasets for monitoring Land use and land cover change (LUCC) over time is also useful to detect reductions or increases in sugarcane areas over time and locate where this is happening, allowing the conversion of biodiverse landscapes to sugarcane to be monitored and limited. Maintenance or expansion of biodiversity within existing sugarcane fields through intercropping or crop rotation can also be monitored using EO (Liu et al., 2023; Wang et al., 2020). This information could be useful to help plan intervention strategies for improving the sustainability of sugarcane production.

There are four environmental and two social sustainability indicators in the Bonsucro Production standard that could be related to the use of EO for classifying sugarcane (Bonsucro, 2023).

Table 2.3. Four environmental and two social sustainability indicators in the Bonsucro Production standard that could be related to Earth observation data (Bonsucro, 2023)

Indicator	Definition
2.1.3	Social: “the operator ensures that all workers on the premises included in the unit of certification have access to safe water and adequate sanitation facilities
2.1.6	Social: “The operator ensures that first aid and an emergency response are available to all workers on the premises included in the unit of certification”
4.1.1	Environmental: “The operator ensures that the biodiversity and natural ecosystems are mapped”
4.1.3.	Environmental: “The operator ensures that areas of legally protected natural ecosystems (nationally or internationally), or areas classified as HCV, have not been converted to sugarcane on or after 1 January 2008”,
4.1.4	Environmental: “The operator ensures that areas of natural ecosystems (either legally protected or not) or areas classified as HCV have not been converted to agriculture on or after 1 January 2021”.
4.2.5	Environmental: “The operator prevents sugarcane tops and leaves from being burned after harvest.”

Classifying sugarcane alongside other crops, natural ecosystems, high conservation value landscapes (HCVs) and water can directly help cane operators help assess indicators 4.1.1, 4.1.4, 4.1.3 and 4.3.1 (Table 2.3). Intercropping cowpea, black and green grams with sugarcane is also important for the management of pest such as Root borer, *Emmalocera depressella* (Mishra et al., 2023). Additionally, classifying sugarcane that is burnt after harvest would help with identifying where sugarcane burning takes place. Often sugarcane mills have GPS derived field boundaries associated with particular farmers, so classifying sugarcane burning and monitoring this over time would help with

identifying field boundaries that are prone to burning, which could be useful for helping to prevent sugarcane tops and leaves of sugarcane from being burned after harvest (Bonsucro indicator 4.2.5).

Identifying the location of sugarcane fields overtime through classification may also provide information about the ratoon year that a particular harvested sugarcane crops supplied to the mill came from, i.e. a first-year ratoon or fifth year ratoon. The type of cane supplied to the mill is a concern for sugarcane managers as virgin cane tend to have higher sucrose content than ratoon cane, with sucrose content tending to decrease with each ratoon year. Information that allows for the discrimination of virgin and ratoon sugarcane crops is therefore useful for sugarcane factory managers. Virgin sugarcane crops are also less susceptible to red-rot disease, a disease that can reduce the quality of sugarcane crops and create huge losses in yield (Figure 2.3), therefore avoiding ratooning the crop can also be a management practice against the disease (Shukla et al., 2022).



Figure 2.3. A sugarcane field affected by red rot disease

There is a general limitation of EO datasets when assessing social sustainability as the social aspects of sugarcane sustainability such as decent employment, food security and land tenure cannot be directly measured using EO data. This

gap could be filled with commercial data, but that would incur cost. Alternatively, by combining classification outputs with different sources of socio-economic data, many different aspects of social sustainability could be assessed.

Determining the proximity of sugarcane fields to urban areas, could provide information about poverty and working conditions. For example, the distance between sugarcane fields, houses, roads, and public transport would provide an indication of how far farmers travel to get to work and their mode of transport e.g. public transport or walking via dirt roads. Studies carried out by Andrade et al. (2011) and Silva et al. (2010) have also suggested a link between sugarcane burning and increased cancer risk. Therefore, the identification of burn scars, from sugarcane burning, and their distance to urban areas (houses) could also provide information about the health of farmers and local residents (El Chami et al., 2020). Child labour is also another social issue surrounding sugarcane production (El Chami et al., 2020), therefore perhaps additional information such as the distance between sugarcane plantations and schools could help assess this.

To limit heat stress, workers must have access to areas of shade, adequate sanitation and access to water in the field, therefore knowing where areas of shade and toilet blocks are in relation to the classified sugarcane fields could help with the assessment of Bonsucro indicator 2.1.3 (Table 2.3). Information relating to road access and the distance of sugarcane fields from hospitals and schools could also indirectly provide information relating to Bonsucro indicators 2.1.6, while an additional, Geographical Information systems (GIS) system, that integrates EO data and social data, could be used to help provide information relating to Bonsucro indicator 2.3.4 (Table 2.3).

2.6.1 Single date image classification

Early EO identification and classification of sugarcane focused on using single-date images with parametric supervised and unsupervised classifiers (Abdel-Rahman and Ahmed, 2008). Supervised classification approaches use

representative training data, where the cover type is known. The reference data are then inputted into the supervised classification algorithm allowing landcover classes to be assigned to unclassified pixels, with some degree of error (Killeen et al., 2015). Supervised classification algorithms like the Maximum Likelihood (ML) algorithm, have been used to distinguish sugarcane from surrounding vegetation (Johnson and Kinsey-Henderson, 1997) and estimate the sugarcane area (Yawson et al., 2018). Vikesh et al. (2010), for example, used EO data from the Advanced Spaceborne Thermal Emission and Radiometer (ASTER) satellite, along with 2500 pixels of ground survey training data and the supervised maximum likelihood (ML) classification algorithm to map sugarcane varieties in the village of Sultanpur, in Haridwar, Uttarkhand, India. They reported that the ML algorithm was more successful in accurately classifying sugarcane and other agricultural land cover types than other classification algorithms. Vikesh et al. (2010) also found the use of spectral indices like Normalized Difference Vegetation Index (NDVI) with the Maximum Likelihood classifier, after atmospheric correction, to be more accurate at classifying sugarcane and other agricultural landcover types compared to the use of raw spectral signatures of different landcover types. However, Vikesh et al. (2010), also noted the impact of spectral mixing and confusion caused by mixed pixels.

Compared to parametric supervised classification approaches, unsupervised clustering is often the favoured approach for classification over large areas (Wulder et al., 2004). This is because, in the absence of training data, unsupervised classification algorithms use clustering to split pixels into separate categories, which can then be transformed into landcover classes using additional datasets or expert knowledge (Olaode et al., 2014). Unsupervised K-means classification algorithms and raw reflectance data obtained from optical Multi-spectral Radiometer MSR5 imagery were used to classify the landcover types of seven fields into bare land, cotton, and sugarcane, demonstrating the K-means algorithm's capability for accurately classifying bare land and cotton within the studied fields from raw reflectance data, as although there were some occurrences of sugarcane being misclassified, the overall classification of sugarcane was found to be relatively good (Shifa et al., 2011). The study by Shifa

et al. (2011) is relevant to the topic of sugarcane mapping over large areas using current EO imagery because it can be replicated using Landsat data as the five reflectance bands of MSR5 are compatible with the spectral reflectance bands of some Landsat instruments (Shifa et al., 2011).

Recently, the use of parametric classification studies has tended to decrease being replaced by machine learning algorithm-based classification studies which can better handle complex datasets (Rogan et al., 2008; Yu et al., 2013). Hence, machine learning algorithms are being used for land cover and land use change mapping, including the mapping of sugarcane (Praveen et al., 2019). Examples of previous studies that have used machine learning to classify sugarcane are outlined below.

The use of machine learning for sugarcane classification and the classification of six other land cover types in Uttar Pradesh and Bihar, India was carried out by Praveen et al. (2019). This research aimed to assess the transferability of three supervised machine learning algorithms to real-world agricultural landcover classification needs using free-to-use datasets like Landsat 8, Landsat 5, night-time light, and SRTM-DEM images available at the GEE platform. These images were captured in 2018, 2009 and 1998. The three machine learning algorithms reviewed by Praveen et al. (2019) were Support Vector Machine with Radial Basis Function (SVM-RBF), Random Forest (RF) and Classification and Regression Tree (CART). SVM was the best model for separating sugarcane from other land cover types in Uttar Pradesh and Bihar.

These findings have since been built upon by other researchers like Villareal and Tongco (2020), who used an Object-Based Image Analysis (OBIA) combined with an SVM algorithm to map sugarcane growth stages in Turburan, Philippines. OBIA is an alternate approach used to classify surfaces from remote sensing data, where surface features are classified using objects that are made up of groups of homogenous pixels (Blaschke, 2010). To classify high-medium and low-elevation objects, Villareal and Tongco (2020) used canopy height information extracted from Light Detection and Ranging (LiDAR) datasets and texture information obtained from red-green-blue (RGB) ortho-photos, which

were merged into one raster dataset collection for image segmentation and rule-development in eCognition software. To identify sugarcane and shrubs the SVM algorithm was used and validated using ground survey data obtained using a portable Global positioning system (GPS) device (Villareal and Tongco, 2020). Villareal and Tongco (2020) estimated the total size of land classified as sugarcane plantations as 13.93 ha with a 98.4% degree of accuracy. The use of LiDAR data also enabled Villareal and Tongco (2020) to estimate the proportion of sugarcane for each growth stage with an accuracy of 88%, 94.4%, 96.3%, and 91.7% for establishment, tillering, yield formation, and ripening, respectively. Their findings demonstrated the advantages of combining SVM and OBIA classifying sugarcane at different growth stages, which other studies can adopt to estimate future yield, at a similar spatial scale.

The studies above demonstrate the capability of the SVM algorithm to successfully classify sugarcane. One thing to note is the dependency of machine learning models on the input data and how the data are used. Attributes like spatial and spectral resolution, for example, influence the amount and type of information obtained from EO imagery. So, to map sugarcane over large areas, enough data must first be available and of high resolution to reduce spectral mixing and capture variation between fields (Cihlar and Jansen, 2001). Both studies use a range of features including height, texture for very high-resolution sensors, spectral reflectance from EO and vegetation indices that can be used to discriminate sugarcane. Praveen et al. (2019), for example, used night-time light to separate built-up and urban areas from vegetated areas and then the median spectral responses in all Landsat 5 and 8 spectral bands for the Kharif and Rabi season for land cover classification in non-built-up areas, in India. Spectrally similar vegetation located at different altitudes was separated using SRTM-DEM data. On the other hand, the study by Villareal and Tongco, (2020) used Lidar data in addition to RGB orthophotographs to classify sugarcane at different growth stages. Hence this method used by Villareal and Tongco (2020) outperformed the methodology used by Praveen et al. (2019) for sugarcane classification.

The above studies show that it is possible to identify, map and estimate the size of sugarcane plantations from EO imagery to a reasonable degree of accuracy, at different growing stages. They also provide useful information for future studies. For example, spectral indices like NDVI (after atmospheric correction) can be used to discriminate sugarcane from other landcover types compared to the use of raw spectral signatures (Vikesh et al., 2010). These studies also highlight the advantages of using unsupervised classification algorithms where sufficient field reference data are difficult to obtain (Shifa et al., 2011). In addition to spectral differences between sugarcane and other land use types, Praveen et al. (2019) and Villareal and Tongco (2020) also demonstrate how additional features of sugarcane like height, texture and elevation could be used to separate sugarcane from spectrally similar crops. They also highlight the advantage of machine learning algorithms for processing large datasets quickly.

The spatial resolution of the EO dataset affects the precision of the land cover classification (Cihlar and Jansen, 2001). The advantages of high spatial resolution sensors to reduce spectral mixing between small fields are also highlighted by Praveen et al. (2019) and Villareal and Tongco (2020). More information and land cover types can be captured, for example, within Landsat images, at a resolution of 30 m, than MODIS images at a resolution of 250 m. Moreover, with a spatial resolution of 10 m, the mapping of sugarcane can be improved further if Sentinel-2 images are used, as there is even less spectral mixing between land cover types. However, there are trade-offs between spatial, spectral, temporal, and radiometric resolution between EO satellite sensors. So, as spatial resolution increases, spectral, temporal, and radiometric resolution often decreases. Optically sugarcane can be detected from the visible blue, green, red and NIR wavelength bands, which can be extracted from Landsat and Sentinel-2 image products. On the other hand, SPOT images, used by Johnson and Kinsey-Henderson (1997) to classify land use in Australian sugarcane regions, only contained a panchromatic band with a spatial resolution of 10 m, as well as the green, red and NIR multi-spectral bands with a spatial resolution of 20 m. The availability of data controlled by the temporal resolution of specific EO imagery, and their cost is also a key factor for the mapping of sugarcane over

large areas. Praveen et al. (2019) for example, used Landsat 5 and 8 datasets because they had high enough temporal resolution to obtain a cloud-free image (or an image with minimal cloud) during the Kharif and Rabi seasons and at a high enough spatial resolution to capture small landcover classes like sugarcane. Multiple scenes are often required to map large areas, so using EO data scenes with purchasing fees can quickly become expensive. Another reason Praveen et al. (2019) used Landsat images was to reduce costs as this dataset can be openly sourced and free to use. The study by Villareal and Tongco (2020), on the other hand, used LIDAR data which is very expensive to obtain and requires specialists to use the instrument.

Although single-date mapping is useful, it is unable to study change over time (Sudmanns et al., 2020) and there can be problems due to cloud cover. Features such as spectral reflectance, texture, shape/size, and height are used to aid sugarcane classification. However, sugarcane is a dynamic crop, so these features change with phenological variations and management practices (e.g., irrigation, planting, and harvest) over time. Without temporal information to establish a baseline for the 'normal relationship' between these features of sugarcane and time, the discrimination of sugarcane from other land cover types and vegetation becomes limited. Clouds can also be a source of missing optical data, therefore multi-temporal mapping which uses multiple images from different points in time, can allow data gaps caused by cloud cover to be interpolated.

2.6.2 Time-series-based classification

Time-series based approaches involve the use of multiple, chronologically ordered EO datasets instead of using images from one point in time. They allow landcover to be assessed continually over many different spatial and temporal resolutions (Morel et al., 2014). The spectral signatures of sugarcane and other upland crop types like paddy rice and maize can be similar. Hence, one option is to distinguish between crops, not based on a single date, but in terms of a distinct signal in the time-series of a crop. Sugarcane, for example, has a single annual cropping cycle whereas paddy rice and maize may be cropped two or

three times per year in India. Knowledge of sugarcane phenology has been used in the past, along with spectral data to map sugarcane fields.

Wang et al. (2020), for example, annually mapped sugarcane in Guangxi province, China at a spatial resolution of 10 m, using a pixel and phenology-based time-series. The time-series used by Wang et al. (2020) was made up of Landsat-7/8, Sentinel-2 and Sentinel-1 images (LC/Sentinel-2 /S1) taken between 2017 and 2019. Wang et al. (2020) first separated the four general land cover classes within the study area (evergreen vegetation, built-up barren land, water-related land cover and croplands) using values of NDVI, Enhanced Vegetation Index (EVI), Land Surface Water Index (LSWI) and Modified Normalized Difference Water Index (mNDWI) and thresholds pre-determined by previous studies. An expert classifier based on the phenological differences between the different crop types was then built to separate the croplands class into sugarcane, paddy rice and maize as this approach overcame the issue of spectral similarity between the crop types. To do this, Wang et al. (2020) created a map for 2018 showing crops with single, double, and triple cropping cycles for pixels grouped as upland crops, using NDVI and LSWI time-series smoothed by a Savitzky–Golay (SG) filter to detect the beginning of the season (green-up date (GUD)), peak (start date of peak season (SDPS)), and end of the season (senescent date (SD)) for each pixel. They then counted the total number of peaks per annual cycle for each pixel in 2018 to classify crops with single, double and triple annual cycles in the study area in 2018. The accuracy of this map was then evaluated using sample reference data. The respective accuracies of the maps depicting crops with single, double, and triple annual cycles in 2018 were 88%, 96% and 96% (Wang et al., 2020). Once single-cycle croplands were isolated, Wang et al. (2020) compared the dates of each phenological metric with the dates of known sugarcane fields, and a decision tree classifier was then used to separate sugarcane from other single-cycle dryland crops. This study defined sugarcane as crops with single annual peaks. Using this definition and the mapped single annual cycle crops, the study estimated the total sugarcane area in Guangxi to be 8940 km². This was ~1% greater than the value projected by the Guangxi Agricultural Statistics Report. The green-up dates for 2019 sugarcane were then

monitored to assist with the sustainable management of sugarcane, following the generation of an annual sugarcane map depicting sugarcane distribution for 2019 (Wang et al., 2020).

Although the study by Wang et al. (2020) highlights the potential of phenological-based time-series methods of sugarcane mapping, one disadvantage was their heavy reliance on optical EO datasets such as the Landsat and Sentinel-2 series, with Radar only being used for the detection of flooded rice fields. A problem with optical data is its inability to observe land cover and land use change where cloud cover is present (Meraner et al., 2020). Cloud cover can therefore cause missing data and gaps within the time-series. This makes it more difficult to map sugarcane using the approach outlined by Wang et al. (2020) in regions where regular occurrences of substantial amounts of cloud cover are common (Meraner et al., 2020). Because of the difficulties in applying this approach within cloudy regions, the reach of this research is limited as on average 55% of the Earth's terrestrial surfaces experience heavy cloud cover, at various times of the year (Meraner et al., 2020).

One way to overcome the cloud is to use RADAR datasets like Sentinel 1A. Unlike optical datasets, the Sentinel-1 mission provides land use and land cover change information even with cloud cover (Meraner et al., 2020). Like with optical data, the findings of (Olfindo et al., 2020) demonstrate the sensitivity of Sentinel-1 A co-polarized (VV) and cross-polarized (VH) backscatter time-series profiles to characteristics and features unique to sugarcane crops. Using this information, Olfindo et al. (2020) was able to use dynamic time warping (DTW) and a collection of known sugarcane plantations to distinguish other sugarcane plantations from spectrally similar crop types based on the shape of and patterns displayed in their VV and VH profiles. To achieve this, they first pre-processed the datasets. After pre-processing Sentinel 1A VH and VV profiles for known sugarcane fields with similar plantation and harvesting dates were produced. The generated profiles were then shortened to represent the time after planting and before harvesting (one growth cycle). Once the profiles had been shortened to one growth cycle, the shape of these profiles was used in conjunction with a DTW algorithm to detect other sugarcane plantations, with an accuracy greater than 92%.

Both methods outlined by Wang et al. (2020) and Olfindo et al. (2020) are new concepts and thus have only been tested in their respective 'test study sites'. One research question is whether both methods can be applied to different areas of the world. Another question is whether the phenological characteristics of sugarcane vary with location. Both Wang et al. (2020) and Olfindo et al. (2020) extracted phenological information from time-series profiles of known sugarcane fields. A third research question could be whether prior knowledge of the location of sugarcane fields is needed to identify sugarcane elsewhere or whether the shape of the profiles Olfindo et al. (2020) and derived phenology Wang et al. (2020) will be sufficient for sugarcane identification. Both sugarcane mapping methods proposed by Wang et al. (2020) and Olfindo et al. (2020) have advantages and disadvantages. It was interesting to see that Olfindo et al. (2020) used field-level knowledge of planting and harvesting dates to isolate the time-series of one sugarcane growing season while Wang et al. (2020) used crop calendar information to extract time-series profiles that represented one growth cycle. This could be used as an argument as to why the methods by Wang et al. (2020) may be more useful for the detection of sugarcane in regions where planting and harvesting information is not easily accessible. However, in India, the management of sugarcane plantations is heterogenous, with each field having a different planting and harvesting time, therefore the approach by Wang et al. (2020) would also be limited in this region. Future research should consider methods for overcoming the reliance on planting and harvesting dates for isolating individual sugarcane growing cycles and combining optical and SAR EO datasets for a more robust sugarcane mapping approach.

2.6.3 Land cover change and biodiversity

Adami et al. (2012), used time-series from MODIS and Landsat with a web tool to calculate the amount of direct land use change (dLUC) from other land use types to sugarcane in the South-central region of Brazil between 2000 and 2009. Using the web tool and 1,046 reference points representing known classes, Adami et al. (2012) classified each component of the Landsat and MODIS images

as either annual crop, citrus, forest, pasture, and signature, for each year. This was then used to determine the percentage change between each year, with the result that the conversion of land from pasture and other crops to sugarcane accounted for >99% of land cover change occurring in South-central Brazil between 2000 and 2009 (Adami et al., 2012). In turn, this approach was used to predict the effect of sugarcane expansion on deforestation and biodiversity. Similarly, it may be possible to determine the effect of changes in sugarcane area on the demand for irrigation water and possible changes in water availability. The use of EO datasets could allow for this analysis to be replicated over different times and sugarcane-growing regions.

2.7 Opportunities and Challenges

The above literature demonstrates how EO time-series information can be used for biodiversity monitoring, yield prediction and classifying sugarcane. By observing the continuous relationship between sugarcane growth stages, yield and socio-economic factors like, weather patterns, water availability, soil nutrient stress, management strategies, pests, and disease, the detection of any anomalies which might indicate stress becomes easier (Mulianga et al., 2013). The identification of early stress warning signs, and the impact of practices like monocropping also becomes easier, allowing strategies to be put in place to reduce damage to yield and sugarcane health (Bégué et al., 2010).

Sugarcane mapping over large areas is useful for the assessment and improved sustainability of sugarcane production. Many single-date mapping supervised and unsupervised parametric classifiers and machine learning algorithms have been used to map sugarcane. However, because these approaches rely on a point measurement, they can have difficulty separating sugarcane from spectrally similar classes. These constraints can be overcome by using EO time-series data.

The above literature review suggests that there is no standardised time-series approach for the mapping of sugarcane making comparisons between regions difficult. The studies by Wang et al. (2020) and Olfindo et al. (2020) are the first

studies to successfully map sugarcane in China and the Philippines respectively using a combination of spectral time-series data and phenological knowledge. The application of these methods for other sugarcane regions and global sugarcane mapping has therefore not yet been explored, creating a knowledge and methodological gap within the current literature. Mapping is one part of assessing sustainable sugarcane production, variables such as the impact of deforestation and biodiversity, water, fertiliser, and pesticide use, monocropping, sugarcane health and yield, cost of production and profits are also important to assess. Despite this, there is to date not one combined method for sugarcane classification mapping and the assessment of sustainable sugarcane production. It must also be noted that not all the elements of sustainability outlined in Table 2.1 can be monitored using EO imagery, leaving knowledge gaps within the study of EO for sustainable sugarcane production. Nevertheless, using EO to monitor some elements of sustainability will help provide valuable information to aid the sustainable production of sugarcane. These elements include soil health, water use, fertiliser pollution risk and some aspects of biodiversity.

Hence there are several research challenges relating to the use of EO data to improve the sustainability of sugarcane production. One challenge is to determine whether sugarcane can be monitored from long-term satellite programmes and new satellite constellations in near real-time, and how such data combined with machine learning tools could improve the social, economic and environmental sustainability of sugarcane production systems.

Although having a centralised system for monitoring sugarcane production would be beneficial to sugarcane mills, it is important to also be mindful of ethical implications. One unintentional negative consequence of a centralised system for monitoring sugarcane production would be that individual farmers would not be able to carry out local deals in turn making the farmers feel as though they are being spied on.

3 AUTOMATING THE DERIVATION OF SUGARCANE GROWTH STAGES FROM OPTICAL EARTH OBSERVATION TIME-SERIES

This chapter is published as Joshi, Neha, Daniel M. Simms, and Paul J. Burgess. 2024. "Automating the Derivation of Sugarcane Growth Stages from Earth Observation Time-series" Remote Sensing 16, no. 22: 4244. <https://doi.org/10.3390/rs16224244>

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>

3.1 Introduction

Sugarcane (*Saccharum officinarum*) is used in the production of 68% of global sugar and is also used for bioethanol (FAO, 2021). 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 (Hess et al., 2013). 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 (Inman-Bamber et al., 2005), 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 (Solomon, 2016). Sugarcane production also employs millions of farmers alongside transport and tertiary roles associated with the industry (Mall et al., 2016; Nandhini and Padmavathy 2017). 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 (Mnisi, 2012). As a result, the production of vast quantities of sugarcane may be unsustainable through widespread use of monocultures and inappropriate pesticide and fertiliser use (Singh and Tiwari, 2018), which negatively impact productivity and degrade soil and water resources to the point where short-term restoration becomes difficult (Gujja et al., 2010). Many businesses seek to be economically, environmentally, and socially sustainable, where profitable sugar production is combined with desirable employment and minimal negative environmental effects (Inman-Bamber et al., 2005; Nandhini and Padmavathy 2017). 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 (Li, 2022), as resource requirements vary with sugarcane growth stage (Mall et al., 2016) 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 (Singh and Tiwari, 2018). 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 (Olfindo et al., 2020). These sample data will often not fully represent the heterogeneous nature of the sugarcane industry, and the resulting information quickly becomes outdated (Wang et al., 2020).

An alternative approach to field surveys is to combine EO data from satellites with biophysical crop models (Kussul et al., 2019). Satellite sensors allow for large areas to be covered quickly and changes in biomass to be studied in near real-time (Fu et al., 2019; Sudmanns et al., 2019). 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 (Fu et al., 2019; Stubbings et al., 2019).

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 (Wang et al., 2020; Dimov et al., 2022; Zheng et al., 2022; Yeasin et al., 2022). 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 (Yeasin et al., 2022; Meraner et al., 2020). 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 (Chen et al., 2004). This is problematic for the development of algorithms, as these parameters vary spatially and temporally, affecting the interpretation of EO time-series (Wilson et al., 2014).

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 (Sakamoto et al., 2010; Sakamoto, 2018; Zhou et al., 2020; Liu et al., 2022). 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 (Sakamoto et al., 2010; Zhou et al., 2020), or, where this information is not available, state-level (Sakamoto, 2018) and national-level statistics (Liu et al., 2022) 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 (Wang et al., 2020; Zheng et al., 2022). 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 (Kooistra et al., 2024). 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.

3.2 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 (Virtanen et al., 2020) (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 (Hunter, 2007).

3.2.1 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 3.1). 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.

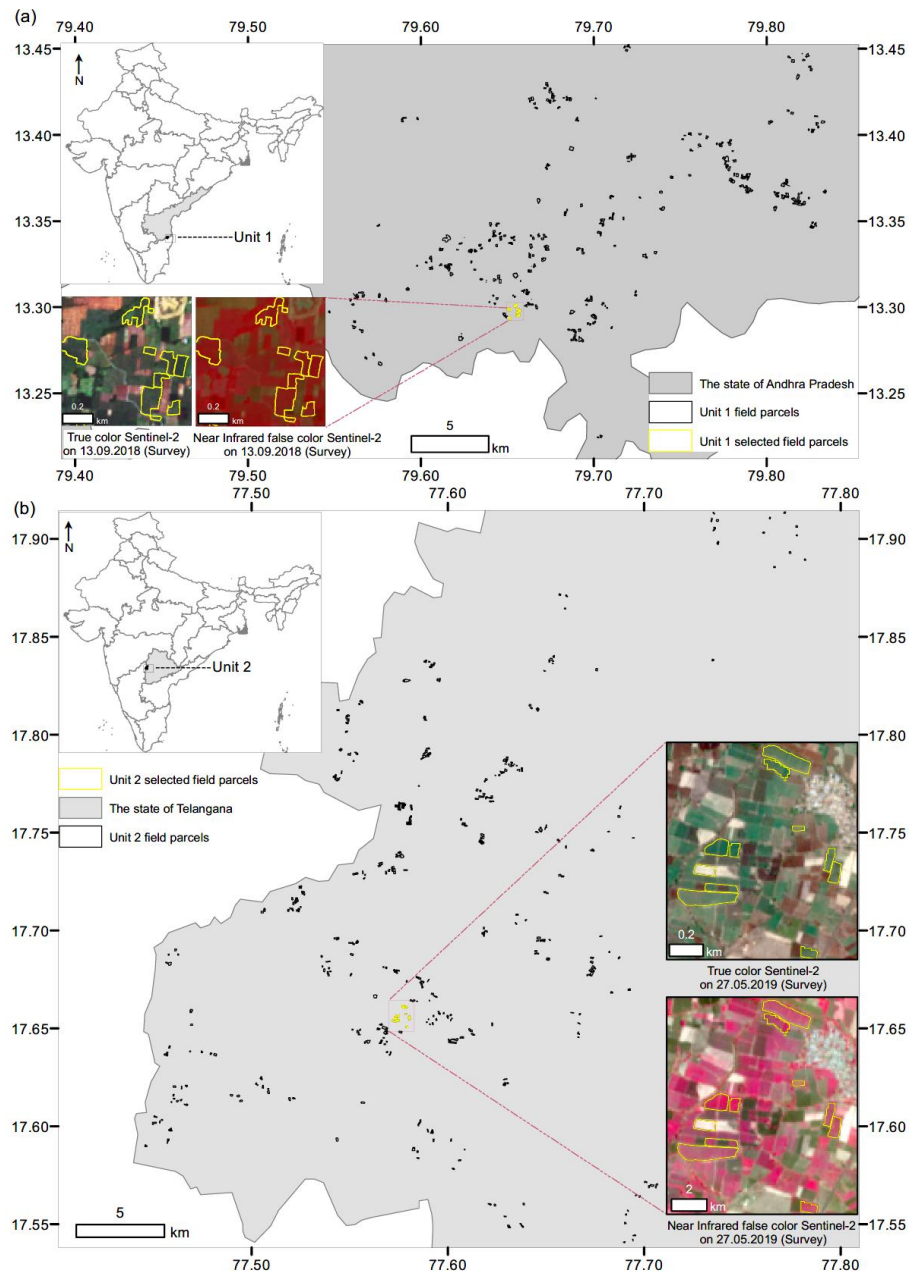


Figure 3.1. Sugarcane fields associated with processing Units 1 and 2 located in southern India. Sugarcane area Unit 1 is situated in (a) Andhra Pradesh, and (b) Unit 2 is in Telangana.

3.2.2 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 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 (Gorelick et al., 2016). 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 (TOA), and Sentinel-2a and Sentinel-2b 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 3.1) (Table 3.1). 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). Landsat 8 with a resolution of 30 m (0.09 ha) and a Sentinel-2 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 3.1), and they were considered as potential EO datasets for deriving sugarcane phenology.

$$NDVI = (NIR - R) / (NIR + R)$$

Equation 3.1

Table 3.1 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. The difference between SR and TOA data is the effects of the atmosphere (Jin et al., 2019). 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 (Wang et al., 2020). Previous studies have found that for the interpretation of temporal data the difference between TOA and SR data is mitigated when vegetation indices, like NDVI, are used. While absolute values for TOA and SR reflectance values were significantly different, the interpretation of the vegetation index profiles calculated with SR and TOA datasets were similar (Jin et al., 2019).

3.2.3 Selecting a Sugarcane Growth and Development Model

The development of sugarcane can be described as four stages (Mall et al. 2016; Doorenbos 1977; Alface et al. 2019) using a trapezoidal model to describe biomass change over time through changes in NDVI or crop evapotranspiration (Figure 3.2). 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% (Mall et al., 2016), typically lasting about 70–75 days (Mall et al., 2016; Doorenbos, 1977). 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 (Mall et al., 2016; Doorenbos, 1977). 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 (Mall et al., 2016; Doorenbos, 1977).

After harvest, a second ratoon crop can grow from the base of the harvested cane (Xu et al., 2021). Sugarcane can grow up to eight ratoon crops from initial planting, but with every ratoon, there is a tendency for reduction in yield (Singh et al., 2014). 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 (Xu et al., 2021). 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 (Doorenbos, 1977).

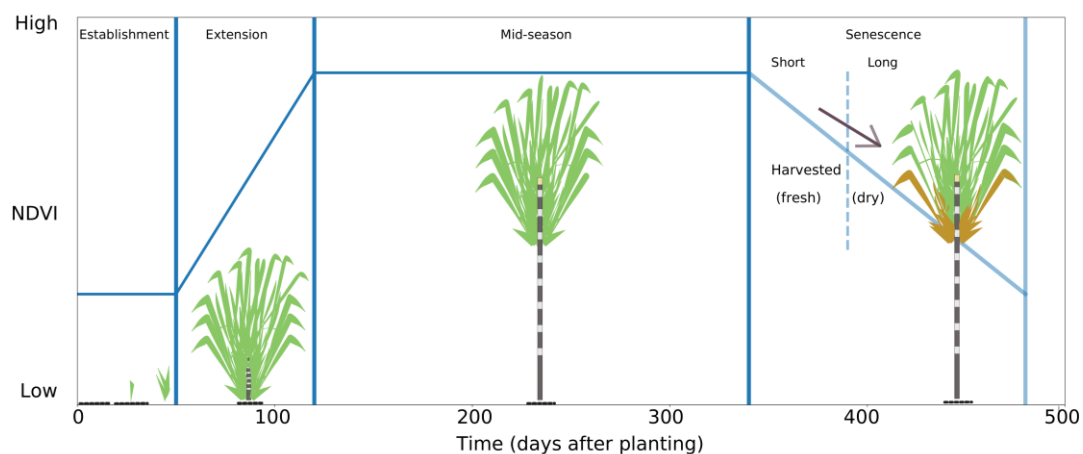


Figure 3.2. A schematic diagram depicting the four stages of sugarcane growth in India based on the FAO model for sugarcane evapotranspiration (Doorenbos 1977).

An understanding of growth stages can help farmers understand how management can minimise the negative impacts of climate change and maintain yields, by combining this information with weather and yield datasets (Goswami et al., 2006; Mahato, 2014; Mall et al., 2016). The simplified trapezoidal model depicted in Figure 3.2 captures key growth stages important for sugarcane management, therefore is useful for providing an insight into effective management strategies under different climate scenarios. Optimal conditions in mid-season include high rainfall (>1000 mm), high temperatures (~30°C), and high humidity levels (~80%), while dry weather is needed in the senescence stage. In-fact, high water levels during senescence can reduce sucrose concentrations and thereby sugar yields (Fageria et al., 2010; Mall et al., 2016). Poorly managed sugarcane can also increase the crop's vulnerability to pests and disease, in turn affecting economic, environmental and social sustainability (Singh, 2018) (Figure 2.1 and Figure 3.2). With the knowledge of the timing of the mid-season and senescence stage, the informed farmer can modify planting, harvesting, fertiliser and pesticide application, and irrigation type (Nandhini and Padmavathy, 2017).

3.2.4 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 3.3), which is detailed in the following sections. A small number of fields with an area less than a Landsat 8 image pixel (900 m²) 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 Landsat 8 TOA and Sentinel-2 TOA time-series for each field was downloaded and processed using the GEE python API.

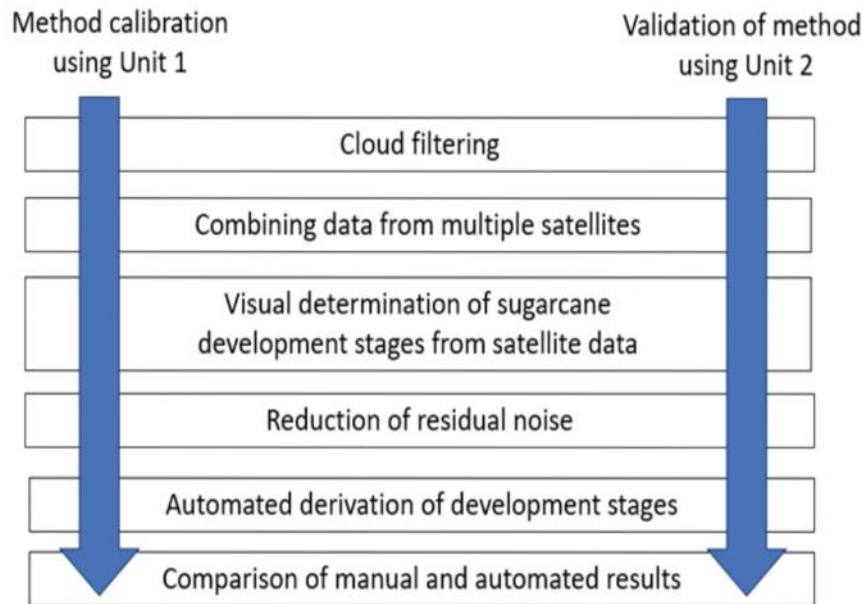


Figure 3.3. The methodology was established using data from Unit 1 and then evaluated using data from Unit 2.

3.2.5 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 3.4). 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 Landsat 8 TOA and Sentinel-2 TOA datasets, respectively (Method 2). With the Sentinel-2 TOA data, Method 1 was able to remove many of the low NDVI values (grey lines in Figure 3.4a), but some erroneous low values remained (drop in NDVI during the mid-season of the red time-series in Figure 3.4a). 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 3.4), following the trapezoidal model by Doorenbos (1977) depicted in Figure 3.2.

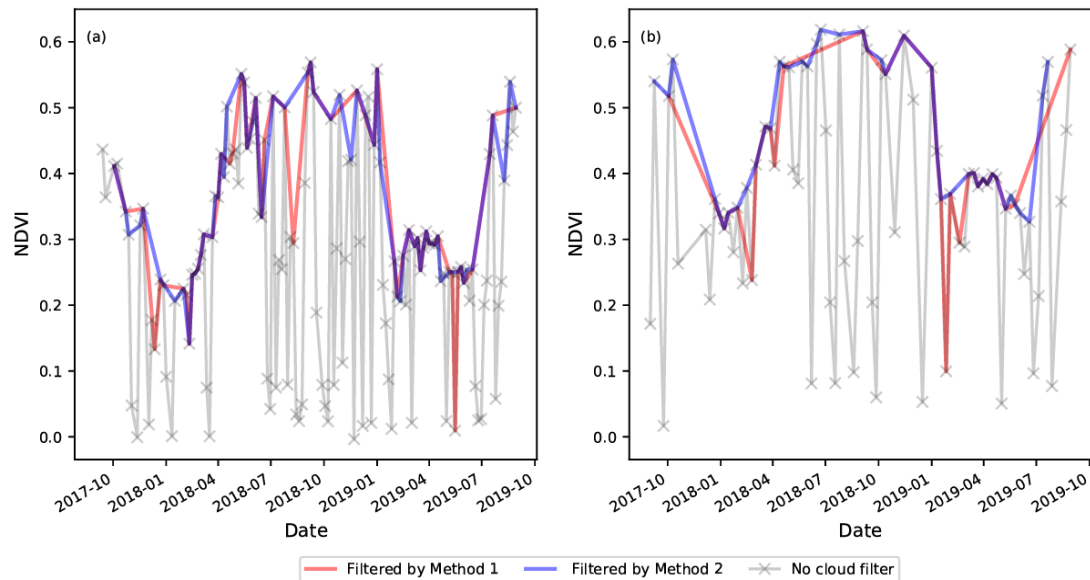


Figure 3.4. Before cloud filtering, cloud cover caused many low values of the Normalised Difference Vegetation Index (NDVI) (grey lines) from the Sentinel-2 (left in (a)) and Landsat 8 (right in (b)) data for sugarcane fields from September 2017 to October 2019. Using Method 1 to filter the clouds removed many of the low NDVI values (red lines), but more low values were removed using Method 2 (blue lines).

3.2.6 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 Landsat 8 TOA and Sentinel-2 TOA data. By using both the Sentinel-2 TOA and Landsat 8 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 Sentinel-2 TOA or only 10–20 observations (one observation every 18–36 days) if only using Landsat 8 TOA (Figure 3.5).

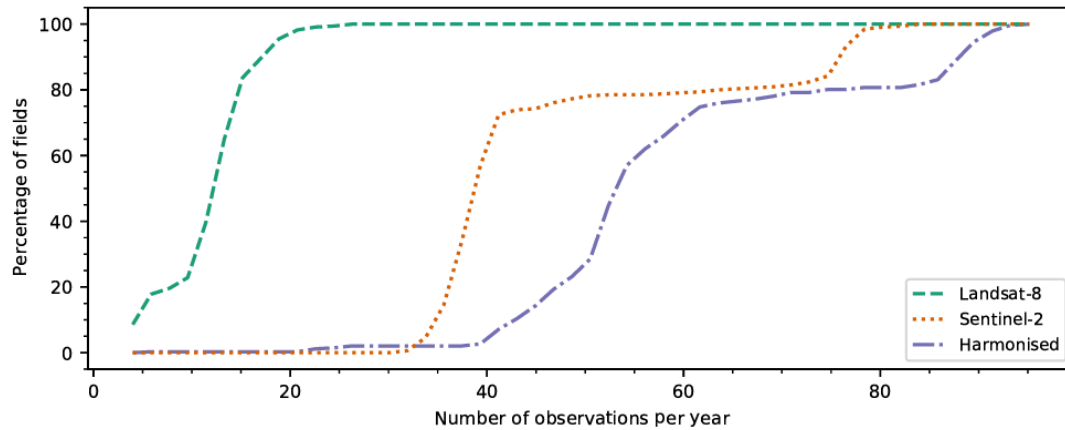


Figure 3.5. A frequency diagram presenting the average number of cloud-free observations received by each field, from three datasets, in one year (1 January 2018–1 January 2019). The number of observations received by the Landsat 8 TOA dataset is presented by the green line, and the number of observations received by the Sentinel-2 dataset TOA is presented by the orange line. The purple line represents the harmonised (Landsat 8/Sentinel-2 TOA) dataset, which is the Landsat 8 and Sentinel-2 dataset combined.

One known issue with combining NDVI data from two satellites is the systemic measurement differences between sensors (Zhang et al., 2018). A correlation of paired NDVI measurements which were collected at the same location and on the same day from Landsat 8 TOA and Sentinel-2 TOA showed a positive bias in the Landsat 8 data compared with Sentinel-2 of 0.16 at low values of NDVI, decreasing as NDVI increases (Figure 3.6). These data were used in a linear line calibration to correct the Landsat 8 values to the Sentinel-2 (Equation 3.2), 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).

$$\text{NDVI}_{\text{Landsat 8}} = 0.78 \text{NDVI}_{\text{Sentinel-2}} + 0.16 \quad \text{Equation 3.2}$$

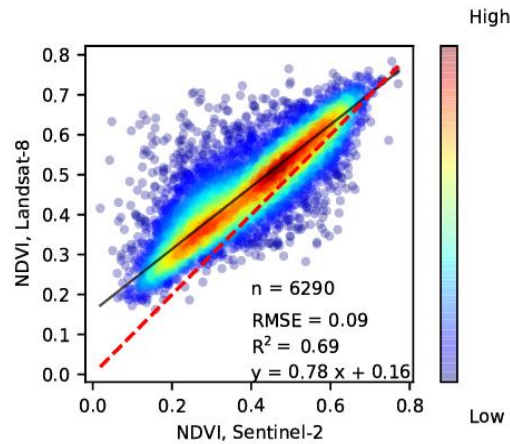


Figure 3.6. Relationship between the Landsat 8TOA NDVI observations and the Sentinel-2 TOA observations for Unit 1. The solid black line shows ordinary least-squares (OLS) regression of the Landsat 8 data against the Sentinel-2 data. Only NDVI values in the range of 0.0 to 1.0 are illustrated. The hashed red line shows a reference 1:1 line, and the colour bar represents the density of observations.

3.2.7 Visual Determination of Sugarcane Phenology

Using the harmonised Landsat 8 and Sentinel-2 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 3.7) using the criteria described in (Table 3.2). 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 (Zhou et al., 2024). Therefore, the SOS in this study refers to the first point of the sugarcane growing season detected from EO imagery (Figure 3.7).

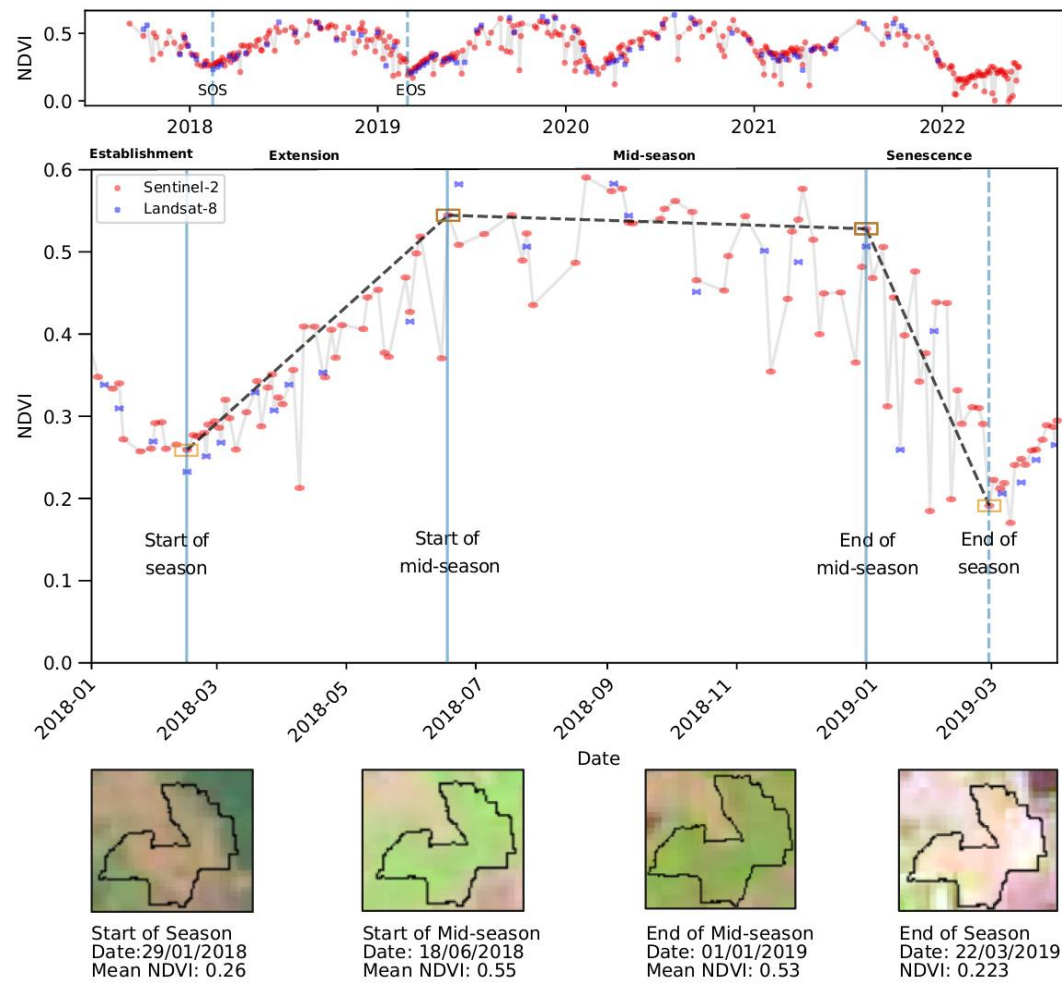


Figure 3.7. An illustration of how field reference data of sugarcane growth stages were manually interpreted using an example sugarcane field in Unit 1 for the 2018 season.

