UCLouvain

AGRO LOUVAIN

Faculté des bioingénieurs

Towards a crop-specific drought impact assessment tool in Chiredzi District, Zimbabwe, using Sentinel-2 and Landsat-8/9 imagery

Auteur :	Marichal Théo
Promoteurs :	Pierre Defourny
	Sophie Bontemps
Lecteurs :	Charles Bielders
	Baptiste Delhez
Année Académique :	2023 - 2024
Bioingénieur :	Sciences et technologies de l'environnement

Acknowledgements

First, I would like to express my deep gratitude to my thesis supervisor, Pierre Defourny, whose advice was crucial in the completion of this master's thesis. His insights were of great importance in challenging me to improve the fundamental aspects of this thesis.

I am very thankful to my co-supervisor, Sophie Bontemps, whose meetings, wise advice, and support greatly contributed to the constant improvement of this thesis. I am also grateful for her support, her numerous pieces of feedback, and the time she dedicated to enlightening me in completing this master's thesis.

Special thanks to Boris Norgaard, for his crucial advice, his ability to help me overcome problems, and his optimism, which have been extremely encouraging throughout this process. His constructive feedback during numerous meetings have been of vital importance.

Finally, my deep appreciation goes to my family, whose encouragement and support have been essential in enabling me to complete this master's thesis. I am grateful for their belief in my abilities and their constant support.

Contents

Acknowledgements ii

Contents iii

List of figures vi

List of tables ix

Acronyms x

1 Introduction 1

2 Literature review 3

2.1 Definition and types of droughts . . . 3

2.2 Remote sensing data for drought monitoring . . . 4

2.2.1 Advent of Remote Sensing . . . 4

2.2.2 Overview of remote sensing techniques and resolutions . . . 4

2.2.3 Monitoring of climatic variables . . . 5

2.2.4 Monitoring of ecological variables : vegetation monitoring . . . 8

2.2.5 EWS for Food Security . . . 12

2.2.6 Global Information and Early Warning System on Food and Agriculture (GIEWS) . . . 12

2.3 Remote sensing data for crop type mapping . . . 13

2.3.1 The advent of Sentinel-2 for crop type classification . . . 13

2.3.2 Satellite images time series . . . 15

2.3.3 Metrics based on SITS of vegetation indices . . . 16

2.3.4 Crop classification in remote sensing . . . 17

3 Objectives 21

4 Materials and method 22

4.1 Study area . . . 23

4.2 High-spatial-resolution VHI . . . 24

4.2.1 High-spatial-resolution VHI methodology and comparative analysis with the FAO's global VHI . . . 25

4.2.2 GEE platform . . . 26

4.2.3 Data processing . . . 27
4.2.4 Cloud masking . . . 27
4.2.5 NDVI and LST time series, temporal filtering and smoothing . 28
4.2.6 Calculation of VCI and TCI . . . 30
4.2.7 Standardising the spatial resolution for both VCI and TCI indices 31
4.2.8 Calculation of the VHI . . . 34
4.2.9 Adjustment of VHI according to the agricultural calendar . . . 34
4.3 Crop Type Classification . . . 35
4.3.1 In situ data preparation . . . 36
4.3.2 Calibration data and validation data . . . 38
4.3.3 Sentinel-2 Data Acquisition . . . 39
4.3.4 Metrics and time series of spectral indices for feature extraction 39
4.3.5 Random Forest classification . . . 41
4.3.6 Majority Filtering . . . 42

5 Results and discussion **44**
5.1 VHI Results . . . 44
5.1.1 Evolution of the VHI at the district level . . . 44
5.1.2 Comparison between FAO's VHI at the district level . . . 45
5.1.3 High spatial resolution for a better monitoring of drought impact on agricultural crops . . . 47
5.2 Classification Results . . . 48
5.2.1 Confusion matrix and accuracy metrics . . . 49
5.2.2 Gini Feature Importance . . . 51
5.3 Integration of the VHI time series and the crop type map . . . 52
5.3.1 Assessing the VHI by crop type . . . 52
5.3.2 Assessing the health of sugarcane at the field scale on targeted parcels . . . 57
5.3.3 Assessing the health of sorghum at the field scale on randomly selected fields . . . 59
5.3.4 General discussion . . . 67

6 Conclusion **68**

List of Figures

1 Types of droughts [14] 4
2 Different types of resolutions in remote sensing [16] 5
3 Link between the depth of penetration and the wavelength[28] 8
4 Spectral signature of green vegetation[30] 9
5 Weekly weather and ecological-related NDVI profiles over multiple years for different types of vegetation, notably: 1) coniferous forest, annual precipitation above 80 inches (in.); 2) mixed coniferous forest and steppe, 10-20 in.; 3) steppe, 10-20 in.; 4) prairie, 20-40 in.; 5) mixed coniferous and broadleaf forest, 20-40 in.; 6) mixed coniferous and broadleaf forest in Maine, 40-60 in.; 7) mixed coniferous and broadleaf forest in California, 40-60 in.; 8) desert and desert shrub, below 10 in.; 9) coniferous forest, 20-40 in., elevation 4000-6000 m; 10) steppe, 10-20 in.; 11) broadleaf forest, 40-60 in.; 12) mixed coniferous and broadleaf forest in Florida, 40-60 in.[35] 10
6 Sentinel-2 spectral bands and their characteristics [46] 14
7 Annual NDVI for a crop pixel: AVHRR raw inputs and dekadal smoothed outputs[53] 15
8 Different methods of time series smoothing and gap-filling : processing window lengths and maximum tolerated gap lengths [54] 16
9 Relevant NDVI metrics for crop classification [51] 17
10 Advantages and Disadvantages of different classification Algorithms [55] 19
11 Location of the study area, Chiredzi District, Zimbabwe, and in situ crop sampled by the FAO 24
12 Comparison of FAO's global VHI versus high-resolution VHI 26
13 Comparison of a cloud-masked Sentinel-2 image processed with Sen2Cor and a true colour image captured on March 3, 2024 28
14 Annual (2022-2023) NDVI time series with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi 29
15 Annual (2023-2024) LST time series of only Landsat 8 data with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi . . . 29

16 Annual (2023-2024) LST time series with L8 and L9 integrated data with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi . . . 30
17 Comparison between global-resolution VCI (FAO) and high-spatial-resolution VCI in December 2023 . . . 31
18 Comparison between the resampled TCI from January 20 to February 5, 2024, at 10-meter spatial resolution with bilinear interpolation method (a), and the original TCI at 10-meter spatial resolution (b), alongside a representation of the study area with a true-colour image from Google Satellite (c) . . . 32
19 Comparison between the original VCI from January 20 to February 5, 2024, at 10-meter spatial resolution (a), and the resampled VCI at 30-meter spatial resolution, considering the PSF of Landsat 8 TIRS sensors (b), alongside a representation of the study area with a true-colour image from Google Satellite (c) . . . 33
20 NDVI profiles of the 2023-2024 agricultural season and historical NDVI profiles since 2018 of a crop located in Chiredzi . . . 34
21 NDVI profiles of the 2023-2024 agricultural season and historical NDVI profiles since 2018 after normalisation of the seasonal shift of a crop located in Chiredzi . . . 35
22 Geospatial polygons created based on a composite cloud-masked Sentinel-2 image from February 2024 and a Google Satellite image, using in situ samples from Chiredzi District, Zimbabwe . . . 37
23 Number of polygons and number of 10-meter by 10-meter pixels for the different classes selected after data preparation . . . 38
24 Mean NDVI time series of different land cover classes . . . 40
25 NDVI time series of different crop type classes . . . 41
26 Majority filtering application . . . 43
27 Evolution of vegetation health throughout the 2023-2024 agricultural season in Chiredzi District, Zimbabwe . . . 45
28 Comparison between the two VHIs (FAO's VHI and the high spatial resolution VHI) in January 2024, corresponding to the agricultural crop growth period, and in March 2024, which is close to the harvest period, in the Chiredzi District . . . 46
29 Comparison between the FAO's VHI and the high-resolution VHI developed in this master's thesis: (a) the FAO's VHI at a spatial resolution of 1 km from 01/01/2024 to 10/01/2024, (b) the VHI at a spatial resolution of 10 meters from 04/01/2024 to 20/01/2024, (c) a true-colour image of the area distinguishing natural vegetation from agricultural zones . . . 48