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 idealised 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 3.2. 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 start of a new growth cycle)

3.2.8 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 (Viovy et al., 1992) 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 (Xu et al., 2017).

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 3.5). Using the manually interpreted growth cycles for each field in Unit 1 for the 2018/19 (Table 3.2 and Figure 3.7) growing season, the value for noise (N) was derived from the trough count (TC) minus 1 (Equation 3.3). This is because the trapezoid model (Doorenbos, 1977) presents one trough at the end of the season, so any additional troughs would be the result of noise (See Figure 3.7).

$$N = TC - 1 \quad \text{Equation 3.3}$$

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 3.4).

$$NR = TC_{sp} - TC_{rd} \quad \text{Equation 3.4}$$

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 Landsat 8 and Sentinel-2 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 3.8); there was minimal effect on increasing the window length to more than 40 days (Figure 3.8). 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 3.8). A BISE sliding window period of 30 days was selected.

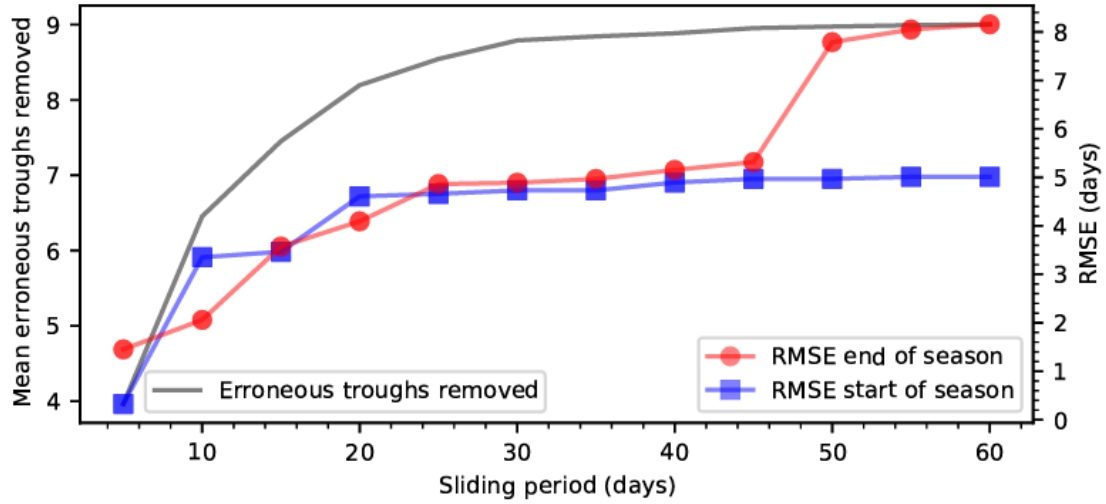


Figure 3.8. The effect of changing the best index slope extraction (BISE) sliding period on the number of additional erroneous troughs removed per field per growth cycle (left-hand scale) and the root mean square error (RMSE) (right-hand scale) in the calculation of the start and end of the sugarcane growth cycle.

The 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) (Zhou et al., 2016), 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 (Li et al., 2021) transforms a time-varying signal $f(t)$ into its magnitude and phase components $F(u)$ in the frequency domain (Equation 3.5), 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.

$$F(u) = \sum_{t=0}^{N-1} f(t) e^{-i\frac{2\pi}{N}ut} / N \quad \text{Equation 3.5}$$

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 3.9). The FFT (over) smoothed profile was retained for further processing to extract the start and end of the mid-season (see Section 3.2.9), then discarded.

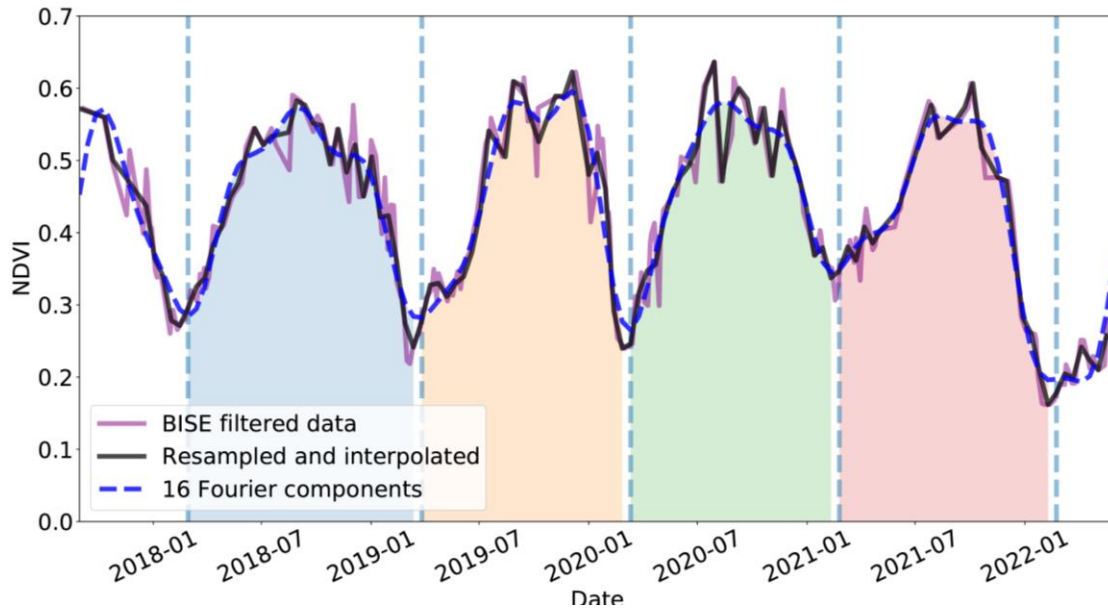


Figure 3.9. Time-series NDVI decomposed into individual seasons using the low-frequency components of a fast Fourier transform for a sugarcane field with four ratoons.

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.

3.2.9 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 3.2.8) 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:

$\lambda p1$ ranging from $\lambda 1$ to $\lambda_{max}+2$, where λ_{max} is the maximum NDVI, and $\lambda p2$ ranging from λ_{mid} to λn .

To calculate the SMS, $\lambda p1$ was first normalised (x and y values ranging from 0 to 1), $\lambda p1_{norm}$, and a knee algorithm, based on the work of Satopaa et al. (2011), was used to locate the index of the value on $\lambda p1_{norm}$ with the largest perpendicular Euclidean distance from the line extended from $\lambda p1_{norm}1$ to $\lambda p1_{norm}n$ (Figure 3.10a,b), which is the index of the point SMS on the 14-day resampled profiles (Figure 3.10e). To calculate the EMS, the order of the values of $\lambda p2$ was reversed and then normalised $\lambda p2_{invnorm}$. The same approach was then used to locate the index of the value of largest perpendicular Euclidean distance from the line $\lambda p2_{invnorm}1$ to $\lambda p2_{invnorm}n$ (Figure 3.10d), which is the reverse index of the second knee on the 14-day resampled profiles (Figure 3.10d).

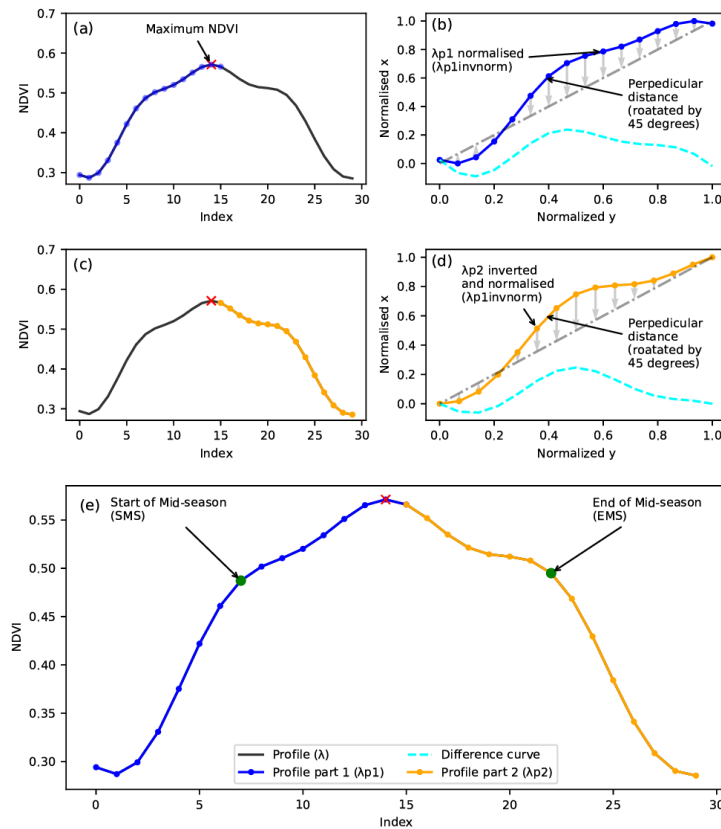


Figure 3.10. A schematic diagram illustrating how the start of the mid-season (SMS) and the end of the mid-season (EMS) were derived.

3.3 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.

3.3.1 Start and End of Season

Manual and automatically derived SOS ranged from 65 to 347 days (Figure 3.11). An initial comparison of the two methods showed an R^2 value of 66% and a RMSE of 26 days (Table 3.3). The relatively low R^2 and high range was related to three outliers with predicted SOS 150 days greater than the manual labelling (Figure 3.11). 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. Un-removed cloud was particularly a problem for smaller fields (<90 ha) as the proportion of field boundary covered by cloud during the mid-season was much higher than larger fields. The three small outlier fields had big impact on the overall R^2 values which increased to 72% from 66%, after the removal of these three points (Table 3.3). RMSE on the other hand, represents the square root of the average of squared differences between the predicted and manually interpreted SOS at a field level, therefore the removal of the three outlier points increased the value for RMSE slightly from 26 days to 27 days (Table 3.3). Providing RMSE values at field level accounts for more variation therefore an RMSE of +/- 27 days may look high, however if the RMSE was provided at larger scales i.e. at the site level there would be a lot less variation and the value for RMSE would be lower.

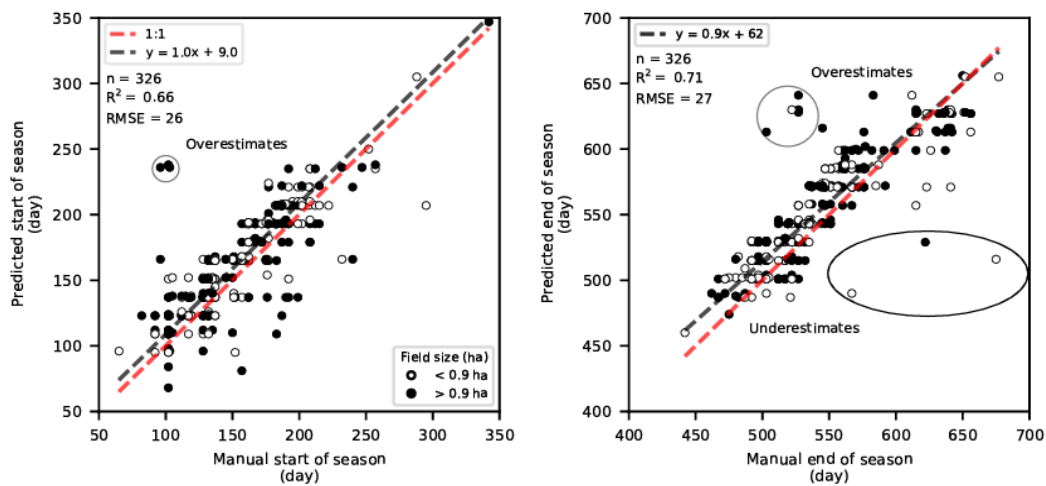


Figure 3.11. Unit 1 sugarcane fields used for algorithm development showing the relationship between automated and manually derived (left) start of the season and (right) end of season, using 14-day resampled time-series, where day is the number of days after 1 September 2017. Equations show the linear relationship before the removal of outliers.

Table 3.3 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	$SOS_p = 1.00 SOS_m + 9$	0.66	26
	323*	$SOS_p = 0.99 SOS_m + 9$	0.72	27
End of season	326	$EOS_p = 0.90 EOS_m + 62$	0.71	27
	305*	$EOS_p = 0.90 EOS_m + 64$	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 3.3). Analysis of Figure 3.11 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 3.3).

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 3.4). There was a cluster of outliers showing overestimation in SOS (Figure 3.12) 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 3.12). The EOS estimates ranged from 422 to 722 days (Figure 3.12) with an R^2 value of 81% and an RMSE of 17 days, with no outliers.

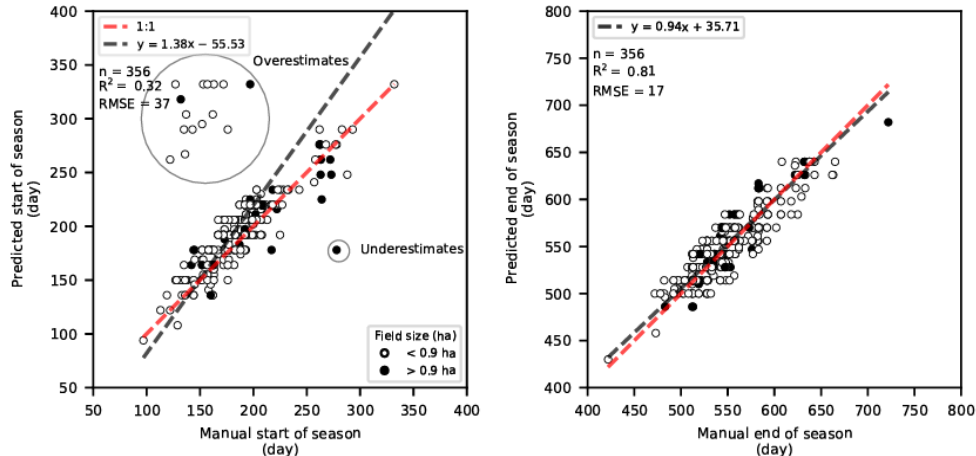


Figure 3.12. Unit 2 sugarcane fields used for validation showing the relationship between the automated and manually derived (left) start of the season and (right) end of season using 14-day resampled time-series, where day is the number of days after 1 September 2017. Results are not affected by field size. Equations show the linear relationship before the removal of outliers.

Table 3.4. 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	Field count	Equation	R ²	RMSE, days
Start of season	356	$\text{SOS}_p = 1.38\text{SOS}_m - 56$	0.32	37
	344*	$\text{SOS}_p = 1.08\text{SOS}_m + 18$	0.78	23
End of season	356	$\text{EOS}_p = 0.94\text{EOS}_m + 36$	0.81	17

3.3.2 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 3.13). The corresponding days for the EMS ranged from 334 to 599 days (Figure 3.13). Prior to the removal of outliers, the R² 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 3.13). 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 3.13). 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 3.5).

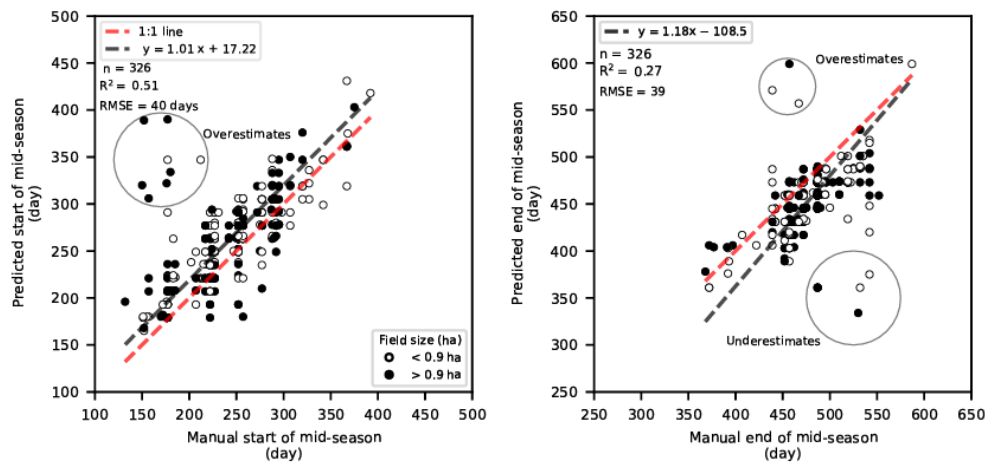


Figure 3.13. Sugarcane fields from Unit 1 used for calibration, showing the relationship between automated and manually derived (left) start of mid-season and (right) end of mid-season, using 14-day resampled time-series, where day is the number of days after 1 September 2017. The equations show the linear relationship after the removal of outliers.

Table 3.5. 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	Equation	R ²	RMSE, days
Start of mid-season	326	SMS _p = 1.01 SMS _m + 17	0.51	40
	318*	SMS _p = 0.99 SMS _m + 17	0.72	29
End of mid-season	326	EMS _p = 1.18 EMS _m - 109	0.27	39
	316*	EMS _p = 0.90 EMS _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 3.14). The corresponding days for the end of the mid-season ranged from 361 to 612 days (Figure 3.14), with the initial analysis showing R² values for SMS and EMS of 65% and 28%, respectively, and an RMSE of 30 days (Table 3.6). 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.

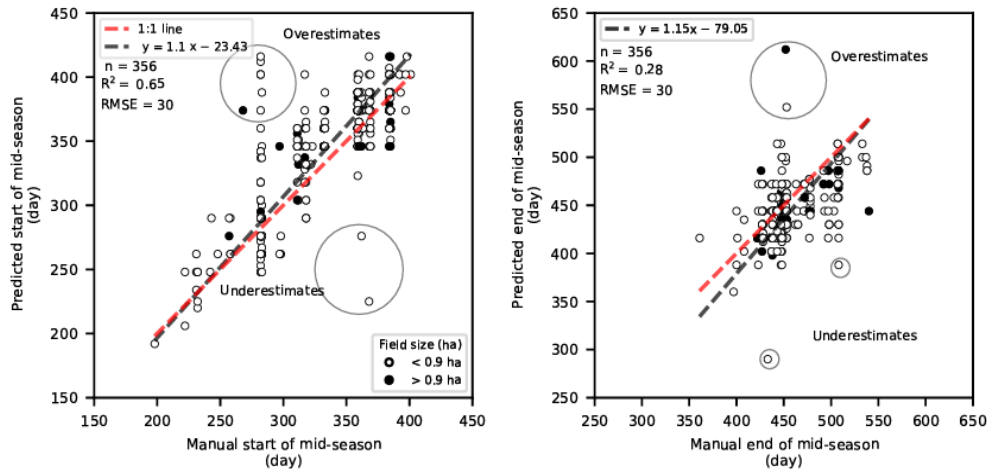


Figure 3.14. Sugarcane fields from Unit 2 used for validation, showing the relationship between automated and manually derived (left) start of mid-season and (right) end of mid-season, using 14-day resampled time-series, where day is the number of days after 1 September 2018. The equations show the linear relationship after the removal of outliers.

Table 3.6. Relationship between the predicted (SMSp) and manually derived (SMSm) date of the start of the mid-season and the predicted (EMSp) and manually derived (SMSm) date of the end of the mid-season for fields in Unit 2, including the effect of removing outliers.

Relationship	Number of data	Equation	R^2	RMSE, days
Start of mid-season	356	$SMSp = 1.10SMSm - 23$	0.65	30
	344*	$SMSp = 1.08SMSm - 17$	0.79	23
End of mid-season	356	$EMSp = 1.15EMSm - 79$	0.28	30
	352*	$EMSp = 0.93EMSm + 20$	0.36	27

3.3.3 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 (Mall et al., 2016) for sugarcane grown in Uttar Pradesh, India and the FAO trapezoid model for sugarcane grown in tropical climates. 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 (Mall et al., 2016) for sugarcane grown in Uttar Pradesh, India and the FAO trapezoid model for sugarcane grown in tropical climates (Doorenbos, 1977) (Table 3.7).

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 3.7). 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 3.7). 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 3.7). The length of the mid-season and senescence stages were similar to the values reported by Mall et al. (2016) for Uttar Pradesh, India and Doorenbos (1977) for the Tropics. The length of the reported manual and automated estimates of the extension stage for both Unit 1 and Unit 2 were much longer than those reported by Mall et al. (2016) and Doorenbos (1977). 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.

Table 3.7. 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. (2016) and Doorenbos (1977). 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. (2016)	45	120	130	110	360
FAO(V)	50	120	220	140	480
FAO(R)	30	80	180	60	320

3.4 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 (Wang et al., 2020; Dimov et al., 2022; Zheng et al., 2022; Yeasin et al., 2022), 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.

3.4.1 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 Sentinel-2 TOA and

Landsat 8 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 3.4).

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 Landsat 8 TOA and Sentinel-2 TOA NDVI values by developing a local regression calibration, which was similar to the regression reported by others (Zhang et al., 2018). The harmonisation worked well, increasing the frequency of cloud-free observations from every 18–36 and 4–9 days, respectively, for Landsat 8 TOA and Sentinel-2 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 3.14). 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 337 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 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 (Olfindo et al., 2020; Vavlas et al., 2020; Bao et al., 2023; Besten et al., 2023). 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.

3.4.2 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 (Wang et al., 2020; Dimov et al., 2022; Yeasin et al., 2022; Zhu et al., 2022). 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 (Zhu et al., 2022), 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. By comparing NDVI timeseries profiles with ground reference photos, 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) (Appendices 9.2 and

9.3). This is important for sustainable sugarcane production because the mid-season stage is particularly sensitive to drought stress with high water demands. Drought stress is less critical during the senescence stage (Doorenbos, 1977). The onset of senescence marks a period when dry weather is advantageous as excess water can reduce sucrose concentrations and thereby sugar yields (Fageria et al., 2010; Mall et al., 2016). Identifying the mid-season and senescence stages is therefore directly linked to the sustainable development goal indicator 2.1.4, specifically the sub-indicator linked to monitoring variation in water quality (Table 2.1).

The end of the mid-season (on-set of senescence) could also be derived, and the derived lengths for this period from Unit 1 Unit 2 sugarcane fields were comparable to previous studies (Doorenbos, 1977; Mall et al., 2016). Monitoring the length of growing stages can also be useful for highlighting possible regional differences in optimal field management. For example, the longer length of the mid-season in Unit-1 sugarcane fields, compared to Unit 2, could indicate that fields should be harvested later to maximise yields. By contrast Unit 2 in the Indian state of Telangana faced extreme drought from 2016-2018, and this could have resulted in a shorter mid-season. In this case, planning earlier harvesting of the cane would be advantageous to reduce the effect of the drought. Using the technology developed in Chapter 3, it is now possible for additional research to look at the interactions between climate and the length of growth seasons, which could provide information for future proofing the industry against global environmental changes. The ability to examine the cropping patterns across multiple seasons and regions can help to determine if management factors, such as irrigation, influence seasonal dynamics. Irrigation is important for monitoring the use of available water resources, for environmental sustainability.

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 (Misra et al., 2022). 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 (Mall et al., 2016; Doorenbos, 1977) 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 (Marin et al., 2021) and is therefore not a sustainable use of resources. Yield is directly linked to SDG indicator 2.4.1, particularly the theme of land productivity, profitability and decent employment (Table 2.1).

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 for enhancing the resilience theme of SDG indicator 2.4.1 (Table 2.1). 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.

3.4.3 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 (Olfindo et al., 2020). 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 EOS resulting from the point of harvest coinciding with a cloud event in turn resulted in the overestimation of the EMS 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 (Morel et al., 2014). 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 (Wang et al., 2020; Birdi and Kale, 2017; Cruz-Sanabria et al., 2020). 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.

3.5 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 (See Table 2.1).

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.

4 DERIVING SUGARCANE GROWTH STAGES FROM SYNTHETIC APERTURE RADAR

4.1 Introduction

Field and factory managers and government officials can use information on the development, growth and harvest of sugarcane to improve the economic, social, and environmental impact of sugarcane production. One way to derive information of crop development is to use remote optical sensors. These have been used to detect changes in levels of chlorophyll and the normalised difference vegetation index (NDVI), which is associated with the level of green biomass. The value of NDVI is calculated using measured Red and Near Infrared (NIR) reflectance values, recorded by optical sensors.

Measures of NDVI are available from medium-resolution sensors such as Landsat and Sentinel-2, and through platforms such as the GEE. A new method for automatically deriving the stages of the FAO trapezoid model for sugarcane growth (Figure 4.1) was developed in Chapter 3, using reflectance data from the Landsat 8 and Sentinel-2 satellites. The method first optimised the balance between the removal of cloud cover and noise with maintaining a high frequency of data. It then automatically decomposed the long time-series profiles into individual sugarcane growing seasons, negating the need for prior knowledge of sugarcane planting and harvesting dates. Once split into individual growing seasons and after isolating the sugarcane growing cycles for two growing seasons in two areas (Unit 1 and Unit 2), a knee function was applied to derive the start and end of the mid-season stage. I observed that the NDVI level stabilises at the beginning of the mid-season (SMS), as the number of new tillers (branches forming from the nodes of internodes at the base of the main shoot) was reduced, and the focus of growth moved towards maturity and yield formation. By contrast, sugarcane maturity and yield formation are not recorded by NDVI as the amount of green biomass at the canopy level does not change during the mid-season. However, as sugarcane enters the senescence phase, the green leaf area and photosynthesis declines as the crop dries and sugars is converted to harvestable sucrose (Mall et al., 2016). To capture the decrease in

NDVI at the end of the mid-season (EMS), the methodology identified the point where the observed values of NDVI in the time-series gradually started to decrease after a period of stability (See Figure 4.1). It was possible to use this approach without prior knowledge of planting and harvesting dates for the splitting of the season. However, the initial screening for cloud cover still left erroneous measurements associated with cloud cover, which in turn led to errors when deriving the SMS and EMS, reducing the relationship between automated assessments of growth stages and manual assessments based on visual observation.

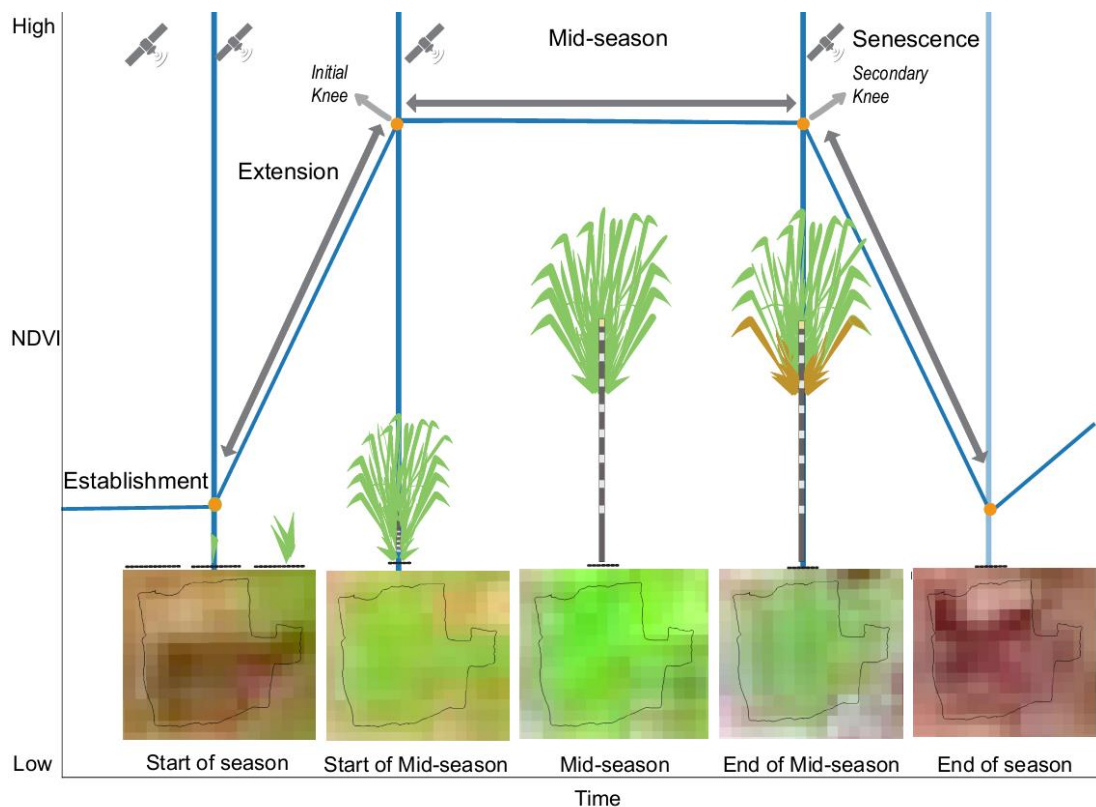


Figure 4.1 A schematic diagram of the sugarcane growth stages that can be automatically derived from processed NDVI time-series using methods developed in Chapter 3. Below the diagram are Sentinel-2 image composites for an example field, with the date of the image corresponding to the date of the derived growth stage.

An alternative to optical images is the use of data recorded by SAR instruments. SAR can also determine chlorophyll content and changes in the structure of vegetation as vegetation grows. SAR instruments are active instruments that transmit electromagnetic waves toward a target, and energy scattered back is recorded by the SAR instrument's receiver. Being active means that SAR instruments are not reliant on sunlight and the data provided are less affected by clouds as the microwave backscatter can penetrate clouds. SAR backscatter is captured at different wavelengths depending on the sensor which provides information about the structure of surface objects. Surface roughness and smoothness is dependent on how much vertical transmitted backscatter is captured by the instrument, which in turn affects how light or dark surfaces appear in SAR derived images. In the context of SAR, soil surfaces, are rougher than vegetation surfaces. Although the energy received by SAR instruments is scattered randomly at various intensities, the vertically transmitted backscatter recorded by SAR instruments from soil is often higher than what is received by vegetation surfaces. The primary scattering mechanism of vegetation is volume scattering with more volume backscatter being directed back at SAR instrument than vertically transmitted energy (Ottinger and Kuenzer, 2020). Many studies have found a sensitivity of SAR backscatter and vegetation indices derived from SAR to changes in vegetative structure and dielectric properties of crops as they grow, therefore SAR data is useful for crop yield and time-series classification studies as the data is not affected by clouds meaning an increased frequency of observations compared to optical datasets (Bao et al., 2023).

Data from Sentinel-1 (an active SAR instrument) is freely available on the GEE platform. Sentinel-1 collects C-band SAR imagery at a variety of different polarisations and resolutions. The data produced by Sentinel-1 is of medium spatial resolution (10 m) and temporal resolution (6-12 days depending on the location). Hence Sentinel-1 data can be used to overcome the limitation of cloud in optical instruments for monitoring vegetation growth over time (Bao et al., 2023; Besten et al., 2023; Olfindo et al., 2020; Vavlas et al., 2020). Sentinel-1 operates in dual polarisation, with one vertical transmitter (V) polarised and two parallel receiver vectors for horizontal (H) or vertical (V) polarisation over the land.

VV backscatter is vertical plane transmitted and received often referred to in literature as co-polarised backscatter. VV backscatter is sensitive to surface scattering (from rough surfaces such as soil) and volume scattering (from within vegetation). Bao et al. (2023) and Vavlas et al. (2020) report that VV backscatter can be associated with above-ground biomass during the initial and planting stages of cereal crops such as wheat, canola, soya, and maize with low and stable scattering during periods of minimal biomass in areas unaffected by weeds. This suggests that VV-backscatter could be used in the seasonal decomposition of long time-series datasets and overcome the limitations of cloud in optical instruments.

VH-backscatter is vertical plane transmitted and horizontal plane received (cross-polarized) and is primarily sensitive to volume scattering. Bao et al. (2023) and Vavlas et al. (2020) report that VH backscatter is more responsive to the increase in leaf biomass and plant density (observed during the extension phase of crops) than VV backscatter, with increased volume scattering and multiple reflection scattering from the smooth surface of crop leaves. As the vegetative growth of wheat and canola crops stabilised, with a thick vegetative canopy and many leaves complexly formed, VH backscatter typically reaches a maximum and remains relatively stable until the onset of senescence and the start of the harvesting season (Bao et al., 2023; Vavlas et al., 2020). During the harvesting season, backscatter typically starts to reduce with a decrease in vegetation cover and increased penetration of electromagnetic energy through crop canopy and exposure of the surface soil (Bao et al., 2023). This suggests that VH backscatter time-series data could be used to identify the mid-season stages and possibly the harvest of sugarcane.

An alternative approach is to monitor the phenology of crops using a combination of VH and VV backscatter. Bao et al. (2023) constructed vegetation indexes using a combination of VH and VV backscatter and Vavlas et al. (2020) used the VH/VV cross-ratio (CR) to offset the effect of double bounce scattering from the combined effect of the soil interaction and double bounce scattering off the stem of cereal crops. This could have potential for sugarcane monitoring. Olfindo et al. (2020) found that the Sentinel-1 A co-polarized (VV) and cross-polarized (VH)

backscatter time-series profiles of known sugarcane fields featured characteristics unique to the sugarcane crop itself and therefore used sugarcane time-series information to classify sugarcane successfully. Besten et al. (2023) also used Sentinel-1 backscatter time-series to understand how VH and VV backscatter respond to sugarcane yield variability and waterlogging.

Each of the above approaches to monitor crop development using SAR datasets depended on known planting and harvesting periods and crop calendars for isolating individual growing seasons within long time-series data. Although the timing of a sugarcane crop is consistent in some areas, the timing of sugarcane operations in India can be very variable and hence methods for monitoring growth over time cannot rely on assumed planting and harvesting dates. Moreover, the extensive areas of sugarcane can make surveys of planting and harvest data impractical and time-consuming.

This paper therefore investigates whether Sentinel-1 data could be used to overcome the challenge of cloud within seasonal decomposition methods for improved estimates of the SOS and EOS and subsequently derive sugarcane growing stages using the trapezoid method developed in Chapter 3. In the process, the relative benefits of the three measures of backscatter (VV, VH, or CR) were also examined. The differences between the Sentinel-1 backscatter data, which records structural differences, and the NDVI data derived from optical satellites, which is a proxy for the changes in green biomass, were also investigated. Once decomposed into seasons, this study also investigated the relationship between the Sentinel-1-derived knees and the start and end of the mid-season derived by the optical data. For the second knee for example, contrary to the low gradual decline in NDVI representing the onset of senescence, we investigated whether the last observation before harvest would present itself as the point before an abrupt decrease in volume scattering after vegetation was cleared, as suggested by Olfindo et al. (2020).

4.2 Materials and Methods

The methodology starts by describing the study areas and field data and the selection of appropriate backscatter datasets. It follows the same format as Chapter 3, whereby the methodology was developed on sugarcane fields located in Unit 1 and evaluated using sugarcane fields in Unit 2. Manual estimates of sugarcane development stages derived from visual interpretation of raw NDVI time-series were used to validate the automated process for seasonal decomposition using Sentinel-1 backscatter datasets and the derived Sentinel-1 knees. The derived Sentinel-1 knees were also compared alongside the start and end of mid-season stages derived from the processed optical data in Chapter 3 and their temporal differences were assessed. In addition to comparing manual and optical estimates of sugarcane growth stages, Sentinel-2 image composites, and photographs were used to validate the findings.

4.2.1 Study areas and field data

This study is a continuation of the research conducted in Chapter 3, therefore the same Indian sugarcane factories located in Chittoor District, Andhra Pradesh (Unit 1) and Zaheerabad in the Sangareddy district of Telangana (Unit 2) that were used in Chapter 3 were also used in this study (Figure 4.2). Both sites have a tropical climate with wet and dry seasons and field sizes range 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. After removing fields with boundary changes before and after harvest, 313 fields in Unit 1 and 355 in Unit 2 were studied for their respective survey year. Unit 1 was surveyed in 2018, and Unit 2 was surveyed in 2019. During these surveys field polygons were digitised by walking the field perimeter with a GPS device and observations of crop height, irrigation method (drip or flood irrigated), and ground photography were also taken. All sugarcane fields in Unit 1 and Unit 2 were irrigated, with 13% drip irrigated in Unit 1 compared to 87% in Unit 2, and the remaining flood irrigated. Sugarcane crops were also a mixture of virgin cane (fields planted from cuttings) from the season surveyed and ratoon cane, which regrows 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.

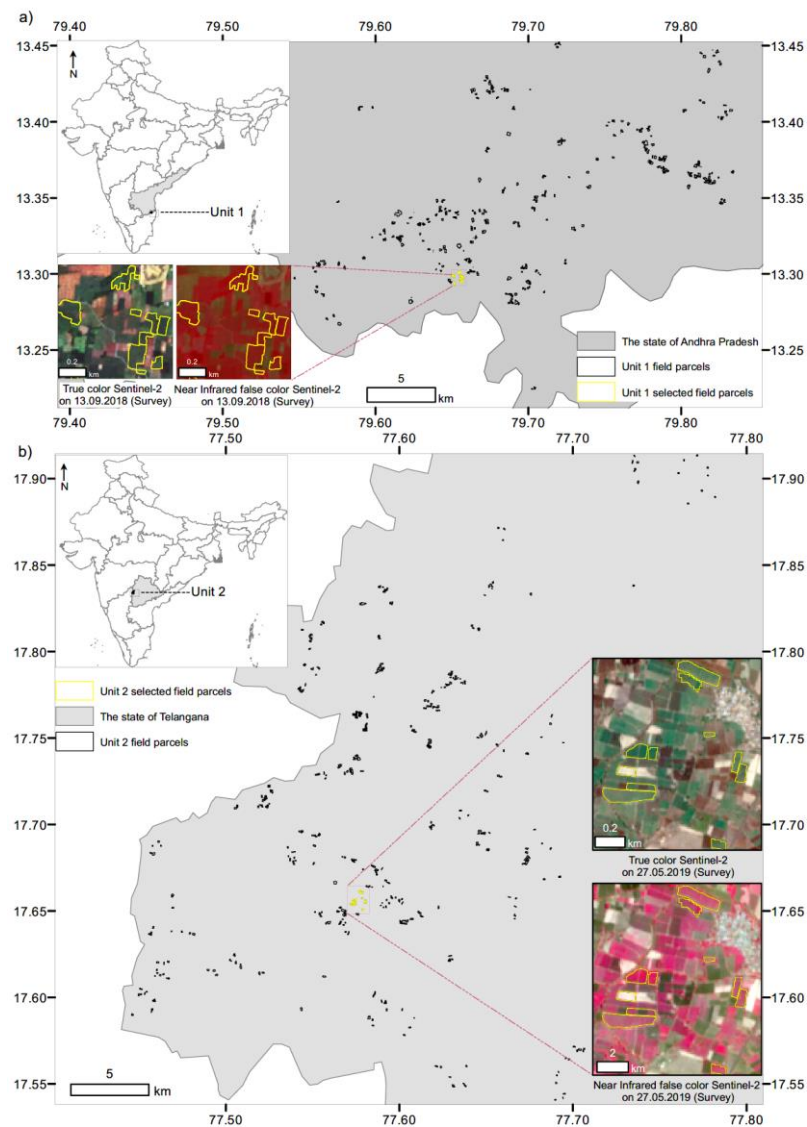


Figure 4.2. The location of sample sugarcane fields in the Telangana and Andhra Pradesh states of India.

4.2.2 Sentinel-1 datasets

In this study, EO descending time-series data from Sentinel-1 A and B were used to monitor sugarcane growth during the 2018-2019 season in Unit 1 and the 2019-2020 season in Unit 2. The Sentinel-1 satellites provide backscatter information at 5.405 GHz (C-band) at 5 x 20 m spatial resolution (Besten et al.,

2023). The Sentinel-1 GRD (Ground Range Detected) C-band GEE product was used to extract the ground backscatter to derive sugarcane growth stage information for Unit 1 and Unit 2, because of its high spatial resolution (10 m) (Vavlas et al., 2020) and because radiometric calibration and ortho-correction are already performed using the Sentinel-1 Toolbox (Demissie et al., 2023). Acquiring GRD images directly from the GEE platform is advantageous as the GEE processing chain converts the Sentinel-1 data to normalised backscatter, correcting for elevation differences in areas greater than 60 degrees latitude, then pre-processes the data by removing thermal and radiometric noise and performs data calibration (Mullissa et al., 2021; Singha and Swain, 2023).

To investigate the backscatter effects of sugarcane, the mean temporal Sentinel-1 GRD backscatter in the single polarisation (VV and VH) was obtained for each sugarcane field located in Unit 1 and Unit 2 from the GEE catalogue. The mean CR was also obtained for each field parcel, and this ratio's applicability for monitoring sugarcane growth over time was also assessed. Figure 4.3 and Figure 4.4 show the average backscatter over five sugarcane growing seasons for one sugarcane field in Unit 1 and Unit 2 respectively.

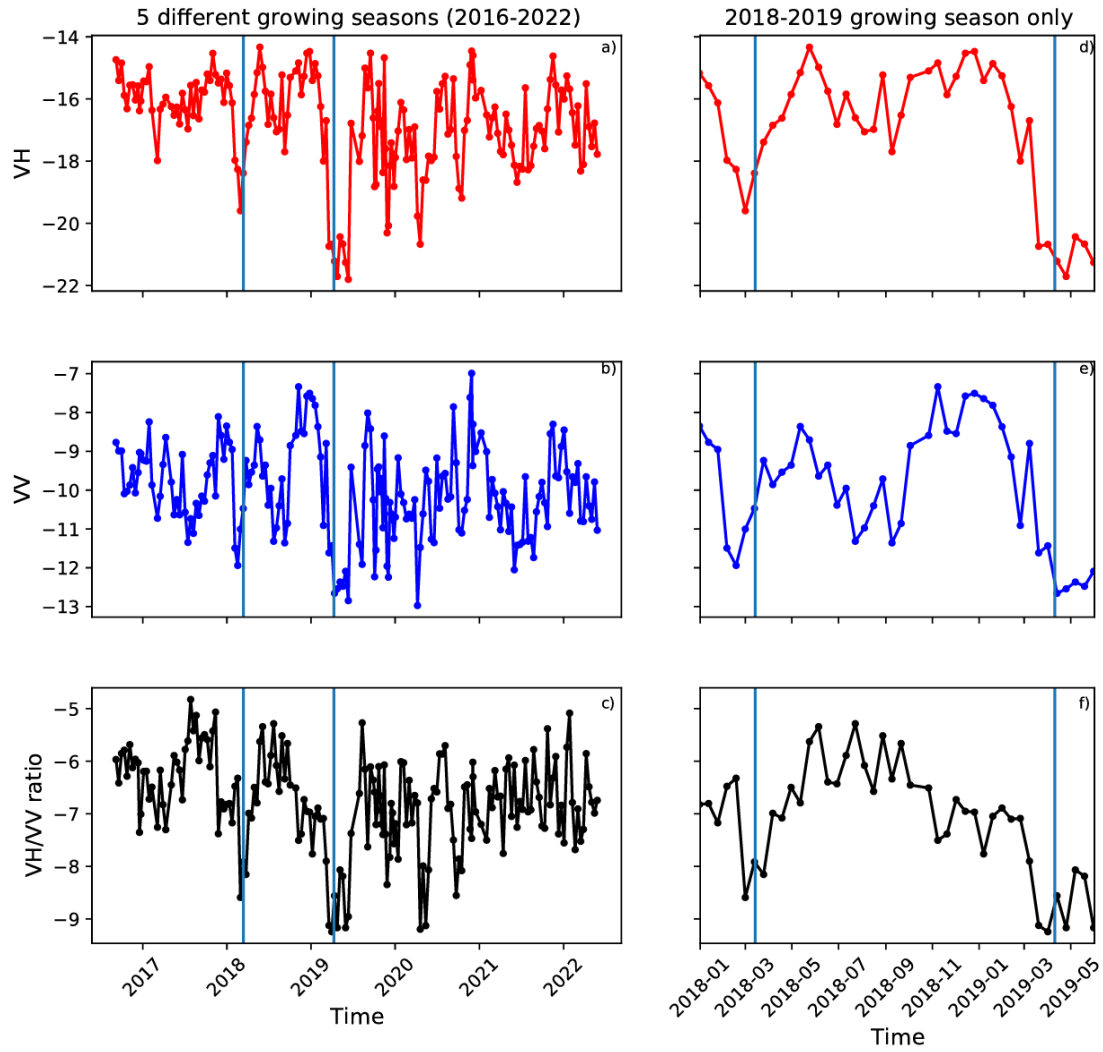


Figure 4.3. An example of the respective mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 for a field in Unit 1 for five seasons (a,b,c), and mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 of the same field for the 2018 2019 sugarcane growing season only (d,e,f). The hashed vertical lines presented the start and end of the season dates previously derived from the optical analysis, for the 2018-2019 growing season.

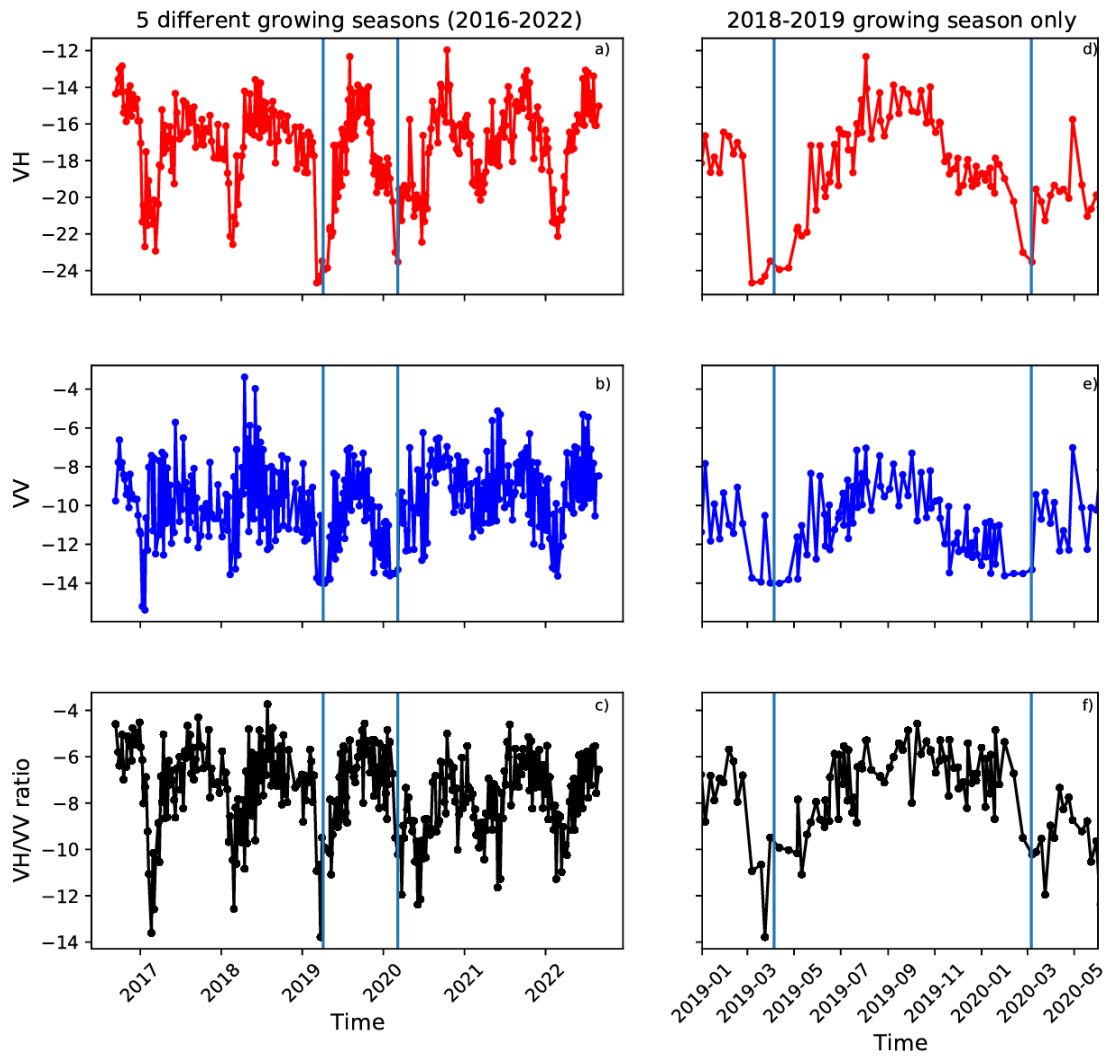


Figure 4.4. An example of the respective mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 for a field in Unit 2 for five seasons (a,b,c), and mean VH, VV, and VH/VV ratio back scatter from Sentinel-1 of the same field for the 2019-2020 sugarcane growing season only (d,e,f). The hashed vertical lines presented the start and end of the season dates previously derived from the optical analysis, for the 2019-2020 growing season.

4.2.3 Establishment of data-processing method

To develop a method to automate the determination of sugarcane development stages from EO data, a technique was established for processing the VV and VH backscatter time-series datasets (Figure 4.5).

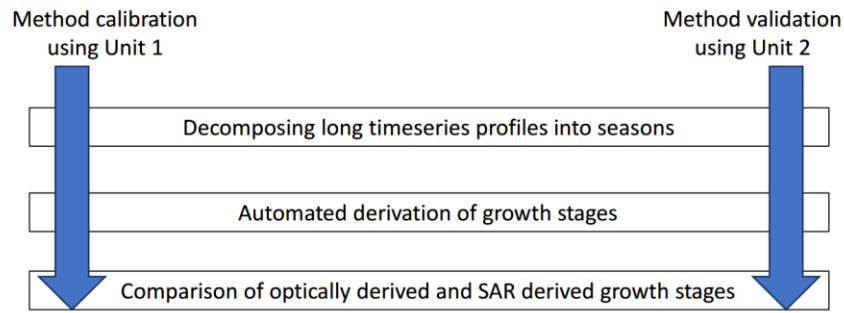


Figure 4.5. The methodology was established using data from Unit 1 and then evaluated using data from Unit 2.

4.2.4 Decomposing the time-series into one season (SAR data only)

To test the benefits of using SAR data for seasonal decomposition over optical datasets, the season-splitting methodology developed in Chapter 3 was applied to VV, VH, and CR backscatter time-series datasets for sugarcane fields located in Unit 1 and Unit 2.

The splitting algorithm was designed to identify periods of low vegetation cover, which in the case of Sentinel-1 data would be represented by periods of low backscatter, at the SOS and EOS for each sugarcane growing season. The Fast Fourier Transform (FFT) algorithm was applied to the mean time-series of each field to remove erroneous periods of low backscatter by transforming the time-series into magnitude and phase frequency components. Once separated into magnitude and phase frequency components, the low-frequency information representing the annual sugarcane growing seasons could be separated from higher-frequency noise (Li et al., 2021). The FFT algorithm is an algorithm that computes the Discrete Fourier Transform (DFT) much faster than the direct use of the DFT algorithm. The FFT first converts the time-series from the time domain to the frequency domain using Equation 4.1, where $F(u)$ is the signal after the Fourier transform with u representing the frequency domain, $f(t)$ is the time-series with t representing the time domain, and N is the length of the time-series.

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

Equation 4.1

Once simplified into sigmoidal curves using the FFT algorithm, a Difference of the Gaussian (DOG) algorithm was used to decompose long time-series data into seasons. The higher frequency components of the FFT were first suppressed using the DOG low-pass function. This suppression was determined by the standard deviation of the Gaussian function with higher standard deviations producing very suppressed (almost flat signals) and lower standard deviations producing less suppressed signals. Higher suppressed signals were then subtracted from the less suppressed signals to create a bandpass filter that removed low-contrast areas. For each sugarcane field, the method subtracted $F(u)$ s with a standard deviation of 9 from $F(u)$ s with a standard deviation of 2. Peaks of the remaining areas of high contrast were extracted to identify the edges of each sugarcane growth cycle, corresponding to periods of low vegetation cover and hence low backscatter. The seasonal decomposition function was programmed with two conditions to minimise the effect of erroneous low backscatter observations on the splitting of the seasons. The first condition was to select only peaks greater than 0. The second condition was one peak per year as the average length of one sugarcane growth cycle is one year. Once peaks were identified the $F(u)$ s were converted back without the higher frequencies using the Inverse Fourier transform (Li et al., 2021). A buffer (± 6 observations) was applied to each of the derived sigmoid curve edges and minimum backscatter within the buffer was extracted and taken as the respective SOS and EOS of each sugarcane growth cycle.

The frequency at which Sentinel-1 backscatter data is captured is dependent on location. Sentinel-1 is a constellation of two satellites, theoretically, that means a potential 6-day revisit time for a studied area. However, the data power and bandwidth available for acquisitions are limited in certain study areas. Study areas closer to the poles receive repeat data every 6 days whereas areas further away receive data every 12 days. Hence, to minimise misinterpretation when converting the time-series to and from the frequency domain, it was important to

resample and interpolate missing data within the time-series to a known frequency with equally spaced intervals before using the FFT. For the Unit 1 dataset, a 12-day resampling interval appeared optimal for the Fourier analysis, as the mode interval in this dataset (84%) with only a few intervals differing from this observation (10% of observations had an interval of 6 days and the remaining 6% were either 1,5,7,24 or 36 days). Where there were no good-quality observations in the 12-day interval period, linear interpolation was used to fill the data gaps, using adjacent good-quality observations, where there were multiple observations within the 12-day interval period, the mean value of all observations was used to represent the observation of that period. The time-series for fields located in Unit 2 were also resampled, however, a 6-day resampling interval was used instead of a 12-day interval as this was the mode interval of the Unit 2 time-series datasets.

The impact of using SAR data to split the season was assessed against manual estimates of the SOS and EOS derived through visual interpretation of the raw NDVI time-series in Chapter 3. The manual SOS was derived by identifying the first point of the time-series before a rapid increase in NDVI, representing the point before a rapid increase in green biomass. The manual EOS was derived by identifying the last low NDVI observation before an increase in NDVI, marking the start of a new season. Comparisons between the SOS and EOS derived from the SAR and processed NDVI time-series were also made. The optically derived SOS and EOS were respectively the point at which the algorithm automatically identified the first point of the time-series before a rapid increase in NDVI and the last point of the time-series before an increase in NDVI. To aid the comparison of the manual, automated optical, and automated SAR-derived estimates of SOS and EOS, the dates at which these development stages fell were normalised as the number of days after 1 September 2017 for Unit 1 and the number of days after 1 September 2018 for Unit 2.

4.2.5 Deriving the knees from Sentinel-1 backscatter

To test the benefits of using SAR data for deriving sugarcane growth stages over optical datasets, the same knee algorithm used to identify the start and end of the mid-season (SMS and EMS respectively) in Chapter 3 from temporal NDVI measurements was applied to temporal observations of VV, VH, and CR backscatter for sugarcane fields located in Unit 1 and Unit 2.

The initial knee describes the point of levelling off, where the amount of recorded photosynthetic vegetation in the canopy no longer increases with increasing values with time (See Equation 4.1). The results of Chapter 3 show that the identification of the initial knee (point of levelling off in the NDVI time-series) can be used to identify the start of the mid-season (SMS) in processed optical time-series (See Equation 4.1). At the SMS, NDVI stabilises with the focus of growth moving from tillering (increase in number of stems) to stem-elongation (an effect not captured by NDVI observations of the sugarcane canopy). The second knee, an inverted knee, was used to capture the last point of stability before a decline in NDVI, to identify the EMS within the processed optical time-series. This is because the EMS marks the onset of senescence, therefore the browning of sugarcane leaves and reduction in green biomass over time, which can be recorded as a decline in NDVI (Figure 4.1).

The study compared the derived knees to the growth stages previously derived from the processed optical data (in Chapter 3) and validated the findings with photographs and Sentinel-2 image composites. The study first cross-referenced the dates of the derived knees with field photographs collected during the 2018 and 2019 survey in Unit 1 and Unit 2 respectfully and Sentinel-2 True and False colour composite images to describe what the derived growth stages represented on the ground. To do this, the study identified the fields where the dates of the derived knees were within ± 3 days of the date of the photograph and used these fields to validate phenology. Where field photography was unavailable, for periods of the derived knees, the study used Sentinel-2 band 4,3,2 True colour composite (TCC) and band 8,4,3 NIR False colour composite (FCC) images for validation. The study acquired multiple Sentinel-2 image composites (before and

after the date of the derived knees) and used them to validate the derived phenological stage.

4.3 Results

4.3.1 Calibration

Manual and automated start of the season (SOS)

When the season splitting algorithms were applied to the Sentinel-1 backscatter time-series datasets for fields located in Unit 1, there was no systematic relationship between the SOS manually interpreted from raw NDVI time-series data and the SOS automatically derived from the VV, VH, and CR backscatter time-series profiles, R^2 values of 0.02 for VV and VH backscatter and 0.00 for CR backscatter (See Figure 4.6).

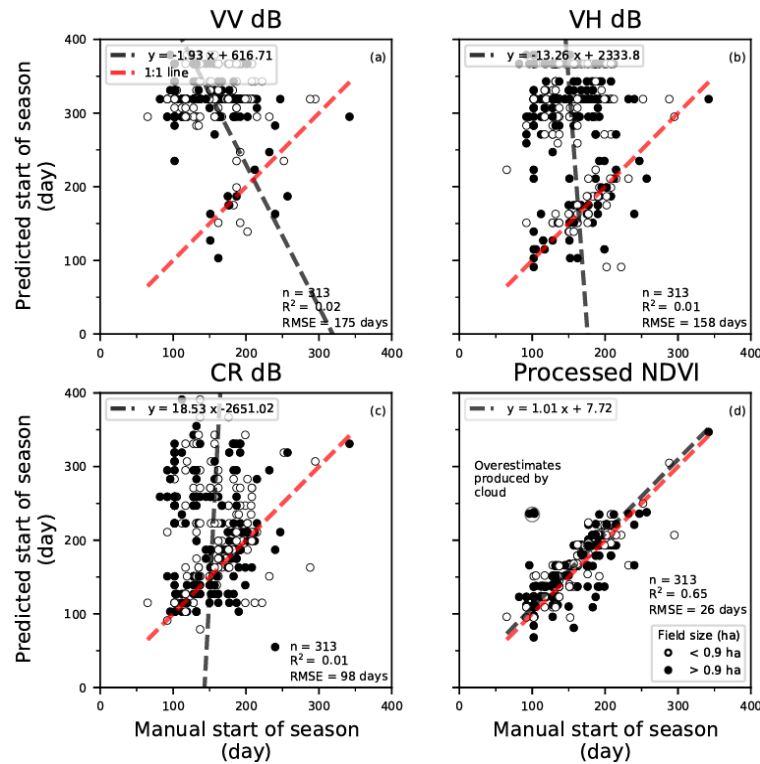


Figure 4.6. Relationship between the automated (p) and manually derived (m) start of the season (SOS) for fields in Unit 1 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults

Manual and automated end of season (EOS)

The relationships between the use of VV, VH, and CR backscatter with manual estimates of EOS were also weak with respective R^2 values of 0.15, 0.43, and 0.20 respectively, compared to the relationship of the predictions derived from the processed optical time-series and manually derived estimates with an R^2 value of 0.78 (See Figure 4.7).

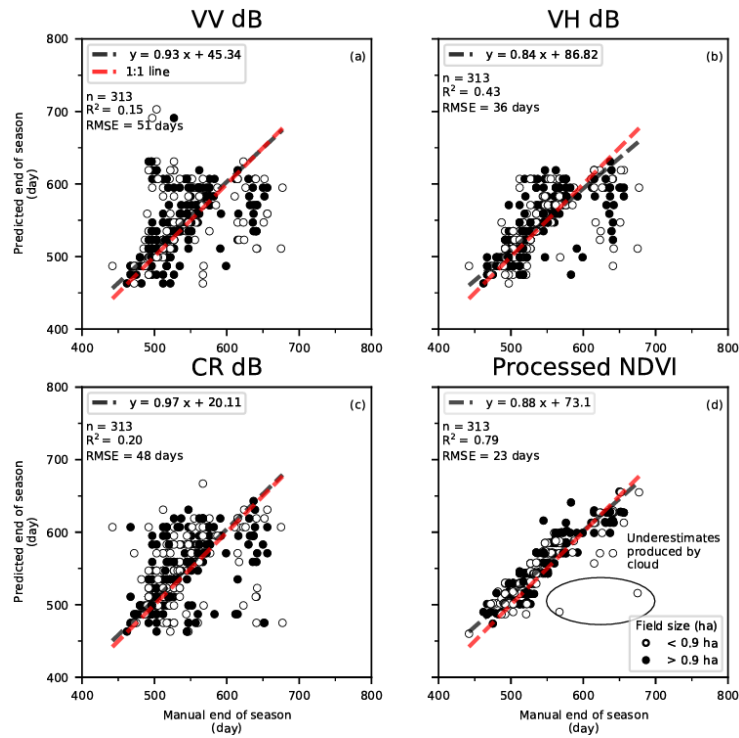


Figure 4.7. Relationship between the automated (p) and manually derived (m) end of the season (EOS) for fields in Unit 1 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.

These results show that, despite the cloud penetrating abilities of SAR instruments such as Sentinel-1, there was no benefit of using Sentinel-1 backscatter time-series datasets for decomposing long time-series into seasons compared to the use of processed NDVI time-series. Upon visualisation of the individual SAR time-series for each sugarcane field, against the optically derived dates for the SOS and EOS, we found a slight offset with the SAR growing season starting before the optically derived growing season (See Figure 4.3 and Figure 4.4). This offset may be because the optically derived SOS and EOS were derived using a time-series with a 14-day interval whereas the SAR time-series had a 12-day interval in Unit 1. To resolve the offset in the seasonal splitting, we used a buffer of six observations before the automated optically derived SOS dates and one observation before the automated optically derived EOS dates and the minimum values within the buffer were derived as the respective SOS and EOS for the SAR time-series of each field.

Manual and automated start of the season (SOS) using buffer

The dates for the SOS automatically derived as the minimum backscatter values from within six observations of the automated dates SOS from processed NDVI time-series using VV, VH, and CR backscatter time-series, were compared with the manual and predicted dates for the SOS derived from NDVI time-series, outlined in Chapter 3. To aid comparison, the dates for the derived automated estimates of SOS for each field were normalised as the number of days after the nominal day of the season (DOS), defined as 1 September 2017.

For the Unit 1 calibration dataset, the predictions for the SOS derived from VH, VV, and CR ranged from 49 to 319 DOS, 67 to 319 DOS, and 43 to 331 DOS respectively, the manually derived SOS ranged from 65 to 342 DOS and the automated SOS derived from processed NDVI time-series ranged from 68 to 347 DOS.

Figure 4.8a-c depicts the relationship between the predicted SOS dates derived from Sentinel-1 backscatter and the manually derived SOS dates, visually derived from raw NDVI time-series data. The predictions for the SOS derived from VH, VV, and CR depicted in Figure 4.8a-c showed a stronger relationship with manual estimates of SOS derived from raw NDVI time-series, R^2 values of 0.49, 0.57 and 0.49, respectively and RMSE values of 35, 28, and 32 days (see Figure 4.8a-c) compared to the VH, VV, and CR estimates of SOS depicted in Figure 4.6. However, Figure 4.8a-c also showed three outliers where the SOS had been overestimated. Subsequent analysis of the three fields with overestimated SOS dates showed that these overestimations resulted from erroneous late estimations of SOS derived using the processed optical data caused by unremoved clouds during the mid-season. These results show that the impact of clouds persisted in VV, VH and CR backscatter time-series estimates of the SOS as the erroneous overestimations for the SOS derived from processed NDVI time-series had a knock-on effect on deriving the minimum backscatter values. After removing these three points, the R^2 for the relationship between VV,

VH and CR-derived SOS dates with manual interpretations increased to 0.68, 0.57 and 0.55 with respective RMSE values of 32, 24 and 29 days.

Figure 4.8d-f shows the relationship between the predicted SOS dates derived from Sentinel-1 backscatter and the predicted SOS dates derived from processed NDVI time-series data. The relationship between the VV, VH and CR-derived SOS dates was higher when compared to the SOS derived from the processed NDVI than the relationship with the manual. The predictions of SOS derived from VH backscatter, and processed NDVI time-series were most similar, giving the most consistent results. The R^2 and RMSE values for the relationship between the predictions of SOS derived from VH backscatter, and processed NDVI time-series were 0.81 and 22 days respectively and the intercept and slope of the regression equation were also closest to 0, compared to the predicted SOS dates derived from VV and CR backscatter time-series, with respective R^2 values of 0.73 and 0.63 and RMSE values of 31 and 27 days.

There is also an offset between the VV, VH and CR-derived SOS dates and the NDVI-derived SOS dates. The negative intercepts in the regression equations depicted in Figure 4.8a-f shows a tendency at the low values for VV, VH and CR-derived SOS dates to result in lower values than the NDVI-derived SOS dates. On the other hand, at higher SOS values the relationship between the VV, VH and CR-derived SOS dates and the NDVI-derived SOS dates converge. This convergence at the high SOS dates is clear in Figure 4.8e.

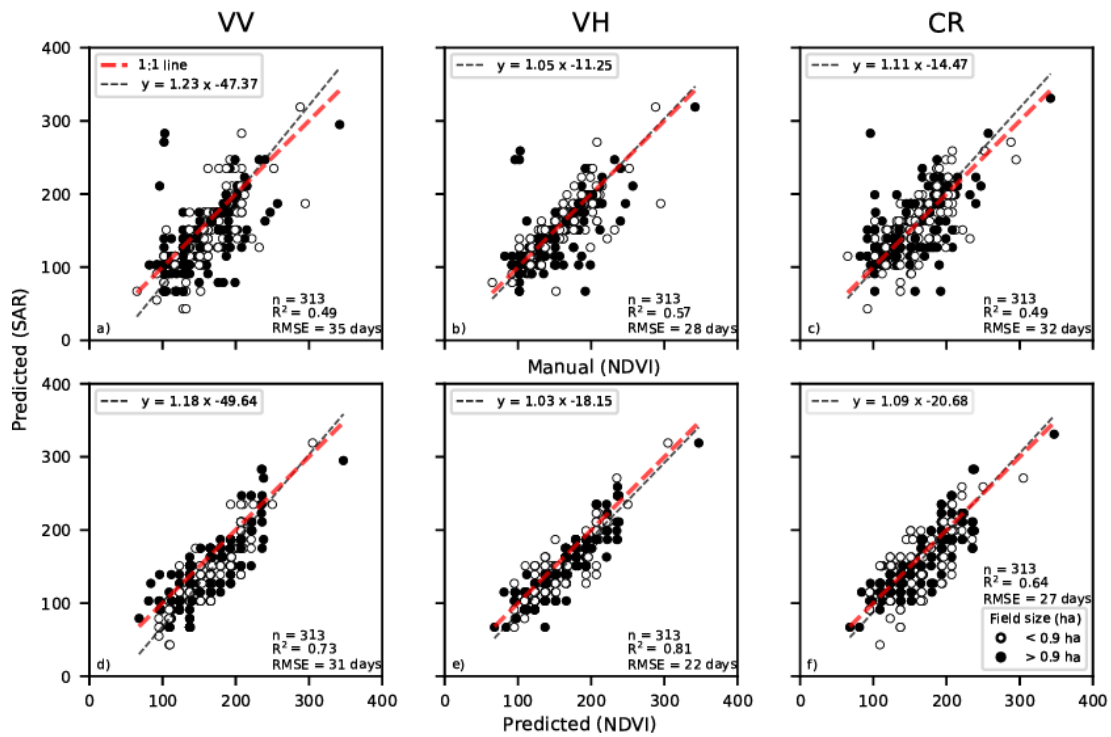


Figure 4.8. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived start of the season (SOS) (y-axis) and SOS manually interpreted from raw NDVI time-series profiles (x-axis) for fields located in Unit1 (a-c). The relationship between the SOS derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the SOS automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (x-axis) (d-f).

Manual and automated end of season (EOS) using buffer

The dates for the EOS automatically derived as the minimum backscatter values from within six observations of the automated dates for the EOS from processed NDVI time-series using VV, VH, and CR backscatter time-series, were compared with the manual and predicted dates for the EOS derived from NDVI time-series outlined in Chapter 3. To aid comparison, the dates for the derived automated estimates of EOS for each field were also normalised as DOS.

For the Unit 1 calibration dataset, the predictions for the EOS derived from VH, VV, and CR ranged from 367 to 631 DOS, 541 to 643 DOS, and 379 to 643 DOS respectively, the manually derived SOS ranged from 422 to 677 DOS and the

automated EOS derived from processed NDVI time-series ranged from 460 to 656 DOS.

Figure 4.9a-c depicts the relationship between the predicted EOS dates derived from Sentinel-1 backscatter and the manually derived EOS dates visually derived from raw NDVI time-series data. The predictions for the EOS derived from VH, VV, and CR are depicted in Figure 4.8a-c and showed a stronger relationship with manual estimates of EOS derived from raw NDVI time-series with respective R^2 values of 0.55, 0.60, 0.58 and RMSE values of 35, 30, 34 days (see Figure 4.9a-c) compared to the VH, VV, and CR estimates of EOS depicted in Figure 4.7.

Figure 4.9d-f depicts the relationship between the predicted EOS dates derived from Sentinel-1 backscatter and automated estimates of EOS derived using processed NDVI time-series data. The estimations of the EOS derived from all three backscatter time-series (VV, VH, and CR) were systematically earlier than the estimates of the EOS derived from the process NDVI time-series. The relationship was stronger for the EOS estimates derived from VH backscatter with the slope often close to one, an R^2 value of 0.81 and RMSE of 24 days compared to the predicted EOS dates derived from VV and CR backscatter time-series, with respective R^2 values of 0.75 and 0.74, and RMSEs of 31 and 30 days.

Figure 4.9c and Figure 4.9f also depicts a cluster of early estimated EOS dates derived from CR backscatter time-series data for sugarcane fields less than 0.9 ha compared to the EOS derived from NDVI time-series. This may be due to an increase in spectral mixing in smaller fields.

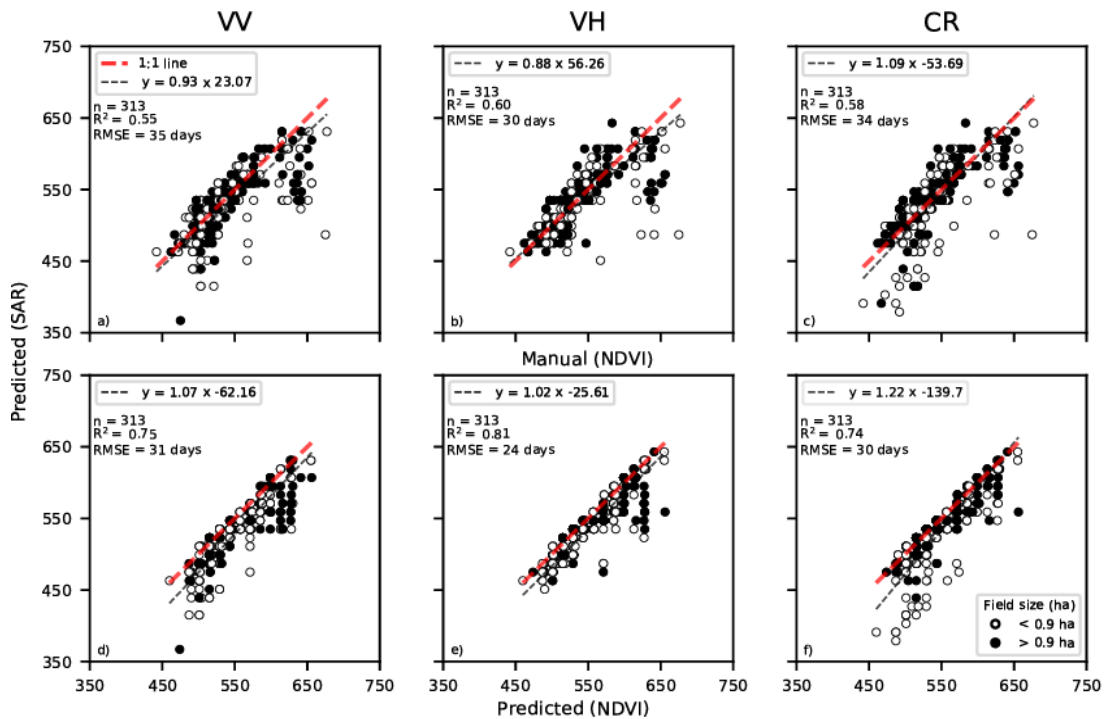


Figure 4.9. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived end of the season (EOS) (y-axis) and EOS manually interpreted from raw NDVI time-series profiles (a-c) (x-axis) for fields located in Unit 1. The relationship between the EOS derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the EOS automatically derived by processed NDVI time-series profiles for fields located in Unit 1 (x-axis) (d-f).

Initial knees from Sentinel-1 backscatter and mid-season (SMS) NDVI time-series

This section compares the results of the optically derived Start of the mid-season (SMS) and the results of the initial knees derived using VV, VH and CR backscatter time-series normalised as DOS.

For the Unit 1 calibration dataset, the predictions for the initial knee derived from VV, VH, and CR ranged from 67 to 427 DOS, 127 to 439 DOS, and 91 to 355 DOS respectively. The manually derived SMS from visual interpretation of raw NDVI time-series ranged from 132 to 392 DOS and the automated SMS derived from processed NDVI time-series ranged from 165 to 431 DOS.

The initial knee derived from VH, VV, and CR depicted in Figure 4.10a-c showed a weak positive linear relationship with manual estimates of SMS derived from visual interpretation of raw NDVI time-series data, R^2 values of 14, 22, and 16, respectively, and RMSE values of 69 and 58 days (see Figure 4.10a-c).

Figure 4.10d-f also shows that the relationship predictions of the initial knee derived from VH backscatter, and processed NDVI time-series were stronger than the relationship between manually derived SMS dates with R^2 values of 0.25, 0.40, and RMSEs of 79, 66, and 67 days. The estimations of the initial knee derived from all three backscatter time-series (VV, VH, and CR) were systematically earlier than the estimates of the SMS derived from the raw and processed NDVI time-series, however, the systematically earlier relationship was stronger for the VH derived initial knee.

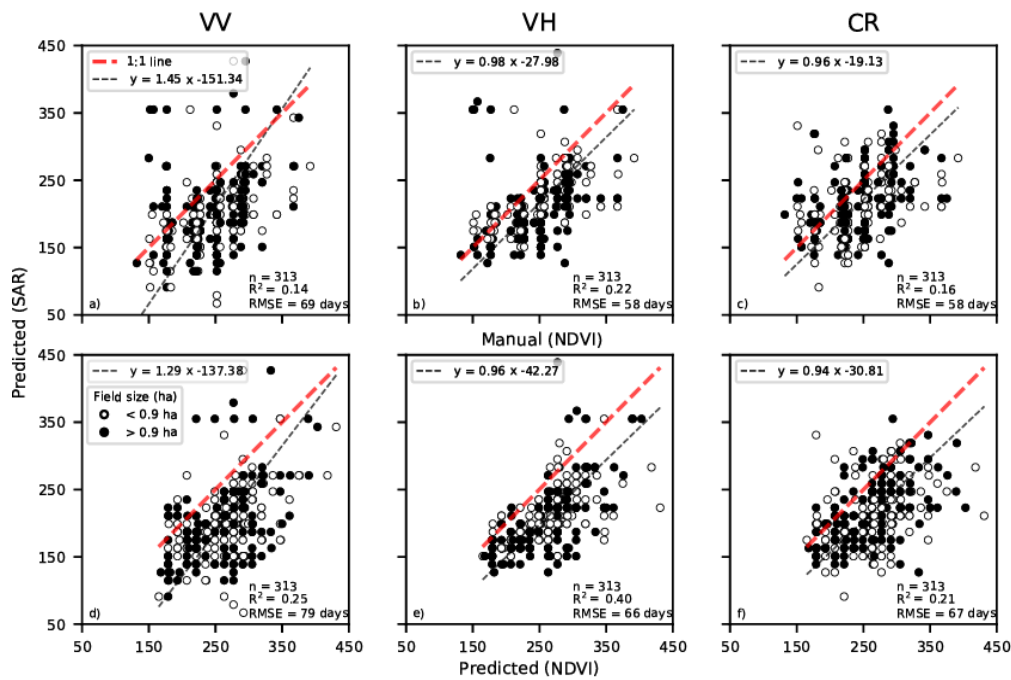


Figure 4.10. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived second knees (y-axis) and the start of mid-season (SMS) manually interpreted from raw NDVI time-series profiles (a-c) (x-axis) for fields located in Unit 1. The relationship between the start of mid-season (SMS) derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the initial knee automatically derived by processed NDVI time-series profiles (x-axis) for fields located in Unit 1 (d-f).

Second knees from Sentinel-1 backscatter and end of mid-season NDVI time-series

This section compares the results of the optically derived end-of-mid-season dates (EMS) with the second knee dates derived using VV, VH, and CR backscatter time-series, normalised as DOS.

For the Unit 1 calibration dataset, the predictions for the second knee derived from VV, VH, and CR ranged from 355 to 583 DOS, 427 to 583 DOS, and 319 to 631 DOS respectively. The manually derived EMS from visual interpretation of raw NDVI time-series ranged from 361 to 587 and the automated EMS derived from processed NDVI time-series ranged from 368 to 587 DOS.

The second knee derived from VH, VV, and CR depicted in Figure 4.11 showed a weak positive linear relationship with manual estimates of EMS derived from visual interpretation of raw NDVI time-series data, with R^2 values of 0.24, 0.28, and 0.20, respectively and RMSE values of 34, 31, and 46 days (see Figure 4.11 a-c).

Figure 4.11d-f shows a stronger relationship between the predictions of the initial knee derived from VV, VH and CR backscatter, and processed NDVI time-series than the relationship between manually derived SMS dates with R^2 values of 0.27, 0.40 and RMSEs of 38, 37 and 52 days. The estimations of the second knee derived from all three backscatter time-series (VV, VH, and CR) were systematically later than the estimates of the EMS derived from the raw and processed NDVI time-series, and the systematically later relationship was strongest for the VH backscatter derived initial knee.

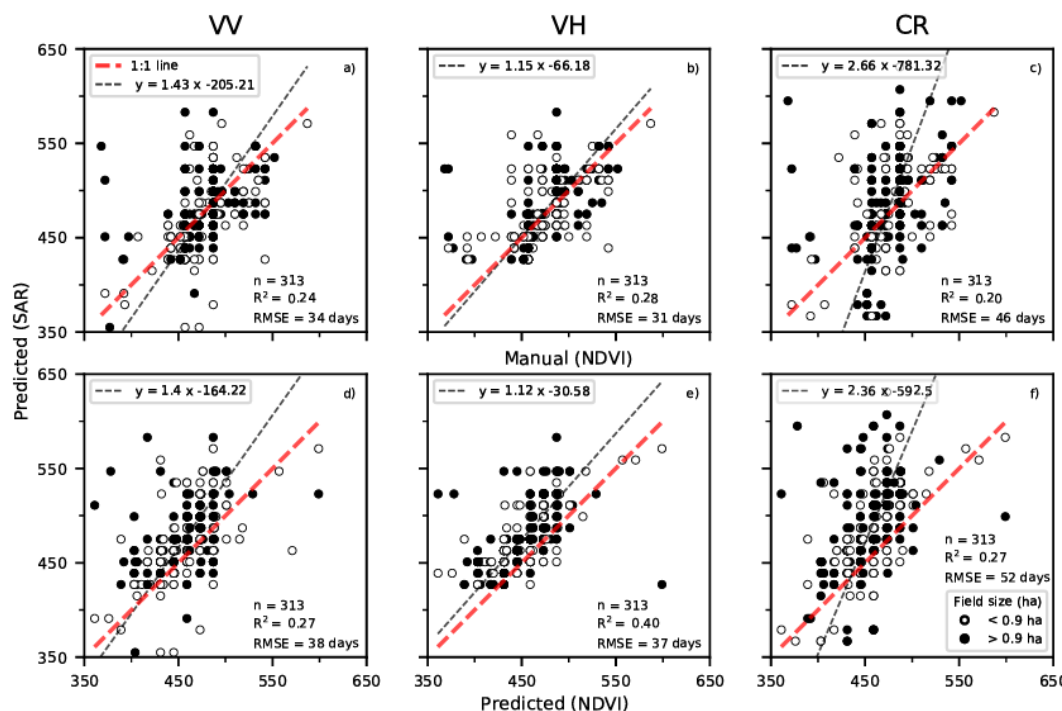


Figure 4.11. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived second knees (y-axis) and the end of mid-season (EMS) manually interpreted from raw NDVI time-series profiles (a-c) (x-axis) for fields located in Unit 1. The relationship between the end of mid-season (EMS) derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the initial knee automatically derived by processed NDVI time-series profiles (x-axis) for fields located in Unit 1 (d-f).

4.3.2 Evaluation

The findings for the Unit 1 calibration dataset were evaluated with fields located in the Unit 2 evaluation dataset. This section compares the results for the evaluation dataset of Unit 2. To aid comparison, the dates for the derived automated estimates of growth stages were normalised as the number of days after the nominal day of the Unit 2 season (Unit 2 DOS), defined as 1 September 2018.

Manual and automated start of season (SOS)

In the Unit 2 evaluation dataset, for the derived SOS, the relationship between the manual and automated estimates was higher when processed optical NDVI time-series were used (after the removal of the fields with changing field boundaries after harvest), compared to when VV, VH, and CR backscatter time-series profiles were used. There was no relationship between the use of VV and VH backscatter for predicting the SOS with R^2 values of 0.01, 0.13, and 0.06 respectively (see Figure 4.12).

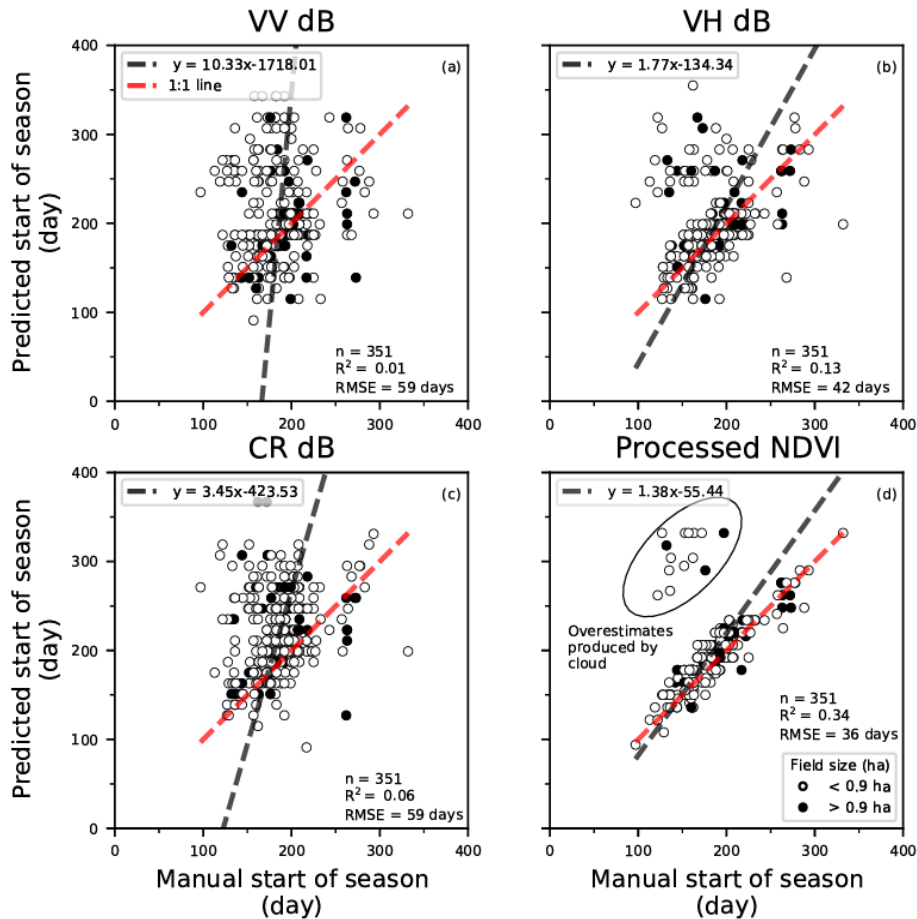


Figure 4.12. Relationship between the automated (p) and manually derived (m) start of the season (SOS) for fields in Unit 2 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.

Manual and automated end of season (EOS)

In the Unit 2 evaluation dataset, for the derived EOS, the relationship between the manual and automated estimates was also higher when processed optical NDVI time-series were used (after the removal of the fields with changing field boundaries after harvest), compared to when VV, VH, and CR backscatter time-series profiles were used. There was a weaker relationship between the use of VV, VH, and CR backscatter with manual estimates of EOS R^2 values of 0.06, 0.24, 0.17 respectively (See Figure 4.13), compared to the relationship of the predictions derived from the processed optical time-series with a R^2 value of 0.50 (See Figure 4.13).

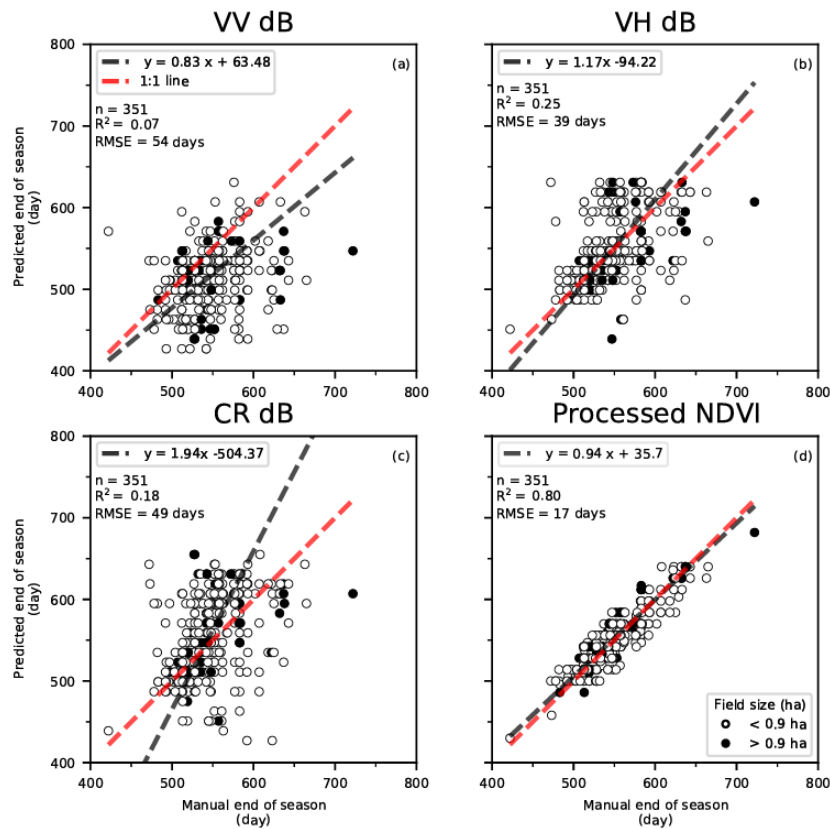


Figure 4.13. Relationship between the automated (p) and manually derived (m) end of the season (EOS) for fields in Unit 2 using VV, VH, CR back scatter (dB) and processed NDVI time-series after the removal of field parcels with boundary faults.

Like Unit 1 these results also show that, despite the cloud penetrating abilities of SAR instruments such as Sentinel-1, for fields located in Unit 2 there was no benefit of using Sentinel-1 backscatter time-series datasets for decomposing long time-series into seasons compared to the use of processed NDVI time-series. Upon visualisation of the individual SAR time-series for each sugarcane field, against the optically derived dates for the SOS and EOS, there was also a slight offset with the SAR growing season starting before the optically derived growing season in Unit 2 (See Figure 4.3 and Figure 4.4). This offset may be because the optically derived SOS and EOS were derived using a time-series with a 14-day interval whereas the SAR time-series had a 6-day interval in Unit 2. To resolve the offset in the seasonal splitting, we used a buffer of six observations before the automated optically derived SOS dates and one observation before the automated optically derived EOS dates and the minimum values within the buffer were derived as the respective SOS and EOS for the SAR time-series of each field.

Manual and automated start of the season (SOS) using buffer

The dates for the SOS automatically derived as the minimum backscatter values from within six observations of the automated dates for the SOS from processed NDVI time-series using VV, VH, and CR backscatter time-series, were compared with the manual and predicted dates for the SOS derived from NDVI time-series outlined in Chapter 3.

For the Unit 2 evaluation dataset, the predictions for the SOS derived from VH, VV, and CR ranged from 91 to 343 Unit2-DOS, 91 to 319 Unit2-DOS, and 55 to 355 Unit2-DOS respectively, the manually derived SOS ranged from 97 to 332 Unit2-DOS and the automated Unit2-DOS derived from processed NDVI time-series ranged from 94 to 332 Unit2-DOS.