30 Crop type map of the study area at the district level for the 2023-2024 agricultural season: displayed in true color (A) and with a higher-resolution (A'), as well as with crop classification (B) and its higher resolution counterpart (B') 49
31 Classification confusion matrix 50
32 Classification confusion matrix 50
33 Hierarchical classification confusion matrix 51
34 Gini index 52
35 VHI time series of in situ points of Chiredzi 53
36 VHI time series of in situ polygons of Chiredzi 53
37 VHI time series of sugarcane classified fields in Chiredzi 54
38 VHI time series of in situ points of Chiredzi 54
39 VHI time series of in situ polygons of Chiredzi 55
40 VHI time series of in situ reclassified sorghum fields in Chiredzi . . . 55
41 Evolution of vegetation health of sesame fields across the 2023-2024 agricultural season 56
42 Evolution of vegetation health of millets fields across the 2023-2024 agricultural season 56
43 Evolution of vegetation health of maize fields across the 2023-2024 agricultural season 56
44 Randomly selected sugarcane fields 57
45 Evolution of the VHI of intensive sugarcane fields from October 1, 2023, to April 29, 2024. 58
46 Comparative study of the 2023-2024 season NDVI time series with the time series from previous seasons for randomly selected sugarcane fields. 59
47 Monthly precipitation time series for the 2023-2024 agricultural season, based on CHIRPS data, covering a 5 km^2 area that includes randomly selected sugarcane fields (figure 45) 59
48 Randomly selected extensive sorghum fields in Chiredzi 60
49 Evolution of VHI of randomly selected sorghum fields throughout the 2023-2024 agricultural season 61
50 Analysis and explanation of the evolution of VHI values (distributed over 11 periods of 16 days) for a sorghum field during the 2023-2024 agricultural season, using a comparative study of the 2023-2024 season NDVI and LST time series with the time series from previous seasons 64
51 Analysis and explanation of the evolution of VHI values for a sorghum field during the 2023-2024 agricultural season, using a comparative study of the 2023-2024 season NDVI and LST time series with the time series from previous seasons. 66

List of Tables

1 Distribution of samples by land category in the study Area, based on FAO's windshield and statistical survey 36

Acronyms

SDG: Sustainable Development Goals
EWS: Early Warning Systems
FAO: Food and Agriculture Organization
SPI: Standard Precipitation Index
MoA: Ministry of Agriculture
ZIMSTAT: National Statistics Office in Zimbabwe
EO: Earth Observation
GEO: Geostationary Earth orbit
LEO: Low Earth orbit
EUMETSAT: European Organisation for the Exploitation of Meteorological Satellites
NASA: National Aeronautics and Space Administration
TRMM: Tropical Rainfall Measuring Mission
GPM: Global Precipitation Mission
CHIRPS: Climate Hazards Group Infrared Precipitation Satellite
PNP: Percentage of Normal Precipitation
ET: Evapotranspiration
MODIS: Moderate Resolution Imaging Spectroradiometer
PSF: Point Spread Function
VIIRS: Visible Infrared Imaging Radiometer Suite
SMAP: Soil Moisture Active and Passive
NIR: Near-infrared
SWIR: Short-wave infrared
VIS: Visible
IR: Infrared
NDVI: Normalized Difference Vegetation Index
VCI: Vegetation Condition Index
NDWI: Normalized Difference Water Index
AVHRR: Advanced Very High Resolution Radiometer
ETM+: Enhanced Thematic Mapper Plus
TIRS: Thermal Infrared Sensor
TCI: Temperature Condition Index
GEOGLAM: Group on Earth Observations for Global Agricultural Monitoring
CM4EW: Crop Monitor for Early Warning
LAI: Leaf Area Index
SITS: Satellite Images Time Series
ESA: European Space Agency

TOA: Top Of Atmosphere
ST: Surface Temperature
BT: Brightness Temperature
Sen2Cor: Sentinel 2 (atmospheric) Correction
ROI: Region of Interest
QGIS: Quantum Geographic Information System
GSD: Ground Sample Distance
MSI: Multispectral Instrument
FWHM: Full Width at Half Maximum
SOS: Start of Season
NOAA: National Oceanic and Atmospheric Administration
GIEWS: Global Information and Early Warning System on Food and Agriculture
LST: Land Surface Temperature
VHI: Vegetation Health Index
ML: Maximum Likelihood
DTs: Decision Trees
RF: Random Forest
CNNs: Convolutional Neural Networks
GEE: Google Earth Engine

Chapter 1

Introduction

Drought, due to its complexity and diversity, ranks among the most devastating natural phenomena worldwide and remains difficult to characterise [1]. It depends on climatic factors such as precipitation, moisture deficit, and temperature, as well as the anthropogenic influence on hydrological processes and the physical environment [2]. There is no global consensus on the definition of drought; however, drought manifests as a water deficiency relative to the usual conditions. Unlike floods, which are characterized by sudden and distinct beginnings and endings, determining the onset, end, and spatial extent of droughts is challenging [3].

Droughts development is slow, they can persist for years, affect large areas, and have significant impacts on human well-being, the environment, and the economy [3]. Over time, the occurrences of droughts will only increase due to climate change [4]. Indeed, climate change has intensified the frequency and severity of drought periods, significantly impacting Southern Africa [5]. El Niño, a climate pattern caused by changes in Pacific Ocean winds, currents, and temperatures, has been occurring more and more frequently than average, affecting the region's food security [6].Due to the vastness of the Pacific Basin, which covers about one third of the Earth's surface, the climate changes that occur there, such as variations in winds and humidity, have global repercussions. This phenomenon results in droughts, which are associated with increasing poverty, insecurity and a mortality rate particularly high in Sub-Saharan Africa [7]. Consequently, it causes a divergence from the Sustainable Development Goals (SDGs), especially SDG1 (no poverty), SDG2 (zero hunger), and SDG3 (good health and well-being) [8]. This situation is particularly alarming for countries like Zimbabwe lacking the resources to address drought damage and the tools for monitoring and anticipating such impacts [9]. Additionally, Zimbabwe still relies on over 60 percent of rain-fed subsistence agriculture [8]. These farmers are directly dependent on climatic conditions. Zimbabwe experienced an extended drought from late January to mid-March 2024, during a crucial period of crop development. This situation threatens the food security of millions of people [10].

It is therefore essential to ensure effective monitoring and mitigation of drought. To

achieve this, various global EWS dedicated to food security have been established since the early 1970s [11]. When combined with effective communication and international collaboration, these systems enhance the ability to assess, mitigate, and even prevent potential threats [12].

These tools utilise remote sensing technology to gather information about objects or areas, such as the Earth, without direct contact. This is achieved using sensors mounted on drones, planes, or satellites, allowing for data collection at various scales [13]. EWS use remote sensing-based indicators to monitor climatic factors and understand the impacts of droughts. However, many indicators rely on low spatial resolution data, which is not suitable for field-scale analysis in agriculture. Low-resolution data makes it difficult to isolate the spectral responses of fields from those of surrounding land covers, resulting in noisy data when assessing drought impacts on agriculture. This study proposes to test the use of higher spatial resolution data in drought monitoring tools in order to assess drought impacts at field-level.

In 2024, the Food and Agriculture Organization (FAO), an United Nations agency, has set up a project with the Geomatics lab of Pr.Defourny, which aims at building the capacity of the Ministry of Agriculture (MoA) and of the National Statistics Office in Zimbabwe (ZIMSTAT) to produce better quality, more complete, disaggregated and timely crop data, as well as improved use of earth observation (EO) data to improve agricultural statistics and provide inputs for measuring some SDGs indicators related to agriculture. One activity of this project is to develop a prototype for an EWS for drought and to integrate it with crop maps for impact assessment. This work is carried out in this context.

Chapter 2

Literature review

This chapter explores the different types of droughts and their effects on agriculture. To assess these impacts, two key elements are necessary: a drought indicator and a crop map, which will be combined to evaluate drought effects on agricultural areas. Initially, the chapter explains the role of remote sensing in creating drought monitoring indicators. Then, it reviews current drought assessment systems highlighting their strengths and limitations. Finally, it highlights the importance of distinguishing not only between cultivated and non-cultivated areas but also between different types of crops through crop mapping, in order to achieve for an efficient drought impact assessment.

2.1 Definition and types of droughts

The progression of drought describes how a lack of precipitation (i.e., below-normal levels) advances through the hydrological cycle. It begins as a meteorological drought and can evolve into a hydrological drought if the situation persists [2]. Generally, three types of drought are distinguished (figure 1). First, there are meteorological droughts, characterised by a lack of precipitation and/or unusually high evaporation and transpiration compared to the historical average for a specific period. They can last from 6 weeks to 6 months and show significant variability in both time and space [14]. Meteorological droughts often lead to agricultural droughts, which occur when soil moisture decreases, causing water stress for plants and resulting in reduced productivity. Finally, if the drought persists, it leads to hydrological drought, characterised by reduced river flows, and lower water levels in reservoirs, groundwater, and lakes. This type of drought affects water resources on a larger scale and over a longer period.These droughts have significant economic, social, and ecological impacts. The focus of this study is on the characterisation of the first two types of drought.