Figure 4.14a-c depicts the relationship between the predicted SOS dates derived from Sentinel-1 backscatter and the manually derived SOS dates, manually interpreted from raw NDVI time-series data. The predictions for the SOS derived from VH, VV, and CR are depicted in Figure 4.14a-c and show a stronger

relationship with manual estimates of SOS derived from raw NDVI time-series, with respective R^2 values of 0.22, 0.27 and 0.26, and RMSE values of 37, 34, 36 (see Figure 4.14a-c) compared to the VH, VV, and CR estimates of SOS depicted in Figure 4.12. However, Figure 4.14a-c also showed outliers where the SOS had been overestimated. Subsequent analysis of the 15 fields with overestimated SOS dates showed that these overestimations were a by-product of erroneous late estimations of SOS derived using the processed optical data caused by unremoved clouds during the mid-season. The impact of clouds persisted in VV, VH and CR backscatter time-series estimates of the SOS as the erroneous overestimations for the SOS derived from processed NDVI time-series had a knock-on effect on deriving the minimum backscatter values. After removing these 15 points, the R^2 for the relationship between VV, VH and CR-derived SOS dates with manual interpretations increased to 0.51, 0.60 and 0.49 with respective RMSE values of 25, 22 and 26 days.

The predictions of SOS derived from VH backscatter and processed NDVI time-series were most similar with a R^2 value of 0.76 and RMSE of 25 days (see Figure 4.14f) compared to the predicted SOS dates derived from VV and CR backscatter time-series, with respective R^2 values of 0.66 and 0.65 and RMSEs of 28 and 26 days. The predictions of SOS dates derived from VH backscatter were also systematically earlier than the estimates of the SOS dates derived from the process NDVI time-series.

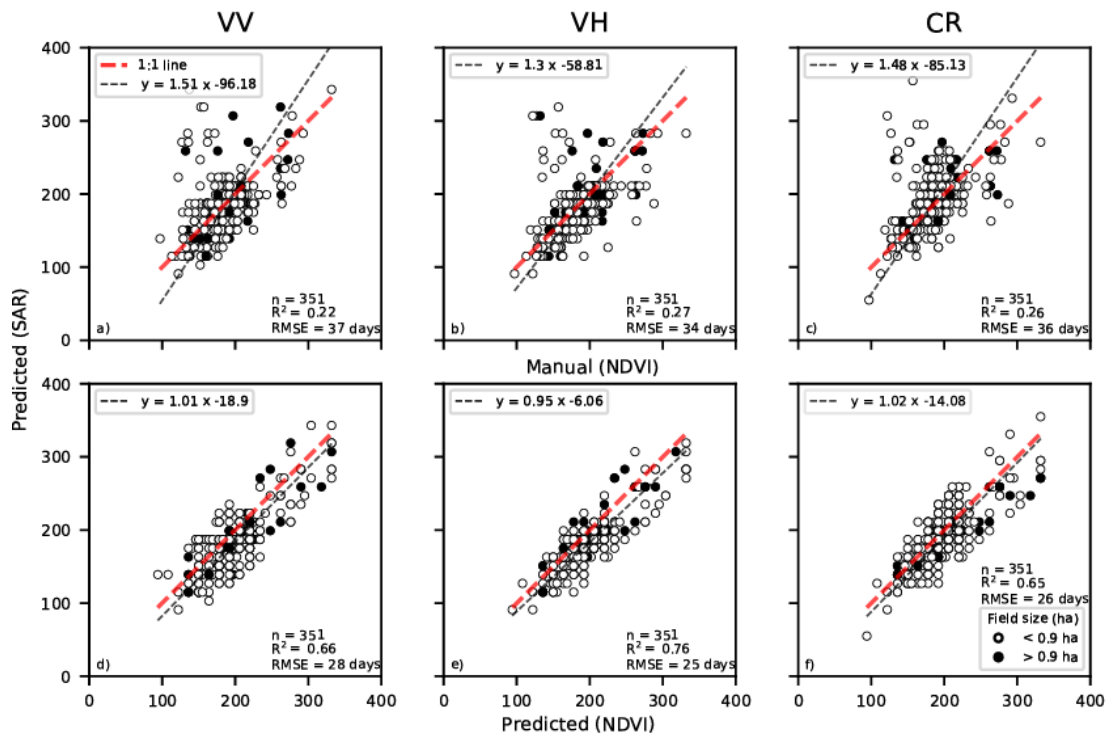


Figure 4.14. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived start of the season (SOS) (y-axis) and SOS manually interpreted from raw NDVI time-series profiles (x-axis) for fields located in Unit2 (a-c). The relationship between the SOS derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the SOS automatically derived by processed NDVI time-series profiles for fields located in Unit 2 (x-axis) (d-f).

Manual and automated end of season (EOS) using buffer

The dates for the EOS automatically derived as the minimum backscatter values from within 10 observations of the automated dates for the EOS from processed NDVI time-series using VV, VH, and CR backscatter time-series for fields located in Unit 2, were compared with the manual and predicted dates for the EOS derived from NDVI time-series outlined in Chapter 3.

For the Unit 2 evaluation dataset, the predictions for the EOS derived from VH, VV, and CR ranged from 403 to 631 Unit 2-DOS, 439 to 631 Unit 2-DOS, and 439 to 643 Unit 2-DOS respectively, the manually derived SOS ranged from 422

to 722 Unit 2-DOS and the automated EOS derived from processed NDVI time-series ranged from 430 to 682 Unit 2-DOS.

Figure 4.15a-c depicts the relationship between the predicted EOS dates derived from Sentinel-1 backscatter and the manually derived EOS dates, visually derived from raw NDVI time-series data. The predictions for the EOS derived from VH, VV, and CR are depicted in Figure 4.15a-c and show a stronger relationship with manual estimates of EOS derived from raw NDVI time-series, R^2 values of 45, 55, 58, respectively and RMSE values of 38, 30, 28 days (see Figure 4.15a-c) compared to the VH, VV, and CR estimates of EOS depicted in Figure 4.13.

Figure 4.15d-f depicts the relationship between the predicted EOS dates derived from Sentinel-1 backscatter and automated estimates of EOS derived using processed NDVI time-series data. The estimations of the EOS derived from all three backscatter time-series (VV, VH, and CR) were systematically earlier than the estimates of the EOS derived from the process NDVI time-series. The systematically earlier relationship was stronger for the EOS estimates derived from CR backscatter with an R^2 value of 0.69 and RMSE of 26 days compared to the predicted EOS dates derived from VV and VH backscatter time-series, with respective R^2 values of 0.55 and 0.65, and RMSEs of 37 and 28 days.

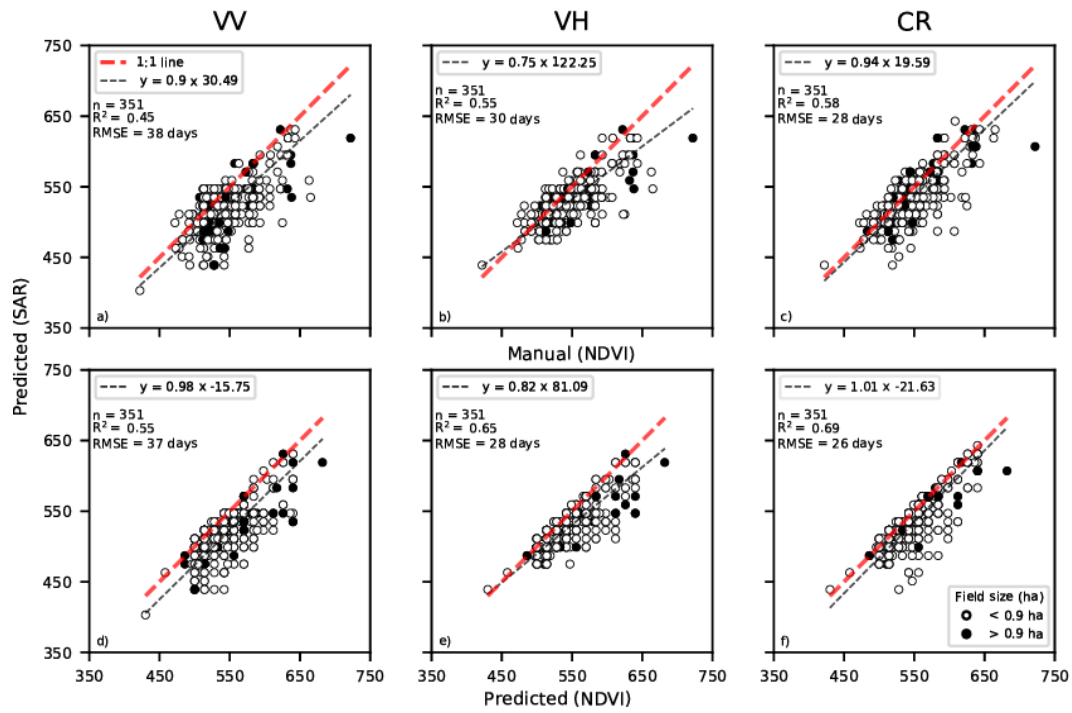


Figure 4.15. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived end of the season (EOS) (y-axis) and EOS manually interpreted from raw NDVI time-series profiles (x-axis) for fields located in Unit2 (a-c). The relationship between the EOS derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the EOS automatically derived by processed NDVI time-series profiles for fields located in Unit 2 (x-axis) (d-f).

Initial knees from Sentinel-1 backscatter and start of mid-season NDVI time-series

This section compares the results of the optically derived start of the mid-season (SMS) and the results of the initial knees derived using VV, VH and CR backscatter time-series for fields located in Unit 2, normalised as Unit 2-DOS.

For the Unit 2 evaluation dataset, the predictions for the initial knee derived from VV, VH, and CR ranged from 127 to 475 Unit 2-DOS, 127 to 403 Unit 2-DOS, and 127 to 427 Unit 2-DOS respectively. The manually derived SMS from visual interpretation of raw NDVI time-series ranged from 198 to 401 Unit 2-DOS and the automated SMS derived from processed NDVI time-series ranged from 192 to 416 Unit 2-DOS.

The initial knee derived from VH, VV, and CR depicted in Figure 4.16a-c show a weak relationship with manual estimates of SMS derived from visual interpretation of raw NDVI time-series data, R^2 values of 0.01 and 0.04 (see Figure 4.16a-c).

Figure 4.16d-f also shows a weak relationship between predictions of the initial knee derived from VV, VH and CR backscatter and processed NDVI time-series with respective R^2 values of 0.02 and 0.04. Interestingly, although there was no strong linear relationship the initial knees derived from VV, VH and CR backscatter time-series profiles were earlier than the SMS derived from NDVI time-series.

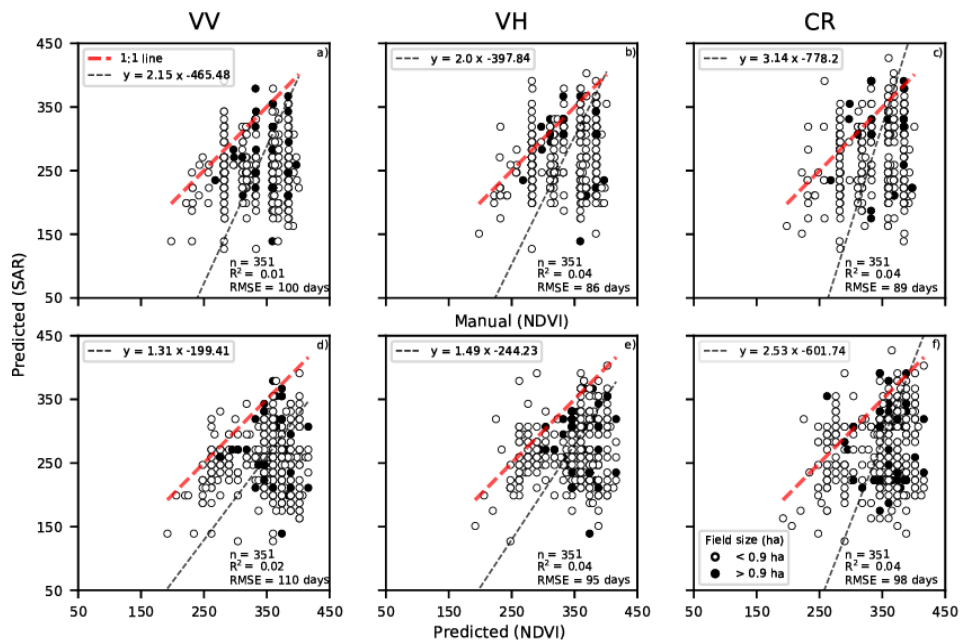


Figure 4.16. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived initial knees (y-axis) and the start of mid-season (SMS) manually interpreted from raw NDVI time-series profiles (a-c) (x-axis) for fields located in Unit 2. The relationship between the start of mid-season (SMS) derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the initial knee automatically derived by processed NDVI time-series profiles (x-axis) for fields located in Unit 2 (d-f).

Second knees from Sentinel-1 backscatter and end of mid-season NDVI time-series

This section compares the results of the optically derived End of the mid-season dates (EMS) with the second knee dates derived using VV, VH and CR backscatter time-series, normalised as Unit 2-DOS.

For the Unit 2 evaluation dataset, the predictions for the second knee derived from VV, VH, and CR ranged from 331 to 607 Unit 2-DOS, 367 to 619 Unit 2-DOS, and 379 to 619 Unit 2-DOS respectively. The manually derived EMS from visual interpretation of raw NDVI time-series ranged from 361 to 540 and the automated EMS derived from processed NDVI time-series ranged from 290 to 514 Unit 2-DOS.

The second knee derived from VH, VV, and CR depicted in Figure 4.17a-c showed a weak positive linear relationship with manual estimates of EMS derived from visual interpretation of raw NDVI time-series data, with R^2 values of 0.15, 0.14 and 0.12, respectively and RMSE values of 52, 46, and 47 days (see Figure 4.17a-c). Figure 4.17d-f shows that the relationship predictions of the second knees derived from VH backscatter, and processed NDVI time-series were stronger than the relationship between manually derived SMS dates with an R^2 value of 0.19 and RMSE of 51 days. The estimations of the second knee derived from all three backscatter time-series (VV, VH, and CR) were systematically later than the estimates of the EMS derived from the raw and processed NDVI time-series, and the systematically later relationship was stronger for the VH derived initial knee.

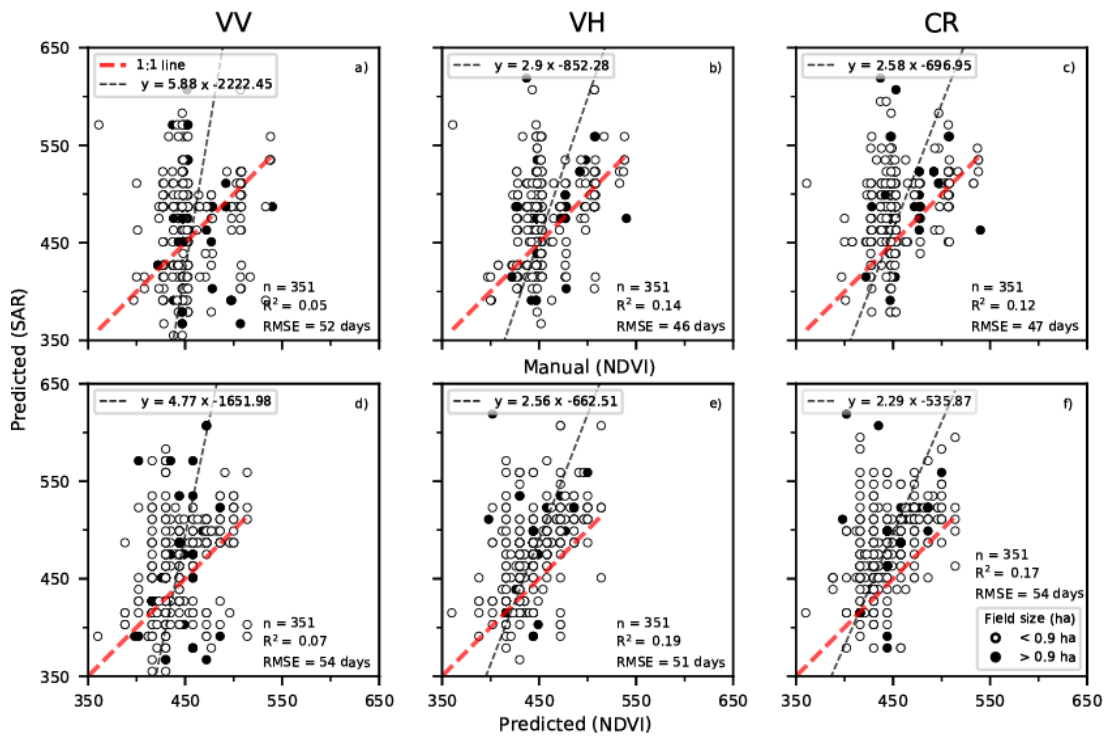


Figure 4.17. The relationship between the Sentinel-1 back scatter (VV, VH, CR) derived second knees (y-axis) and the end of mid-season (EMS) manually interpreted from raw NDVI time-series profiles (a-c) (x-axis) for fields located in Unit 2. The relationship between the end of mid-season (EMS) derived by Sentinel-1 back scatter (VV, VH and CR) (y-axis) and the initial knee automatically derived by processed NDVI time-series profiles (x-axis) for fields located in Unit 2 (d-f).

4.4 Discussion

4.4.1 Seasonal decomposition

The study first investigated the advantages of using Sentinel-1 backscatter data over processed optical data in Chapter 3 for splitting the seasons for overcoming cloud. Despite its ability to penetrate clouds, the results showed no benefit of using only Sentinel-1 backscatter time-series datasets for decomposing long time-series into seasons over processed NDVI time-series datasets.

When the season splitting was applied to the Sentinel-1 backscatter time-series datasets only for fields located in both Unit 1 and Unit 2, there was no systematic relationship between the SOS manually interpreted from raw NDVI time-series

data and the SOS automatically derived from the VV, VH and CR backscatter time-series profiles. Instead, most Sentinel-1 back scatter predictions for the SOS were significantly later than the manually derived values for the SOS. One reason could be the variability of the SAR time-series data compared to the processed optical time-series. The splitting algorithm developed in Chapter 3 was dependent on the presence of sigmoid curves. The FFT in the methodology was used as a tool to simplify the time-series profiles for each field into sigmoid or bell-shaped curves, where there is an increase during the initial phase, followed by a levelling off in the crop development phase and a gradual decrease during senescence. However, the relative height of the backscatter profiles, relating to small structural differences in the sugarcane crops over time, varies more frequently over time, compared to the relative height of the NDVI time-series which relates to less frequent larger biomass differences over time, making the FFT algorithm unsuitable for smoothing the highly variable SAR signal, affecting the shape of the time-series profile and subsequent steps in the seasonal decomposition methodology. The Sentinel-1 backscatter time-series of fields in Unit 1 and Unit 2 is interrupted by fluctuations in backscatter during the mid-season (Figure 4.3 and Figure 4.4). These fluctuations in backscatter behave the same way as clouds in the processed optical time-series, whereby erroneous low values of NDVI during the crop development stage resulted in overestimates for the SOS for the processed NDVI time-series.

Unlike the Sentinel-1 automated estimates for SOS, there was a relationship between the EOS manually interpreted from raw NDVI time-series data and the EOS automatically derived from the VV, VH and CR back scatter time-series profiles. The relationship between manual estimates of EOS and those predicted using processed NDVI time-series was better than the EOS predictions derived from Sentinel-1 backscatter time-series. Like with the predicted SOS, one of the reasons for the weaker relationship between manual estimates of the EOS and Sentinel-1 derived estimates of the EOS could also be the fluctuations in the SAR time-series data.

Combining NDVI time-series profiles and Sentinel-1 data overcame the challenges of using only Sentinel-1 datasets for seasonal decomposition.

Combining NDVI and VH-backscatter time-series data produced the most improved estimates of SOS and EOS for Unit 1 and SOS for Unit 2. However, the combined use of NDVI and CR backscatter produced the most improved estimates of EOS for Unit 2, although this was only slightly better than the use of NDVI and VH with an R^2 value of 0.69 compared to 0.65. An offset between the SOS and EOS derived from NDVI only and the SOS and EOS derived from combined NDVI and Sentinel-1 backscatter was found. The SOS derived from combined NDVI and Sentinel-1 backscatter was systematically earlier than the SOS derived from only NDVI. EOS derived from combined NDVI and Sentinel-1 backscatter was systematically later than the EOS derived from NDVI only. These were different points in the temporal profile than the SOS and EOS derived from only NDVI data.

The results for splitting of the seasons highlight the difference between the type of data recorded using SAR and optical satellites and hence the importance of optical data for splitting the seasons. The differences in the SOS and EOS between the two methods are related to the properties of the backscatter and NDVI signal. Sentinel-1 backscatter signal is sensitive to changes in geometrical structure and texture changes in vegetation canopy and soil surfaces, which are detected earlier than changes in biomass (reflection characteristics of vegetation canopy and soil surfaces) from the Landsat-8 and Sentinel-2 signals. The NDVI signal is also pre-processed whereby small changes over time have been smoothed out and the signal is resampled to a 14-day interval.

However, the combined use of Sentinel-1 backscatter did not overcome the challenge of clouds as the splitting of the season was still affected by clouds during the mid-season. As a result, the RMSE values of the dates when the growing stages transition into the next stage are more than 23 days, which is the best that could be achieved. An RMSE of +/-23 day could limit the usability of this information, particularly if this method is applied to short cycle crops like wheat, as the window of error would account for a large proportion of the growth cycle (~ 3 months). The period of growth for sugarcane, on the other hand, is much longer (~12 months), the window of error accounts for a much smaller proportion

of the growth cycle therefore the limitations of using the automated growth stages are much less than they would be for short cycle crops.

For improved estimates of the SOS and EOS, future studies should investigate alternative methods that do not rely on the FFT algorithm and use NDVI time-series data for decomposing long Sentinel-1 time-series datasets. This is especially important for cases where short cycle crops are studied.

4.4.2 Comparison of SAR-derived and NDVI initial and second knees

Initial knee

As shown in Figure 4.18, it is apparent that the initial knee derived from the SAR data is different from the initial knee derived using the NDVI data. Figure 4.10 and Figure 4.16 also indicate an earlier VH-derived initial knee compared to the SMS derived from process NDVI time-series profiles for fields in Unit 1 and Unit 2. The field photographs presented in Figure 4.18 shows that the VH-derived initial knee and SMS derived from process NDVI time-series data represent different stages of the sugarcane growth cycle. The initial knee derived from the VH dB time-series represents a growth stage where the leaves of the sugarcane crop are developing (see Figure 4.18a). The leaves at this initial knee, derived from VH dB, are relatively short, the soil between sugarcane rows is exposed, and each sugarcane plant can be identified. By contrast, the start of the mid-season stage (SMS) derived from processed NDVI time-series data coincides with a time when the leaves of the sugarcane crop are fully developed and much longer than the leaves of the sugarcane crop at the initial knee derived from VH dB. The soil between sugarcane rows is no longer exposed at the SMS stage as the long leaves for each sugarcane crop combine to create a complete canopy (see Figure 4.18b).

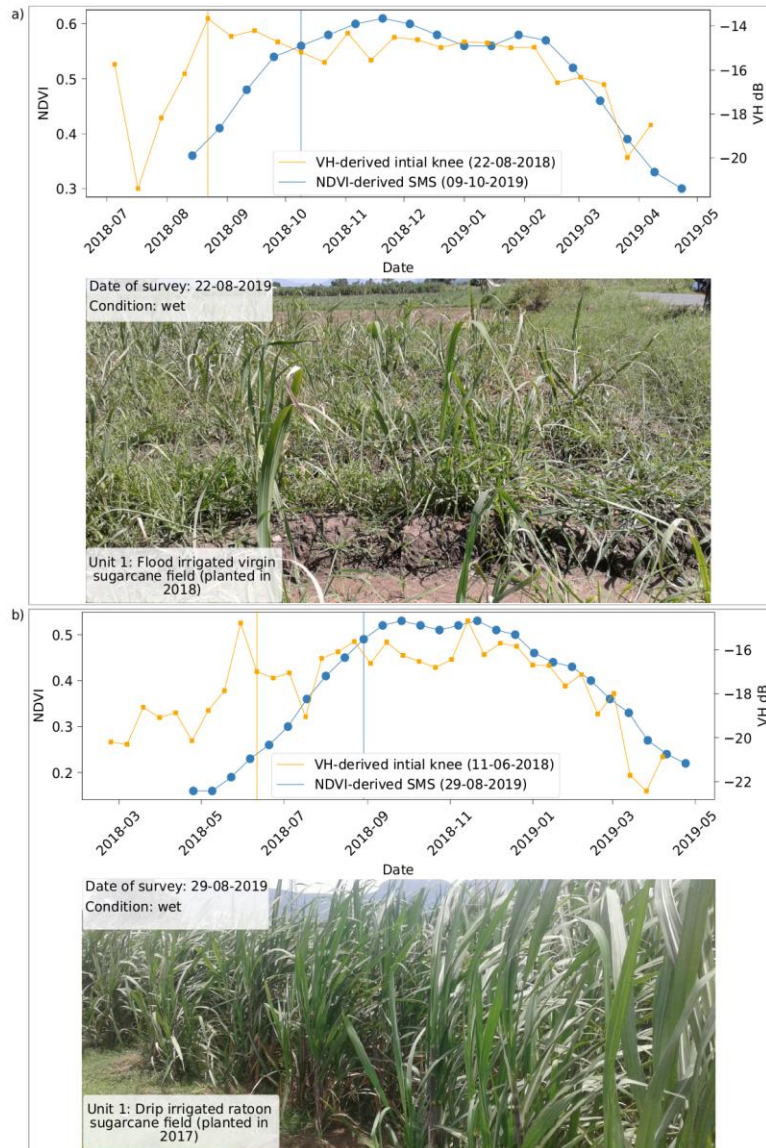


Figure 4.18. Mean VH from processed Sentinel-1 Ground Range Detected (GRD) time-series profiles of an example sugarcane field during the 2018 season, together with pre-processed mean NDVI time-series profiles. The primary y-axis on the left displays NDVI (in blue) and the date of the Start of the Mid-season (SMS) derived from the mean NDVI time-series profile is displayed as the solid blue vertical line. The secondary y-axis displays VH backscatter (in orange) on a decibel scale. The date of the initial knee derived from the mean VH profile is displayed as the solid orange vertical line. The photograph below a) was taken on the same day that the initial knee was derived from the mean VH profile and the photograph below b) was taken on the same day that the SMS was derived from the mean NDVI profile.

Second knee

In a similar way to the first knee, the second knee derived from VH dB time-series data seems to identify a different crop stage than the EMS derived from processed NDVI time-series data. Observation of Figure 4.19 suggests that whereas the second knee (EMS) derived from the processed NDVI represents the onset of senescence, the second knee derived from VH dB time-series data indicates the onset of harvest (last observation before harvest). As shown in Figure 4.19, the VH backscatter values tend to remain high even as there is a drop in NDVI after the derived EMS, suggesting that the sugarcane senesces before it is harvested. The decline in NDVI could be related to reduced green biomass after the EMS but the continued stability of VH dB suggests the maintenance of a similar crop structure. By contrast the second knee derived from the VH dB time-series is followed by abrupt decrease suggesting a dramatic structural change such as crop harvest. The differences in growth stages between the EMS derived from processed NDVI time-series data and the second knee derived from VH dB time-series data are supported by Sentinel-2 False colour composite (FCC) reference images. In Figure 4.19a and Figure 4.19b (Sentinel-2 FCC image composites for before and after the NDVI-derived EMS), the field boundaries are red indicating the presence of vegetation, despite a reduction in NDVI. At Figure 4.19c (Sentinel-2 FCC image composite after the VH-derived second knee), a green field boundary indicates bare soil (vegetation removal after harvest).

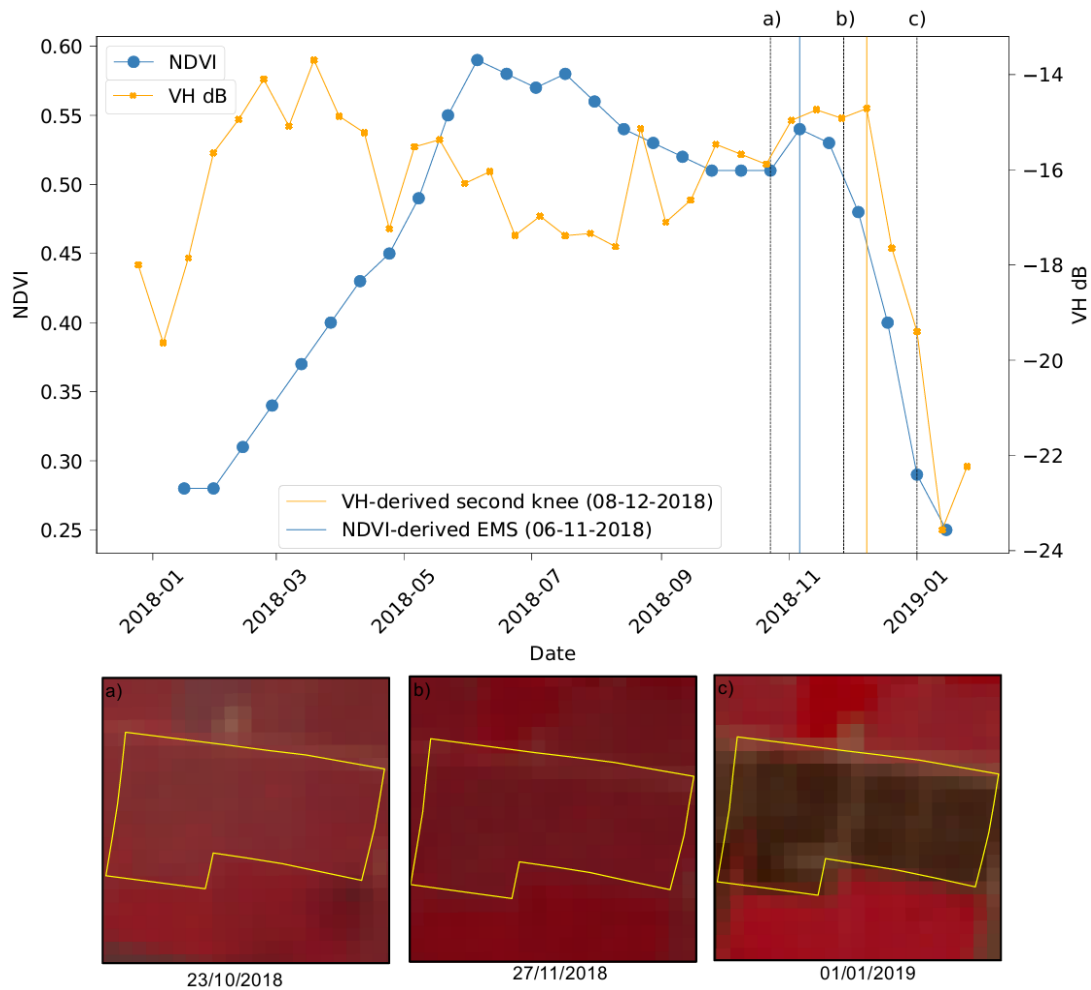


Figure 4.19 Mean VH from processed Sentinel-1 Ground Range Detected (GRD) time-series profiles of an example sugarcane field during the 2018 season, together with pre-processed mean NDVI time-series profiles. The primary y-axis on the left displays NDVI (in blue) and the date of the end of the Mid-season derived from the mean NDVI time-series profile is displayed as the solid blue vertical line. The secondary y-axis displays VH backscatter (in orange) on a decibel scale. The date of harvest derived from the mean VH profile is displayed as the solid orange vertical line. The date of three consecutive cloud free Sentinel-2 reference images a, b, and c are indicated as vertical dotted lines, the respective scenes for these dotted lines are displayed below in False colour (8,4,3), with red indicating the presence of high vegetative cover and green indicating bare soil.

4.4.3 Combining two models of sugarcane phenology

The two models of sugarcane phenology, one using NDVI and one using SAR, are presented together in Figure 4.20. The durations of the derived growth stages are also compared to those for sugarcane reported by Molijin et al. (2019) in Israel, Mall et al. (2016) in Uttar Pradesh, India, and a generic relationship reported by FAO (Doorenbos, 1977) (Table 4.1).

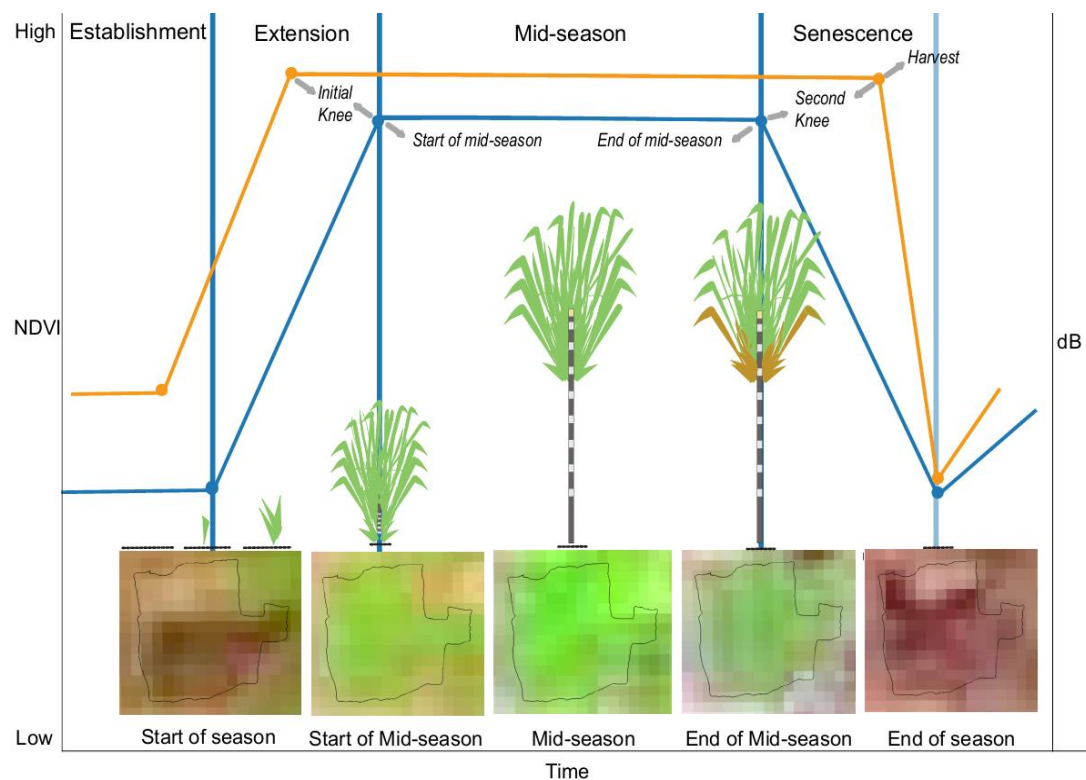


Figure 4.20. A schematic diagram demonstrating the different aspects of sugarcane growth and development derived from the processed NDVI (blue line) and VH dB (orange line) time-series data. The lower part of the diagram are Sentinel-2 image True Colour composites for an example field, with the date of the image corresponding to the date of the derived growth stage.

In Unit 1, the mean duration of the establishment stage to the initial knee derived from VH dB time-series data was 66 days for virgin and 57 days for ratoon cane (Table 4.1). In Unit 2, the corresponding mean durations were 85 days for virgin

and 98 days for ratoon cane. The values derived in Unit 1 were similar to the duration from establishment to the tillering stage reported by Molijin et al. (2019).

Table 4.1. Calculated duration (days) of five sugarcane growth stages for virgin (v) and ratoon (R) crops Unit 1 and Unit 2 in days, together with standard deviations. Values from Molijin et al (2019), Mall et al. (2016) and Doorenbos (1977) are also presented.

	Establishment to VH-derived initial knee	VH-derived initial knee to start of Mid-season	Mid-season	End of mid-season to VH-derived second knee	Harvest to End of season	Total growth cycle
Unit 1(v)	66± 31	57± 37	179±43	28±27	45±24	377±42
Unit 1 (R)	57± 25	51± 40	197±44	24±27	52±33	383±49
Unit 2 (V)	85± 44	73± 59	99±53	33±42	51±29	343±41
Unit 2 (R)	98± 57	66± 73	102±54	22±38	61±35	351±55

	Establishment to tillering	Tillering to Mid-season	Mid-season	Senescence to Harvest Harvest to End of season	Total growth cycle
Molijin et al., 2019	60	90	90	120	360

	Establishment to extension	Mid-season	Senescence to End of Season	Total growth cycle
Mall et al., 2016	120	130	110	360
FAO (V)	120	220	140	480
FAO (R)	80	180	60	320

The mean duration of the initial VH knee to the mid-season stage was also shorter in Unit 1 (51-57 days) than in Unit 2 (66-73 days). By contrast, the predicted duration of the mid-season in Unit 1 (179-197 days) was substantially longer than that (99-120 days) in Unit 2. The length from the EOS (derived using the NDVI data) to the second VH knee was similar in Units 1 and 2 (22-33 days) and in a similar way the duration between the second VH knee and the End of the season was broadly consistent between Unit 1 and Unit 2. The derived total length of the growth cycle for Unit 1 (377-383 days) and Unit 2 (343-351 days) were also broadly similar.

4.4.4 Time-series knees and management

In Unit 1 and Unit 2, the initial knees predicted using VV, VH, and CR back scatter were derived earlier than those derived from NDVI (see Table 4.1). The earlier knees in the backscatter profiles compared to the NDVI could be related to the hyper-sensitivity of the backscatter signals to soil moisture before the leaf development stage (Molijn et al., 2019). As the sugarcane canopy develops, the sensitivity of Sentinel-1 backscatters to soil moisture decreases, and the signal becomes more responsive to volumetric scatter caused by structural changes in the vegetative canopy (Molijn et al., 2019). NDVI, on the other hand, is only sensitive to photosynthetic activity in the sugarcane canopy (Zeng et al., 2022), therefore the signal saturates later than the Sentinel-1 backscatter signal at the SMS as this is the stage where the canopy becomes fully developed.

Between the two Units, the initial knees derived from Unit 1 were earlier than those derived from Unit 2. One reason could be the differences in weather patterns between Unit 1 and Unit 2. The time-series for Unit-1 refers to the 2018-2019 growing season for sugarcane fields located in Andhra Pradesh, while the time-series for Unit-2 refers to the 2019-2020 growing season for sugarcane fields in Telangana, which are likely to have experienced different growing conditions. As outlined above, Sentinel-1 backscatter is sensitive to changes in soil moisture before the leaf development stage, therefore the time-series of fields exposed to rainfall during this stage have been found to display an earlier

saturation (initial knee) than fields not exposed to rainfall during this stage (Molijn et al., 2019).

In addition to the initial knees being derived earlier than the SMS, the relationship between the derived SMS and the initial knees from VV, VH and CR backscatter time-series data was stronger for fields located in Unit 1 with respective R^2 values of 0.25, 0.40 and 0.21 compared to Unit 2, with respective R^2 values of 0.02 and 0.04. In Unit 2, it took longer for the mean backscatter time-series of each field to reach its peak than in Unit 1 and there was no period of stability after the peak, instead, backscatter decreased shortly after the maximum backscatter had been reached. The differences between Unit 1 and Unit 2 could be related to management. Unit 1 contains mostly flood-irrigated fields and Unit 2 contains mostly drip-irrigated fields. Findings from previous studies suggest that flood irrigation events in Unit 1 could mimic the appearance of rainfall events in the backscatter time-series. Previous studies have found that flood irrigation events have the same effect on Sentinel-1 backscatters as rainfall events, with increased soil moisture changing the dielectric properties of the soil which produces increased backscattering (Gao et al., 2018), unlike drip-irrigation events which are harder to detect (Ouaadi et al., 2021). Rather than depicting a rainfall event, the initial knee for Unit 1 could be depicting an irrigation event, before the onset of tillering. Determining the occurrence of irrigation events is important for the management and allocation of resources like water, especially in instances of drought, therefore, future research should use weather data and irrigation-related field data to investigate this. Where rainfall events are known, these results could be used to predict irrigation events equally where irrigation events are known, these results could be used to predict rainfall (Molijn et al., 2019), which is useful for monitoring the water use theme of indicator 1.2.4 for SDG 2 (Table 2.1).

In addition to moisture, findings by Besten et al. (2023) also suggest that the differences in the shape of the profiles between Unit 1 and Unit 2, could also be attributed to yield with fields located in Unit 2 having lower yields than fields located in Unit 1. Besten et al. (2023) found that the time-series of high-yielding sugarcane fields displaying a period of stability, from the initial knee until harvest,

in the VH and VV backscatter time-series, while VH and VV backscatter gradually declined after the initial knee in low yielding fields.

Previous studies have found a relationship between ratooning month, biomass accumulation and sucrose yield (Besten et al., 2023), affecting the shape of the backscatter time-series. Due to seasonal variation, the relationship between biomass and intercepted radiation is affected by ratooning month. This means the radiation use efficiency for sugarcane varies depending on the ratooning month (Donaldson et al., 2008). Yield is indeed affected by ratooning month. The findings of Besten et al. (2023) show how ratoon month (or season planted) can affect yield and therefore shape of sugarcane time-series profiles, however after separating the time-series for fields planted in both Unit 1 and Unit 2 and plotting their average profiles, there seemed to be no evidence of season planted affecting the shape of the time-series profile. (Besten et al., 2023) also found a delayed maximum in the VV, VH, CR, and NDVI time-series of waterlogged fields. However, there was no waterlogging in the photographs for fields within Unit 2. The time-series for Unit 1 and Unit 2 are for two completely different locations with different management practices (only 13% of fields in Unit 1 are drip irrigated compared to 87% in Unit 2) and growing years (2018-2019 and 2019-2020). Future research should therefore consider the impact of the different management practices on the shape of the backscatter time-series data and the ability to derive the initial knee. Monitoring and improving yields by adapting management is important for sustainable sugarcane production, particularly its links to SDG indicator 2.4.1 and its themes of land productivity, profitability and decent employment (Table 2.1).

4.4.5 Harvest

At the end of the mid-season the drying of the sugarcane results in a decrease in both NDVI and Sentinel-1 backscatter (Molijn et al., 2019). Compared to the gradual reduction in NDVI at the end of the mid-season, the analysis of the second knee, derived from Sentinel-1 backscatter, found abrupt changes in backscatter indicating harvest. The harvest is a different point in the temporal

profile, that was systematically later than the EMS derived from NDVI time-series data. The differences between these points in the sugarcane growth profiles are related to the properties of the sensors.

NDVI is sensitive to the reflection characteristics of vegetation canopy and soil, such as changes to the colour of the sugarcane canopy as the crop starts to senesce which causes a reduction in NDVI. Changes in NDVI at harvest are smoothed out with pre-processing and resampling at a 14-day interval. The Sentinel-1 backscatter signal was not smoothed and was more sensitive to changes in surface structure and texture changes such as vegetation removal after harvest. VH backscatter was the most sensitive to harvest as the signal is sensitive to only volumetric scattering from vegetation canopies. VV and CR backscatter, on the other hand, continue to receive backscatter from the soil after the vegetation is removed.

Identifying the harvest is important for sustainable management. Harvesting should be completed within 2-3 months after flowering (Pavani et al., 2023) to avoid losing sugar. Table 4.1 shows that sugarcane is harvested ~30 days after EMS, in Unit 1 and Unit 2. This is an important finding for sustainability, particularly SDG indicator 2.4.1 (see Table 2.1) and its themes of land productivity, profitability and decent employment, as it shows that the fields in Unit 1 and Unit 2 were not harvested too early, with farmers waiting for senescence (the yellowing of leaves), which allows for the conversion of fructose and glucose to sucrose therefore higher yields (Mall et al., 2016). A harvesting time of ~30 days after EMS also shows that Unit 1 and Unit 2 were not harvested too late, to avoid reduced sucrose concentration in the stalks. The CR backscatter at harvest can also indicate yield. Low CR backscatter at the point of harvest has been found to correlate to sugarcane crops with higher yields, and higher sucrose concentration in the stalks (Besten et al., 2023). Sugarcane must be transported to the factories less than 24 hours after harvest to reduce post-harvest loss as the cane weight and juice sucrose content reduce considerably beyond this time, impacting both growers and the sugar industry. Monitoring sugarcane harvest dates in near-real time would be useful for managers of sugarcane factories

helping to plan transport routes and schedule the allocation of resources for crushing operations to avoid post-harvest losses (see Table 2.1).

To evaluate the use of SAR data for deriving sugarcane growth stages, this study regressed the predicted SAR growth stages against the NDVI derived growth stages and compared its slope and intercept values, which is a common approach for evaluating models (Piñeiro et al., 2008). However, there is no consensus to which values should be on the X and Y axis and switching which values are placed on either the X or Y axis could produce different results, in turn changing the conclusions of the model (Piñeiro et al., 2008). While this is an issue, the orthogonal regression was used in place of linear regression to minimise any effect.

4.5 Conclusion

This chapter highlights the benefits of combining optical and SAR time-series information for monitoring sugarcane growth and management over time. The different properties of the optical and SAR instruments provide complementary information on sugarcane growth and management. Biomass signals from optical sensors were stronger than those from SAR, but SAR was able to provide additional information from structural differences in the time-series data. We can see this in the difference between the “knees” or inflexion points derived from Sentinel-1 and NDVI time-series datasets. Sentinel-1 backscatter data can provide information about the timing of tillering growth stage and water availability where weather, irrigation, and yield information is available. The results also demonstrate that sugarcane harvest date can be observed from the sharp decline in VH backscatter, improving the determination of harvest date compared to optical datasets. Combining the information from SAR and optical sensors improves sugarcane growth monitoring and can be used as a management tool to aid sustainable sugarcane production by monitoring several themes of SDG indicator 2.4.1 including water use, land productivity and profitability (Table 2.1). Additionally, it is important to note that the optical and SAR data used in this study are freely available from the GEE platform, these methods can be applied more widely for different study areas and other crops.

5 FINDING THE SWEET SPOT: AUTOMATING SUGARCANE DETECTION FROM LONG-TIME-SERIES DATA

5.1 Introduction

Sustainable production requires information on where sugarcane cultivation occurs over large areas and multiple years. This information when used when combined with other datasets, allows sugarcane operators to make informed choices about sustainable production by comparing productivity and yield with different environmental conditions and management practices across study areas and years. Comparing management over time and space is critical for managing resources like water, fertilisers and pesticides (Mall et al., 2016), based on the timing of planting and harvesting. Cross-referencing this information with yield datasets of regional averages is critical to assessing whether sustainability interventions are working.

Classification of EO images is one method for determining where sugarcane production occurs. Once the locations of sugarcane fields are identified, factory managers and government planners may be able to identify the extent of sugarcane production and additional image interpretation could be used to predict the size of potential harvests. The ability to monitor large areas quickly over large spatial and temporal scales also allows for assessments of land use change, with potential implications for changes in biodiversity and soil carbon storage. The method may also be used to determine the use of crop rotations in between sugarcane growing cycles. Many sugarcane classification algorithms focus on the spectral aspect of remote sensing imagery using single-day and sometimes multi-day images. However, these approaches can perform poorly when separating sugarcane from spectrally similar classes such as paddy, rice and maize (Liu et al., 2023; Wang et al., 2020).

Time-series based classification algorithms can detect changes over time, as the algorithms involve the use of multiple chronologically ordered EO datasets instead of using images derived from one point in time, allowing landcover to be assessed continually over many different spatial and temporal resolutions (Morel et al., 2014). The phenology of sugarcane means that it can have a distinct

pattern in satellite time-series. Therefore, time-series classification-based approaches could be more robust than single-day imagery for sugarcane classification.

The capacity of machine learning classification for handling complex datasets such as time-series data is far greater than simple parametric classification methods such as the maximum likelihood (ML) classifier, therefore machine learning models are often used for time-series based classification (Rogan et al., 2008; Yu et al., 2013). Machine learning approaches for the classification of sugarcane and other crops, using time-series data, include shallow learning models such as expert decision trees (DT) and Dynamic time warping (DTW). These approaches rely on crop calendar information for isolating individual sugarcane growing seasons and subsequently deriving thresholds based on phenology or the shape of the isolated time-series signal, therefore they work well for classification in the Philippines and China where the start of sugarcane production occurs at a similar time each year (Ofindo et al., 2020; Wang et al., 2020). However, calendar information is unavailable for India as the start of the sugarcane season can be variable, which limits the application of these threshold-based classification approaches.

Recent studies for the classification of crops in the United States of America, and France have demonstrated that deep learning using convolutional neural networks (CNNs) in remote sensing-based classification studies can outperform shallow learning models such as DT, DTW but also Random Forest (RF) and Support Vector Machines (SMV) for classification tasks. Hence, there is interest in developing temporal classification algorithms not based on pre-calculated temporal features and thresholds extracted from vegetation index time-series, but the use of deep learning algorithms which automatically learn features from training datasets. However current temporal CNN models developed for the classification of crops are also inappropriate for sugarcane classification in India. They too require single season input data, with each input time-series labelled as belonging to one class (Pelletier et al., 2019), limiting their approach to study regions where crop calendar information is available. In India and other areas

where management is variable, the locations of sugarcane fields are needed to derive known growth information.

Additionally, using single-season data for classification has limited application for rapid monitoring biomass change over time with multiple time-series inputs and outputs needed. If sugarcane can be classified from long-continuous time-series data over many seasons, studies such as those by Nandhini and Padmavathy (2017) could be adapted to observe continuous changes in sugarcane production continuously over a sugarcane growth cycle, as opposed to at one point during sugarcane growth, and at a field level as opposed to Statewide. This is important as the study by Nandhini and Padmavathy (2017), found irrigation to be the key to increased sugarcane productivity, but was unable to determine the best times to irrigate the fields (at which growth stages) because of its limited spatial and temporal resolution of the data.

Automatically identifying sugarcane from long-time continuous series data, is highly desirable to overcome the reliance on crop calendar data. Hence, the objective of this study is to develop a model that would allow for sugarcane and other crops to be localised and classified from long NDVI time-series data in one step and test its robustness.

5.2 Materials and methods

5.2.1 Convolutional neural networks in remote sensing

Convolutions can filter data by producing an output that is the weighted average of the input data. The filter is called a kernel and acts like a moving average. The kernel moves across the input data and takes the dot product of two matrixes. To do this, values of the input matrix within the defined kernel, are multiplied by the kernel size and this value is saved into a new matrix (feature map) by taking the sum (Zhao et al., 2024) (Figure 5.1).

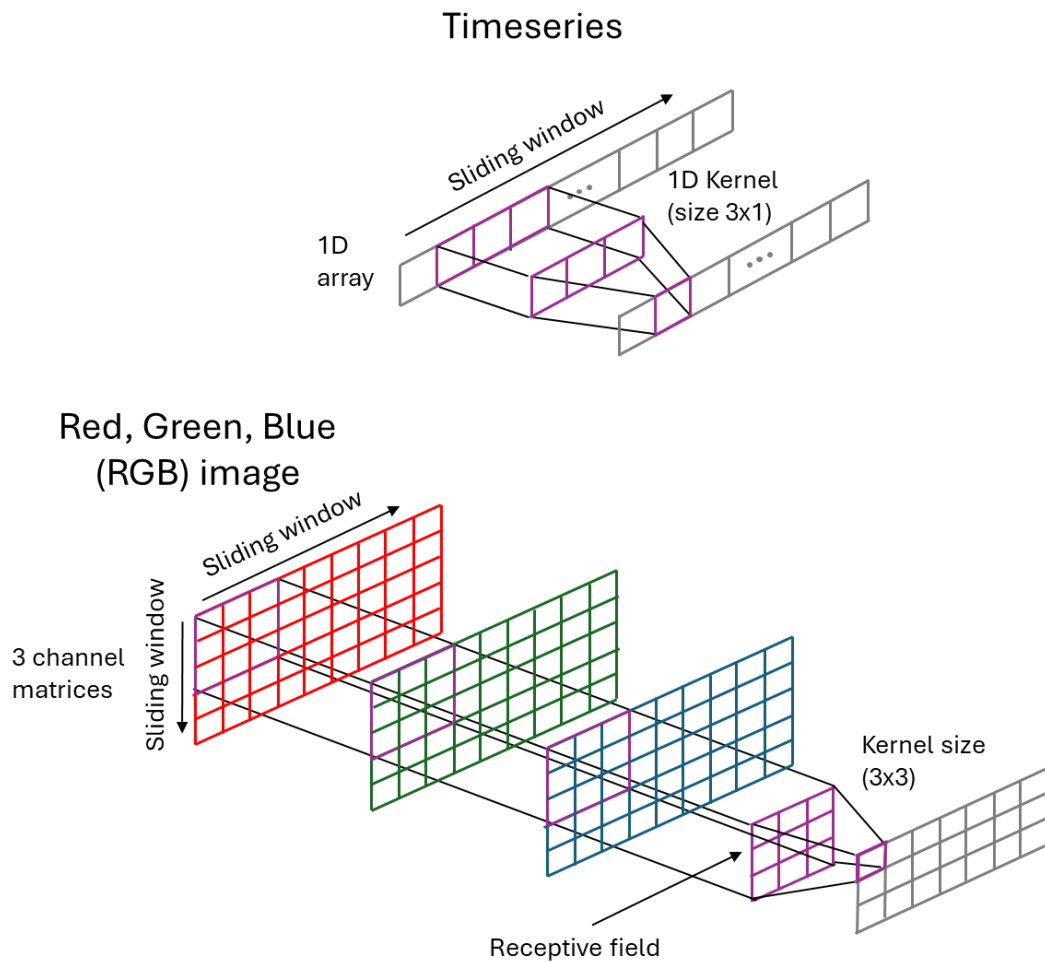


Figure 5.1. A schematic diagram demonstrating the process of convolution in 1-dimensional time-series data and red, green, and blue (RGB) images.

CNNs are a deep learning approach that automatically learn features from training data hierarchically (Lima and Marfurt, 2020). CNN have several layers that are important for feature extraction and classification. For feature extraction, the following layers are important: input, convolution, batch normalisation, activation and pooling.

Not all input data is structured the same. Data within natural photographs and EO images are organised as 2-dimensional arrays (2D) with multiple channels, while time-series data is structured as a one-dimensional (1D) array. CNN models based on natural photographs have 2D inputs with 3 channels, as different features within images are often displayed as varying intensities of red, green

and blue (RGB). Each red, green and blue component has different values, which are taken as the input. Input data can be split into batches for efficient training therefore, the input data can also be referred to as a batch. Time-series data is structured as a one-dimensional (1D) array, hence 1D CNN models are designed to handle 1D ordered arrays as input data and, therefore carry out tasks such as classification within time-series data. In 1D CNN models the kernel moves in one direction, rather than two, to derive key features from the input data to train the model (Figure 5.1).

Convolution occurs within the convolutional layers. CNN models can have one convolutional layer but depending on the complexity of the classification task, layers can be added to deepen and increase the complexity of the model. Using images as an example, CNN models tend to learn general features (textures, lines, edges) within the convolutional layers closest to the input data, learning more complex features (complete shapes) in later convolutional layers (Pires de Lima and Marfurt, 2020). Increasing the convolutional layers in CNN models can enhance model performance, but this is met with a trade-off of increased run time and training (Zhao et al., 2024). Each kernel in a convolutional layer can detect a different feature and produce its feature maps. This means the convolutional layer can perform several convolutions simultaneously, resulting in many linear activations (Goodfellow et al., 2016).

Batch normalisation layers normalise the activations across each feature map separately. This is an important step as it prevents the model from overfitting the training data (Goodfellow et al., 2016).

Before model weights are trained, each activation is run through a nonlinear activation function, such as the rectified linear activation function (ReLU) to introduce non-linearity into the model (Goodfellow et al., 2016). This allows the model to learn complex patterns in the data and can reduce the sum of errors in a classification task (Pires de Lima and Marfurt, 2020).

Pooling layers can simplify the feature maps by reducing their size, whilst keeping important information useful for the discrimination of classes (summary statistics for specified portions of the convolved input). Max pooling reduces the resolution

of the feature maps by taking the maximum value in an area of the convolved array (specified by the pooling size), to produce an output feature map with only the maximum values of the original feature map (Goodfellow et al., 2016). Stride pooling, on the other hand, reduces the resolution of the data by missing out certain values. Pooling is useful as simplifying the feature map can help to reduce the time it takes to run the model.

Once feature extraction is complete, the fully connected layer takes all the extracted information, flattens it and assigns it to a class to produce a final prediction in the output layer.

5.2.2 Model selection

The method used, was to develop a CNN model to perform sugarcane classification from 6 years of processed time-series data from India without dependence on field data and crop calendars. The developed model was based on the approach described by Kensert et al. (2021) for the automated detection of peaks in reversed-phase liquid chromatography using the Python package Tensorflow. The approach by Kensert et al. (2021) is itself based on the You Only Look Once (YOLO) model developed by Redmon et al. (2015) for rapidly detecting objects of interest from natural images and artwork. The YOLO model works by taking complete images as an input and outputting predicted class probabilities for an object of interest, e.g. a cat, for 2D portions of an image called bounding boxes (0 being no and 1 being yes). Unlike the original YOLO model, the model implemented by Kensert et al. (2021), did not define exact bounding boxes (x-coordinates and widths). Instead, the model took a training set of 1,000,000 chromatograms each with 8192 observations as the input data to produce predictions for the x-coordinate location of the peaks, associated peak probabilities (0 being no and 1 being yes) and peak area. I chose to adapt this model to classify sugarcane because of the similarity in behaviour between the unique trapezoid shape of sugarcane within time-series datasets, and peaks in chromatogram datasets. The idea is that if sugarcane can be automatically detected from long time-series data by adapting the model developed by Kensert

et al. (2021), localisation and classification can be carried out in the same step, overcoming the limitations of classification approaches that require crop calendar information to isolate patterns in time-series data before classification. To our knowledge, this is the first time this approach has been used for EO data.

5.2.3 Study area

The Indian sugarcane factories located in Chittoor District, Andhra Pradesh (Unit 1) and Zaheerabad in the Sangareddy district of Telangana (Unit 2) that were used in Chapters 3 and 4 were used in this study (Figure 5.2 and Figure 5.3). Both Indian States have a tropical climate with wet and dry seasons. Unit 1 was surveyed in 2018, and Unit 2 was surveyed in 2019. During these surveys field polygons were digitised by walking the field perimeter with a GPS device and observations of crop height, irrigation method (drip or flood irrigated), and ground photography were also taken. All sugarcane fields in Unit 1 and Unit 2 were irrigated, with 13% drip irrigated in Unit 1 compared to 87% in Unit 2, and the remaining flood irrigated. Sugarcane crops were a mixture of virgin cane (fields planted from cuttings) from the season surveyed and ratoon cane, which regrows 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. The sugarcane field sizes range 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. In addition, fields located north of the sugarcane fields, where other crops were grown, were also included in this study. In this chapter 246 sugarcane field polygons in Unit 1 and 103 in Unit 2 were included in this study. During the 2018 and 2019 surveys, photographs of the fields north of each sugarcane field in Unit 1 and Unit 2, now referred to as the northern fields, were also taken. Where the photographs showed evidence of other crops being grown in the northern fields, field boundaries were digitised using the digitisation tool in ArcGIS Pro. In total 78 and 103 northern field boundaries were digitised for Unit 1 and Unit 2, with field sizes ranging from 0.11 ha to 10.87 ha respectively (Figure 5.2 and Figure 5.3).

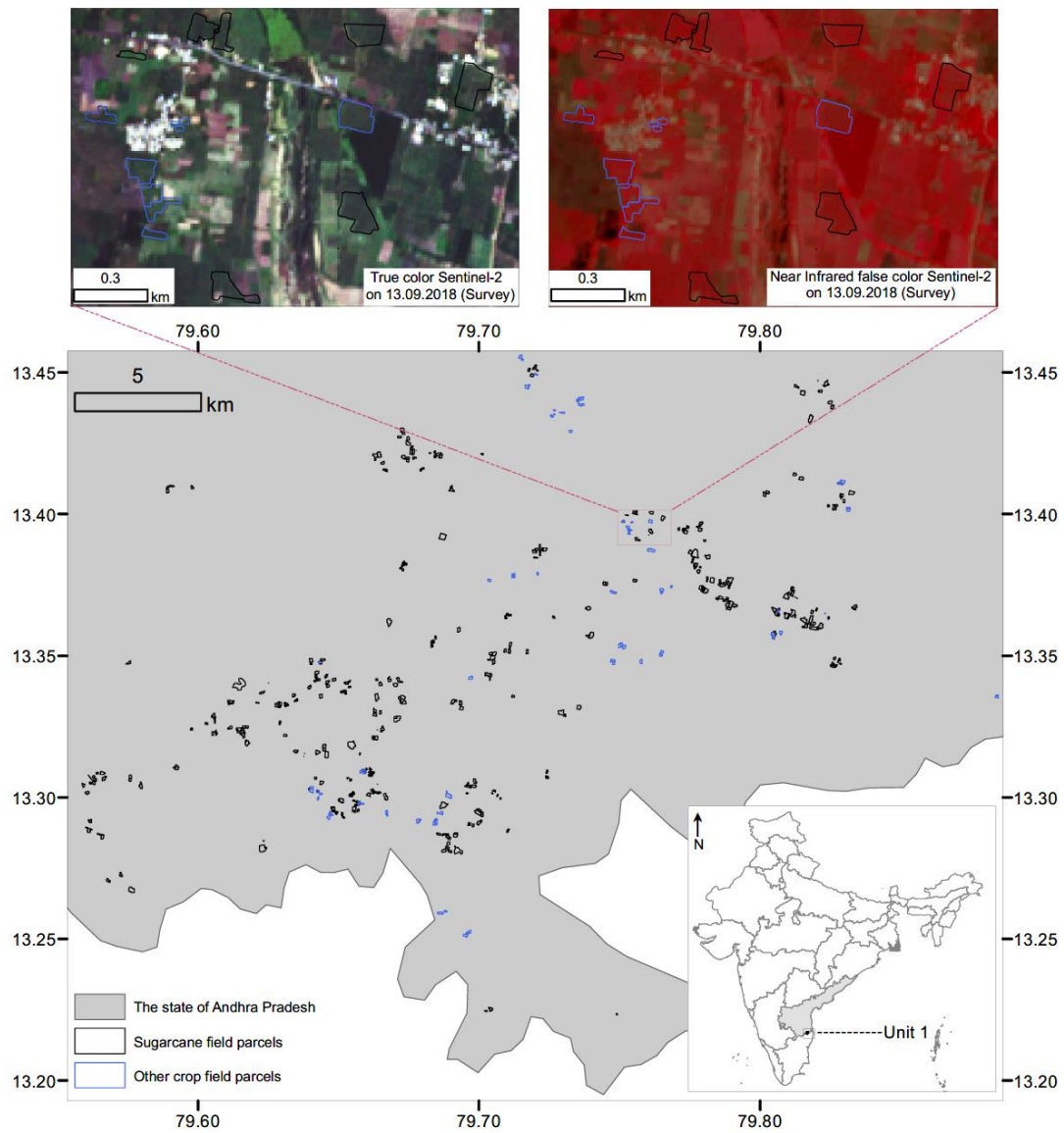


Figure 5.2. Sugarcane and other crop fields associated with processing Unit 1 located in the southern Indian state of Andhra Pradesh.

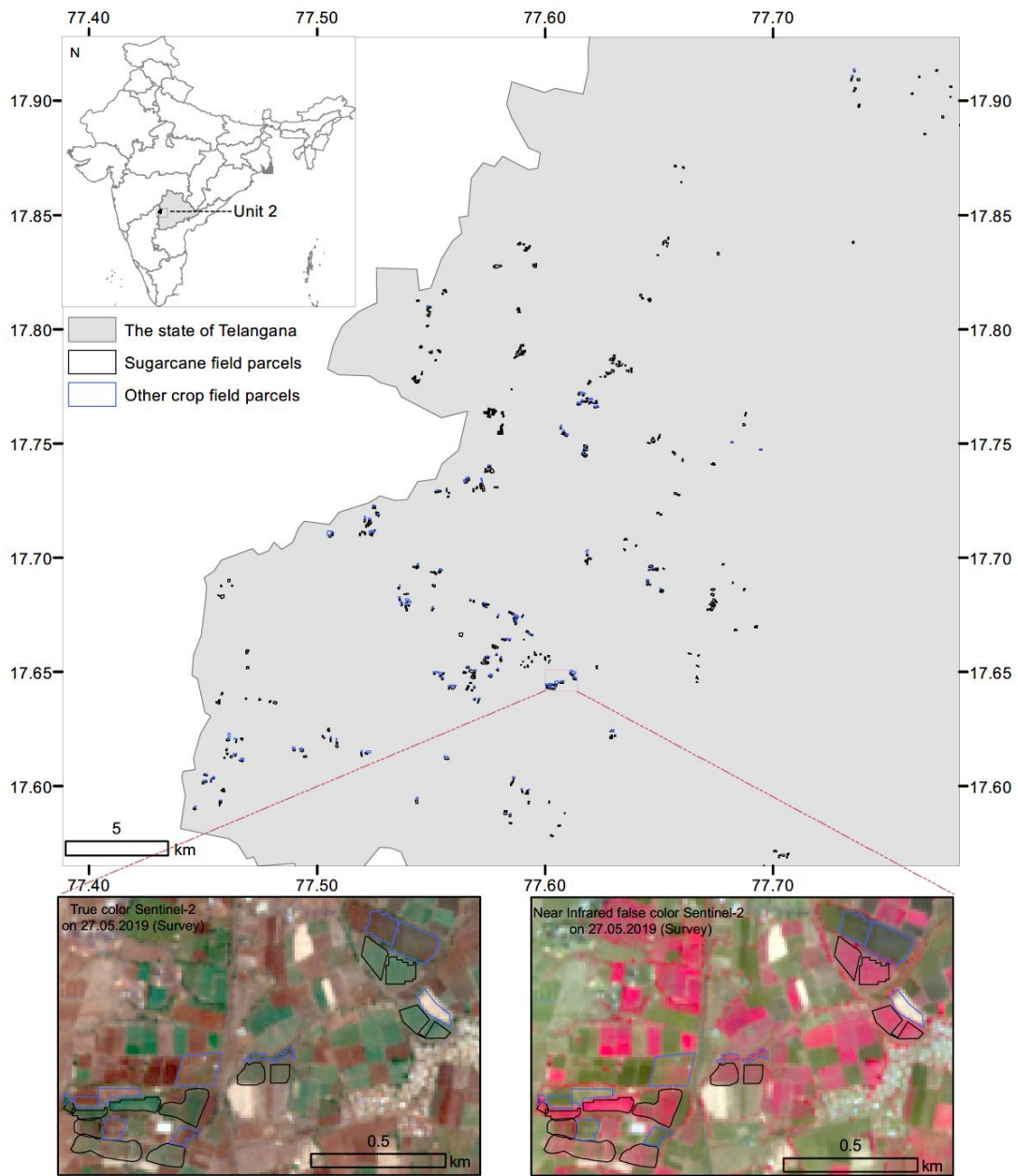


Figure 5.3. Sugarcane and other crop fields associated with processing Unit 2 located in the southern Indian state of Telangana.

5.2.4 Source of NDVI data

To develop an approach for automatically classifying sugarcane and other crops from long NDVI time-series data, the mean reflectance values for the red and Near-infrared (NIR) electromagnetic wavelength bands were obtained for each

sugarcane and northern field located in Unit 1 and Unit 2, from Landsat 8 and Sentinel-2 TOA satellite datasets using the GEE API in python. NDVI was then calculated using Equation 5.1 which was followed by cloud filtering, sensor harmonisation, and BISE filtering to remove additional noise and re-sampling to 14 days using the pre-processing methodology developed in Chapter 3. To maximise the number of observations used for training and evaluating the developed classification model, the study period was extended from one year (2018 or 2019) to over 6 years (September 2015 to May 2022), which meant that the mean time-series for each field contained just over six growing seasons worth of information.

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

Equation 5.1

5.2.5 Model implementation

The model designed by Kensert et al. (2021), took each 1-dimensional (1D) array with 8192 datapoints (size 8192 x 1) representing whole liquid chromatograms as inputs and produced 3-dimensional (3D) output predictions with 256 data points each with an assigned value for peak location, probability and area (size 256 x 3). Each mean NDVI processed time-series for fields located in Unit 1 and Unit 2, contained 176 data points (size 176 x 1), representing 6+ years of information. These 1D array datasets were much simpler than the input chromatograms used by Kensert et al. (2021), so to account for these differences, the model was modified. The first modification was changing the size of the output array. Because the input NDVI time-series for each sugarcane field was much shorter than the input chromatograms, the input time-series was reduced from an array with 176 columns and 1 row, to produce 1D output arrays with 44 columns and 3 rows using two convolutional blocks with a kernel size of 3 for both convolutional layers. Rather than an output that contained prediction values for peak probability, area, and size, each row in the output produced by the CNN model of this study had prediction values for each data point concerning whether

it belonged to the no-data, sugarcane or other crop class, with values of 0 (representing not belonging) or 1 (representing belonging) for each class. The two convolutional blocks also contained a batch normalisation layer taking batches of 10-time-series profiles, a rectified linear activation layer (ReLU), and a max pooling to reduce the input of each layer by a factor of 2. To create the 44 x 3 output with predicted probabilities for no-data, sugarcane and other crops, the last convolutional block applied three 1 x 1 filters. Like the model implemented by Kensert et al. (2021), the classification predictions as to whether an observation in the time-series belonged to the no-data, sugarcane and other crop class, were also passed through a sigmoidal function, not shown in Figure 5.4.

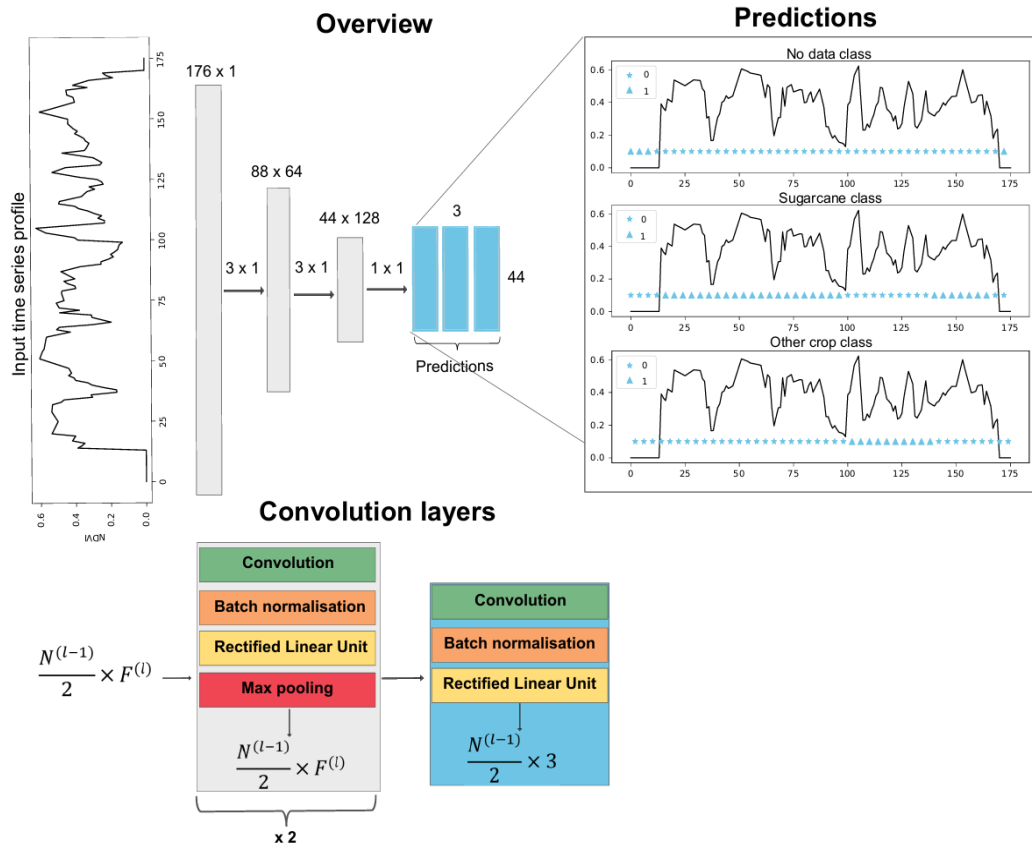


Figure 5.4. The architecture of the initial convolutional neural network (CNN) model for sugarcane detection in long-time-series data. The model starts with an input of 6-year processed NDVI time-series for each field (size 176 x 1), which is then forward propagated through the convolutional blocks (depicted in grey) to produce an output array (size 44 x 3) with predictions for “no data”, “sugarcane” and “other crop” classes. Each convolutional block contains a convolutional layer, rectified linear unit activation, batch normalisation layer and a max pooling layer (used to simplify the time-series). Note, that the dimensions written next to the arrows between each convolutional block are the kernel size (3 x 1) for each convolutional layer. The final convolutional block depicted in blue also contains a convolutional layer, a rectified linear unit activation, and a batch normalisation layer, but does not contain a max pooling layer. The dimension written next to each convolutional block in the overview represents the dimension of the data. The final convolutional block produces output predictions for whether observations in the simplified time-series belong to the no data, sugarcane and other crop classes, which can be seen in blue with (1: yes, 0: no). Adapted from (Kensert et al., 2021).

5.2.6 Model training and evaluation for identifying sugarcane

To train the CNN model, each 6-year time-series profile was manually labelled as either belonging to a “no-data”, “sugarcane” or “other crop” class using guidelines outlined in Figure 5.5.

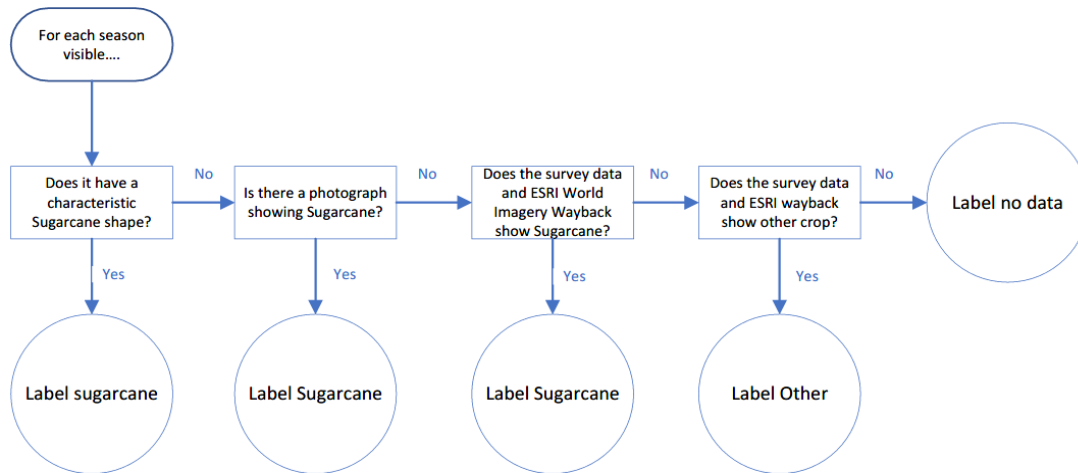


Figure 5.5. The set of guidelines used to manually label the training data as either belonging to the no-data, sugarcane or other crop class. Note: a characteristic sugarcane shape refers to a year-long trapezoid (See Chapter 3) with an increase in NDVI, followed by a period of plateau and then a period of decline in NDVI until harvest; photographs are derived from the 2018 and 2019 survey for Unit 1 and Unit 2 respectively and ESRI World Imagery Wayback was used as a form of high-resolution imagery where photographs were unavailable.

When CNN models are trained it is common for the training data to be passed through the network multiple times, with more information being learnt after each epoch (complete pass of the reference data through the network during training). To reduce computational complexity, the reference used to train the model was split into batches of 10.

Loss functions are important for measuring the agreement between the training data and the model prediction. Gradient-based optimisation algorithms ensure minimal loss by modifying the layer weights for each batch of reference data fed into the CNN. After many epochs, there comes a point where there is no longer any significant decrease in the loss, and at this point, training usually stops. The

most common gradient-based optimisation algorithm used for training CNN models is the adaptive moment estimation (Adam) algorithm because its tailored design towards deep neural networks allows for models to be trained quickly (for computational efficiency) while storing lots of data (because of its large memory) (Alzubaidi et al., 2021). Hence, Adam was also used, to minimise the negative log-likelihood between true and predicted class observations in the CNN model of this study (Kensert et al., 2021).

The chromatogram data used to train the model developed by Kensert et al. (2021) were normalised to standardise the input chromatogram training data (to a range between 0 and 1). However, NDVI already has a range between 0 and 1, therefore NDVI time-series data used in this chapter were not normalised, as further normalisation would reduce the meaning in the magnitude of data which is important for crop classification. For example, some crops have higher NDVI values than others which is an important feature in the dataset that could help discriminate it from another crop (Pelletier et al., 2019).

Ablation experiments are often used when developing or modifying CNN models for a specific role. Within ablation experiments, different components of the model architecture are changed and even removed. The impact of removing or changing these components on the performance of the overall model is assessed. This study conducted four ablation experiments to optimise the model for sugarcane and other crop classifications. The first experiment looked at the effect of balancing the data in the input classes, by increasing the number of observations of the class with the least number of observations (hereon referred to as the minority class). The second experiment looked at the effect of changing the size of the kernel within the convolutional layers and to determine the most appropriate kernel size in each layer, this was followed by a third experiment which looked at optimising the number of convolutional layers. The fourth experiment tested how well the model would generalise for classifying sugarcane and other crops in a different study area.

To evaluate the performance of the trained models for identifying periods of sugarcane or other crops in long time-series data, validation datasets containing

processed NDVI time-series profiles were used to produce a confusion matrix for the computation of recall, precision and F1 for both the sugarcane and other crop classes (Figure 5.6).

		<u>Confusion matrix for sugarcane</u>			<u>Confusion matrix for other crops</u>		
		<i>Classes predicted by the model</i>			<i>Classes predicted by the model</i>		
		No data	Sugarcane	Other crops	No data	Sugarcane	Other crops
<i>Original interpretation of classes</i>	No data	TN	FP	TN	TN	TN	FP
	Sugarcane	FN	TP	FN	TN	TN	FN
	Other crops	TN	FP	TN	FN	FN	TP

Figure 5.6. A schematic diagram demonstrating how recall, precision and F1 scores were calculated for sugarcane and other crops using the confusion matrixes for each crop.

Recall is a measure of the ability of a model to identify all instances of a particular class, while precision is the ratio of correctly classified data points belonging to a specific class out of all instances of the model's predictions of that class, and the F1 score is a measure of both recall and precision (Figure 5.6). Recall, precision and F1 can be calculated using Equation 5.2, Equation 5.3, and Equation 5.4 respectively where: TP_{class} is the number of observations that are predicted as the class of interest that were the class of interest in the originally interpreted dataset; FN_{class} is the number of observations that are not predicted as the class of interest that was the class of interest in the originally interpreted dataset; and FP_{class} is the number of observations that are predicted as the class of interest that were not the class of interest in the originally interpreted dataset.

$$Recall_{class} = \frac{TP_{class}}{(TP_{class} + FN_{class})} \quad \text{Equation 5.2}$$

$$Precision_{class} = \frac{TP_{class}}{(TP_{class} + FP_{class})} \quad \text{Equation 5.3}$$

$$F1_{class} = \frac{2 (TP_{class})}{2 (TP_{class} + FN_{class} + FP_{class})}$$

Equation 5.4

5.2.7 Experiment 1: the effect of balancing the data in the input classes

The training NDVI data from Unit 1 comprised of 233 profiles. Of these “sugarcane” was the most common class for 159 profiles while “other crops” was the most common class in the remaining 74 profile. Overrepresentation of one class and underrepresentation of another class in the training data can result in convergence during training, harming the model’s performance (Buda et al., 2018). Therefore, an initial study was completed to determine the effect of class imbalance on the classification of sugarcane and other crops. Hence the performance of the initial model, outlined in Figure 5.4, using an imbalanced training dataset was compared with a balanced training dataset. The proportion of profiles with the most common class being “sugarcane” and “other crops” were balanced by randomly oversampling the class with the least number of observations or minority class (“other crops”) in the training data. The number of profiles with the “other crops” class as their most common class was increased from 76 in the original imbalanced training dataset to 159 in the balanced training set, which is the same amount as the number of profiles that have “sugarcane” as their most common class, in the training and test data (Table 5.1). As sugarcane is the class with the most observations (majority class) the number of sugarcane profiles in the training data did not change (Table 5.1). Randomly oversampling the minority class (least common class) was chosen over randomly under-sampling the majority class (most common class), as it ensured enough data was retained in the training data without overfitting, and has been cited as being a robust method for producing high-performance classification models (Buda et al., 2018).

Table 5.1. The number of profiles in the training and test profiles for sugarcane and other crops with an imbalanced and balanced weight. N majority class refers to the number of observations in the dominant class (“sugarcane”) in the time-series profiles. N minority class refers to the number of observations in the uncommon class (“other crops”) in the time-series profiles.

Datasets	N Majority sugarcane profiles		N Majority other crop profiles	
	Training	Test	Training	Test
Imbalanced	159	74	67	23
Balanced	159	74	159	23

5.2.8 Experiment 2: finding appropriate kernel sizes for Layer 1 and Layer 2

Kernel size is an important part of the convolutional layer within CNN models and optimising kernel size (stride) is important to improving the performance of the model, as the computational speed increases as kernel size increases, but the ability to extract features decreases (Zhao et al., 2024). Hence, this study also investigated the effect of kernel size model performance. The kernel sizes investigated started from three increasing by increments of one up to eight. A value of eight was chosen as the maximum kernel size as more than eight kernels exceeded one year (one sugarcane growing cycle). The study wanted to investigate at what point in the time-series sugarcane could be classified, which would indicate whether this approach could be used for within-season classification.

5.2.9 Experiment 3: optimising the number of convolutional layers

The number of convolutional layers in a CNN model also has an impact on the model’s performance, with the model’s ability to extract high-level features and comprehensive features increasing with an increased number of layers, but this can also decrease the computational speed of the model, taking longer to run

(Zhao et al., 2024). This study therefore investigated the effect of adding convolutional layers to the performance of the CNN model for classifying sugarcane and other crops.

5.2.10 Experiment 4: generalisation

Assessing generalisation is important for evaluating the model's overall performance by checking for overfitting, as it provides an understanding of how useful the model would be for real-world applications where new training data for a new geographic area would be difficult to collect (Maxwell et al., 2021). This study assessed how well the model developed and trained on the six-year time-series for fields in Unit 1, performed for the time-series of fields located in Unit 2 (a different study area, unseen by the model). The study first compared the results of the model trained and evaluated on Unit 1 time-series data with the model trained on Unit 1 and evaluated on Unit 2 time-series data. The study then compared the results of the model trained and evaluated on Unit 2 time-series data.

5.3 Results

5.3.1 Effect of balanced classes

Balancing the training data by increasing the amount of "other crop" data, reduced the model's performance when classifying sugarcane but enhanced it when classifying other crops. Balancing the training dataset by increasing the number of "other crop" observations in the training data resulted in an increased number of false negatives (FNs) for the "sugarcane" class and an increased number of true positives (TP) for the "other crop" class (Table 5.2). Table 5.2 shows that after balancing the training data, the number of FNs for the "sugarcane class", for example, increased from 534 to 1095, resulting from an increase in the number of "other crops" being incorrectly classified as sugarcane. For the sugarcane class, increasing the amount of "other crop" data to balance the training data

reduced the recall from 0.74 to 0.47 but increased the precision from 0.79 to 0.85 (Table 5.3). By contrast, the effects were the opposite for the “other crop” class, where the inclusion of additional “other crop” data increased the recall from 0.73 to 0.89 but reduced the precision from 0.66 to 0.54. Increasing the amount of “other crop” data to balance the training data resulted in the number of TPs in the “other crop” class increasing from 1084 to 1311 (Table 5.2). In both cases (the classification of sugarcane and other crops), the F1 value decreased by balancing the data. After balancing F1 scores from for the prediction of the “sugarcane” class decreased from 0.76 to 0.60, and for the “other crop” class F1 decreased from 0.70 to 0.76 (Table 5.3).

Table 5.2. A confusion matrix depicting the effect of balancing the training data on the classification of sugarcane and other crops

Original interpretation of classes	Total	Classes predicted by the model					
		Imbalanced training data			Balanced training data		
		No data	Sugarcane	Other	No data	Sugarcane	Other
No data	694	654	13	27	657	5	32
Sugarcane	2051	5	1517	529	3	956	1092
Other	1479	7	388	1084	3	165	1311
Total	4224	666	1918	1640	663	1126	2435

Table 5.3. Recall, precision, and F1 scores of the convolutional neural network (CNN) model for classifying sugarcane and other crops for imbalanced and balanced training data

		Recall	Precision	F1
Sugarcane class	<i>Imbalanced training data</i>	0.74	0.79	0.76
	<i>Balanced training data</i>	0.47	0.85	0.60
Other crops class	<i>Imbalanced training data</i>	0.73	0.66	0.70
	<i>Balanced training data</i>	0.89	0.54	0.67

5.3.2 Impact of kernel size on model performance

Increasing the kernel size to 8 in both layers one and two (kernel sizes 8,8) produced the highest precision, recall and F1 scores for both the “sugarcane” and “other crop” classes, for the model run using imbalanced training data (Table 5.4). Table 5.4 shows that running the model with imbalanced data and kernel sizes 8,8 produced respective precision, recall and F1 scores of 0.86, 0.87, and 0.86 for the “sugarcane class”. Similar results can be seen for the “other crop” class with precision, recall and F1 scores of 0.81.

For the model run using the balanced training data, increasing the kernel size in the first layer to 6 and the kernel size in the second layer to 8 (kernel sizes 6,8) produced a higher recall but a lower precision for the sugarcane class than using a kernel sizes 8,8 (Table 5.5). Table 5.5 shows that the respective recall for kernels 6,8 and 8,8 with values of 0.76 and 0.75, and the respective precision with values of 0.92 and 0.93 were relatively similar for the sugarcane class which was reflected in their F1 scores. F1 scores for the sugarcane class for the model ran using imbalanced training data were the same for kernel sizes 6,8 and 8,8 with a value of 0.83 (Table 5.5). For the other crop class kernel sizes 8,8 produced the highest recall, precision, and F1 scores for the model ran using balanced training data. Hence, kernel sizes 8,8 were deemed the most optimal for the model ran using both imbalanced and balanced training data.

Table 5.4. The impact of changing the kernel size (from 3 to 8 in increments of one), for the first and second convolutional layer, on the recall (R), precision (P), and F1 scores (F1) of the convolutional neural network (CNN) model for classifying sugarcane and other crops with imbalanced data

		Kernel size in convolutional layer 2																	
		3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8
		R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1
Kernel size in convolutional layer 1	Sugarcane																		
	3	0.74	0.79	0.76	0.77	0.81	0.79	0.81	0.82	0.81	0.82	0.83	0.82	0.83	0.83	0.83	0.84	0.83	0.84
	4	0.76	0.81	0.78	0.80	0.81	0.80	0.81	0.82	0.81	0.82	0.83	0.83	0.86	0.83	0.84	0.85	0.83	0.84
	5	0.77	0.80	0.79	0.80	0.82	0.81	0.81	0.83	0.82	0.83	0.83	0.83	0.84	0.83	0.84	0.87	0.84	0.85
	6	0.79	0.81	0.80	0.81	0.82	0.82	0.82	0.83	0.83	0.84	0.83	0.84	0.86	0.84	0.85	0.86	0.85	0.86
	7	0.80	0.82	0.81	0.81	0.82	0.81	0.84	0.83	0.83	0.85	0.83	0.84	0.85	0.85	0.85	0.86	0.85	0.85
	8	0.80	0.82	0.81	0.83	0.82	0.82	0.83	0.84	0.83	0.86	0.84	0.85	0.86	0.85	0.86	0.87	0.86	0.86
	Other crops																		
	3	0.73	0.66	0.70	0.75	0.69	0.72	0.76	0.73	0.75	0.77	0.75	0.76	0.77	0.76	0.77	0.77	0.77	0.77
	4	0.75	0.68	0.71	0.75	0.72	0.73	0.75	0.73	0.74	0.78	0.75	0.76	0.77	0.78	0.78	0.77	0.78	0.77
	5	0.74	0.69	0.72	0.77	0.73	0.75	0.77	0.73	0.75	0.78	0.75	0.76	0.77	0.77	0.77	0.78	0.80	0.79
	6	0.76	0.71	0.73	0.76	0.73	0.75	0.78	0.75	0.76	0.77	0.76	0.77	0.78	0.79	0.78	0.80	0.80	0.80
	7	0.76	0.72	0.74	0.76	0.73	0.75	0.77	0.76	0.76	0.77	0.77	0.77	0.79	0.78	0.79	0.80	0.79	0.80
	8	0.76	0.72	0.74	0.77	0.75	0.76	0.79	0.75	0.77	0.78	0.79	0.78	0.80	0.80	0.80	0.81	0.81	0.81

Table 5.5 The impact of changing the kernel size (from 3 to 8 in increments of one), for the first and second convolutional layer, on the recall (R), precision (P), and F1 scores (F1) of the convolutional neural network (CNN) model for classifying sugarcane and other crops with balanced data

		Kernel size in convolutional layer 2																	
		3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8
		R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1
Kernel size in convolutional layer 1	Sugarcane																		
	3	0.47	0.85	0.60	0.58	0.85	0.69	0.64	0.86	0.73	0.65	0.88	0.75	0.70	0.89	0.79	0.70	0.90	0.79
	4	0.54	0.85	0.66	0.56	0.86	0.68	0.66	0.87	0.75	0.71	0.89	0.79	0.68	0.90	0.78	0.74	0.91	0.82
	5	0.63	0.84	0.72	0.63	0.87	0.73	0.67	0.88	0.76	0.68	0.88	0.77	0.73	0.90	0.81	0.73	0.93	0.82
	6	0.56	0.86	0.68	0.67	0.86	0.75	0.71	0.88	0.79	0.72	0.89	0.80	0.74	0.91	0.81	0.76	0.92	0.83
	7	0.60	0.87	0.71	0.65	0.88	0.74	0.67	0.90	0.77	0.70	0.90	0.79	0.77	0.91	0.83	0.71	0.92	0.80
	8	0.61	0.88	0.72	0.62	0.87	0.72	0.70	0.89	0.78	0.75	0.90	0.82	0.74	0.92	0.82	0.75	0.93	0.83
	Other crops																		
	3	0.89	0.54	0.67	0.86	0.59	0.70	0.86	0.62	0.72	0.88	0.63	0.74	0.88	0.67	0.76	0.90	0.67	0.77
	4	0.87	0.57	0.69	0.87	0.58	0.70	0.87	0.64	0.74	0.89	0.68	0.77	0.90	0.66	0.76	0.90	0.71	0.79
	5	0.83	0.61	0.71	0.88	0.62	0.73	0.88	0.65	0.75	0.88	0.66	0.75	0.90	0.70	0.79	0.93	0.70	0.80
	6	0.87	0.58	0.70	0.85	0.64	0.73	0.87	0.67	0.76	0.88	0.68	0.77	0.91	0.71	0.79	0.91	0.72	0.81
	7	0.89	0.60	0.72	0.88	0.63	0.74	0.90	0.65	0.76	0.90	0.67	0.77	0.90	0.73	0.80	0.93	0.69	0.79
	8	0.89	0.62	0.73	0.87	0.61	0.72	0.88	0.67	0.76	0.89	0.71	0.79	0.92	0.71	0.80	0.93	0.72	0.81

Table 5.6. A confusion matrix depicting the results of the initial model ran using; kernel sizes 8,8 (in the first and second convolutional layer) and imbalanced training data; balanced training data and kernel sizes 6,8 and 8,8 (in the first and second convolutional layer).