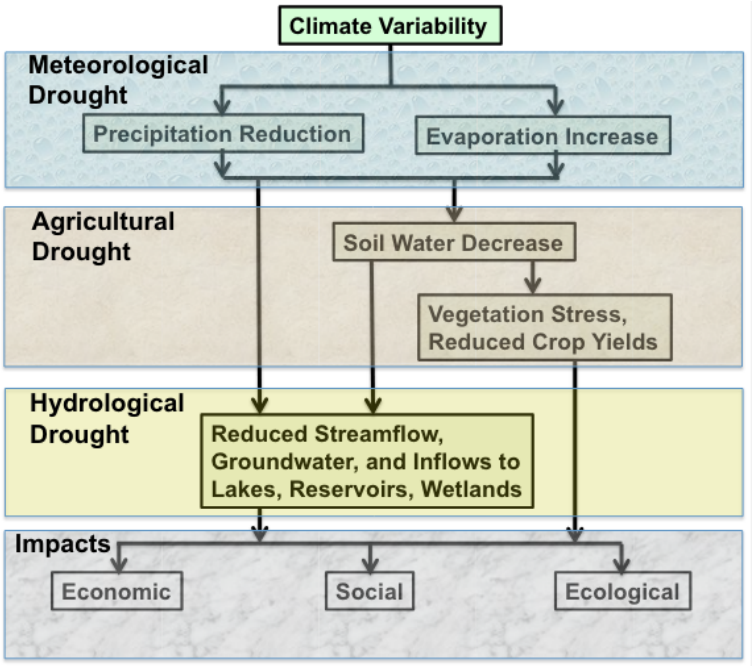


Figure 1: Types of droughts [14]

2.2 Remote sensing data for drought monitoring

2.2.1 Advent of Remote Sensing

Before the advent of remote sensing, drought monitoring relied on on-site observations. Instruments such as rain gauges were used to measure precipitation, and probes were used to assess soil moisture. Although these techniques provided accurate data for specific locations, their spatial coverage was limited. It was challenging to extrapolate these measurements to large areas, and a large number of measurement points were required. Over the past five decades, there has been a shift from traditional on-site measurements to remote techniques that allow for drought monitoring by collecting data on distant objects or areas. This change is particularly relevant in Zimbabwe, where weather stations are poorly distributed and do not provide regular data, both spatially and temporally [15].

2.2.2 Overview of remote sensing techniques and resolutions

Remote sensing technology uses sensors to capture energy emitted or reflected by surfaces, based on their physical or chemical characteristics, across different bands of the electromagnetic spectrum. This helps identifying the spectral signature of the observed surface. Depending on the spectrum range studied, remote sensing can be classified into optical, thermal, and microwave remote sensing.

Optical remote sensing covers the visible (450-700nm) and infrared (750-3000nm) spectra, from near-infrared (NIR) to short-wave infrared (SWIR). Thermal remote

sensing focuses on thermal infrared (3000-15000nm). Microwave remote sensing uses the microwave spectrum (1mm-1m).

The collected data are characterised by various types of resolution (figure 2). Choosing a specific resolution often requires trade-offs. It is essential to find an optimal balance suited to the user's objectives.

Spatial resolution refers to the size of the ground object represented by a single pixel (figure 2).Higher spatial resolution makes it easier to identify small objects but increases storage space requirements and often comes with a lower revisit frequency, leading to a lower temporal resolution. Temporal resolution is the frequency at which images of a specific Earth location are updated, also known as the satellite revisit rate. For example, monitoring different crop growth phases requires a specific revisit interval to effectively observe and assess agricultural conditions over time. Radiometric resolution is a satellite sensor's ability to detect and differentiate different levels of radiation from the Earth's surface. The sensor breaks down these radiation levels into discrete values using bits. The number of bits used affects how well the sensor can capture the intensity and shades of colour. Spectral resolution is the sensor's ability to distinguish small intervals within the electromagnetic spectrum.

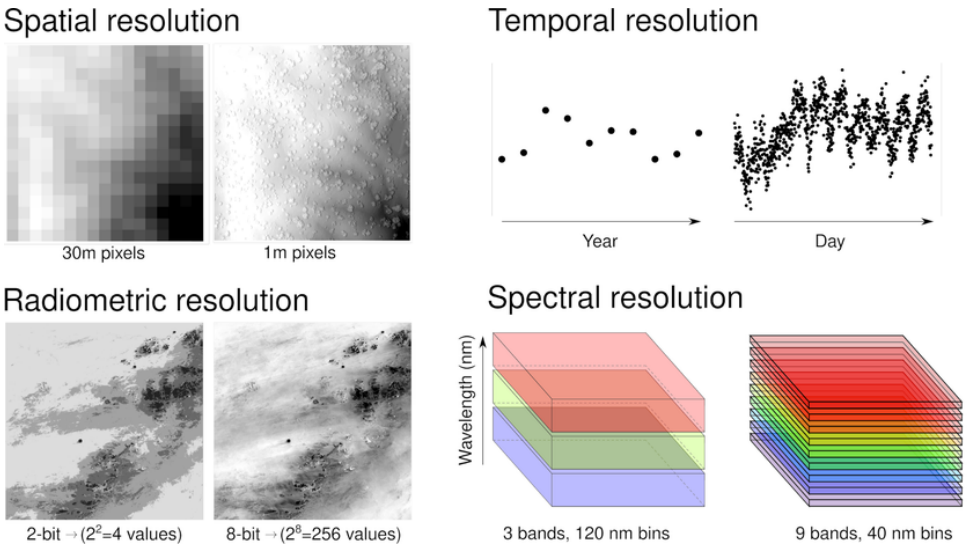


Figure 2: Different types of resolutions in remote sensing [16]

2.2.3 Monitoring of climatic variables

Remote sensing data now cover nearly all aspects of drought progression, significantly enhancing our understanding of this phenomenon [11]. However, data for each climatic and environmental component are collected at varying spatial and temporal resolutions.

Precipitation monitoring

Monitoring precipitation is crucial for identifying meteorological drought episodes. Precipitation can be monitored over a larger area with coarser resolution or over a smaller area with finer resolution. Understanding the differences between geosynchronous Earth orbit (GEO) satellites and Low Earth Orbit (LEO) satellites is essential for this process.

GEO satellites are positioned at a high altitude of approximately 36,000 km. They are located above the Earth's equator and rotate in alignment with the Earth's rotation, providing continuous coverage of a specific area. In contrast, LEO satellites operate at altitudes ranging from 100 to 1,600 km. Unlike GEO satellites, LEO satellites are not constrained to a fixed trajectory, and their orbital plane can be inclined, allowing them to follow various paths around the Earth. Their proximity to the Earth's surface enables them to capture images with higher spatial resolution.

Optical sensors, operating in the visible (VIS) and infrared (IR) spectrum, are deployed on both GEO and LEO satellites. For example, the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) has launched Meteosat satellites 8-11 (GEO), providing continuous coverage over Africa and Europe. Notably, Meteosat 11 delivers data with a temporal resolution as frequent as every 5 minutes[17]. These IR/VIS sensors typically provide more frequent data but with higher uncertainties, and they offer estimates of rainfall volume [18].

Conversely, microwave sensors are only deployed on LEO satellites. These sensors do not rely on sunlight, allowing for data collection under any weather conditions, including cloudy weather or at night. For instance, National Aeronautics and Space Administration's (NASA) Tropical Rainfall Measuring Mission (TRMM) satellite (1997-2015) and its successor, the Core Observatory satellite of the NASA Global Precipitation Mission (GPM) since 2014, are equipped with microwave sensors. The successor satellite features a spatial resolution of 5 km and a temporal resolution of 3 hours [19]. Satellites operating in the microwave spectrum, such as the Climate Hazards Group Infrared Precipitation Satellite (CHIRPS), utilise an integrated approach that combines satellite data with ground observations to enhance data accuracy. This methodology is particularly valuable because satellite imagery often provides only regional averages that may not fully capture the intensity of extreme precipitation events. In contrast, precipitation data derived from weather station records are less accurate in rural areas due to the sparse distribution of stations. This integrated approach of satellite data with ground-based observations achieves thus more precise measurements of such events [20]. CHIRPS features a spatial resolution of 5 km and a daily temporal resolution.

The most used precipitation index for drought monitoring is the standard precipitation

index (SPI) [21]. The Standardised Precipitation Index (SPI) is a method used to measure precipitation anomalies over a given period, such as a month or several months. The SPI analyses the frequency of precipitation over different time periods. The precipitation data is then converted into a standard normal distribution for easy comparison. The SPI indicates whether a period is dry (negative SPI) or wet (positive SPI). In addition to the SPI, percentiles and the Percentage of Normal Precipitation (PNP) are also used to assess whether a period is wetter or drier than the average recorded precipitation [22].

Evapotranspiration monitoring

Evapotranspiration (ET)refers to the process where water transitions from liquid to vapour, occurring both through surface evaporation and plant transpiration into the atmosphere. This phenomenon plays a critical role in climate regulation, water resource management, and ecosystem functions [23]. Since ET fluxes cannot be directly observed, empirical models have been developed to estimate them. Some examples of satellites estimating ET are the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments on NASA's Terra and Aqua satellites. Terra, launched in 1999, orbits from north to south across the equator, while Aqua, launched in 2002, orbits from south to north. MODIS is equipped with two 250-meter bands, five 500-meter bands, and twenty-nine 1-kilometer bands. These instruments capture images of the entire Earth every 1 to 2 days, providing regular, comprehensive views of the planet [24]. Meanwhile, the Visible Infrared Imaging Radiometer Suite (VIIRS) on the national Oceanic and Atmospheric Administration's (NOAA) Suomi National Polar-orbiting Partnership satellite offers a spatial resolution of 0.5 km and a temporal resolution of one day [25]. There still remains a need for finer-scale data to enhance the accuracy of drought monitoring through ET and to refine the development of more sophisticated models [26].

Soil moisture monitoring

Soil moisture significantly influences the distribution of rainwater by either facilitating its infiltration into the ground or causing runoff. This process is crucial for the exchanges of water and energy between the atmosphere, the surface, and the underlying soils [27]. Soil moisture content is closely linked to its dielectric permittivity, which is the soil's response to an applied electric field. By measuring this permittivity, one can accurately determine the moisture levels in the soil. Microwaves, which are highly sensitive to changes in dielectric permittivity, are used to measure this characteristic at specific frequencies of the electromagnetic spectrum. Since water has a much higher dielectric permittivity than soil, an increase in soil moisture leads to enhanced wave propagation. Furthermore, the relationship between wave penetration depth and wavelength allows for the measurement of moisture at different soil depths, from the surface down to deeper layers (figure 3).

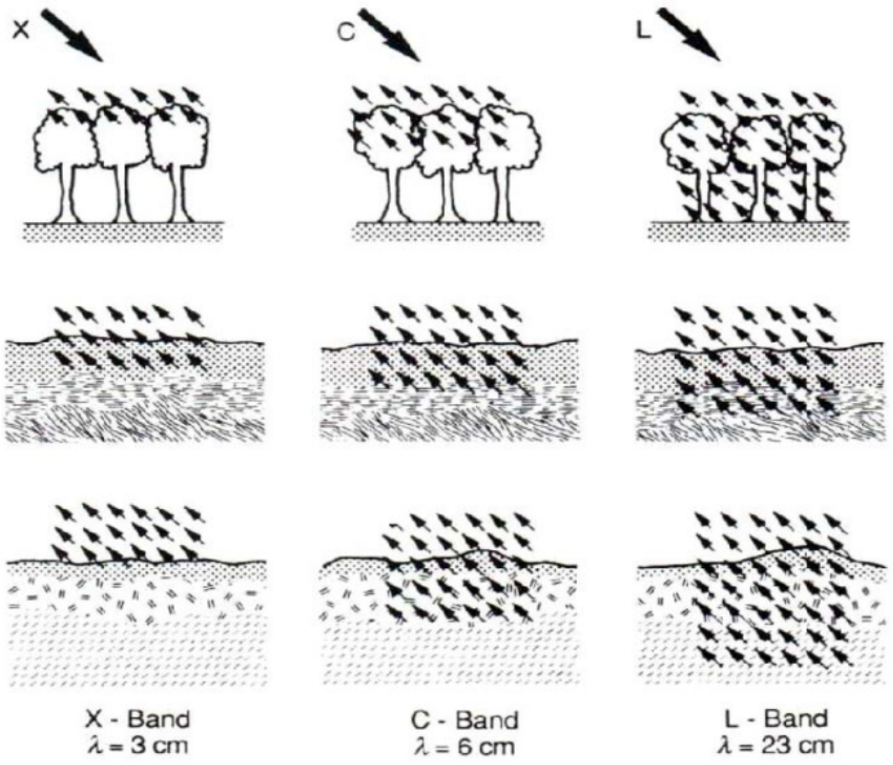


Figure 3: Link between the depth of penetration and the wavelength[28]

To capture these measurements, sensors can be either passive, receiving naturally emitted microwave energy from the Earth's surface like radiometers, or active, emitting microwaves to illuminate a target, such as radars.

Various satellites equipped with these sensors operate across different bands of the electromagnetic spectrum in both active and passive modes. For instance, the Soil Moisture Active and Passive (SMAP) satellite, operational since 2015, employs an L-band radiometer with a spatial resolution of 30 km and had also a synthetic aperture radar component which failed in 2015 [29]. Additionally, there is, for instance, the Sentinel-1 satellite which is outfitted with high-resolution radars that deliver soil moisture data with a resolution of 10 meters and a temporal frequency of every six days.

2.2.4 Monitoring of ecological variables : vegetation monitoring

Vegetation is closely connected to various aspects of the water cycle. When precipitation decreases, drying out the soil and root zone, the effects on vegetation quickly become apparent.

Monitoring vegetation health is mainly done using optical remote sensing, which includes both visible and infrared light. Plants are identified by a significant difference in how they reflect visible and near-infrared wavelengths. This characteristic is typ-

ical of green vegetation, whereas other surfaces, such as soil and water, reflect these wavelengths differently (figure 4).

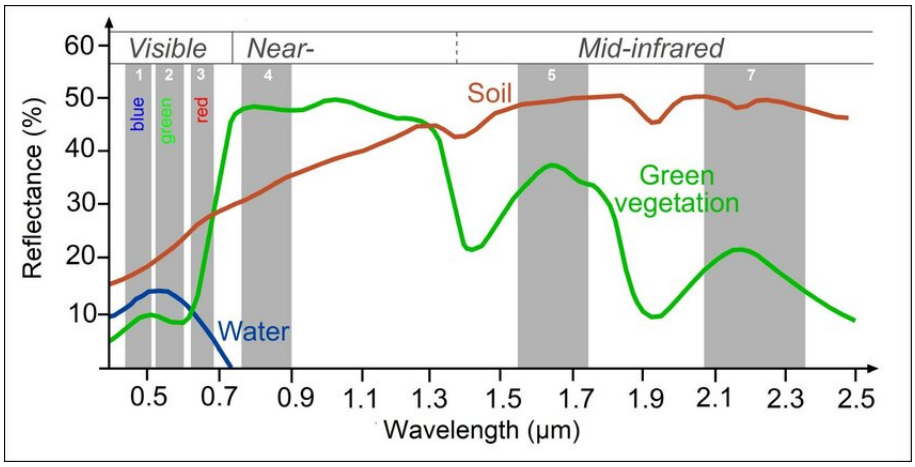


Figure 4: Spectral signature of green vegetation[30]

The Normalised Difference Vegetation Index (NDVI) is the most widely used vegetation index and serves as the primary tool for drought monitoring. This index assesses vegetation health by examining how chlorophyll absorbs red light and how the foliar mesophyll reflects infrared light [31]. The NDVI values range from -1 to 1, with low values indicating vegetation under moisture stress and higher values reflecting a greater density of healthy, green vegetation [32]. This indicator is particularly relevant for monitoring agricultural drought, as droughts are the main cause of moisture stress, impacting thus the health and growth of vegetation. [33] The index is calculated as follows:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \tag{2.1}$$

The Normalised Difference Water Index (NDWI) is another common used drought index utilised to emphasise the presence and extent of water in observed areas [34]. It is calculated as follows using the near infrared (NIR) and the short-wave infra-red (SWIR) bands and range from -1 to 1.

$$NDVI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \tag{2.2}$$

The Vegetation Condition Index (VCI), which scales NDVI values between their historical minimum and maximum, is also widely used for drought monitoring. The VCI is particularly useful for assessing the impact of droughts on vegetation by isolating weather-related effects from ecological factors. Ecological factors include the inherent and stable characteristics of the ecosystem where the vegetation develops.

This encompasses soil quality, nutrient availability, vegetation type, and biological interactions such as competition between plants or symbiosis. For example, a type of forest may have a specific period of leaf growth due to its genetic adaptations and interactions with other organisms in its ecosystem. These factors are often constant over long periods, thus influencing the life cycle of the vegetation in a predictable and measurable way year after year. Additionally, the NDVI curve is affected by meteorological factors, such as changes in weather conditions including rainfall or droughts. This influence is evident in how vegetation responds to environmental changes, and the VCI enables one to focus solely on those weather-related factors. Figure 5, from a US study [35], shows how ecological and meteorological factors impact the NDVI profiles of different vegetation types.