Original interpretation of classes	Total	Classes predicted by the model					
		Imbalanced training data			Balanced training data		
		No data	Sugarcane	Other	No data	Sugarcane	Other
No data	694	650	21	23	656	12	26
Sugarcane	2051	7	1778	266	1	1531	519
Other	1479	2	275	1202	0	102	1377
Total	4224	659	2074	1491	657	1645	1922

5.3.3 Impact of the number of convolutional layers

Increasing the number of convolutional layers affected recall, precision, and overall F1 scores for the models with kernel sizes 8,8, run using balanced and imbalanced training data. The results depicted in Table 5.7 show that the model trained using imbalanced training data with kernel sizes 8, 8 and convolutional layers 64, 128 (Model 1), and the model trained using balanced training data with kernel sizes 8,8 and convolutional layers 64, 128, 256, 512 (Model 2) were most optimal for predicting the probability of the “sugarcane” and “other crop” class within long time-series data. Model 1 produced the highest recall and F1 for the prediction of the “sugarcane” class, with respective values of 0.87 and 0.86 (Table 5.7), while Model 2 produced the highest precision for the predicted “sugarcane” class with a value of 0.93 and the highest recall and F1 for the predicted “other crop” class with respective values of 0.93 and 0.82 (Table 5.7). Figure 5.7 and Figure 5.8 depict the architectures of the two most optimal models for predicting the probability of sugarcane and other crop presence within long-time-series data depending on user needs.

Table 5.7. The impact of the number of convolutional layers on the recall, precision and F1 scores of the convolutional neural network model for classifying sugarcane and other crops using both balanced and imbalanced training data and kernel sizes 8,8.

	Convolutional layers	Imbalanced			Balanced		
		Recall	Precision	F1	Recall	Precision	F1
Sugarcane	64, 128	0.87	0.86	0.86	0.75	0.93	0.83
	64, 128, 256	0.85	0.86	0.86	0.79	0.92	0.85
	64, 128, 256, 512	0.86	0.85	0.86	0.78	0.93	0.85
	64, 128, 256, 512, 1024	0.85	0.87	0.86	0.77	0.92	0.84
	64, 128, 256, 512, 1024, 2048	0.85	0.85	0.85	0.68	0.93	0.78
Other	64, 128	0.81	0.81	0.81	0.93	0.72	0.81
	64, 128, 256	0.82	0.79	0.81	0.92	0.75	0.82
	64, 128, 256, 512	0.81	0.80	0.80	0.93	0.74	0.82
	64, 128, 256, 512, 1024	0.84	0.79	0.81	0.92	0.73	0.81
	64, 128, 256, 512, 1024, 2048	0.81	0.79	0.80	0.95	0.67	0.79

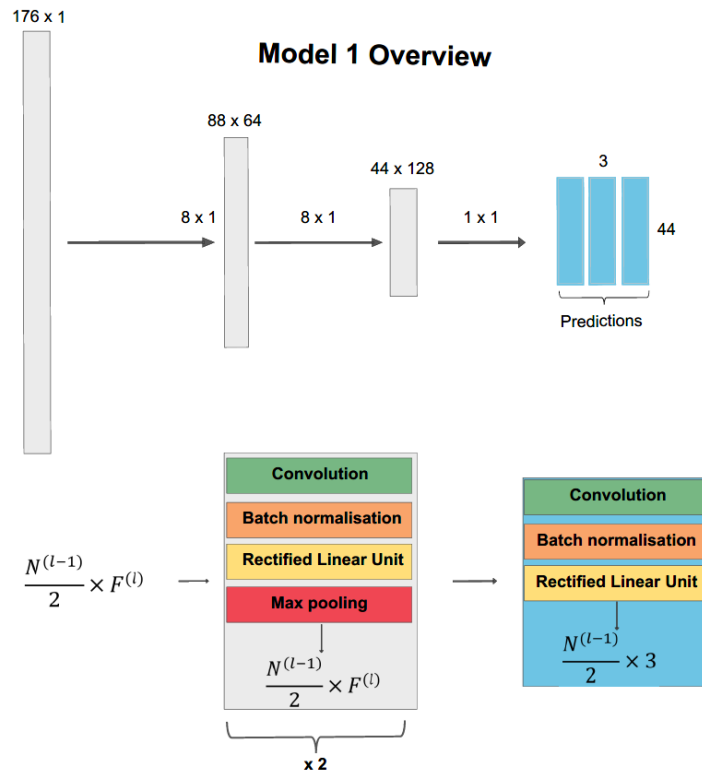


Figure 5.7. The architecture of Model 1 for sugarcane detection in long-time-series data. The model starts with a 6-year processed NDVI time-series for each field (size 176 x 1) as the input, which is then forward propagated through the convolutional blocks (depicted in grey and blue) to produce an output array (size 44 x 3) with predictions for the no data, sugarcane and other crop class. Each convolutional block contains a convolutional layer, rectified linear unit activation and batch normalisation layer. The grey blocks also contain a max pooling layer, the kernel size of the grey layers is 8 x 1. The dimension written next to each convolutional block in the overview represents the dimension of the data. The final convolutional block produces output predictions for whether observations in the simplified time-series belong to the no data, sugarcane and other crop classes, which can be seen in blue.

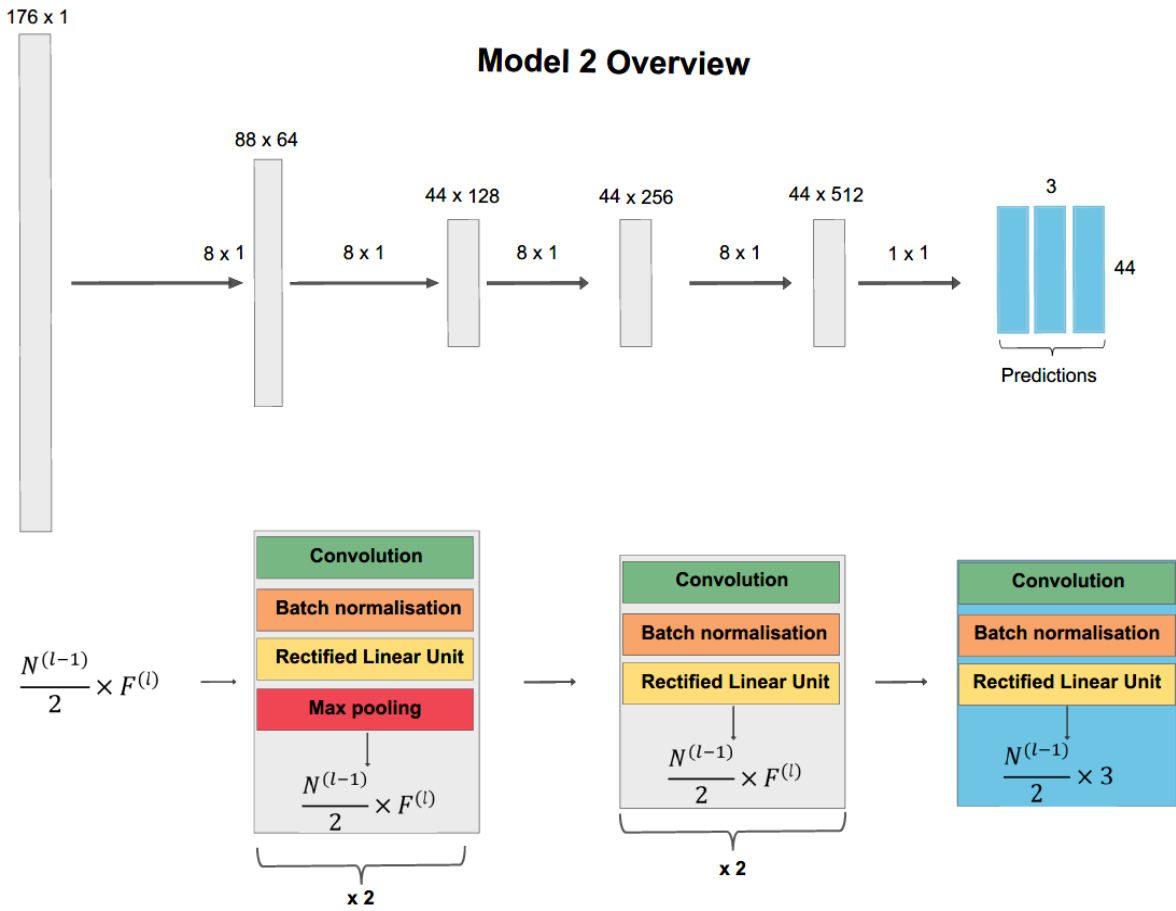


Figure 5.8. The architecture of Model 2 for sugarcane growth detection in long-time-series data. The model starts with a 6-year processed NDVI time-series for each field (size 176 x 1) as the input, which is then forward propagated through the convolutional blocks (depicted in grey and blue) to produce an output array (size 44 x 3) with predictions for the no data, sugarcane and other crop class. Each convolutional block contains a convolutional layer, rectified linear unit activation and batch normalisation layer. The first two grey blocks also contain a max pooling layer and use a kernel size of 8 x 1. The dimension written next to each convolutional block in the overview represents the dimension of the data. The final convolutional block produces output predictions for whether observations in the simplified time-series belong to the no data, sugarcane and other crop classes, which can be seen in blue.

5.3.4 Generalisation

To assess for generalisation, the performance of Model 1 and Model 2 trained using Unit 1 data were also evaluated using an evaluation dataset of time-series profiles for fields located in Unit 2. The results of this experiment are depicted in Table 5.8 and Table 5.9. For the classification of sugarcane, Model 1 performed better than Model 2 with a higher number of true positive observations (1982 compared to 1522) (Table 5.8), and higher recall, precision and F1 values (0.69, 0.82 and 0.75 compared to 0.60, 0.53 and 0.56) (Table 5.9). For the classification of other crops, the performance of Model 2 was better than Model 1 in terms of the number of true positive values (1106 compared to 965) (Table 5.8) and recall (0.70 compared to 0.60), yet precision and F1 scores were slightly higher for Model 1 than Model 2 with respective values of 0.53 and 0.56 compared to 0.45 and 0.55 (Table 5.9).

Table 5.8. A confusion matrix depicting the performance of Model 1 and Model 2 trained using Unit 1 data and evaluated using Unit 2 data

Original interpretation of classes	Total	Classes predicted by the models					
		Model 1			Model 2		
		No data	Sugarcane	Other	No data	Sugarcane	Other
No data	4144	4133	7	4	4123	18	3
Sugarcane	2886	40	1982	864	22	1522	1342
Other	1594	201	428	965	138	350	1106
Total	8624	4374	2417	1833	4283	1890	2451

Table 5.9. Model 1 and Model 2 performance for identifying sugarcane fields located in Unit 2. Both models were trained using Unit 1 data and evaluated using Unit 2 data.

		Recall	Precision	F1
Sugarcane class	<i>Model 1</i>	0.69	0.82	0.75
	<i>Model 2</i>	0.53	0.80	0.64
Other crops class	<i>Model 1</i>	0.60	0.53	0.56
	<i>Model 2</i>	0.70	0.45	0.55

5.3.5 Training and evaluating Model 1 and Model 2 using Unit 2 data

The analysis was repeated using training and testing data from Unit 2 (70%, 30% split). Both Model 1 and Model 2 were retrained using training data from Unit 2 and their recall, precision and F1 scores are displayed in Table 5.10 and Table 5.11. The results show an increase in recall, and F1 values for the classification of sugarcane using both Model 1 and Model 2 trained and evaluated using Unit 2 data compared with training using Unit 1 data and evaluating using Unit 2 data (See Table 5.9 and Table 5.11). The respective recall and F1 values of Models 1 and 2 trained using Unit 1 data and evaluated using Unit 2 data for the classification of sugarcane were 0.69, 0.75 and 0.53, 0.64, while the respective recall and F1 values of Models 1 and 2 both trained and evaluated using Unit 2 data the classification of sugarcane using were 0.97, 0.87 and 0.94, 0.89. The precision on the other hand was slightly better for Model 1 trained using Unit 1 data and evaluated using Unit 2 data with a value of 0.82 (See Table 5.9 and Table 5.11). For the other class, Model 1 performed better when trained and evaluated using Unit 2 data than when trained using Unit 1 data and evaluated using Unit 2 data with respective precision, recall and F1 scores, with values of 0.87, 0.75, 0.81 compared to 0.53, 0.60, 0.56 (Table 5.9 and Table 5.11). Training and evaluating Model 2 using only Unit 2 data also resulted in higher recall and F1 scores for the other class than training the model using Unit 1 data and evaluating using Unit 2 data with respective precision, recall and F1 values of 0.87, 0.75, 0.81 (Table 5.11) compared to 0.45, 0.70, 0.55 (Table 5.9).

Table 5.10 A confusion matrix depicting the performance of Model 1 and Model 2 trained and evaluated using Unit 2 data (70%, 30% split)

Original interpretation of classes	Total	Classes predicted by the models					
		Model 1			Model 2		
		No data	Sugarcane	Other	No data	Sugarcane	Other
No data	1278	1238	37	3	1238	37	3
Sugarcane	845	0	817	28	0	776	49
Other	473	5	171	297	2	117	354
Total	2596	1243	1025	328	1240	930	406

Table 5.11. Model 1 and Model 2 performance for identifying sugarcane fields located in Unit 2. Both models were trained and evaluated using Unit 2 data (70%, 30% split).

		Recall	Precision	F1
Sugarcane class	<i>Model 1</i>	0.97	0.80	0.87
	Model 2	0.94	0.84	0.89
Other crops class	Model 1	0.63	0.91	0.74
	Model 2	0.75	0.87	0.81

5.4 Discussion

Factory managers and government planners could benefit from the capacity to locate sugarcane fields using EO. Once the locations of sugarcane fields are established, additional image interpretation can be used to determine crop growth stages and potential harvested yields. Temporal remote sensing datasets are useful for crop classification as they allow for separating spectrally similar crops and deep learning architecture of the CNN models are useful for reducing dependency on arbitrary thresholds enabling them to perform better than RF and SVM algorithms.

YOLO-based CNN models are useful for quickly and efficiently detecting objects of interest within (natural) images. Originally designed for the identification of

objects within natural images and artwork, the application of YOLO has become widespread, with modified versions being applied to the medical sector for the detection of Melanoma from photographs (Nie et al., 2019) and kidney localisation within CT scans (Lemay, 2023), but also the agricultural sector with the YOLO-Tomato model developed for the detection of tomatoes from photographs. The YOLO-Tomato model is a great example of how a CNN model originally designed for one purpose (detecting objects within artwork) can be modified and successfully implemented within a different sector (agriculture). By outperforming the Faster R-CNN model developed by Ren et al. (2017), the results of Lawal (2021) suggested the potential use of the YOLO-Tomato model for real-time applications such as programming harvesting robots.

Based on the concept of YOLO models, Kensert et al. (2021) built a CNN model for automatically identifying peaks within 1D liquid chromatogram datasets. Given the unique trapezoid shape sugarcane displays within NDVI time-series data (discovered in Chapter 3), in this chapter, the model developed by Kensert et al. (2021) was modified for automatically identifying sugarcane within continuous NDVI time-series data. To ensure that the developed CNN model was fit for classifying sugarcane, a set of ablation experiments was performed to fine-tune three key hyperparameters of the model. This section discussed the results of the ablation experiments and how well the modified approach worked for EO data.

The first experiment explored the effect of balancing the “sugarcane” and “other crop” classes in Unit 1 training data. This was important because the overrepresentation of one class and underrepresentation of another class in the training data can result in convergence during training, harming the model’s performance (Buda et al., 2018), while balancing the classes in the training data can improve the performance of the algorithm (Valova et al., 2020). Real-world problems involve naturally imbalanced dataset classes (Dablain et al., 2024), so it is advantageous to have CNN models that can handle imbalanced data (Valova et al., 2020). This first experiment of our study showed that balancing the training data, by increasing the proportion of “other crop” data (the minority class), reduced the overall performance of the model for sugarcane classification whilst increasing the model’s overall performance of classifying other crops with an

increase in the number of FN values for the sugarcane class and true positive values for the other crop class (Table 5.2).

The impact of balanced versus imbalanced training data has also been studied by Dablain et al. (2024) who used CNN models to classify popular image datasets and Valova et al. (2020), for the classification of 25 architectural features from photographs in their first experiment and identification of boats in their second experiment. The findings of Dablain et al. (2024) and Valova et al. (2020) support the findings of this chapter, by showing that balancing training data also increased the accuracy of the model for predicting the minority class and reduced overfitting. This is important, as it suggests that if the balancing for the model developed in this chapter occurred by increasing the sugarcane class (majority class), the opposite effect would have been seen. Both Dablain et al. (2024) and Valova et al. (2020) agreed that balancing training data may not always be the right approach for optimising the performance of a CNN model, noting the importance of considering the specific task at hand before balancing the classes.

The advantages of balancing the “sugarcane” and “other crop” classes in the training data for this study are also dependent on the task the user wants to achieve and how important each class is to that task. In the area of medicine, balancing the training dataset by increasing the number of observations belonging to the minority class can be important with users interested in accurately locating cancerous cells (minority class) amongst healthy cells (majority class) (Dablain et al., 2024). However, balancing the training dataset by increasing the number of observations in the minority class may not always be helpful for agricultural applications. For example, training the CNN model developed in this study using the imbalanced dataset would be more useful for cane operators interested in knowing where sugarcane is located so subsequent growth stages can be derived for management. However, if cane operators already knew the location of sugarcane crops and were interested in locating other crops, to monitor crop rotation between harvests, then training the model using the balanced dataset would be more useful.

The second experiment explored the effect of kernel size on the performance of the CNN model for classifying sugarcane and other crops. The results found for both convolutional layers in the models trained using balanced and imbalanced data, from a choice of kernel sizes between 3 and 8, a kernel size of 8 produced the highest precision, recall and F1 scores. Kernel size is related to the scale of the temporal profiles. An optimal kernel size of 8 suggests that the classification of sugarcane is strongest when one season's data is in convolution. This is because the mean processed NDVI time-series, for each field, used in this study was resampled to a 14-day interval. After pooling, the input of the last convolutional layer was 44 observations for the 6-year time-series, which roughly equates to seven observations per year. If seven observations represent one year, then a kernel size of 8 would mean that over one year's worth of observations (one season), were being used for convolution. This suggests that using the model developed in this chapter, it would be difficult to classify sugarcane from partial seasons or within a season as sugarcane is still growing. These results show that the shape of full sugarcane growth cycles for its classification, which further justifies using multi-seasonal time-series data for the classification of sugarcane over using methods that require isolating single growing seasons. This is because the testing inputs of a CNN must be the same size as the training data. Padding the time-series of shorter-cycle crops to the length of crops with longer growing seasons could distort the input training data, and therefore the classification of different crops from the time-series data.

A closer inspection of the prediction output for kernel sizes less than 8, highlighted the effect of cloud events during the mid-season, which created a drop in NDVI leading to sugarcane being mistaken for two short crop growth cycles. This effect was also seen in Chapter 3. This further supports the argument for classifying sugarcane from long-time-series data where crop calendars and planting and harvesting dates are unavailable, as a whole season's data is required for classifying sugarcane to allow for full growth cycles to be captured.

Convolutional layers are important components of CNN models, used to extract key features of the input data which are then used to train the model weights for classification. The third study considered the trade-off between increasing the

model's ability to extract high-level features (by increasing the number of convolutional blocks) and the time it takes for the model to run. This chapter attempted to optimise the number of convolutional layers in the model by assessing the effect of adding convolutional layers to the initial CNN model (Figure 5.4) on its performance for classifying sugarcane and other crops. The sub-objective was to find a compromise between precision, recall and F1 values for classified sugarcane, computational complexity and run-time. The initial CNN model (Figure 5.4) had two convolutional layers. The initial model trained using an imbalanced dataset, containing more observations of the "sugarcane" class than the "other crop" class, produced the highest recall and F1 scores with respective values of 0.87 and 0.86 for the classification of sugarcane and the highest precision for the classification of other crops with a value of 0.81. Adding two convolutional layers to the initial model and training the model using balanced data produced the highest precision for the classification of sugarcane with a value of 0.93 and the highest recall and F1 scores for the classification of other crops with respective values of 0.93 and 0.82 (Table 5.7). The addition of convolutional layers did not impact run-time as the dataset used in this study was small.

The above results show that modifying the approach developed by Kensert et al. (2021) worked well for classifying sugarcane and other crops from EO time-series datasets. The outcome of this study was two optimal models for sugarcane classification, depending on user needs. Model 1 produced high recall and Model 2 produced high precision for the classification of sugarcane. The differences in recall and precision were to do with the distribution of the data with more observations in the time-series for sugarcane in the unbalanced dataset as a proportion of the data compared to the other class, which meant that more instances of sugarcane were correctly predicted as sugarcane out of all sugarcane observations compared to the balanced dataset which had a higher number of correctly classified sugarcane out of all instances of predicted sugarcane. The similarity in F1 scores between Model 1 and Model 2 meant that the best model for sugarcane classification depends on user needs. The training dataset used in this study was relatively small compared to other machine

learning models, therefore the increased depth of Model 2 compared to Model 1 did have minimal effect on run-time. However, the application of Model 2 to other studies with larger training datasets may be limited compared to Unit 1, based on its longer run-time. This is something that future research should consider, especially with the addition of more specific crop classes, rather than a generic “other crop” class.

Furthermore, the high recall, precision and F1 scores of both Model 1 and Model 2 trained using Unit 1 data for sugarcane and other crop classification in Unit 2, showed that both models could successfully generalise for an unseen study area. A good generalisation for the classification of sugarcane from both models shows that the time-series profiles for sugarcane were similar in both Unit 1 and Unit 2, despite different management practices. Out of the two models, the higher precision, recall and F1 scores for Model 1 compared to Model 2, suggest that Model 1 would be more suitable for classifying sugarcane in the Unit 2 region. On the other hand, similar F1 scores and large differences in precision and recall for the classification of other crops in Unit 2, suggest that selecting an optimal model for other crop classification would depend on user needs. Furthermore, compared to the classification of sugarcane in Unit 1 and Unit 2 and the other crop class in Unit 1, the precision, recall and F1 scores for the classification of other crops in the Unit 2 evaluation dataset was lower using both Model 1 and Model 2. Lower precision, recall and F1 scores for the classification of other crops in the Unit 2 evaluation dataset compared to the classification of sugarcane in Unit 1 and Unit 2 and the other crop class in Unit 1 suggests that the types of other crops in Unit 1 differed from the types of crops in Unit 2, and that there was more variation of crops in Unit 2 compared to Unit 1. Many other hyperparameters not explored in this study could also affect the performance of the derived CNN model. For example, input dimension, batch size and max pooling sizes are affected by a trade-off between providing enough information for training the model and reducing the computational complexity and therefore time to run the model. In future work, the optimisation of these additional hyperparameters could be investigated to improve the performance of the developed CNN model for sugarcane classification.

Retraining both Model 1 and Model 2 using Unit 2 data improved precision, recall, and F1 values for the classification of both sugarcane and other crops located in Unit 2 compared to training using Unit 1 data, which suggests that training and evaluating both models using data tailored to each specific study area would be desirable where data is available. For the classification of sugarcane in Unit 2, the performance of Model 1 and Model 2 was similar in terms of F1 scores, with a difference of 0.02 favouring Model 2. Model 2 also had a higher precision than Model 1 with a difference of 0.04 whereas Model 1 had a higher recall than Model 2 with a difference of 0.03. These results suggest that like for the classification of sugarcane in Unit 1, choosing an optimal model for classifying sugarcane in Unit 2 would also be dependent upon user needs, with Model 1 being more optimal if the user needs involved not missing sugarcane profiles, whilst Model 2 would be more appropriate if the user needs required correct predictions. Compared to the classification of sugarcane, there was a greater difference in the precision, recall and F1 scores of Models 1 and Model 2 for the classification of the other crop class. The respective differences between recall and F1 scores for classifying other crops using Model 1 compared to Model 2 were 1.2 and 1.14, favouring Model 2. Precision, on the other hand, was higher for Model 1 compared to Model 2, with a difference of 0.04. This suggests that, if the user needs required accurate classification of other crops, it would be more appropriate to use Model 1 compared to Model 2. Model 2 would therefore also be recommended if user needs were dependent on accurate classification of both sugarcane and other crops.

The identification and planning of fallowing would be one application of the above models for sustainable sugarcane production. Fallowing is where land is set aside before cultivation for one or more vegetative cycles. Updated knowledge about crop rotation can help regenerate natural resources, such as soil and water after exploitation and improving methods for crop type identification has been flagged in previous research as an area that needs attention (Mueller and Warren et al., 2015).

5.5 Conclusions

The classification of sugarcane from EO can be improved by using time-series to separate sugarcane from spectrally similar classes. These results demonstrated the advantages of modifying an approach developed for peak detection in chemical chromatograms to identify sugarcane from long continuous NDVI time-series data containing multiple seasons of growth. The developed model localised and classified sugarcane in the same step, overcoming limitations of previous time-series approaches which required crop calendar information to isolate individual sugarcane growing seasons. Such approaches can be useful for producing large-scale global maps of sugarcane at a field level.

Identifying the presence of sugarcane from long time-series data is important for biodiversity monitoring, for example monitoring the conversion of biodiverse landscapes to sugarcane plantations and maintenance or expansion of biodiversity within existing sugarcane plantations through intercropping or crop rotation. The identification and planning of fallowing is also a potential application of the developed model. Updated knowledge about crop rotation can help with the regeneration of natural resources, such as soil and water after exploitation. Once the location of sugarcane fields is known, management practices such as crop rotation and intercropping at a field-level could be monitored over time (Liu et al., 2023; Wang et al., 2020) and cross-referenced with additional information relating to soil health, water table levels, and regional averages for yield to assess the effectiveness of these management practices for sustainable sugarcane production (Dimov et al., 2022). This is important for monitoring SDG indicator 2.4.1 and its themes of biodiversity, water use, soil health, resilience, land productivity and profitability (Table 2.1).

The utility of the developed CNN model is not limited to sugarcane and other crop. It has much wider applications for sustainability, such as monitoring the classification of specific crops (rather than the generic “other crop” class) and monitoring land cover change. This is important as improving methods for crop type identification has been flagged in previous research as an area that needs attention (Mueller and Warren et al., 2015). Additionally, although trash burning

was not directly investigated in this study, there is a potential to include “trash burning” as a class in the CNN model allowing for future research to detect burning of sugarcane “trash” or waste after harvest by investigating the presence and absence of burn scars over time (Pnvr and Bandaru, 2023). This is important as pollution from trash burning can be related to the environmental and social aspects of sustainable sugarcane production (Andrade et al., 2011).

6 SYNTHESIS

6.1 Introduction

Sustainable sugarcane production is the act of cultivating sugarcane in a way that meets the economic, social and environmental needs of today's generation; without hindering the ability for future generations to fulfil their needs. In India, as elsewhere, the economic needs include the capacity for farmers to grow a profitable crop. Social sustainability includes fair wages for labour and environmental considerations include the appropriate use of water for irrigation and fertiliser to prevent nutrient stress. The work for this thesis started in 2020, as the world was entering a series of lockdowns caused by COVID-19, which restricted social movement, created economic hardship, and led to mixed environmental effects. A report by Solomon et al. (2020), outlined how the restricted movement of people led to a shortage of labour and a delay in harvest for millions of tons of sugarcane which reduced sugarcane yields. Reduced port operations and courier services also restricted the transportation of sugarcane by-products like sugar and ethanol. This, coupled with reduced global and domestic demands for sugar and ethanol, produced by lower fuel consumption, disruption to sugary drink production, and closures to hotels and retail catering, all negatively impacted the social, economic and environmental sustainability of sugarcane production. This created a surplus of unsold sugar and ethanol within sugar mills making it difficult for mills to pay farmers and increasing waste (Solomon et al., 2020). One unforeseen favourable environmental impact of the pandemic was improved air quality (Solomon et al., 2020). Improving the sustainability of sugarcane production is important for future-proofing the industry against global environmental changes such as the increased frequency of droughts and extreme weather. This is a requirement of the United Nations (UN) Sustainable Development Goals (SDGs) by 2030, and therefore important for sustainable sugarcane production.

Information and communications technology (ICT) has been proposed as a means of improving the sustainability of agriculture, with some researchers branding this age of agricultural research as the 4th Industrial Revolution (Lele

and Goswami, 2017). Solomon et al. (2020) suggested that artificial intelligence (AI) in sugarcane management could be used for crop and soil health monitoring and predictive crop analysis, to overcome the challenges faced by the pandemic. This information can help with the planning of irrigation at critical growth stages to improve sustainable use of available water resources (Solomon and Swapna, 2022), and to make more effective use of farm machinery in sugarcane production thereby reducing labour costs (Solomon and Swapna, 2022).

The management of Indian sugarcane is spatially and temporally heterogeneous. Hence there are advantages from collecting management information for large areas at different stages of sugarcane growth. The launch of EO satellites such as the Landsat series in 1972, Sentinel-2 satellites (a and b) in 2015 and 2017, and Sentinel-1a satellite in 2014 provide opportunities to monitor sugarcane plantations over large areas and timescales quickly. There is a known trade-off between the spatial, spectral and temporal resolution of different EO sensors, and the impact of cloud and noise makes interpreting EO data difficult without preprocessing (Chapter 2).

The selection of datasets and pre-processing parameters is subjective and left to the experience of the data analyst. This lack of guidance for selecting appropriate sensors and pre-processing methods for removing cloud noise makes replicability of methods for different study areas difficult. The second limitation associated with the use of EO time-series data for monitoring sugarcane growth over time is the dependency of prior knowledge of growth stages from field surveys or crop calendar. To address these two limitations a pre-processing methodology for extracting sugarcane growth information from EO time-series data was developed in Chapters 3 and 4 (Objective 2).

The classification of sugarcane from EO time-series datasets is important for monitoring sugarcane management over large areas and timescales, by separating sugarcane from spectrally similar crops. However, despite the progress in automatic classification, relating it to how well methods generalise to different study areas need to be overcome. Previous sugarcane classification approaches are also limited by dependency of prior knowledge of growth stages

derived from crop calendars and field surveys. These limitations, prevent automated approaches being used at scale, as crop calendars are only available for study areas where the start of sugarcane production occurs at a similar time each year. Elsewhere in regions such as India, where management is variable, the locations of sugarcane fields are needed to derive known growth information. Chapter 5 explored a novel method to automate the process through learning and join up the temporal information on growth with classification of sugarcane crops (Objective 3).

Through investigating the above research gaps, this work has determined how new EO imagery and machine learning tools can help increase resource efficiency (land, labour, nutrients, water) for sustainable sugarcane production. This chapter contains a summary of the thesis. Methodologies for extracting sugarcane growth information and classifying sugarcane from EO are evaluated using sustainable sugarcane production frameworks (Objective 4). The chapter then closes with concluding thoughts about the practical applications of the developed approaches for monitoring sugarcane using EO and how they can be implemented into frameworks for sustainable sugarcane (Objective 5).

6.2 Monitoring sugarcane production for improved sustainability

Sustainability of sugarcane production could support several of the United Nations Sustainable Development Goals (SDGs), and sub-goals, set for 2030. Target 2.4 under the goal “Zero Hunger”, for example, aims to “ensure sustainable food production systems and implement resilient agricultural practices”. As indicated in Chapter 2, there are at least 11 attributes that can be associated with sustainability including soil health, water, pesticide and fertiliser use (Table 2.1). In terms of the economy, improving the sustainability of sugarcane production could involve improving farm outputs per hectare, profitability and resilience to financial hardships. Improving environmental sustainability with respect to sugarcane production could involve improving the health of soils by minimising the impact of erosion, reduced fertility, waterlogging,

salinization or leaching of pesticide and fertilisers. It could also involve improving biodiversity and access to water. Societal improvement to sugarcane production could involve improving employment, food security and tenure rights to land. The UN (2024) recommends that the above sustainable attributes can be monitored using methods including agricultural and household surveys, administrative data, remote sensing and GIS models. The literature review in Chapter 2, grouped the different methods for monitoring sugarcane production into two categories: methods that use surveys and data collected from field visits, and those that use EO data.

6.2.1 Survey and field data

The first part of the review focused on how data collected from field studies and government surveys have been used in the past to monitor different aspects of sugarcane production and their direct and indirect link to the different components of Target 2.4 (See Table 2.1).

Decision-making, based on location and the timing of sugarcane growing stages, is often carried out using existing information collected from field studies such as those by Narayanamoorthy (2004) or secondary data such as government-published statistics and information published in books and journals (Nandhini and Padmavathy, 2017). However, the collection of field data is labour-intensive and can put a strain on labour resources (Olfindo et al., 2020) or limit the usefulness of the data.

Chapter 2 highlighted that field survey datasets are useful for determining sugarcane growing stages, addressing both spatial (Gujja et al., 2009; Nandhini and Padmavathy, 2017) and temporal (Mall et al., 2016) variations. Optimal conditions in mid-season include high rainfall (>1000 mm), high temperatures (~30°C), and high humidity levels (~80%), while dry weather is needed in the late season senescence stage. In-fact, high water levels during the late season can reduce sucrose concentrations and thereby sugar yields (Fageria et al., 2010; Mall et al., 2016). Poorly managed sugarcane can also increase the crop's

vulnerability to pests and disease, in turn affecting economic, environmental and social sustainability (Singh, 2018) (Table 2.1). With the knowledge of growth stages, the informed farmer can modify planting, harvesting, fertiliser and pesticide application, and irrigation type (Nandhini and Padmavathy, 2017). An understanding of growth stages can also help farmers understand how management can minimise the negative impacts of climate change and maintain yields (Goswami et al., 2006; Mahato, 2014; Mall et al., 2016).

6.2.2 Earth observation data

The second part of the literature review found that using EO to monitor sugarcane growth over large areas and timescales is supported by previous work carried out by Wang et al. (2020), Yeasin et al. (2020), Dimov et al. (2022), and Zheng et al. (2022). These studies demonstrated that changes in NDVI time-series data could be used to derive sugarcane phenological stages to predict yields or to classify sugarcane areas.

The literature review highlighted that the current EO approaches for deriving sugarcane growth stages which is useful for management typically rely on field surveys, crop calendar derived planting and harvesting dates to first derive individual sugarcane growing seasons from long-time-series data (Liu et al., 2023; Wang et al., 2020; Ofindo et al., 2020). However, planting and harvesting date information is difficult to derive without expensive, lengthy and labour-intensive surveys. With the different components of Target 2.4 in mind, the literature review identified two research gaps. These were i) deriving the timing of sugarcane growth stages from EO datasets and ii) classifying sugarcane from long-time-series EO datasets without prior knowledge of sugarcane planting and harvesting dates (Figure 6.1).

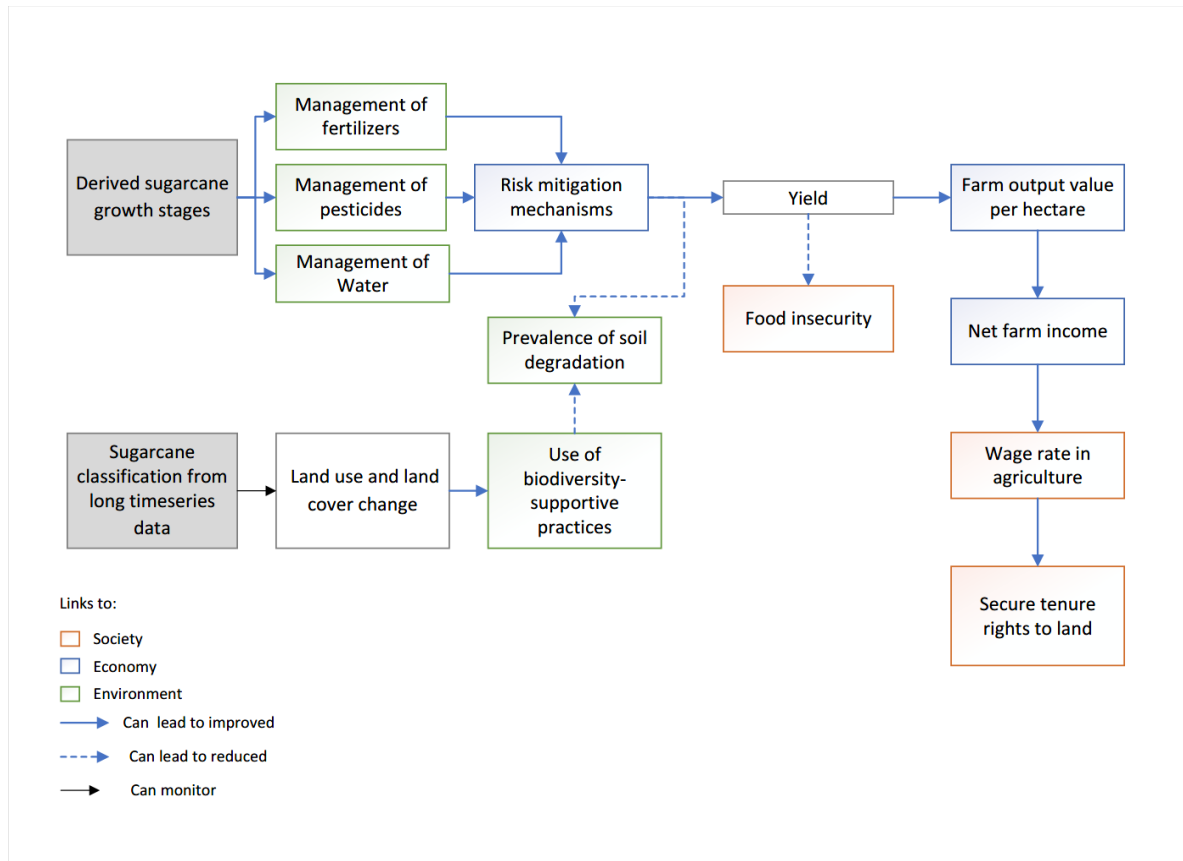


Figure 6.1 Thesis outputs, shaded in grey, and their direct and indirect links to eleven components of the Target 2.4 of the UN Sustainable Development Goals related to society (red outline), economy (blue outline), and the environment (green outline).

6.3 Deriving growth stages from Earth Observation time-series

To meet Objective 2, algorithms were developed to preprocess and decompose long time-series data into individual growing seasons and derive subsequent growth stages important to management. The next sections discuss the outcomes of developing algorithms for preprocessing and decomposing long time-series into individual growing seasons to derive subsequent growth stages important to management.

6.3.1 Growth stages from optical sensors

The methodology developed in Chapter 3 overcame the challenge of selecting appropriate sensors for monitoring sugarcane at a field level over time. The results of Chapter 3 showed that the resolution of Sentinel-2 and Landsat 8 images can be used for monitoring sugarcane growth over time in India. This is because there is minimal spectral mixing between the small field parcels (from 0.14 to 11.5 ha) compared to coarser, yet higher frequency, MODIS-derived NDVI time-series (250 m). These findings are supported by Cihlar and Jansen (2001) and Praveen et al. (2019), who also found less spectral mixing when respectively observing sugarcane fields located in Lebanon and the Indian States of Uttar Pradesh and Bihar using Landsat and Sentinel-2 imagery compared to the use of coarser resolution imagery. Another advantage of using Landsat and Sentinel-2 datasets to monitor sugarcane growth over time is that these datasets are open-access and readily available on the GEE platform with no cost for data download and processing, which provides opportunities for the developed algorithm to have global and near real-time applications.

The cloud filtering and harmonization algorithms developed in Chapter 3 were instrumental in overcoming the challenge of disjointed data caused by clouds, for example, erroneous low estimates of optical data, also confirmed by Zhang et al. (2018). Filtering methods and parameters for temporal smoothing are subjective and left to the experience of the data analyst and local pre-processing parameters for harmonising and removing noise, which has not previously been established for sugarcane fields located in India. A methodology was therefore developed to remove clouds and additional residual noise while maximising the number of cloud-free observations per field per growing season. The cloud and noise removal algorithms developed in Chapter 3 were able to remove more cloud than using the data provided by existing Landsat 8 BQA and Sentinel-2 QA60 cloud masks. Local harmonization and BISE smoothing also reduced the impact of high cloud cover during the monsoon season by reducing the effect of gaps within the cloud-free time-series for this period. The results highlighted the benefits of using localised regression coefficients and BISE sliding periods over the use of

previously published coefficients and sliding period parameters. The derived 30-day sliding period for optimal noise removal from sugarcane field time-series profiles differed from the optimal sliding period for removing noise within Pathfinder AVHRR land time-series datasets for forests, wheat, the Nullabor plain in southern Australia and tropical study sites in Australia (Lovell and Graetz, 2001), and the pre-processing parameters also differed between sugarcane growing study areas. For example, the optimised 14-day resampling interval derived in Chapter 3 differed from the 10-day resampling interval used by Wang et al. (2020) for monitoring sugarcane crops using Landsat 8 and Sentinel-2 imagery in China. After testing the effect of applying local regression parameters for harmonising the Landsat 8 and Sentinel-2 datasets between Unit 1 and Unit 2, the results of Chapter 3 showed that in addition to country specificity, pre-processing should also be carried out at individual study sites, with calibration being site specific despite their proximity (neighbouring states).

The need for local calibration of cloud filtering, harmonisation and BISE filtering parameters to specific study sites can be considered a limitation as the algorithms cannot be directly applied globally. However, the methodology and algorithms developed in Chapter 3 enable others to derive and obtain local optimal pre-processing parameters for any study site using the approach designed in this work.

To address the limitation regarding the lack of transferability between optical EO derived phenological metrics and growth stages on the ground, where planting and harvesting information is absent, the FAO model for sugarcane growth was used to link EO land surface phenology to growth stages used by agronomists. The algorithms took advantage of the levelling off period in the optical reflectance data with reflectance stabilising at the start of the mid-season as the number of new tillers (new shoots) is reduced and the focus of growth moves to stem-elongation. A benefit of the algorithms developed in Chapter 3 over previously used simplified models such as double logistic curves, is that the data is not oversmoothed and important information relating to key sugarcane growth stages important for managing water use is retained (Figure 6.1).

6.3.2 Growth stages from Sentinel-1 data

Overall, the methodology and algorithms developed in Chapter 3 are suitable for monitoring key sugarcane growth stages, however the occurrence of persistent cloud cover during the monsoon affected seasonal decomposition of time-series and subsequent growth stages in some fields. To overcome the limitations the use of Sentinel-1 backscatter time-series data was investigated in Chapter 4. The cloud penetrating abilities of Sentinel-1 is one reason why it was chosen for monitoring sugarcane over time, other reasons being, regular collection of data (6-12 day repeat), free access to processed data using GEE platform as well as providing additional information about structural differences over time and rather than providing information about differences in biomass over time (like optical EO data).

In addition to cloud, the developed algorithms in Chapter 4 also overcame the limitations of requiring known planting and harvesting dates prior for isolating individual sugarcane growth cycles (Olfindo et al., 2020) by combining the optical and Sentinel-1 EO time-series data.

The difference between biomass and structural information meant that additional growth stages different to those derived from optical EO data in Chapter 3 could be derived e.g. the tillering stage and harvest. The identification of harvest was found to be much more precise which was useful as it relates to a high demand for transport and people. Sugarcane needs to be transported to the mills for crushing as soon as it is harvested to minimise post-harvest losses, and knowing harvesting times could help sugar mills improve the scheduling of cane crushing operations. In practice, harvest management, sucrose content and yields are interlinked (Figure 6.1). However, sucrose content and yield were not directly assessed in this study due to constraints in data availability and so the estimations of yield are uncalibrated. Future studies should therefore sample across the fields and compare the findings to those outlined in Chapter 4.

The use of Sentinel-1 backscatter time-series data also gave insights into how different management techniques affected the shape of the backscatter time-series, providing an indirect indication of yield. For example, waterlogging from

incorrectly installed drip irrigation systems which is something that future studies should investigate with field visitation for validation (Bao et al., 2023; Besten et al., 2023).