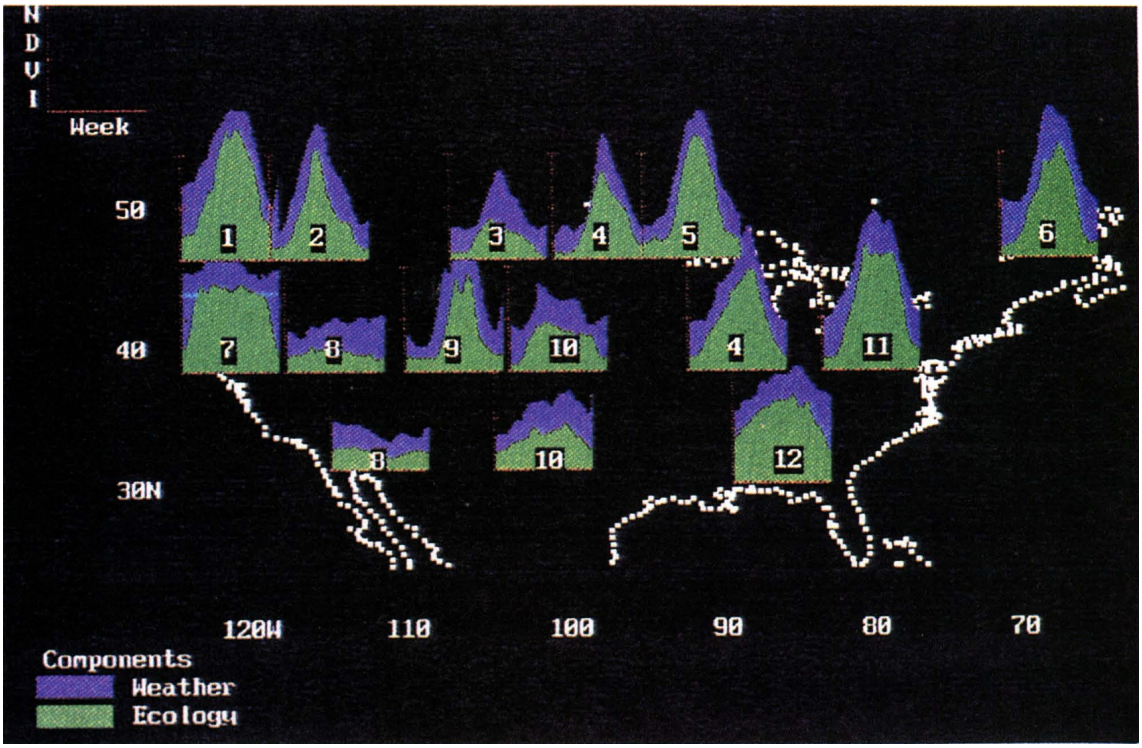


Figure 5: Weekly weather and ecological-related NDVI profiles over multiple years for different types of vegetation, notably: 1) coniferous forest, annual precipitation above 80 inches (in.); 2) mixed coniferous forest and steppe, 10-20 in.; 3) steppe, 10-20 in.; 4) prairie, 20-40 in.; 5) mixed coniferous and broadleaf forest, 20-40 in.; 6) mixed coniferous and broadleaf forest in Maine, 40-60 in.; 7) mixed coniferous and broadleaf forest in California, 40-60 in.; 8) desert and desert shrub, below 10 in.; 9) coniferous forest, 20-40 in., elevation 4000-6000 m; 10) steppe, 10-20 in.; 11) broadleaf forest, 40-60 in.; 12) mixed coniferous and broadleaf forest in Florida, 40-60 in.[35]

The VCI ranges from 0 to 1. A low VCI, resulting from a lower NDVI compared to previous years, suggests less dense or less healthy vegetation, indicating unfavourable conditions which could signify drought. In this situation, VCI values approach 0. Conversely, a high NDVI reflects more favourable conditions, and in such cases, the VCI tends toward 1. The VCI is calculated as follows:

$$VCI = \frac{NDVI_i - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \tag{2.3}$$

- $NDVI_i$ is the smoothed NDVI over a specified time period,
- $NDVI_{\min}$ is the historical minimum NDVI over the specified time period,
- $NDVI_{\max}$ is the historical maximum NDVI over the specified time period.

Kogan demonstrated that the Temperature Condition Index (TCI), which scales temperature values between their minimum and maximum, enhances drought monitoring accuracy when combined with the VCI [36]. This index assesses the influence of temperature on drought development. Initially developed using top-of-atmosphere brightness temperatures (BT), the use of land surface temperature (LST) has since helped to minimise atmospheric effects and reduce the influence of emissivity variability on the TCI [37]. LST is also crucial for understanding thermal energy flux exchanges between the soil and the atmosphere, especially under drought conditions. It is influenced by soil moisture, vegetation cover, and the thermal properties of surface materials [38]. Higher LST mid-season suggests unfavourable conditions or drought, causing TCI values to approach 0. Conversely, lower LST indicates more favourable conditions, which are reflected by TCI values tending toward 1 [35]. The TCI is calculated as follows:

$$TCI = \frac{LST_{\max} - LST_i}{LST_{\max} - LST_{\min}} \tag{2.4}$$

- LST_i is the smoothed LST over a specified time period,
- $LST_{\min}$ is the historical minimum LST over the specified time period,
- $LST_{\max}$ is the historical maximum LST over the specified time period.

The combination of these two indices thus enables the monitoring of drought across different windows of the electromagnetic spectrum: VCI, based on the visible and near-infrared parts, is a direct measure of vegetation moisture status, while TCI, based on the thermal infrared window, is a direct measure of stress driven by temperature conditions [39].

The complementary nature of these two indices has led to the development of a single index, the Vegetation Health Index (VHI), which combines VCI and TCI [40]. It is calculated as follows:

$$\text{VHI} = \alpha\text{VCI} + (1 - \alpha)\text{TCI} \tag{2.5}$$

The contributions of VCI and TCI to vegetation health vary depending on the type of vegetation and climate at a specific location and time, with an effective approximation for α equal to 0.5 [35].

Sensors designed to study the VHI have been available for a long time, particularly since the launch of the first Advanced Very High Resolution Radiometer (AVHRR) in 1979. This technology allowed for vegetation monitoring with a coarse resolution compared to modern sensors, but it offered the advantage of high temporal frequency. [41]. Instruments such as MODIS on NASA's Terra and Aqua satellites, along with the VIIRS on NOAA's Suomi NPP satellite, provide moderate-resolution imaging data with a temporal resolution as frequent as daily [25].

However, significant advancements have now been made, allowing for higher spatial resolution monitoring. The Landsat 8 and 9 series offer a spatial resolution of 30 meters for vegetation monitoring in the optical spectrum and 100 meters in the infrared spectrum, although with a limited temporal resolution of 16 days. Additionally, there have been significant advancements in high-resolution optical data with the Sentinel-2 series, comprising Sentinel-2A (2015) and Sentinel-2B (2017) [42]. These satellites provide 10-meter resolution for vegetation monitoring while maintaining a high temporal resolution of maximum 5 days. This development has opened numerous opportunities for improved agricultural monitoring at a local scale. Both of those missions will be further explained below.

2.2.5 EWS for Food Security

There are several EWS that use an integrated approach to monitor various climatic and environmental variables. This enhances decision-making in agricultural policies and improves agricultural production. However, these EWS may differ slightly due to stakeholders prioritising certain data inputs or methodologies over others. To address this, an initiative called the Group on Earth Observations (EO) for Global Agricultural Monitoring (GEOGLAM) was established, based on existing EWS. GEOGLAM launched the Crop Monitor for Early Warning (CM4EW), a consensus tool designed to reconcile the differences among the various EWS.[43]. However, GEOGLAM emphasises that further progress is needed within these EWS. This includes the necessity for more in situ data, improved calibration and validation of data and the use of more operational high-resolution satellite imagery.[44] .

2.2.6 Global Information and Early Warning System on Food and Agriculture (GIEWS)

The GIEWS is integrated into the CM4EW but has not yet developed a method to assess the impact of droughts using high-resolution optical imagery. Currently, GIEWS utilises optical and thermal bands from the AVHRR sensor on the Metop satellite for vegetation monitoring, achieving a spatial resolution of 1 km. However, despite the availability of operational satellites that could offer higher spatial resolutions, this approach has not yet been adopted. Satellites like Landsat, with its Thermal Infrared Sensor (TIRS) for thermal bands, and Sentinel-2, with its Multispectral Instrument

(MSI) sensor for optical bands, could significantly improve the spatial resolution of FAO's observations.

2.3 Remote sensing data for crop type mapping

To effectively assess the impact of drought on agricultural crops, it is essential to identify where different crops are cultivated. Remote sensing technology for crop type mapping enables the identification of various crop types across specific regions. This crucial information helps determine which crops are at risk of drought based on their location and inherent drought sensitivity.

2.3.1 The advent of Sentinel-2 for crop type classification

The Copernicus program, initiated by the European Space Agency (ESA), aims to monitor the Earth and its components by deploying its seven Sentinel missions. Each mission consists of a constellation of two satellites, each focusing on specific areas such as land, ocean, and atmospheric observation [45]. The Sentinel-2 mission, primarily used for land monitoring, consists of two polar-orbiting satellites that follow the same orbit but are positioned 180 degrees apart. This configuration allows for a revisit time at the equator of 10 days per satellite, effectively reducing to 5 days for the pair, and this can decrease to 2 to 3 days at mid-latitudes. Sentinel-2 delivers high-resolution optical imagery across a multispectral range. It includes 13 bands that span visible and SWIR wavelengths, featuring four bands with a 10-meter resolution, six bands at 20 meters, and three at 60 meters. Additionally, it incorporates three atmospheric bands: Band 1, Band 9, and Band 10 (figure 6).

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

]

Figure 6: Sentinel-2 spectral bands and their characteristics [46]

Those spectral bands are widely used for crops type classification excluding the atmospheric bands: coastal aerosols (band 1), water vapour (band 9) and SWIR cirrus (band 10) that were found to be inappropriate for classifying crop types [47, 48]. The blue (band 2), green (band 3), red (band 4), vegetation red edge (band 5), vegetation red edge (band 6), vegetation red edge (band 7), NIR (band 8), narrow NIR (band 8A), SWIR (band 11) and SWIR (band 12) were used in plenty of studies which found out different importance of those bands as inputs for their classification [49, 50].

These high spatial resolution bands can be used individually or combined to obtain vegetation indices and biophysical variables. Vegetation indices, such as the NDVI and the NDWI, are commonly utilised in drought monitoring and crop classification. Biophysical variables like the Leaf Area Index (LAI) are also important for crop classification. LAI measures the plant canopy by assessing the one-sided green leaf area relative to the ground surface area, with values ranging from 0 for bare soil to 10 for very dense vegetation. Changes in LAI over time are closely linked to the phenological stages of crop growth, making LAI time series essential for distinguishing different crop types.

The brightness index, calculated as follows, has also been used for crop classification. [51]:

$$Brightness = \sqrt{Green^2 + Red^2 + NIR^2 + SWIR^2} \tag{2.6}$$

2.3.2 Satellite images time series

Satellite images time series (SITS) correspond to successive recordings of the same region of interest (ROI) over time. Spectral bands and spectral indices allow for differentiation of vegetation types, but capturing the temporal evolution of these bands and indicators enhances crop discrimination even further. Indeed, while different crops may show similar spectral characteristics at a given moment, their temporal evolution can distinguish them. For example, they may have a peak in vegetation at different times or have different growth cycles. SITS are also valuable for differentiating the growth phases of crops, from germination to harvest, and for assessing how vegetation recovers from stress. Maponya et al. (2020) demonstrated the impact of using multiple images over time to achieve better classification than with a single image [49].

Satellite images are frequently affected by various types of noise, which can arise from factors such as cloud cover, atmospheric aerosols, or sensor malfunctions. These disruptions can lead to missing or erroneous data, making the analysis and interpretation of SITS challenging.

To address this issue, various interpolation methods have been developed. The choice of method depends on the amount of missing data and the desired level of accuracy or smoothing for the reconstructed time series. For example, the FAO uses a weighted least-squares method to smooth its NDVI data. Since the factors causing contamination generally reduce NDVI values, the weighting factor for temporal NDVI smoothing is biased towards peak values over sloping or valley points. A final step ensures that all peak NDVI values are preserved [52, 53](figure 7).

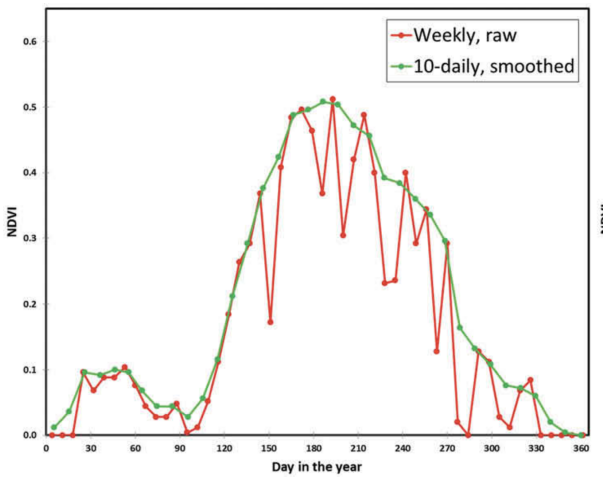


Figure 7: Annual NDVI for a crop pixel: AVHRR raw inputs and dekadal smoothed outputs[53]

Other methods like Savitsky-Golay have been used to smooth time series. This algorithm divides the signal into equal-length intervals, applying a polynomial regression that minimises the least squared error within each interval. The Whittaker smoother has also been used and requires an initial sinusoidal signal to operate effectively. It

has the same purposes as the Savitsky-Golay. S. Kandasamy et al. (2012) made a comparison of different methods of smoothing and gap-filling (figure 8).

Abrev.	Method	Principles	Processing window length (days)	Maximum gap length (days)	Reference
ICSSA	Iterative caterpilar singular spectrum analysis	Decomposition into EOFs using Eigenvalue decomposition	Whole	–	Golyandina and Osipov (2007)
EMD	Empirical mode decomposition	Decomposition into IMFs by "sifting"	Whole	128	Huang et al. (1998)
LPF	Low pass filtering	Fitting a harmonic curve to the series. Followed by 2-pass filtering of the residuals	Whole	128	Bacour et al. (2006b)
Whit	Whitaker	Penalized least square Regression – Smoothness is governed by a parameter value	Whole	–	Eilers (2003)
SGF	Adaptive Savitzky–Golay filter of the series	Savitzky–Golay filter with iterations to fit the upper envelope	72–112	–	Chen et al. (2004)
TSGF	Temporal smoothing and gap filling	Savitzky–Golay filtering with flexible window and linear interpolation for gaps	48–128	128	Verger et al. (2011)
AGF	Asymmetric Gaussian filter	Fitting asymmetric Gaussian Function to seasons	60–300	72	Jönsson and Eklundh (2002, 2004)
Clim	Climatology for each 8-day period	Inter-annual median	Whole	128*	Baret et al. (2011)

* The maximum gap length applies on the yearly reconstructed time series, not on the original time series.

Figure 8: Different methods of time series smoothing and gap-filling : processing window lengths and maximum tolerated gap lengths [54]

2.3.3 Metrics based on SITS of vegetation indices

Once a reliable representation of the evolution of vegetation indices over time is obtained, it becomes possible to evaluate various metrics to further differentiate different types of crops. Basic statistics such as the mean, minimum, maximum, and standard deviation of indices like NDWI and brightness can be used. Other metrics on NDVI have been considered relevant for distinguishing between types of vegetation (figure 9).

Notation	Description
$\mathbf{NDVI}_{\mathbf{max}}$	Maximum value
$\mathbf{NDVI}_{\mathbf{mean}}$	Mean value
$\mathbf{NDVI}_{\mathbf{std}}$	Standard deviation value
$\mathbf{NDVI}_{\mathbf{Dif-max}}$	Maximum difference in a sliding temporal neighborhood having a size *w* (default value *w*=2)
$\mathbf{NDVI}_{\mathbf{Dif-min}}$	Minimum difference in a sliding temporal neighborhood having a size *w* (default value *w*=2).
$\mathbf{NDVI}_{\mathbf{Dif-Dif}}$	Difference between $\mathbf{NDVI}_{\mathbf{Dif-max}}$ and $\mathbf{NDVI}_{\mathbf{Dif-min}}$ value estimating the transition jump
$\mathbf{NDVI}_{\overline{\mathbf{max}}}$	Maximum mean value in a sliding temporal neighborhood having a size *w* (default value *w*=2).
$\mathbf{NDVI}_{\overline{\mathbf{max}}\mathbf{-Lg}}$	Length of the flat interval containing the peak area associated to $\mathbf{NDVI}_{\overline{\mathbf{max}}}$ value.
$\mathbf{NDVI}_{\overline{\mathbf{max}}\mathbf{-Sur}}$	Surface of the flat interval area containing the peak value associated to $\mathbf{NDVI}_{\overline{\mathbf{max}}}$
$\mathbf{NDVI}_{\mathbf{Pos-Sur}}$	Maximum surface of the first positive derivative interval
$\mathbf{NDVI}_{\mathbf{Pos-Lg}}$	Length of the first positive derivative interval associated to $\mathbf{NDVI}_{\mathbf{Pos-Sur}}$ value
$\mathbf{NDVI}_{\mathbf{Pos-Rt}}$	Rate of the first positive derivative interval associated to $\mathbf{NDVI}_{\mathbf{Pos-Sur}}$
$\mathbf{NDVI}_{\mathbf{Neg-Sur}}$	Maximum surface of the first negative derivative interval
$\mathbf{NDVI}_{\mathbf{Neg-Lg}}$	Length of the first negative derivative interval associated to $\mathbf{NDVI}_{\mathbf{Neg-Sur}}$ value.
$\mathbf{NDVI}_{\mathbf{Neg-Rt}}$	Rate of senescence of the first negative derivative interval associated to $\mathbf{NDVI}_{\mathbf{Neg-Sur}}$
$\mathbf{NDVI}_{\mathbf{Pos-Tr}}$	Flag detecting if there is a bare soil transition before the positive derivative
$\mathbf{NDVI}_{\mathbf{Neg-Tr}}$	Flag detecting if there is a bare soil transition after the negative derivative

Figure 9: Relevant NDVI metrics for crop classification [51]

2.3.4 Crop classification in remote sensing

Two types of classification can be performed depending on whether labelled data is provided. Firstly, supervised classification involves providing samples of different land cover classes. The classification algorithm uses these data to learn the characteristics of each class. Once these characteristics are understood, the algorithm can classify the entire image. Secondly, unsupervised classification does not involve labelled data. In this case, the algorithm creates clusters based on the similarities and differences among the image data. It is then up to the user to label these clusters. This type of classification is not being used in this study and it will not be explained in further depth.

Main classification algorithms

Some machine learning supervised classification algorithms are described below:

The Maximum Likelihood (ML) method is founded on probabilistic principles. It utilises training data to statistically represent target classes through their multivariate probability density functions [55]. These functions indicate the likelihood that

certain spectral colours and textures correspond to one type of crop rather than another. This approach allows for complete control over the land cover classes used in the final classification. However, it has certain limitations, such as the assumption that the input data follows a Gaussian distribution. Additionally, it applies the same set of features to all classes and requires a significant amount of computational resources to perform the classification

Decision Trees (DTs) consist of nodes representing variables, branches representing value choices, and leaves indicating the probability of belonging to a particular class. Random Forest (RF) goes a step further by combining multiple DTs. Each tree makes a decision, and the final class is determined by the majority vote of all the trees. The advantages of the RF classifier are its stability and efficiency in achieving a high level of accuracy. Moreover, it can handle a large amount of spectral and temporal features.