6.4 Using machine learning to classify sugarcane

Seasonal decomposition algorithms produced in Chapters 3 and 4 are dependent on information regarding the year of the survey. To overcome this dependency on survey data, an approach was developed in Chapter 5 to automatically classify sugarcane from long time-series data using Convolution Neural Networks (CNN). The CNN model used in Chapter 5, was based on an approach originally designed for the medical sector to detect peaks in chemistry chromatogram datasets (Kensert et al., 2021). This approach was specifically chosen as it allowed for localisation and classification to be carried out in the same step, which overcame the need for isolating patterns in time-series data prior to classification, a limitation of previous classification approaches.

The classification algorithm developed in Chapter 5 has the potential to produce large-scale global maps of sugarcane but also allows for monitoring crop rotation and intercropping at a field-level (Liu et al., 2023; Wang et al., 2020), which is important for measuring biodiversity, soil health and water use (Figure 6.1). The use of EO time-series data for classifying sugarcane amongst spectrally similar vegetation types such as paddy, rice and maize, was limited to threshold-based classification approaches that required the separation of individual sugarcane growing seasons using known planting and harvesting dates or crop calendars which were not appropriate for India as management is variable (Liu et al., 2023; Wang et al., 2020; Ofindo et al., 2020). The model developed in Chapter 5 overcame these challenges for monitoring direct land use change (dLUC) (expansion and reduction of sugarcane plantation areas), by classifying sugarcane from within continuous time-series data, rather than assessing LUC one image per year, e.g. Adami et al. (2012).

It is important to note that the method developed in Chapter 5 is dependent on the pre-processing methodology developed in Chapter 3 and, therefore subject

to the same limitations of the need to localise harmonization and cloud filtering parameters for each study area (not a global model). However, once these localisations have been completed, the method can be applied to hundreds of fields in the same study areas.

The classification methodology developed in Chapter 5, is also not completely independent of field data as the derived field polygons were captured using GPS. Currently, where GPS derived field polygon information is unavailable, the method would have to run at the pixel level, which assumes fields are large enough to give strong phenology signals at 30 m (Landsat), which is computationally expensive as pulling out and processing a profile for each pixel would take a long time. Future studies should therefore investigate methods for grouping pixels or investigate other models that can segment objects as part of the training.

Although the model was able to generalise, the results of Chapter 5 show the model sensitivity to different study areas. Model sensitivity to different phenology and shapes of profiles suggests that there is also a geographical limitation, as training was carried out in only two areas for sugarcane and other crops. Further training and many for examples profiles would be required for a generalised model for classification by profile shape.

6.5 Practical Applications of monitoring sugarcane development from space: Challenges, Opportunities, and System Design

The process of producing sugar and its by-products from raw sugarcane is best understood as a supply chain involving multiple stages, variables and stakeholders (Figure 6.2). This supply chain reflects a forward flow of materials (cane, sugar, ethanol, bagasse) and a backward flow of revenue, policy, and information between farmers, mills, transporters, processors, and regulators (Patil et al., 2024).

The sugarcane supply chain is typically made up of four key stages. The first stage is cultivation. Sugarcane is grown over 10–18 months depending on region and variety (Shukla et al., 2017). The second stage is harvest, where sugarcane can either be manually or mechanically cut when mature. After harvest, the third stage is transportation, where sugarcane must be transported quickly to avoid sucrose deterioration (Misra et al., 2022). Once the cane has been successfully transported to the mills, processing takes place. Processing is the fourth step where sugarcane is crushed to produce sugar and other by-products like ethanol (Patil et al., 2024).

The different stages of the sugarcane supply chain are influenced by a different set of stakeholders, each with the aim of fully optimising their individual processes by managing specific variables outlined in Figure 6.2. For example, farmers aim to maximise their profits by harvesting early in the season, while millers prefer a longer crushing period to improve capacity utilisation and reduce costs (Patil et al., 2024). However, these distinct stakeholder goals and conflicting objectives at each stage of the supply chain can lead to inefficiency, suboptimal coordination and potential financial loss (Singh et al., 2021).

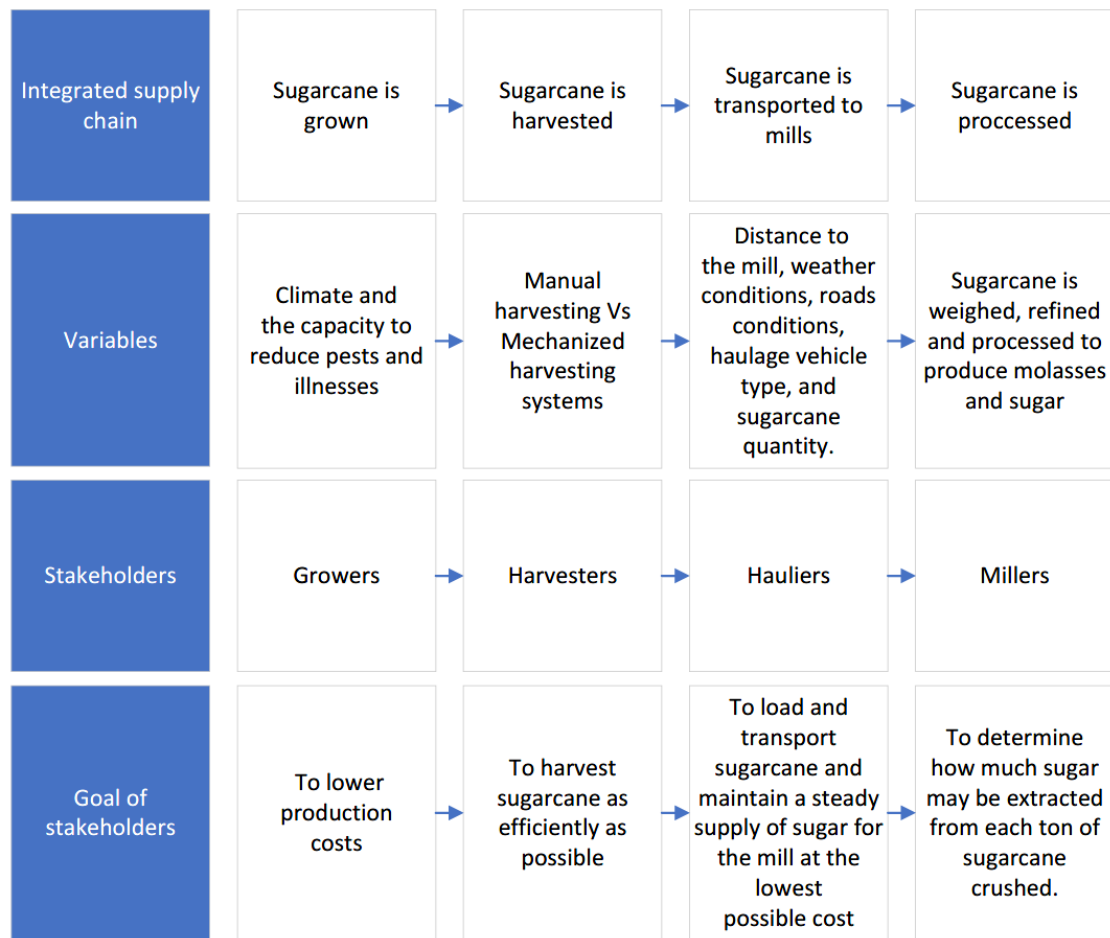


Figure 6.2. The sugarcane supply chain, variables, stakeholders and goals at each stage. Sugarcane growth stages and Production Standard indicators

6.5.1 India's Sugarcane Supply Chain: Structure and Challenges

India follows a broadly similar supply chain structure as other major producers, but its system is uniquely shaped by environmental complexities, institutional regulation and stakeholder diversity, which can give rise to coordination inefficiencies and recurring systemic challenges (Singh et al., 2021).

The growth cycle of Indian sugarcane is between 10-18 months, with both planted and ratoon crops (Shukla et al., 2017). Harvesting decisions are highly sensitive to environmental factors such as temperature, rainfall, pest incidence, and cane maturity with improper timing can result in sucrose degradation and reduced sugar recovery (Patil et al., 2024; Singh et al., 2021; Marin et al., 2021). The long

growing season and the sensitivity to changes in weather causes harvests to be sensitive to yield losses, and reduced sugar recovery in years with irregular rainfall or pest/disease outbreaks (Patil et al., 2024). The use of water-intensive farming practices, high groundwater withdrawal and inefficient irrigation during the sugarcane cultivation period have also caused severe environmental stress in India (Anjaly et al., 2024)

Once harvested, cane must be rapidly transported to nearby mills, for processing, typically within 24 to 48 hours to minimise post-harvest losses (Misra et al., 2022). These downstream products are crucial for India's energy, food, and industrial sectors but the urgency of transport and processing means that delays in logistics or mill shutdowns can cause sucrose deterioration, lowering recovery rates and farmer payments (Patil et al., 2024).

The sugarcane economy is also heavily influenced by government policy. The state sets minimum cane prices (Fair and Remunerative Price, or FRP), allocates procurement zones, enforces ethanol blending mandates and exports are tightly regulated, unlike liberalised markets such as Brazil (Singh et al., 2021). These policies shape planting decisions, payment structures, and mill operations (Patil et al., 2024). However, due to strict regulation and control of government, farmers and mills have limited flexibility to negotiate directly leading to payment delays and inefficient supply allocation (Singh et al., 2021).

The supply chain involves a variety of stakeholders including smallholder farmers, cooperative and private mills, transport contractors, ethanol distilleries, and government agencies, each with distinct and often conflicting objectives (Patil et al., 2024; Singh et al., 2021). For instance, farmers seek to harvest early in the season for higher prices, while millers prefer a steady supply over a longer period to reduce fixed costs per unit (Patil et al., 2024). Conflicting stakeholder goals and mismatches in harvest and crushing schedules result in idle capacity, underutilisation of labour, and poor resource planning (Patil et al., 2024). Many mills also operate below capacity due to aging equipment and poor coordination with growers (Patil et al., 2024).

These mismatches contribute to cyclical crises in the industry including years of overproduction followed by mill closures; unharvested cane due to labour shortages; payment arrears due to cash flow bottlenecks; and unsustainable pressure on water resources. The lack of transparency and coordination between stakeholders across the supply chain further exacerbates these challenges (Patil et al., 2024; Singh et al., 2021).

Given these persistent inefficiencies, there is an urgent need for systems that can deliver real-time, spatially comprehensive, and objective information to support better decision-making and coordination.

As demonstrated in this thesis, EO datasets enables timely, large-scale insights into sugarcane growth stages, harvest readiness, irrigation and environmental stress.

In this context, remote sensing provides a promising solution by delivering objective, near real-time data that could help synchronise decisions across stakeholders for improving visibility, reducing uncertainty thus enabling more responsive supply chain management (Figure 6.3). By integrating EO into supply chain management, mills could better schedule crushing operations, farmers could receive timely agronomic advice, and government agencies could forecast production and manage procurement and price policies proactively (Singh et al., 2021). The following section explores in detail how the tools developed in this thesis could be implemented into the Indian sugarcane supply chain to address specific supply chain gaps for sustainable production.

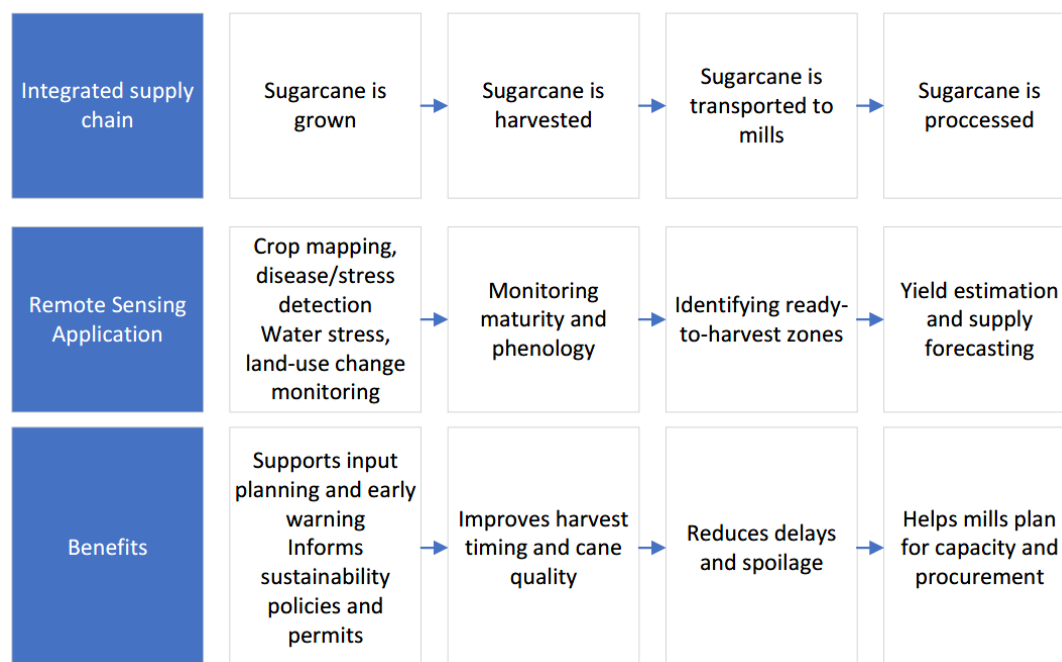


Figure 6.3. Applications of remote sensing at different stages of the sugarcane supply chain and it's benefits.

6.5.2 Practical Applications of EO for sustainable sugarcane growth

One huge advantage of EO is the ability to map crops over large areas. The time-series classification approach developed in Chapter 5, was able to classify sugarcane and other crops from EO long-time-series data over two large Indian states. This is useful for the assessment of direct land use change, to and from sugarcane, and can help with the assessment of four environmental and two social sustainability indicators (Bonsucro, 2023).

Firstly, the developed classification approach could allow planners, who have information on field boundaries (derived GPS data), to use long continuous time-series data to determine in what years sugarcane was grown and for how many consecutive years. If sugarcane was present in the 2018 growing season and again in 2019 for the same field parcel it may indicate that the sugarcane harvested in 2019 is a ratoon of the 2018 sugarcane, as it is possible to plant sugarcane after a previous sugarcane crop. There were also differences in the timing and length of growth stages between virgin and ratoon cane (Chapters 3

and 4), hence, with appropriate training data, there is the potential to use such information to identify which areas may be virgin and ratoon sugarcane.

The discrimination of virgin and ratoon sugarcane crops could be useful for sugarcane factory millers as virgin cane tend to be higher than the yields of ratoon cane, with sucrose content tending to decrease with each ratoon. By looking at how many years sugarcane has been grown in that field parcel, it might be possible to identify what year ratoon the sugarcane supplied to the mill is coming i.e. a first-year ratoon or fifth year ratoon, which could affect its sucrose content, which is a concern of cane managers in India (Figure 6.3). Virgin sugarcane crops are also less susceptible to red-rot disease, a disease that can reduce the quality of the sugarcane crops and create huge losses in yield (Figure 2.3), therefore avoiding ratooning the crop can also be a management practice against the disease (Shukla et al., 2017).

Replacing the other crop class, with more specific classes such as natural ecosystems and high conservation values landscapes (HCVs) could help assess Bonsucro indicators 4.1.1 “the operator ensures that the biodiversity and natural ecosystems are mapped” ; 4.1.4 “the operator ensures that areas of natural ecosystems (either legally protected or not) or areas classified as HCV have not been converted to agriculture on or after 1 January 2021” or just check that there are no sugarcane fields within designated areas and 4.1.3 “the operator ensures that areas of legally protected natural ecosystems (nationally or internationally), or areas classified as HCV, have not been converted to sugarcane on or after 1 January 2008” (see Table 2.3).

Sentinel-2a data was launched in 2015 and Sentinel-2b was launched in 2017 therefore in order to monitor dLUC from natural ecosystems and HCVs to sugarcane in periods where Sentinel-2 data is unavailable, the developed classification approach would also have to be modified to deal with low frequency time-series data (Landsat only), in addition to changing the other crop class to natural ecosystems and HCVs. Additional classes for water, potential weeds, sugarcane affected by pests and sugarcane affected by diseases could also help

with Bonsucro indicator 4.3.1 “the operator identifies the main water resources and catchment areas” (see Table 2.3).

There is a general limitation of EO datasets when assessing social sustainability. This gap could be filled with commercial data, but that would incur cost. Alternatively, the classification model developed in Chapter 5, could be applied to areas where other data on social sustainability is available. By combining classification outputs with different sources of socio-economic data, many different aspects of social sustainability can be assessed. For example, to limit heat stress during the ‘sugarcane growth’ component of the supply chain (Figure 6.3), workers must have access to areas of shade, adequate sanitation and access to water in the field, therefore knowing where areas of shade and toilet blocks could be in relation to the classified sugarcane fields could help with the assessment of indicator 2.1.3 “the operator ensures that all workers on the premises included in the unit of certification have access to safe water and adequate sanitation facilities”. Information relating to road access and the distance of sugarcane fields from hospitals and schools could also indirectly provide information relating to indicators 2.1.6 “the operator ensures that first aid and an emergency response are available to all workers on the premises included in the unit of certification”. Additionally, a Geographical Information systems (GIS) system, that integrates EO data and social data, could be used to help provide information relating to 2.3.4 “the operator ensures that there is no child labour, for all persons on the premises included in the unit of certification, regardless of whether they are employed by the operator” (see Table 2.3).

In addition to crop mapping and the assessment of direct land use change, the outputs of chapters 3-5 can also be useful for managing irrigation and water use, yields and sugar content over large areas quickly (Bonsucro Indicator 3.1.2). Water is the main limiting factor during the mid-season (Carr and Knox, 2011; Mall et al., 2016). The methodology developed in Chapter 3 can be used to derive the mid-season from optical remote sensing datasets and the Initial knee can be derived from passive remote sensing data time-series (specifically VH-backscatter) using the methodology developed in Chapter 4. The EO data used to develop these methods are open access and free to use, therefore combining

these methods provides growers the opportunity to monitor flood irrigation events before the onset of tillering when used with weather data and irrigation-related field data. If yield data is also available, the impact of flood irrigation before the tillering versus no flood irrigation on yield could be assessed to help inform future management practices for improved yields and irrigation productivity (Mall et al., 2016). Root borer, *Emmalocera depressella* infestations, for example, can be managed by flooding the sugarcane fields, therefore the methodology developed in Chapter 4 can provide growers with the potential opportunity to monitor the impact of flood irrigation events on managing infestations and its impact on yield (Figure 6.3). Limiting the supply of water to sugarcane crops is also recommended during the late season, so deriving this stage is useful for informing farmers on when to limit water supply (Mall et al., 2016).

Root borer, *Emmalocera depressella* infestations can also be managed by intercropping cowpea, black and green grams with sugarcane (Mishra et al., 2023). The results of Chapter 3 showed that fields with multiple crops being grown in the field boundary, did not display the idealised sugarcane trapezoid shape. This therefore suggests that the methodology developed Chapter 3, could also be used to monitor intercropping for managing Root borer infestations, with the shape of the intercropped fields differing from non-intercropped sugarcane fields.

The use of agrochemicals such as fertilisers and pesticides is also a common management strategy during the 'sugarcane growth' component of the sugarcane supply chain (Figure 6.3). The optimal use of nutrients and pesticides can be related to specific growth stages therefore, the derivation of key crop growth stages, using methods developed in Chapter 3, alongside weather datasets could be useful for monitoring the most appropriate time for "earthing up" and application of fertilisers for managing Shoot Borer infestations.

The methodologies developed in Chapters 3-5 were dependent on using a complete growth cycle of observations for classification and deriving key sugarcane growth stages that are important for irrigation management. This can only be carried out after a sugarcane growth cycle has been completed.

Nevertheless, the use of historic time-series data to monitor stages over large areas and time scales is useful for establishing baselines for irrigation management where yields and intervention practices are very variable. This is important for improving sustainability in the ‘sugarcane growth’ component of the sugarcane supply chain (Figure 6.3).

6.5.3 Practical Applications of EO for sustainable harvest

Harvesting is the second component of the sugarcane supply chain (Figure 6.3). Modifying the time-series classification approach developed in Chapter 5 by including an additional class for sugarcane that is burnt after harvest would help with identifying areas where sugarcane burning takes place. Each GPS derived field boundary is associated with a particular farmer therefore classifying sugarcane burning would help with identifying field boundaries that are prone to burning, which could be useful for Bonsucro indicator 4.2.5 “the operator prevents sugarcane tops and leaves from being burned after harvest” (see Table 2.3).

6.5.4 Practical Applications of EO for sustainable transport and processing

The length of the mid-season and late-season can be derived from optical data (Chapter 3) and harvest derived from VH-backscatter data (Chapter 4). One recommendation to maximise sugar yields is to gather the sugarcane within 24 hours after harvest to reduce post-harvest losses (Misra et al., 2022). Theoretically, the use of near-real-time EO data could help identify the time delay between harvest and collection for use in the factory, and it could be used to help plan factory routes and allocate resources for crushing. This highlights a limitation of the algorithms developed in Chapters 3 and 4, growth stages could not be derived within season without information on the location of fields, as a whole season of data was needed to derive the subsequent growth stages and classification. However, the use of historical data, such as those used in this thesis can be useful in providing baselines and predicting the timings of sugarcane harvests based on the harvests of previous years (Moore et al., 2022). Sugarcane factories in India, for example, are only run for 4-7 months a year, therefore predicted harvests may help to run factories at full capacity for the

shortest time possible as this would allow hauliers to plan optimal routes for cane collection (Figure 6.3).

6.5.5 An ideal sugarcane supply chain

Currently India's sugarcane sector operates within a complex framework of low mechanisation, seasonal variability, water stress, and delayed farmer payments (Patil et al., 2021). Hence, the research carried out in this thesis is just the first step to improving the sustainability of the current Indian supply chain. An ideal EO system for India's sugarcane supply chain should be multi-tiered, integrating satellite data, digital tools, and institutional collaboration. The use of EO data from satellites such as Sentinel-1, Sentinel-2, and Landsat 8/9 would be at the heart of the system. These datasets would be used alongside the methods developed in Chapters 3-5 to first locate sugarcane and then monitor key growth stages across large geographic areas. The next step would be to feed data from these sources into a centralised web-based analytical dashboard. This dashboard should be accessible to sugar mills, government agencies, and agronomic advisors, providing real-time insights into crop health, expected yields, and regional trends. For farmers, a mobile application would serve as a user-friendly interface offering timely alerts on harvest readiness, irrigation scheduling, and potential disease outbreaks. To guarantee coordinated data sharing, policy support, and on-ground implementation the system would also need strong institutional links such as state agriculture departments, and sugar mill cooperatives. Together, these components would form a robust decision-support system, enabling more efficient, transparent, and responsive management of the sugarcane supply chain (Patil et al., 2024).

However, this type of technological advancement is not without its limitations. Smallholder farmers may struggle to access or interpret remote sensing tools creating a digital divide between stakeholders. Over-reliance on EO data without human validation could further compound errors in agricultural planning with misinterpretation of data leading to incorrect decisions around yield and logistics. Automation may also result in job displacement and there are privacy concerns, and cost barriers associated with high-resolution satellite data for smaller fields.

It is therefore recommended to include investing in farmer education and digital literacy, offering subsidies and enforcing strong data governance frameworks to protect farmer privacy. With thoughtful implementation and inclusive policy support, incorporating EO into the Indian sugarcane supply chain has the potential to significantly enhance productivity, sustainability, and farmer livelihoods.

7 CONCLUSIONS AND FURTHER WORK

In this section, the contribution to knowledge made during this thesis are summarised, in relation to each objective and recommendations about how information from EO can improve the sustainability of Indian sugarcane cultivation (Objective 5)

7.1 Conclusions

The first contribution of knowledge from this thesis, is the development of a method for identifying optimal pre-processing parameters for cloud and noise removal, which allowed for parameters specifically tailored for the time-series analysis of sugarcane fields in India to be derived (Objective 2). Monitoring sugarcane growth over time is critical for managing resources. Optimising agricultural practices, for example fertilisation, irrigation, and pest and disease management, requires information about the seasonal development of sugarcane as management varies with sugarcane growth stage. 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. Local calibration of pre-processing parameters is important, for removing optimal levels of cloud and noise, however prior to this thesis existing methods for choosing pre-processing parameters for cloud and noise removal, was subjective and left to the experience of the data analyst (Objective 1). The optimisation methodology for pre-processing time-series data, developed in this thesis overcame the above challenges and can be followed and applied to other study regions for selecting pre-processing parameters, which couldn't be done before as frameworks for this simply did not exist (Objective 2). Furthermore, to address high cloud coverage during the monsoon season, two different EO datasets from different optical satellites were combined in this thesis, by developing a local regression calibration, to increase the frequency of cloud-free observations from every 18–36 and 4–9 days, respectively, for the first and second optical sensor time-series datasets, to every 3–6 (Objective 2).

This work also demonstrated that records of growth stages can be derived from existing crop models. The second contribution of knowledge was the use of the FAO crop growth model to help develop a method for automatically deriving the timings of key sugarcane growth stages, such as the start of the season, mid-season and EOS, from optical EO time-series, without prior knowledge of planting and harvesting dates. The novel technology developed in this thesis, was also able to provide more information about growth stages important for management than previously used simplified models like double logistic curves. This is because key stages were not removed through over-smoothing and were able to provide information regarding differences in the timing and length of growth stages between virgin cane, ratoons, and different management practices (Objective 2). Prior to this thesis, the current methods used to derive information about the timing of key sugarcane growth stages required the use of field survey or crop calendar derived planting and harvesting dates, however the heterogenous nature of sugarcane management in India makes obtaining planting and harvesting date information difficult with crop calendar information being unavailable and the use of field surveys being expensive, lengthy and labour-intensive.

To add to the above contributions, this thesis, found benefits to combining the use of passive optical and active remote sensing time-series data for overcoming the research gaps that existed around using EO time-series data for monitoring sugarcane growth and management over time (Objective 2). The occurrence of persistent cloud cover during the monsoon affected seasonal decomposition of long optical time-series and the identification of subsequent growth stages in a handful of fields, when optical data was used alone. While the results of Chapter 4 showed no advantage of using active EO data alone for overcoming cloud to decompose long-time-series data into seasons, as the biomass signal from the optical sensors were stronger than the structural differences in the active EO data, combining the passive optical and active EO time-series data produced successful temporal decomposition of active EO time-series into individual seasons. The structural information provided by the active EO datasets also gave different information about growth and insights into how different management

techniques affected the shape of the backscatter time-series, providing an indirect indication of yield, than the biomass data provided by optical EO datasets (Objective 2). The technology developed in this thesis can also be used alongside additional climate datasets to examine the relationship between climate and the length of growth seasons (Objective 1).

Time-series-based methods for sugarcane classification is the only way to identify sugarcane amongst spectrally similar crops, using existing EO data, over wide areas (Objective 1). In this thesis, a novel approach for classifying sugarcane and other crops from long continuous time-series data was developed by modifying a CNN model initially designed for the medical field (Objective 3). The developed model contributed to knowledge as it allowed for localisation and classification to be carried out in the same step, which overcame the need for isolating patterns in time-series data prior to classification which was a limitation of previous classification approaches (Objective 3). The classification of sugarcane using the method developed in this thesis, therefore, has the potential to produce large-scale global maps of sugarcane and can monitor sustainable and unsustainable sugarcane production practices such as mono-cropping or crop rotation. This is important, as the ability to monitor stages over large areas and time scales is advantageous for establishing a baseline for best practices for management where yield is high and intervention procedures for improved sustainability where yield is low.

7.2 Recommendations for further work

Individually the algorithms derived in Chapters 3-5 (to meet Objectives 2 and 3) provide decision support over large areas and automate the process of information retrieval for sustainable resource management, yield, and biodiversity, but together can be used as a strong tool that can be implemented into existing sustainability frameworks for improving the sustainability of sugarcane production. The algorithms derived in Chapters 3-5 can provide information for measuring the indicators of the Bonsucro Production Standard

improving the sustainable management of sugarcane production for large areas efficiently and by providing timely information for rapid appraisal.

The systematic monitoring and certification of sustainable sugarcane production requires the integration of information from a range of sources including EO, yield, and socio-economic data. Within this thesis, the use of freely available EO data has been used to identify and classify sugarcane and monitor growth and development, in the context of environmental sustainability.

For sustainable development, cane operators can use their existing field data to create a baseline for historical production, combining field data on growth stages, management and the location of sugarcane fields with any data from the factory, including yields, records of cane being affected by pests or disease, and social data relating to working conditions. They could then use this information for planning future operations or sustainability campaigns to target resources efficiently. Examples of this include the use of data relating to growth stages and yields to plan irrigation for sustainable water use. Additional historical records of pests like shoot Borer can also be used alongside growth stage and yield data to manage application of fertilisers. Additionally, derived harvest alongside yield datasets could be used to help plan factory routes and allocate resources for crushing, by helping cane operators identify time delay between harvest and collection for use in the factory and its impact on yield. The success of any investment in sustainability can also be assessed, for example, the effect of replacing flood irrigation with drip irrigation, crop rotation and intercropping on yield and vulnerability to disease like red rot and infestations of pests like Root borer.

For environmental sustainability, historical information on the location of sugarcane fields and high conservation values (HCVs) landscapes, can be used to identify the conversion of sugarcane to HCVs. Historical information regarding the location of sugarcane fields, alongside those where trash burning is prone, can be used to assess the effect of trash burning on yield and vulnerability to pests.

Finally, the addition of social data e.g. the distance to urban areas (houses) to sugarcane fields where trash burning is common could help managers to assess social sustainability by providing information about not only the health of farmers associated with those fields but also residents (as trash burning is associated with cancer risk). Distance of sugarcane fields to sanitation facilities can also be used to assess social sustainability.

8 REFERENCES

- Abdel-Rahman, E. M., Ahmed, F. B. (2008). The application of remote sensing techniques to sugarcane (*Saccharum spp. hybrid*) production: A review of the literature. *International Journal of Remote Sensing*, 29(13), 3753–3767. <https://doi.org/10.1080/01431160701874603>
- Adami, M., Rudorff, B. F. T., Freitas, R. M., Aguiar, D. A., Sugawara, L. M., Mello, M. P. (2012). Remote Sensing Time-series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil. *Sustainability*, 4(4), 4. <https://doi.org/10.3390/su4040574>
- Alface, A.B., Pereira, S.B., Filgueiras, R., Cunha, F.F. (2019). Sugarcane spatial-temporal monitoring and crop coefficient estimation through NDVI. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 23(5), 330–335. <https://doi.org/10.1590/1807-1929/agriambi.v23n5p330-335>.
- Anjaly, V., Sindhu, V. K., & Singh, K. (2024). Agronomical interventions for improving sugarcane water productivity: A review. *Sugar Tech*, 26(4), 1053–1067. <https://doi.org/10.1007/s12355-024-01459-0>
- Bao, X., Zhang, R., Lv, J., Wu, R., Zhang, H., Chen, J., Zhang, B., Ouyang, X., Liu, G. (2023). Vegetation descriptors from Sentinel-1 SAR data for crop growth monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 203, 86–114. <https://doi.org/10.1016/j.isprsjprs.2023.07.023>
- Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., and Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing*, 31(20), 5391–5407. <https://doi.org/10.1080/01431160903349057>
- Birdi, P.K., Kale, K. (2017, October 23–27). Identification of growth stage of sugarcane crop using decision tree for Landsat 8 data [Conference Paper]. Proceedings of 38th Asian conference on remote sensing, New Delhi, India.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Bonsucro. (2023, July). Bonsucro Production Standard v5.2. Last accessed May 5, 2025, from <https://bonsucro.com/production-standard/>.
- Buda, M., Maki, A., Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>.
- Chelali, M., Kurtz, C., Puissant, A., Vincent, N. (2021). Deep-STaR: Classification of image time-series based on spatio-temporal representations. *Computer Vision and Image Understanding*, 208–209, 103221. <https://doi.org/10.1016/j.cviu.2021.103221>
- Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based

on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91(3–4), 332–344. <https://doi.org/10.1016/j.rse.2004.03.014>.

Cihlar, J., and Jansen, L. J. M. (2001). From Land Cover to Land Use: A Methodology for Efficient Land Use Mapping over Large Areas. *The Professional Geographer*, 53(2), 275–289. <https://doi.org/10.1111/0033-0124.00285>.

Cruz-Sanabria, H., Sanchez, M.G., Rivera-Caicedo, J.P., Avila-George, H. (2020, October 21-23). Identification of phenological stages of sugarcane cultivation using Sentinel-2 images [Conference Paper]. Proceedings of the 2020 9th International Conference on Software Process Improvement (CIMPS), Mazatlan, Sinaloa, Mexico. <https://doi.org/10.1109/cimps52057.2020.9390095>.

Dablain, D., Jacobson, K. N., Bellinger, C., Roberts, M., Chawla, N. V. (2024). Understanding CNN fragility when learning with imbalanced data. *Machine Learning*, 113(7), 4785–4810. <https://doi.org/10.1007/s10994-023-06326-9>.

de Andrade, S. J., Varella, S. D., Pereira, G. T., Zocolo, G. J., de Marchi, M. R. R., Varanda, E. A. (2011). Mutagenic activity of airborne particulate matter (PM10) in a sugarcane farming area (Araraquara city, southeast Brazil). *Environmental Research*, 111(4), 545–550. <https://doi.org/10.1016/j.envres.2011.03.004>

Demissie, B., Vanhuysse, S., Grippa, T., Flasse, C., Wolff, E. (2023). Using Sentinel-1 and Google Earth Engine cloud computing for detecting historical flood hazards in tropical urban regions: A case of Dar es Salaam. *Geomatics, Natural Hazards and Risk*, 14(1), 2202296. <https://doi.org/10.1080/19475705.2023.2202296>

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. (2023). Understanding Sentinel-1 backscatter response to sugarcane yield variability and waterlogging. *Remote Sensing of Environment*, 290, 113555. <https://doi.org/10.1016/j.rse.2023.113555>

Dimov, D., Uhl, J.H., Löw, F., Seboka, G.N. (2022). Sugarcane yield estimation through remote sensing time-series and phenology metrics. *Smart Agricultural Technology*, 2, 100046. <https://doi.org/10.1016/j.atech.2022.100046>.

Donaldson, R. A., Redshaw, K., Rhodes, R., Antwerpen, R. (2008). Season effects on productivity of some commercial South African sugarcane cultivars, II: Trash production. *Proceedings of the South African Sugarcane Technology Association*, 81, 517-527.

Doorenbos, J., Pruitt, W.O. (1977). Crop water requirements. FAO irrigation and drainage paper (24). FAO. 1–144.

Dou, P., Shen, H., Li, Z., Guan, X. (2021). Time-series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102477. <https://doi.org/10.1016/j.jag.2021.102477>.

- El Chami, D., Daccache, A., El Moujabber, M. (2020). What are the impacts of sugarcane production on ecosystem services and human well-being? A review. *Annals of Agricultural Sciences*, 65(2), 188–199. <https://doi.org/10.1016/j.aoas.2020.10.001>.
- Fageria, N. K., Baligar, V. C., Jones, C. A. (2010). Growth and mineral nutrition of field crops (3rd ed.). CRC Press. <https://doi.org/10.1201/b10160>.
- FAO. (2021). World Food and Agriculture—Statistical Yearbook 2021. Last accessed May 5, 2025, from <https://www.fao.org/3/cb4477en/cb4477en.pdf>
- Fu, W., Ma, J., Chen, P., Chen, F. (2020). Remote Sensing Satellites for Digital Earth. In Guo, H., Goodchild, M.F., Annoni, A (Eds.), *Manual of Digital Earth* (1st ed). 55-123. Springer. https://doi.org/10.1007/978-981-32-9915-3_3
- Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. *Genetic Programming and Evolvable Machines* 19, 305–307 (2018). <https://doi.org/10.1007/s10710-017-9314-z>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2016). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., Xavier, P. K. (2006). Increasing trend of extreme rain events over India in a warming environment. *Science*, 314(5804), 1442–1445. <https://doi.org/10.1126/science.1132027>
- Gujja, B., Loganandhan, N., Vinod Goud, V., Agarwal, M., Dalai, S. (2010). Sustainable Sugarcane Initiative (SSI) Improving Sugarcane Cultivation in India. (1st ed.). ICRISAT. Last accessed May 5, 2025, from <http://oar.icrisat.org/id/eprint/5701>
- Hess, T., Aldaya, M., Fawell, J., Franceschini, H., Ober, E., Schaub, R., Schulze-Aurich, J. (2013). Understanding the impact of crop and food production on the water environment—Using sugar as a model. *Journal of the Science of Food and Agriculture*, 94(1), 2–8. <https://doi.org/10.1002/jsfa.6369>
- Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Inman-Bamber, N., Bonnett, G., Smith, D., Thorburn, P. (2005). Sugarcane physiology: Integrating from cell to crop to advance sugarcane production. *Field Crops Research*, 92(2-3), 115–117. <https://doi.org/10.1016/j.fcr.2005.01.011>
- Jin, Z.N., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B. (2019). Smallholder maize area and yield mapping at national scales with Google earth engine. *Remote Sensing of Environment*, 228, 115–128.
- Johnson, A. K. L., Kinsey-Henderson, A. E. (1997). Satellite-based remote sensing for monitoring land use in the sugar industry. *Heliyon*, 9(8), e18982. <https://doi.org/10.1016/j.heliyon.2023.e18982>
- Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C.,

- and Waske, B. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing*, 8(1), 70. <https://doi.org/10.3390/rs8010070>
- Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., Chuvieco, E. (2020). Towards delivering on the Sustainable Development Goals using Earth observations. *Remote Sensing of Environment*, 247, 111930. <https://doi.org/10.1016/j.rse.2020.111930>
- Kensert, A., Collaerts, G., Efthymiadis, K., Van Broeck, P., Desmet, G., Cabooter, D. (2021). Deep convolutional autoencoder for the simultaneous removal of baseline noise and baseline drift in chromatograms. *Journal of Chromatography A*, 1646, 462093. <https://doi.org/10.1016/j.chroma.2021.462093>
- Kooistra, L., Berger, K., Brede, B., Graf, L. V., Aasen, H., Roujean, J.-L., Machwitz, M., Schlerf, M., Atzberger, C., Prikaziuk, E., Ganeva, D., Tomelleri, E., Croft, H., Reyes Muñoz, P., Garcia Millan, V., Darvishzadeh, R., Koren, G., Herrmann, I., Rozenstein, O., Belda, S., Rautiainen, M., Rune Karlsen, S., Figueira Silva, C., Cerasoli, S., Pierre, J., Tanır Kayıkçı, E., Halabuk, A., Tunc Gormus, E., Fluit, F., Cai, Z., Kycko, M., Udelhoven, T., Verrelst, J. (2024). Reviews and syntheses: Remotely sensed optical time-series for monitoring vegetation productivity. *Biogeosciences*, 21(2), 473–511. <https://doi.org/10.5194/bg-21-473-2024>
- Kussul, N., Lavreniuk, M., Kolotii, A., Skakun, S., Rakoid, O., & Shumilo, L. (2019). A workflow for Sustainable Development Goals indicators assessment based on high-resolution satellite data. *International Journal of Digital Earth*, 13(2), 309–321. <https://doi.org/10.1080/17538947.2019.1610807>
- Lawal, M.O. (2021). Tomato detection based on modified YOLOv3 framework. *Scientific Reports*, 11(1), 1447. <https://doi.org/10.1038/s41598-021-81216-5>
- Lele, U., Goswami, S. (2017). The fourth industrial revolution, agricultural and rural innovation, and implications for public policy and investments: A case of India. *Agricultural Economics*, 48(S1), 87–100. <https://doi.org/10.1111/agec.12388>
- Lemay, A. (2019). Kidney Recognition in CT Using YOLOv3. *ArXiv*, <https://doi.org/10.48550/arXiv.1910.01268>
- Li, S., Xu, L., Jing, Y., Yin, H., Li, X., and Guan, X. (2021). High-quality vegetation index product generation: A review of NDVI time-series reconstruction techniques. *International Journal of Applied Earth Observation and Geoinformation*, 105, 102640. <https://doi.org/10.1016/j.jag.2021.102640>
- Li, Y.R. (2022). Growth and development of sugarcane (*Saccharum* spp. Hybrid) and its relationship with environmental factors. In *Agro-Industrial Perspectives on Sugarcane Production under Environmental Stress*. (1st ed.). Springer Nature Singapore. 1–11. https://doi.org/10.1007/978-981-19-3955-6_1
- Liu, L., Cao, R., Chen, J., Shen, M., Wang, S., Zhou, J., He, B. (2022). Detecting crop phenology from vegetation index time-series data by improved

- shape model fitting in each phenological stage. *Remote Sensing of Environment*, 277, 113060. <https://doi.org/10.1016/j.rse.2022.113060>
- Liu, Y., Ren, C., Liang, J., Zhou, Y., Xue, X., Ding, C., and Lu, J. (2023). A Robust Index Based on Phenological Features to Extract Sugarcane from Multisource Remote Sensing Data. *Remote Sensing*, 15(24), 24. <https://doi.org/10.3390/rs15245783>.
- Lovell, J. L., Graetz, R. D. (2001). Filtering Pathfinder AVHRR Land NDVI data for Australia. *International Journal of Remote Sensing*, 22(13), 2649–2654. <https://doi.org/10.1080/01431160116874>
- Mahato, A. (2014). Climate Change and Its Impact on Agriculture. *Agricultural and Food Sciences, Environmental Science*, 4(4), 6.
- Mall, R. K., Sonkar, G., Bhatt, D., Sharma, N. K., Baxla, A. K., Singh, K. K. (2016). Managing impact of extreme weather events in sugarcane in different agro-climatic zones of Uttar Pradesh. *Mausam*, 67(1), 233-250. <https://doi.org/10.54302/mausam.v67i1.1187>
- Marin, F.R., Rattalino Edreira, J.I., Andrade, J.F., Grassini, P. (2021). Sugarcane yield and yield components as affected by harvest time. *Sugar Tech*, 23(4), 819–826. <https://doi.org/10.1007/s12355-020-00945-5>
- Maxwell, A. E., Warner, T. A., Guillén, L. A. (2021). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sensing*, 13(13), 13. <https://doi.org/10.3390/rs13132591>
- Mello, M. P., Atzberger, C., Formaggio, A. R. (2014, July 13-18). Near real time yield estimation for sugarcane in Brazil combining remote sensing and official statistical data [Conference Paper]. *IEEE Geoscience and Remote Sensing Symposium*, Quebec City, QC, Canada. <https://doi.org/10.1109/IGARSS.2014.6947635>
- Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M. (2020). Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 333–346. <https://doi.org/10.1016/j.isprsjprs.2020.05.013>.
- Mishra, A. K., Kurre, D., Mishra, S. K., Tiwari, B. (2023). Major Insect Pests of Sugarcane (*Saccharum officinarum*) Crop and Their Integrated Management Approaches. *Just Agriculture*, 4(3), 160-166.
- Misra, V., Mall, A., Solomon, S., Ansari, M.I. (2022). Post-harvest biology and recent advances of storage technologies in sugarcane. *Biotechnology Reports*, 33, e00705. <https://doi.org/10.1016/j.btre.2022.e00705>.
- Mnisi, M.S. (2012). The concept of sustainable sugarcane production: Global, African and south African perceptions. *African Journal of Agricultural Research*, 7(31), 4337-4343. <https://doi.org/10.5897/ajar12.446>.
- Molijn, R. A., Iannini, L., Vieira Rocha, J., and Hanssen, R. F. (2019). Sugarcane Productivity Mapping through C-Band and L-Band SAR and Optical

Satellite Imagery. *Remote Sensing*, 11(9), 9.
<https://doi.org/10.3390/rs11091109>

Moore, E. K., Kriesberg, A., Schroeder, S., Geil, K., Haugen, I., Barford, C., Johns, E. M., Arthur, D., Sheffield, M., Ritchie, S. M., Jackson, C., and Parr, C. (2022). Agricultural data management and sharing: Best practices and case study. *Agronomy Journal*, 114(5), 2624–2634.
<https://doi.org/10.1002/agj2.20639>

Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J.F., 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 Sensing*, 6(7), 6620–6635. <https://doi.org/10.3390/rs6076620>.

Mulianga, B., Bégué, A., Simoes, M., and Todoroff, P. (2013). Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI. *Remote Sensing*, 5(5), 5. <https://doi.org/10.3390/rs5052184>

Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N., Reiche, J. (2021). Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine. *Remote Sensing*, 13(10), 10.
<https://doi.org/10.3390/rs13101954>

Nandhini, T. S. K. D., Padmavathy, V. (2017). A study on sugarcane production in India. *International Journal of Advanced Research in Botany*, 3(2), 13-17.
<https://doi.org/10.20431/2455-4316.0302003>

Narayanamoorthy, A. (2004). Impact Assessment of Drip Irrigation in India: The Case of Sugarcane. *Development Policy Review*, 22 (4), 443–62.
<https://doi.org/10.1111/j.1467-7679.2004.00259.x>

Nie, Y., Sommella, P., O’Nils, M., Liguori, C., Lundgren, J. (2019). Automatic detection of melanoma with yolo deep convolutional neural networks. *E-Health and Bioengineering Conference (EHB)*, Iasi, Romania, 2019, 1-4, doi: 10.1109/EHB47216.2019.8970033. Olaode, A., Naghdy, G., Todd, C. (2014). Unsupervised Classification of Images: A Review. *International Journal of Image Processing*, 8(5), 325-342.