Deep learning is a subfield of machine learning that mimics the neural workings of the human brain using multilayered neural networks. This approach enables the modeling of complex relationships within data.Neural networks are characterised by their architecture, which includes an input layer at the beginning and an output layer at the end. In the field of classification, Convolutional Neural Networks (CNNs) are particularly used. They utilise a series of filters, such as pooling layers to reduce dimensions and convolutional layers to detect patterns like textures, to ultimately produce an output, such as the label of a crop type.

Algorithm	Strengths/characteristics	Weaknesses
Maximum Likelihood (Parametric)	• Simple application • Easy to understand and interpret • Predicts class membership probability	• Parametric • Assumes normal distribution of data • Large training sample necessary
Artificial Neural Networks (Non-parametric)	• Manages large feature space well • Indicates strength of class membership • Generally high classification accuracy • Resistant to training data deficiencies – requires less training data than Decision Trees (DTs)	• Needs parameters for network design • Tends to overfit data • Black box (rules are unknown) • Computationally intense • Slow training
Support Vector Machines (Non-parametric)	• Manages large feature space well • Insensitive to Hughes effect • Works well with small training data set • Does not overfit	• Needs parameters: regularization and kernel • Poor performance with small feature space • Computationally intense • Designed as binary, although variations exist
Decision Trees (Non-parametric)	• No need for any kind of parameter • Easy to apply and interpret • Handles missing data • Handles data of different types (e.g. continuous, categorical) and scales • Handles non-linear relationships • Insensitive to noise	• Sensitive to noise • Tends to overfit • Does not perform as well as others in large feature spaces • Large training sample required
Random Forests (Non-parametric)	• Capacity to determine variable importance • Robust to data reduction • Does not overfit • Produces unbiased accuracy estimate • Higher accuracy than DTs	• Decision rules unknown (black box) • Computationally intense • Requires input parameters (#trees and #variables per node)

Figure 10: Advantages and Disadvantages of different classification Algorithms [55]

Performance metrics

The in situ data is divided into calibration data, from which features for classification are extracted, and the remaining data is used for validation. The confusion matrix has actual classes in the rows and predicted classes in the columns. From this confusion matrix, four types of values can be obtained: True Negatives (TN), False Negatives (FN), True Positives (TP), and False Positives (FP).

Using this information, we can calculate accuracy, which represents the percentage of correct predictions out of the total predictions made for a certain class. Recall is the ratio of correctly identified samples to the total number of samples actually belonging to a class, and specificity indicates the proportion of correctly identified negative samples for each class.

The metrics can be calculated as follows:

$$\text{Accuracy} = \frac{\sum TP}{\sum TP + \sum FP} \tag{2.7}$$

$$\text{Recall} = \frac{\sum TP}{\sum TP + \sum FN} \tag{2.8}$$

$$\text{Specificity} = \frac{\sum TN}{\sum TN + \sum FP} \tag{2.9}$$

From these metrics, it is possible to evaluate the F-score, which is particularly useful as in cases of imbalanced data classification and is given as follows:

$$\text{F-score} = 2 \cdot \frac{\text{Accuracy} \cdot \text{Recall}}{\text{Accuracy} + \text{Recall}} \tag{2.10}$$

Finally, the Overall Accuracy (OA) is given as follows:

$$\text{OA} = \frac{\sum TP + \sum TN}{\sum TN + \sum TN + \sum FP + \sum FN} \tag{2.11}$$

Chapter 3

Objectives

In a context where Southern Africa has been severely impacted by droughts in recent years, significantly threatening agricultural production in these countries, it is crucial to have tools that can guide policies to mitigate the impacts of droughts.

EWS have already developed drought indicators but at a coarse spatial resolution. This creates challenges because, at this scale, a pixel meant to represent agricultural fields may include various types of land cover such as forests, houses, or roads, thus biasing the true situation of the health of the fields. By improving the spatial resolution, it becomes possible to better distinguish agricultural surfaces from other types of surfaces, thus reducing interference and allowing a more accurate assessment of vegetation health. This first enables us to provide a more accurate understanding of the actual cultivated area. Moreover, this improvement helps distinguish between different types of crops, thereby facilitating more precise and effective agricultural management.

The main objective of this study consists in developing a robust VHI at a crop-specific level in the district of Chiredzi in Zimbabwe, in response to droughts.

From this objective, specific objectives are derived:

1. Testing the added value of high resolution in a drought model.
    - By evaluating the impact of high resolution in a drought model through visual interpretation of the landscape.
2. Obtain a VHI adapted to the agricultural calendar.
3. Obtain a high-resolution crop type classification.
4. Cross the VHI product with the crop map to evaluate the added value of a high-resolution drought model

Chapter 4

Materials and method

This chapter aims to provide a clear and rigorous methodological framework of the study. To facilitate a better understanding and contextualisation of the work, the content is structured into several sections, each addressing specific aspects of the research process, from the description of the study area to the presentation of data processing techniques and analysis methods.

It begins with an introduction of the study area, focusing on the Chiredzi District in Zimbabwe. The reasons for selecting this region are explained, emphasising its agro-ecological characteristics and vulnerability to drought. This section is essential for establishing the geographical and climatic context of the study.

Next, the development of a high-resolution drought index is discussed. This first pillar of the study focuses on creating a VHI that accurately reflects the health of vegetation. This index allows for precise evaluation of drought conditions at a local scale.

Following this, the methods for drought modelling and monitoring at high-spatial-resolution are detailed, including a comparison between the FAO's global VHI model and the high-resolution model used in this study, highlighting the differences in satellite data and collection periods.

A description of the data processing procedures follows, explaining the necessary steps to transform raw satellite images into usable data, taking into account atmospheric corrections, cloud masking, and temporal interpolation of NDVI and LST time series.

The methodology then introduces the creation of a crop type classification map, the second pillar of the study. This involves identifying the different types of crops present in a given region through a classification map. The ability to distinguish between crop types is crucial for understanding how drought specifically affects each type of crop.

The final section of the methodology combines the two pillars. The information pro-

vided by the crop type classification map is combined with the drought index. By overlaying the VHI on areas identified by the classification map, it becomes possible to observe and analyse the specific impact of drought on different types of crops. This combination allows for a more precise overview of crop health.

Concerning the software tools utilised, the Google Earth Engine (GEE) platform was used for the development of the drought index. The classification map was created using Python, specifically utilising the geopandas package for vector data analysis, the rasterio package for raster data analysis, and the Scikit-learn package for machine learning. Lastly, Quantum Geographic Information System (QGIS) was employed for visual inspection and mapping purposes.

4.1 Study area

The district under study is Chiredzi, located in Masvingo Province, Zimbabwe (figure 34). A Sentinel-2 tile covering the district's main agricultural areas has been used as the ROI. Zimbabwe is divided into five agro-ecological regions based on the average annual rainfall and soil types of each geographical zone. Chiredzi falls within agro-ecological regions 4 and 5, which receive the lowest annual precipitation (less than 500 mm), making the region highly susceptible to drought. Furthermore, Chiredzi is an agricultural district known for cultivating a variety of crops, including sorghum, millet, maize, sesame, and sugarcane. Samples of land cover and crop types were collected by the FAO through surveys and shared as part of this study(figure 34). This extensive agricultural land is a key factor in the evaluation of drought conditions in the area.