Olfindo, N. T., de la Cruz, R. M., Borlongan, N. J. B., Marciano., J. J. S., Olalia, L. C. (2020). Sugarcane plantation mapping using dynamic time warping from multi-temporal Sentinel-1a RADAR images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 519–524.
<https://doi.org/10.5194/isprs-annals-V-3-2020-519-2020>

Ottinger, M., Kuenzer, C. (2020). Spaceborne L-Band Synthetic Aperture Radar Data for Geoscientific Analyses in Coastal Land Applications: A Review. *Remote Sensing*, 12(14), 14. <https://doi.org/10.3390/rs12142228>

Patil, S.M., Prathapan, K., Patil, S.B., Jagtap, S. and Chavan, S.M. (2024). Critical issues and challenges in sugarcane supply chain management: A global perspective. *Sugar Tech*, 26(4), 1033-1052. <https://doi.org/10.1007/s12355-024-01456-3>

- Pavani, G., Malhotra, P. K., Verma, S. K. (2023). Flowering in sugarcane- insights from the grasses. *3 Biotech*, 13(5), 154. <https://doi.org/10.1007/s13205-023-03573-4>
- Pelletier, C., Webb, G. I., Petitjean, F. (2019). Temporal Convolutional Neural Network for the Classification of Satellite Image Time-series. *Remote sensing*, 11(5), 523. <https://www.mdpi.com/2072-4292/11/5/523>
- Pires de Lima, R., Marfurt, K. (2020). Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sensing*, 12(1), 86. <https://doi.org/10.3390/rs12010086>
- Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M. (2008). How to Evaluate Models: Observed vs. Predicted or Predicted vs. Observed? *Ecological Modelling*, 216(3), 316–22. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>
- Pnvr, K., Bandaru, V. (2023). Mapping sugarcane residue burnt areas in smallholder farming systems using machine learning approaches. *Smart Agricultural Technology*, 6, 100347. <https://doi.org/10.1016/j.atech.2023.100347>
- Praveen, B., Mustak, S., Sharma, P. (2019). Assessing the Transferability of Machine Learning Algorithms Using Cloud Computing and Earth Observation Datasets for Agricultural Land Use/cover Mapping. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 423, 585–592. <https://doi.org/10.5194/isprs-archives-XLII-3-W6-585-2019>
- Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–49. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., and Roberts, D. (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment*, 112(5), 2272–2283. <https://doi.org/10.1016/j.rse.2007.10.004>
- Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., Arkebauer, T.J. (2010). A Two-Step Filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sensing of Environment*, 114(10), 2146–2159. <https://doi.org/10.1016/j.rse.2010.04.019>
- Sakamoto, T. (2018). Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 176–192. <https://doi.org/10.1016/j.isprsjprs.2018.02.011>
- Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behaviour. *Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops*, Minneapolis, MI, USA. 166-171. <https://doi.org/10.1109/icdcs.2011.20>

- Shukla, P., Kumar, A., Kumar, M., Raghuvanshi, H. R., Kumar, A., Mishra, P. (2022). Red Rot Disease (*Colletotrichum falcatum*) of Sugarcane and their Integrated Management. *Just Agriculture*, 3(4), 1-4.
- Silva, F. S., Cristale, J., André, P. A., Saldiva, P. H. N., Marchi, M. R. R. (2010). PM2.5 and PM10: The influence of sugarcane burning on potential cancer risk. *Atmospheric Environment*, 44(39), 5133–5138. <https://doi.org/10.1016/j.atmosenv.2010.09.001>
- Singh, P., Rai, R.K., Suman, A., Srivastava, T.K., Singh, K.P., Arya, N., Yadav, R.L. (2014). Soil-root interface changes in sugarcane plant and ratoon crops under subtropical conditions: Implications for dry-matter accumulation. *Communications in Soil Science and Plant Analysis*, 46(4), 454–475. <https://doi.org/10.1080/00103624.2014.997385>.
- Singh, S., Srivastava, S. K., & Jangirala, S. (2021). System Dynamics Analysis of Sugarcane Supply Chain in Indian Sugar Industry. *Global Business Review*, 26(2), 468-480. <https://doi.org/10.1177/0972150921999521>
- Singh, P., Tiwari, A. K. (2018). Sustainable Sugarcane Production. (1st ed.). Apple Academic Press. <https://doi.org/10.1201/9781351047760>
- Singha, C., and Swain, K. C. (2023). Rice crop growth monitoring with sentinel 1 SAR data using machine learning models in google earth engine cloud. *Remote Sensing Applications: Society and Environment*, 32, 101029. <https://doi.org/10.1016/j.rsase.2023.101029>.
- Smith, W. K., Nelson, E., Johnson, J. A., Polasky, S., Milder, J. C., Gerber, J. S., West, P. C., Siebert, S., Brauman, K. A., Carlson, K. M., Arbutnot, M., Rozza, J. P., and Pennington, D. N. (2019). Voluntary sustainability standards could significantly reduce detrimental impacts of global agriculture. *Proceedings of the National Academy of Sciences*, 116(6), 2130–2137. <https://doi.org/10.1073/pnas.1707812116>
- Solomon, S. (2016). Sugarcane production and development of sugar industry in India. *Sugar Tech*, 18(6), 588–602. <https://doi.org/10.1007/s12355-016-0494-2>
- Solomon, S., Rao, G. P., Swapna, M. (2020). Impact of COVID-19 on Indian Sugar Industry. *Sugar Tech*, 22(4), 547–551. <https://doi.org/10.1007/s12355-020-00846-7>
- Solomon, S., Swapna, M. (2022). Indian Sugar Industry: Towards Self-reliance for Sustainability. *Sugar Tech*, 24(3), 630–650. <https://doi.org/10.1007/s12355-022-01123-5>
- Stubbings, P., Peskett, J., Rowe, F., Arribas-Bel, D. (2019). A hierarchical urban forest index using street-level imagery and deep learning. *Remote Sensing*, 11(12), 1395. <https://doi.org/10.3390/rs11121395>.
- Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Baraldi, A., Blaschke, T. (2019). Big earth data: disruptive changes in earth observation data management and analysis? *International Journal of Digital Earth*, 13(7), 832–850. <https://doi.org/10.1080/17538947.2019.1585976>

- Toth, C., and Józków, G. (2016). Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 22–36. <https://doi.org/10.1016/j.isprsjprs.2015.10.004>
- Townshend, J., Justice, C., Li, W., Gurney, C., and McManus, J. (1991). Global land cover classification by remote sensing: Present capabilities and future possibilities. *Remote Sensing of Environment*, 35(2), 243–255. [https://doi.org/10.1016/0034-4257\(91\)90016-Y](https://doi.org/10.1016/0034-4257(91)90016-Y)
- Valova, I., Harris, C., Mai, T., and Gueorguieva, N. (2020). Optimization of Convolutional Neural Networks for Imbalanced Set Classification. *Procedia Computer Science*, 176, 660–669. <https://doi.org/10.1016/j.procs.2020.09.038>.
- Vavlas, N.C., Waive, T. W., Meersmans, J., Burgess, P. J., Fontanelli, G., and Richter, G. M. (2020). Deriving Wheat Crop Productivity Indicators Using Sentinel-1 Time-series. *Remote Sensing*, 12(15), 15. <https://doi.org/10.3390/rs12152385>
- Vikesh, K., Vinod, K., and Kamal, J. (2010). Development of Spectral Signatures and Classification of Sugarcane using ASTER Data. *International Journal of Computer Science and Communication*, 1(1), 245–251.
- Villareal, M. K., and Tongco, A. F. (2020). Remote Sensing Techniques for Classification and Mapping of Sugarcane Growth. *Engineering, Technology and Applied Science Research*, 10(4), 4. <https://doi.org/10.48084/etasr.3694>
- Viovy, N., Arino, O., Belward, A.S. (1992). The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, 13(8), 1585–1590. <https://doi.org/10.1080/01431169208904212>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wang, J., Xiao, X., Liu, L., Xiaocui, W., Qin, Y., Steiner, J. L., Dong, J. (2020). Mapping sugarcane plantation dynamics in Guangxi, China, by time-series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sensing of Environment*, 247, 111951. <https://doi.org/10.1016/j.rse.2020.111951>
- Wilson, A.M., Parmentier, B., Jetz, W. (2014). Systematic land cover bias in Collection 5 MODIS cloud mask and derived products—A global overview. *Remote Sensing of Environment*, 141, 149–154. <https://doi.org/10.1016/j.rse.2013.10.025>.
- Wulder, M. A., Kurz, W. A., and Gillis, M. (2004). National level forest monitoring and modeling in Canada. *Progress in Planning*, 61(4), 365–381. [https://doi.org/10.1016/S0305-9006\(03\)00069-2](https://doi.org/10.1016/S0305-9006(03)00069-2)

- Xu, F., Wang, Z., Lu, G., Zeng, R., Que, Y. (2021). Sugarcane ratooning ability: Research status, shortcomings, and prospects. *Biology*, 10(10), 1052. <https://doi.org/10.3390/biology10101052>
- Xu, X., Conrad, C., Doktor, D. (2017). Optimising phenological metrics extraction for different crop types in Germany using the moderate resolution imaging spectrometer (MODIS). *Remote Sensing*, 9(3), 254. <https://doi.org/10.3390/rs9030254>
- Yawson, D. O., Adu, M. O., and Osei, K. N. (2018). Spatial assessment of sugarcane (*Saccharum spp. L.*) production to feed the Komenda Sugar Factory, Ghana. *Heliyon*, 4(11). <https://doi.org/10.1016/j.heliyon.2018.e00903>
- Yeasin, M., Haldar, D., Kumar, S., Paul, R.K., Ghosh, S. (2022). Machine learning techniques for phenology assessment of sugarcane using conjunctive SAR and optical data. *Remote Sensing*, 14(14), 3249. <https://doi.org/10.3390/rs14143249>.
- Yu, L., Wang, J., and Gong, P. (2013). Improving 30 m global land-cover map FROM-GLC with time-series MODIS and auxiliary data sets: A segmentation-based approach. *International Journal of Remote Sensing*, 34(16), 5851–5867. <https://doi.org/10.1080/01431161.2013.798055>
- Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J. M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J., Asrar, G. R., Chen, M. (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nature Reviews Earth and Environment*, 3(7), 477–493. <https://doi.org/10.1038/s43017-022-00298-5>
- Zhang, H.K., Roy, D.P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., Roger, J.C. (2018). Characterization of Sentinel-2A and Landsat 8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sensing of Environment*, 215, 482–494. <https://doi.org/10.1016/j.rse.2018.04.031>
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57(4), 99. <https://doi.org/10.1007/s10462-024-10721-6>
- Zheng, Y., Li, Z., Pan, B., Lin, S., Dong, J., Li, X., Yuan, W. (2022). Development of a Phenology-Based Method for Identifying Sugarcane Plantation Areas in China Using High-Resolution Satellite Datasets. *Remote Sensing of Environment*, 14(5), 1274. <https://doi.org/10.3390/rs14051274>
- Zhou, J., Jia, L., Menenti, M., Gorte, B. (2016). On the performance of remote sensing time-series reconstruction methods – a spatial comparison. *Remote Sensing of Environment*, 187, 367–384. <https://doi.org/10.1016/j.rse.2016.10.025>
- Zhou, M., Ma, X., Wang, K., Cheng, T., Tian, Y., Wang, J., Zhu, Y., Hu, Y., Niu, Q., Gui, L., Yue, C., Yao, X. (2020). Detection of phenology using an improved shape model on time-series vegetation index in wheat. *Computers and Electronics in Agriculture*, 173, 105398. <https://doi.org/10.1016/j.compag.2020.105398>

Zhou, Q., Guan, K., Wang, S., Hipple, J., Chen, Z. (2024). From satellite-based phenological metrics to crop planting dates: Deriving field-level planting dates for corn and soybean in the U.S. Midwest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 216, 259–273.

<https://doi.org/10.1016/j.isprsjprs.2024.07.031>

Zhu, W., He, B., Xie, Z., Zhao, C., Zhuang, H., Li, P. (2022). Reconstruction of vegetation index time-series based on self-weighting function fitting from curve features. *Remote Sensing*, 14(9), 2247. <https://doi.org/10.3390/rs14092247>

9 APPENDICES

9.1 CURES form



15 January 2021

Dear Miss Joshi ,

Reference: CURES/12584/2021

Title: Earth observation for sustainable sugarcane production

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/12584/2021 has been reviewed. You may now proceed with the research activities you have sought approval for.

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

9.2 Sugarcane NDVI and VH-backscatter time-series profiles, with available photographs corresponding to the Start of the mid-season

The following figures show mean VH from a processed Sentinel-1 Ground Range Detected (GRD) time-series profile of a sugarcane field during the 2019 season, together with a pre-processed mean NDVI time-series profile (left). The primary y-axis on the left displays NDVI (in blue) and the date of the Start of the Mid-season (SMS) derived from the mean NDVI time-series profile is displayed as the solid blue vertical line. The secondary y-axis displays VH backscatter (in orange) on a decibel scale. The date of the initial knee derived from the mean VH profile is displayed as the solid orange vertical line. The photograph (right) was taken 10 days after the SMS was derived from the mean NDVI profile.

Table 9.1. The profile and photograph displayed in this table are of a 1st year ratoon sugarcane field, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 3ft and wet when surveyed on 27th April 2019

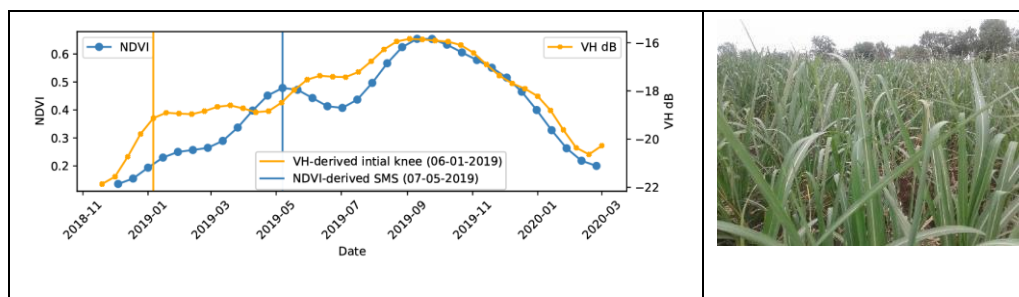
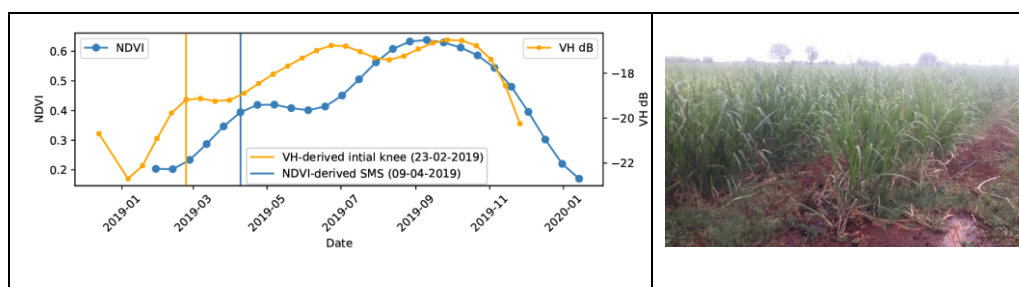


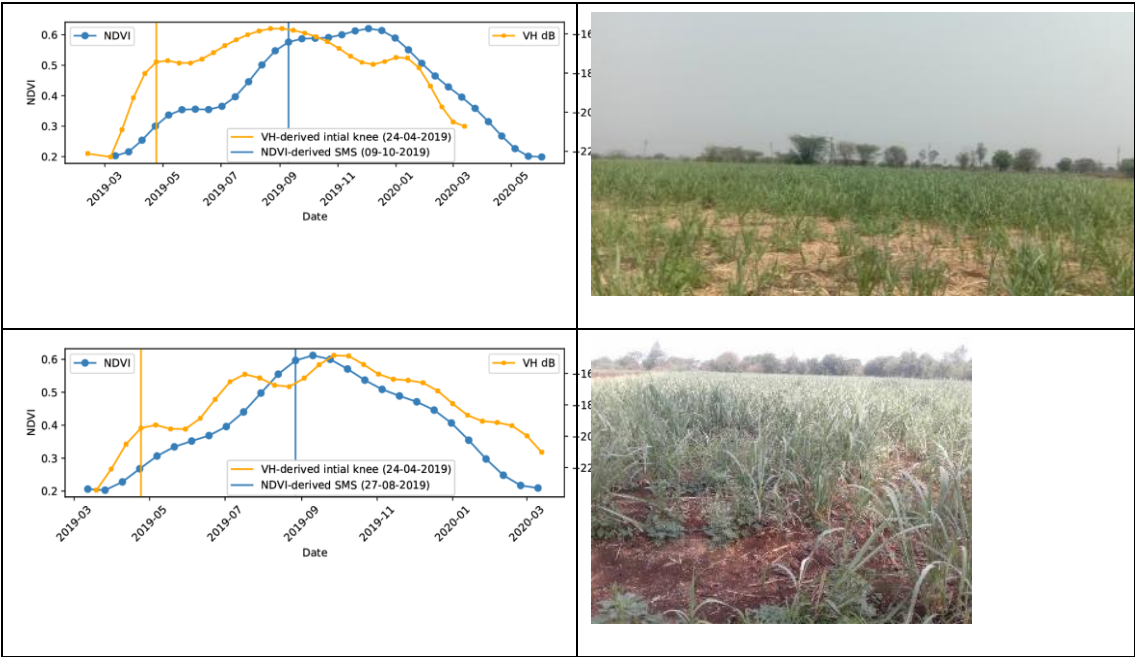
Table 9.2. The profile and photograph displayed in this table are of a virgin sugarcane field, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 4ft and wet when surveyed on 27th April 2019



9.3 Sugarcane NDVI and VH-backscatter time-series profiles, with photographs corresponding to the VH-backscatter derived initial knee

In the following figures mean VH from processed Sentinel-1 Ground Range Detected (GRD) time-series profiles of sugarcane fields during the 2019 season, together with pre-processed mean NDVI time-series profiles (left). The primary y-axis on the left displays NDVI (in blue) and the date of the Start of the Mid-season (SMS) derived from the mean NDVI time-series profile is displayed as the solid blue vertical line. The secondary y-axis displays VH backscatter (in orange) on a decibel scale. The date of the initial knee derived from the mean VH profile is displayed as the solid orange vertical line. The photographs (right) were taken on the same day that the initial knee was derived from the mean VH profile.

Table 9.3 The profiles and photographs displayed in this table are of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 2ft and wet when surveyed on 24th April 2019.



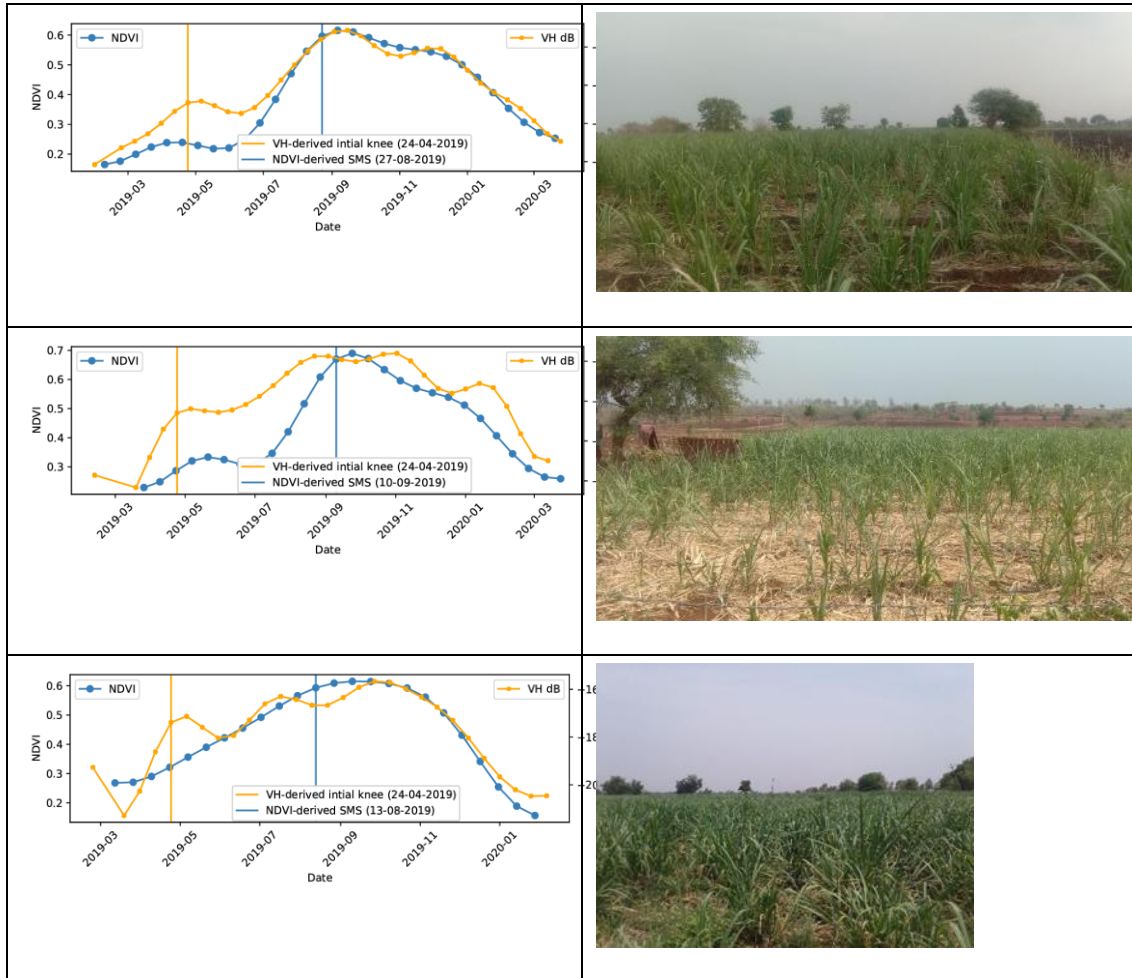


Table 9.4 The profile and photograph displayed in this table is a virgin sugarcane field, first planted in 2019, located in Unit 2. This field was first drip irrigated, and the conditions of the field was reported as being 2ft and dry when surveyed on 24th April 2019.

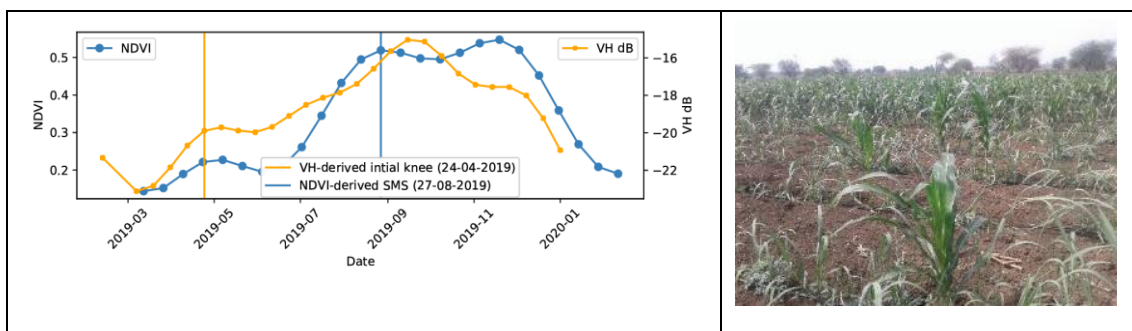
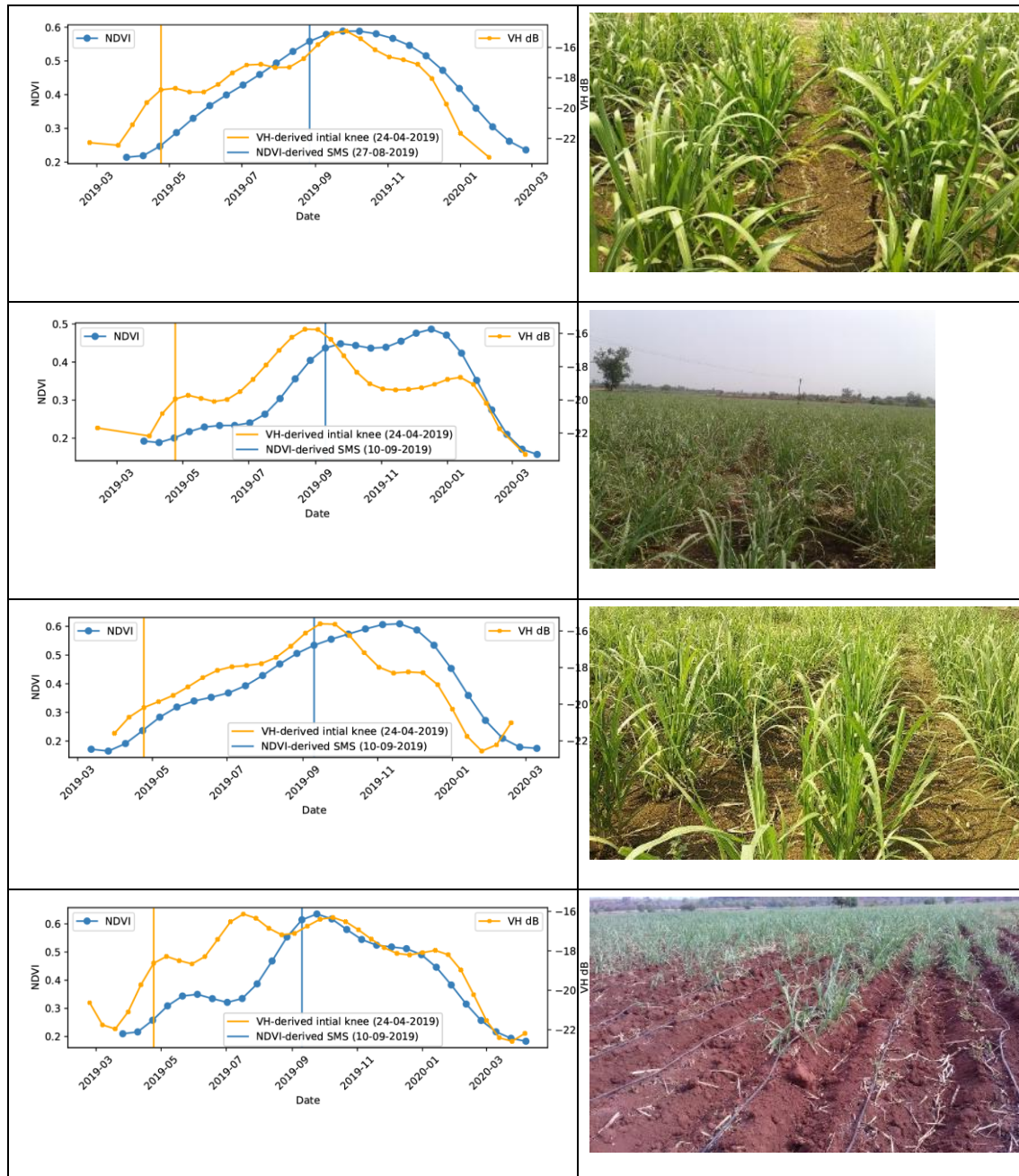
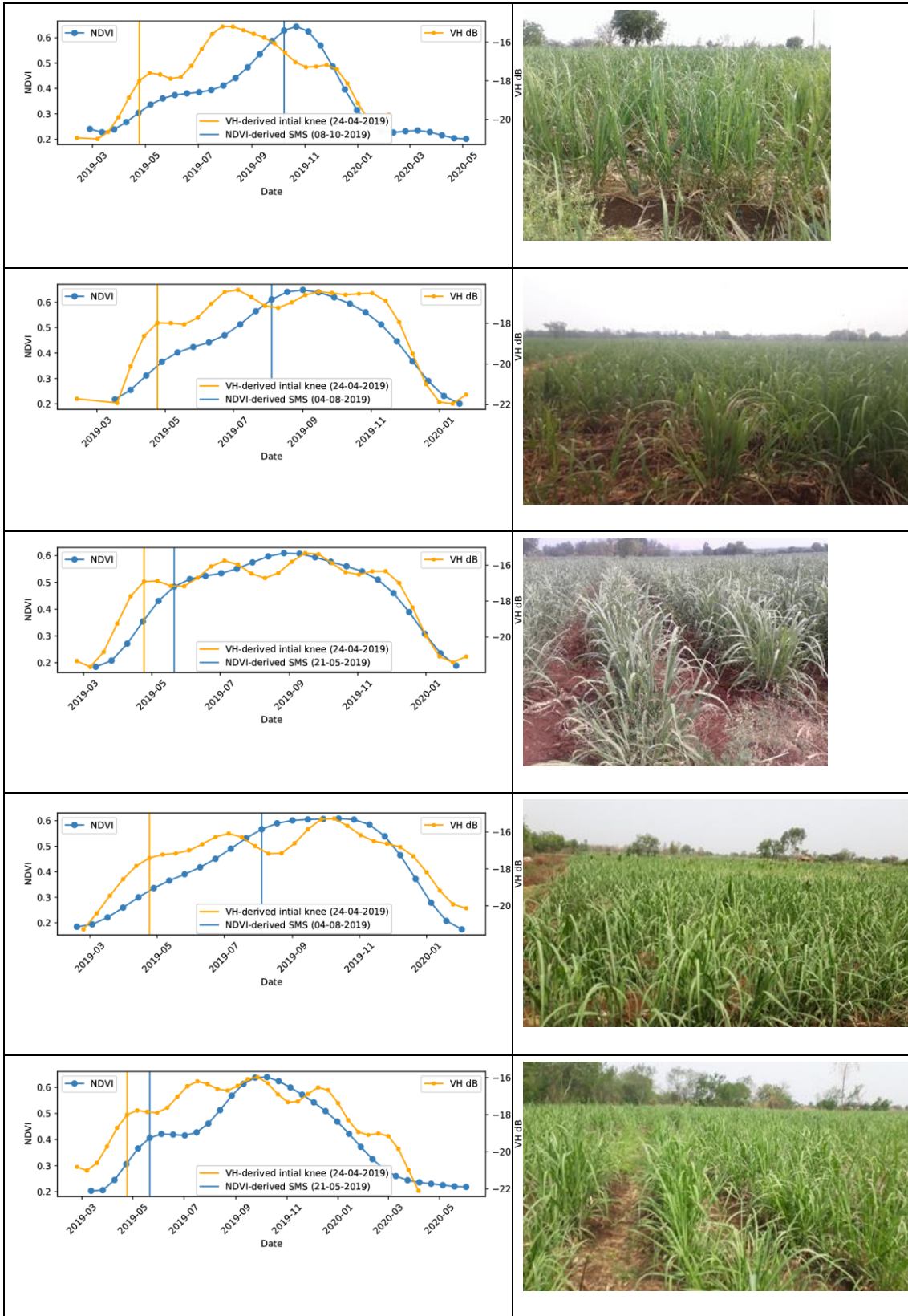


Table 9.5. The profiles and photographs displayed in this table are of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 3ft and wet when surveyed on 24th April 2019.





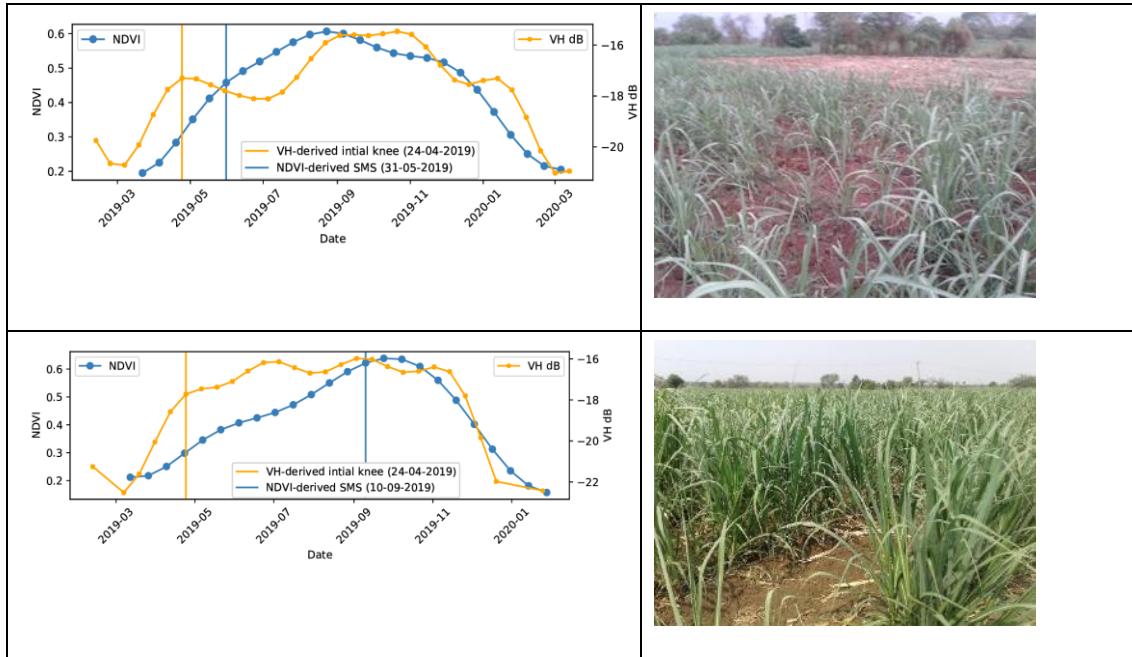


Table 9.6. The profile and photograph displayed in this table is a virgin sugarcane field, first planted in 2019, located in Unit 2. This field was first drip irrigated, and the conditions of the field was reported as being 3ft and dry when surveyed on 24th April 2019.

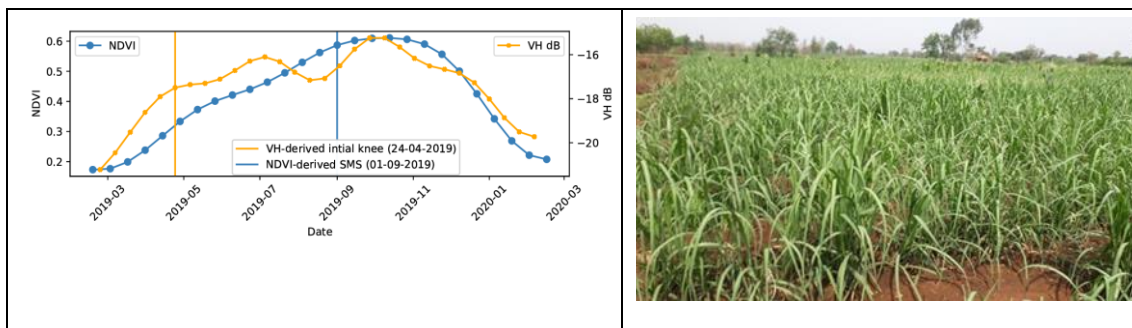


Table 9.7 The profiles and photographs displayed in this table are of virgin sugarcane fields, first planted in 2019, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 4ft and wet when surveyed on 24th April 2019.

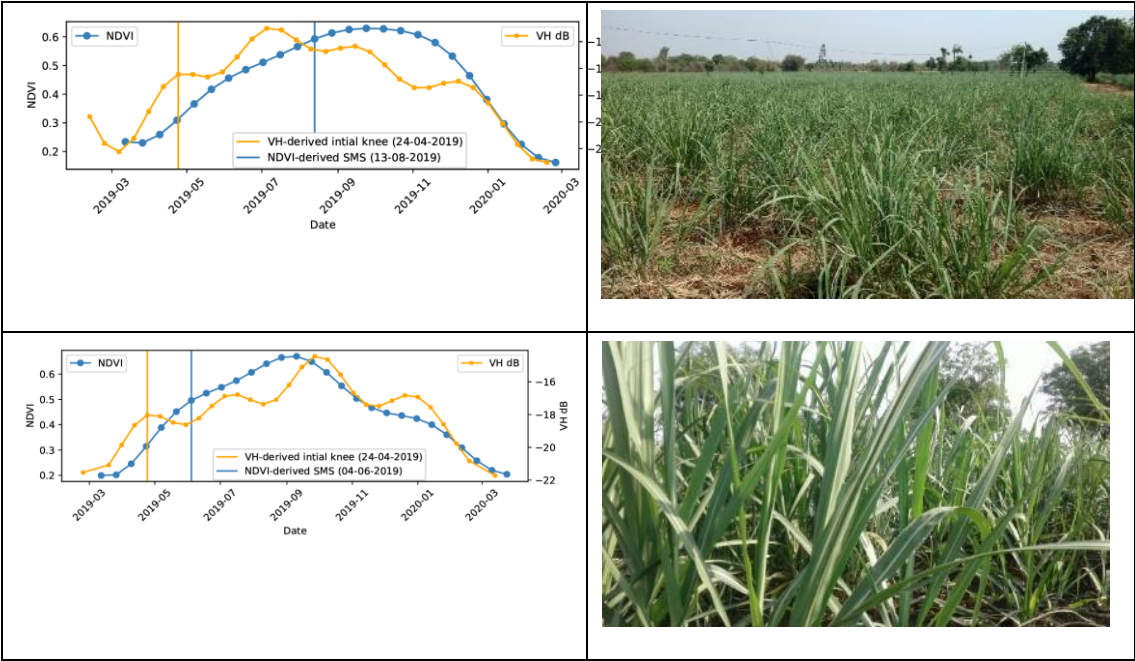
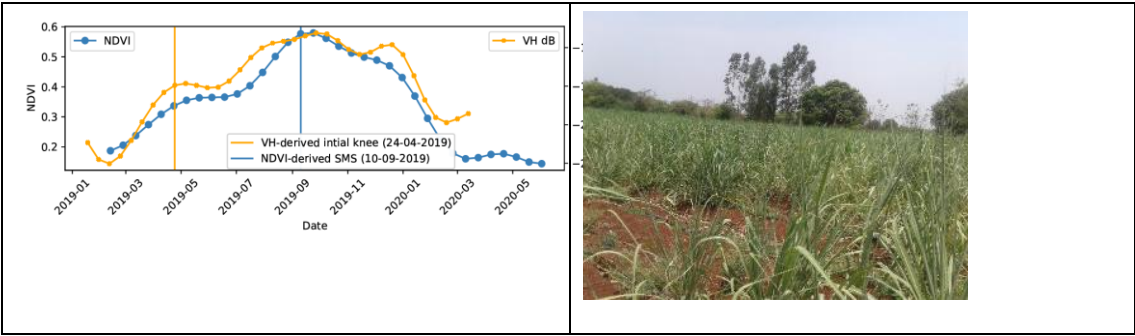


Table 9.8. The profiles and photographs displayed in this table are of 1st year ratoon sugarcane fields, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 2ft and wet when surveyed on 24th April 2019.



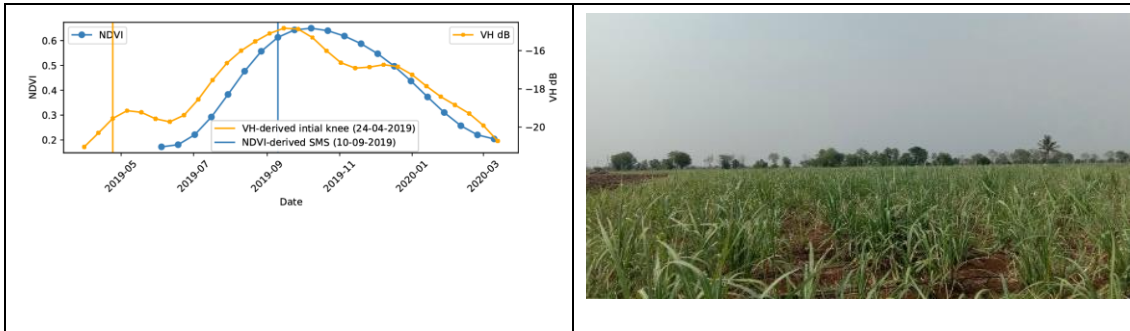


Table 9.9 The profile and photograph displayed in this table are of a 1st year ratoon sugarcane field, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 3ft and wet when surveyed on 24th April 2019.

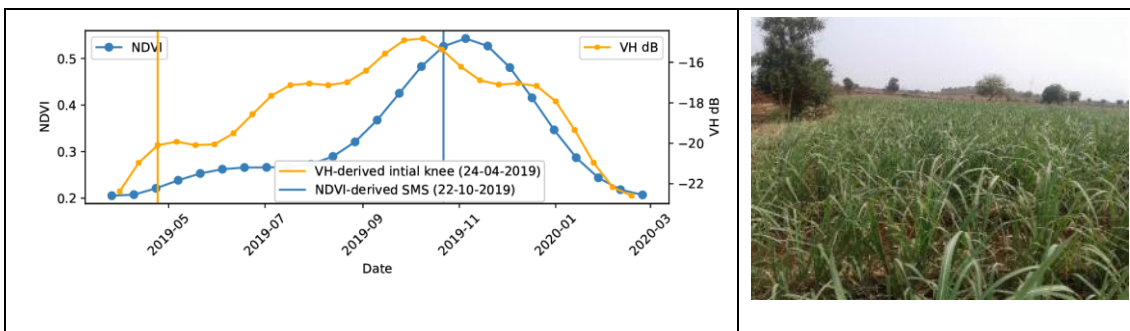
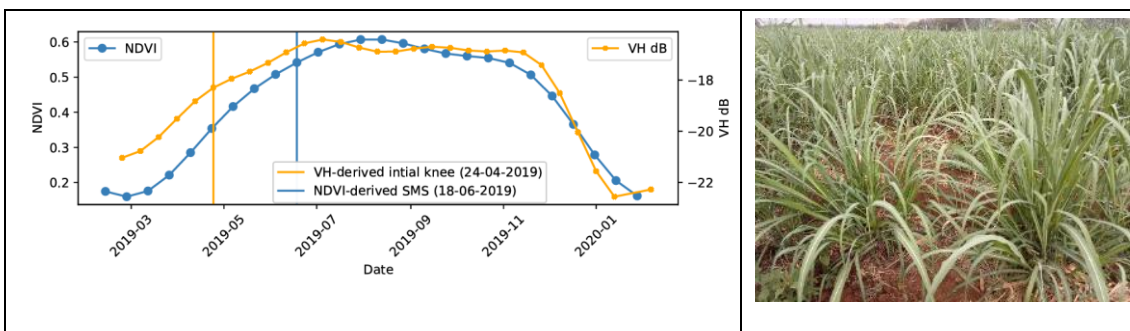


Table 9.10 The profile and photograph displayed in this table are of a 1st year ratoon sugarcane field, first planted in 2018, located in Unit 2. These fields were first drip irrigated, and the conditions of the fields were reported as being 5ft and wet when surveyed on 24th April 2019.



9.4 Data Sources

This research used a combination of remote sensing and ground-based data to monitor and classify sugarcane growth across the study area. The key data sources are outlined below:

9.4.1 Field Data

Ground reference points (GRP coordinates), field photographs, and survey data were collected through collaboration with the industry partner. These data were used for model validation and calibration but are subject to commercial restrictions and cannot be shared publicly.

9.4.2 Remote Sensing Data

Landsat 8 imagery was obtained via the Google Earth Engine (GEE) API, providing medium-resolution data for long-term monitoring. Sentinel-2 imagery (10–20 m resolution) was used for vegetation monitoring and detecting growth stages. Sentinel-1 radar data complemented optical data for temporal classification and analysis under cloudy conditions. All satellite imagery was accessed through the Google Earth Engine cloud-based platform.

9.4.3 Publicly Available Dataset

Chapter 3 of this thesis was published as a standalone study, and in compliance with the UKRI open access policy, the associated dataset specifically the derived sugarcane growth stage information is openly available via Cranfield Online Research Data at: <https://doi.org/10.57996/cran.ceres-2631>. All other data used in this research have been anonymised where necessary, and any commercially sensitive details have been removed. Anonymised data supporting other parts of the thesis are hosted in a private GitHub repository. Access can be granted upon reasonable request. Interested researchers may contact the author directly via email to request access.

9.5 Thesis code description

Custom Python code was developed to process satellite imagery, analyse time series data, and apply classification techniques relevant to the sugarcane production cycle. The full codebase is hosted in a private GitHub repository: <https://github.com/centa-neta-sugar/phd-thesis-code>. Access can be granted upon request. Interested researchers may contact the author at **n.joshi@cranfield.ac.uk** to request access.

The GitHub repository is organised by chapter as follows:

- **GEE_code**

This folder contains scripts demonstrating how the Google Earth Engine (GEE) Python API was used to download and process Landsat-8, Sentinel-1, and Sentinel-2 imagery used throughout the thesis.

- **phenology_automation**

This folder contains methods developed in Chapters 3 and 4. Python scripts were used to monitor key phenological stages in the sugarcane growth cycle. These notebooks illustrate how Sentinel-2 and Landsat-8 time-series data were pre-processed, and how key sugarcane growth stages were identified.

- **Timeseries_classification**

This folder includes the classification pipeline developed in Chapter 5, used to distinguish sugarcane fields from other land covers using multi-temporal imagery. The workflow involves time-series feature extraction, classification model development, and accuracy assessment.

The codebase makes use of several Python libraries, including:

- geemap and ee (Google Earth Engine Python API)
- pandas, numpy, and matplotlib for data manipulation and visualisation
- scikit-learn for machine learning and classification tasks