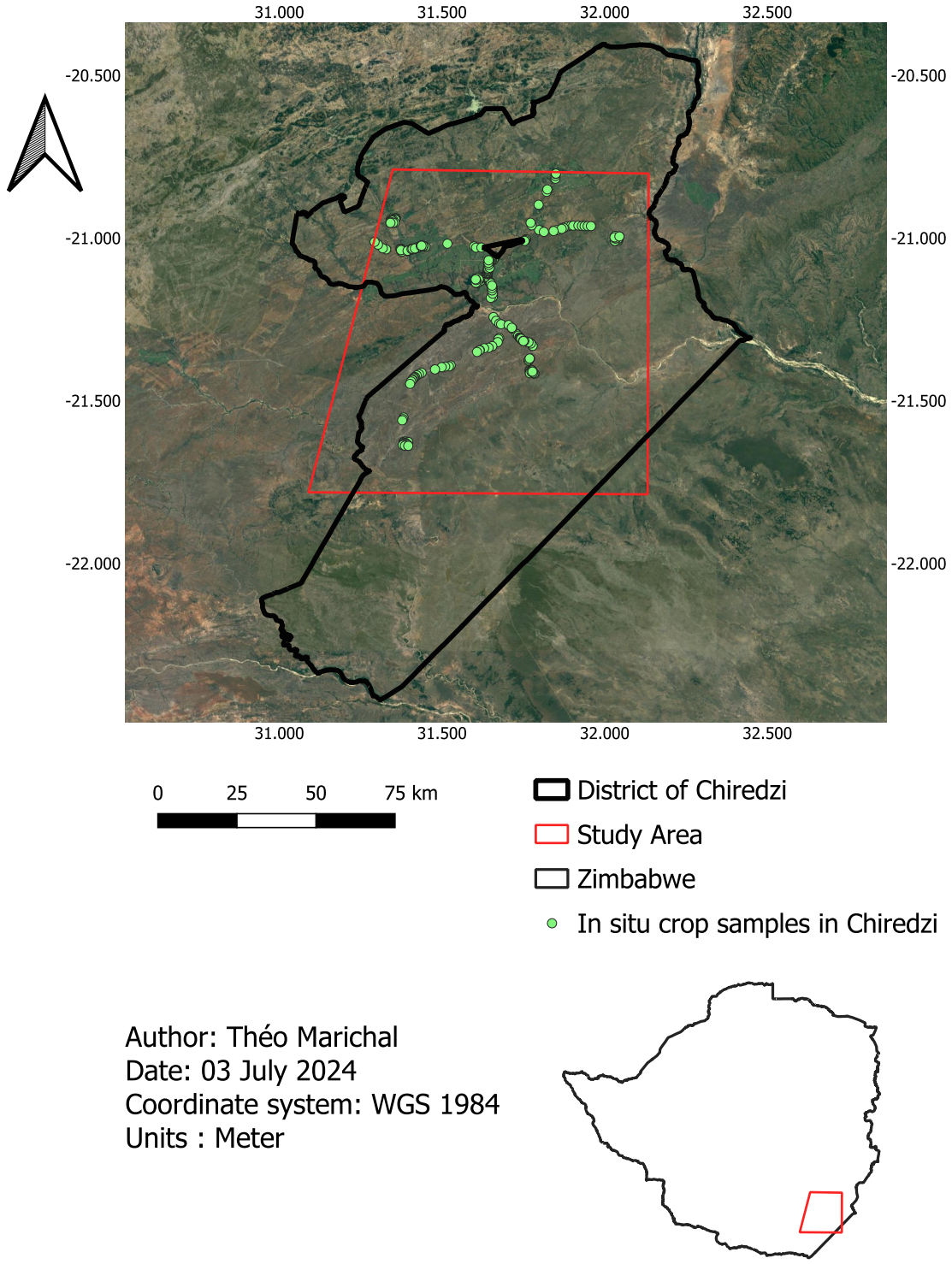


Figure 11: Location of the study area, Chiredzi District, Zimbabwe, and in situ crop sampled by the FAO

4.2 High-spatial-resolution VHI

In this section, an overview of the overall methodology to obtain a high-resolution VHI and a comparative analysis with the FAO's global VHI model is given. The data

processing steps of cloud masking and calculating smoothed and gap-filled NDVI and LST time series are then detailed. Once these time series are obtained, the VCI and TCI can be calculated to ultimately create the VHI index. Finally, this index can be adjusted to align with the agricultural calendar.

4.2.1 High-spatial-resolution VHI methodology and comparative analysis with the FAO's global VHI

The method for obtaining a high-resolution VHI is generally similar to that used for a lower-resolution VHI (figure 29). The key differences lie in the satellite data collected and the collection periods. For the FAO model, both optical (VCI) and thermal (TCI) data from 2007 to 2024 are sourced from the Metop-AVHRR sensor, with a resolution of 1 kilometer. The data from 1984 to 2006, are sourced from the NOAA-AVHRR sensor, which has a spatial resolution of 16 kilometers. However, these earlier data have been resampled to a 1-kilometer resolution to ensure consistency in spatial resolution across all the historical data. The high-resolution model developed in this master thesis uses optical data from the MSI sensor on Sentinel-2, collected at a 10 m resolution. This data has been available since 2018. The model also uses thermal data collected by the TIRS sensor on Landsat 8 at a 100 m resolution, available since 2013. Additionally, thermal data from the same sensor on Landsat 9 at a 100 m resolution is used, which has been available since 2022. Another distinction is in the type of thermal data used for the two models. The FAO model utilises BT, whereas the high-spatial-resolution model employs LST. Finally, due to climate and weather variations, the start of agricultural seasons can vary. To ensure consistent comparisons from year to year, the seasons are normalised. This allows for the comparison of the same periods in the agricultural cycle each year, thus enabling the use of a VHI based on an agricultural calendar rather than a year calendar.

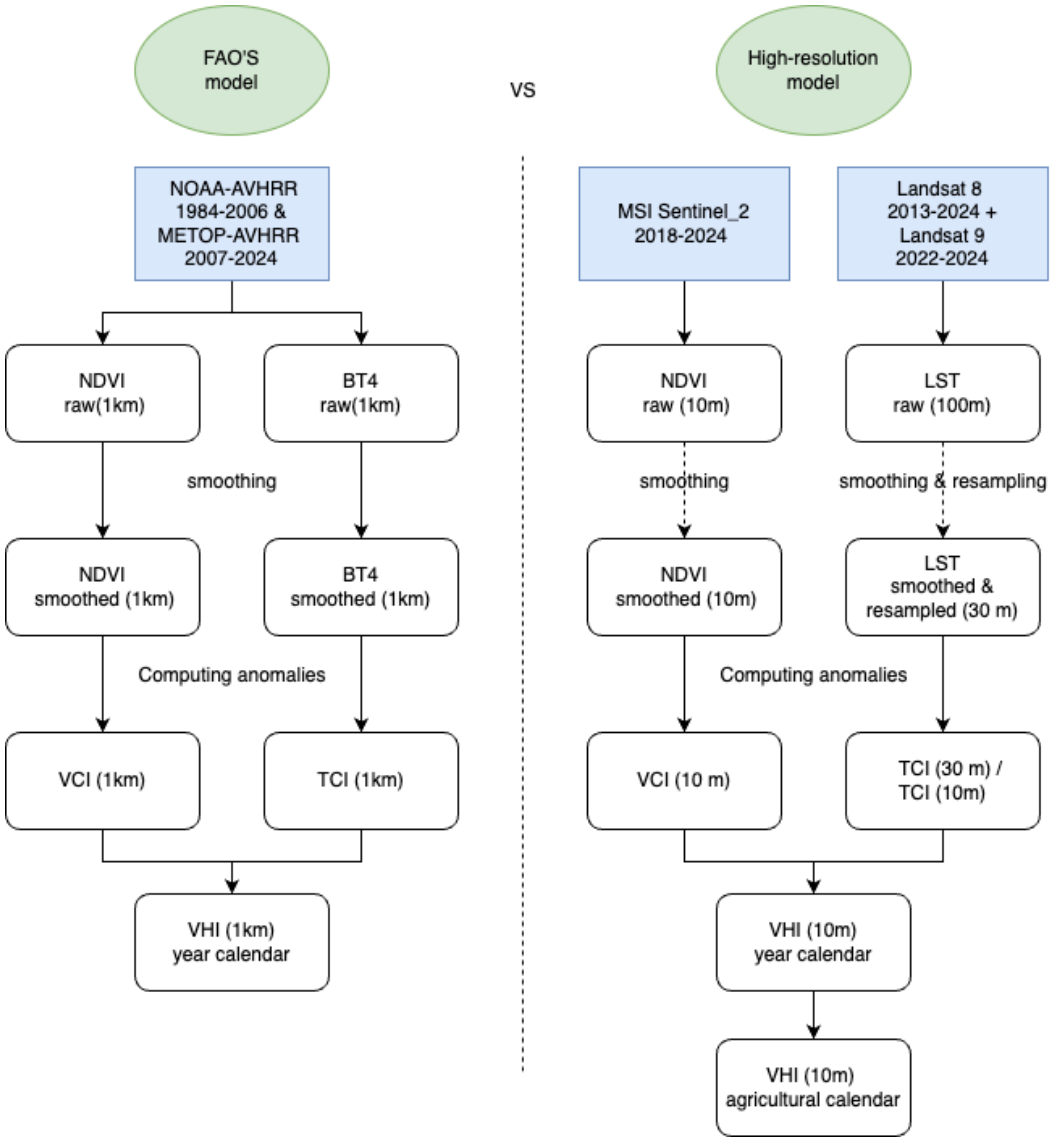


Figure 12: Comparison of FAO's global VHI versus high-resolution VHI

4.2.2 GEE platform

GEE is a cloud-based platform that provides access to a wide range of geospatial data for analysis and visualisation. It integrates multiple data sources, including a comprehensive historical archive and numerous sensors, and supports extensive data processing algorithms without the need for a local installation. The platform's strength lies in its extensive collection of high-resolution historical data, crucial for tasks such as obtaining high-resolution VHI in this study. The data required, sourced from Landsat and Sentinel, is easily managed on GEE without needing local storage resources. However, one significant issue with GEE is that all data processed or downloaded through the platform are controlled by Google, posing potential concerns regarding data ownership and privacy.

4.2.3 Data processing

Raw images require certain processing steps to be correctly interpreted subsequently. For Sentinel-2, raw images are first orthorectified and georeferenced, producing top of atmosphere (TOA) images at Sentinel-2 Level 1C. These images can further be processed to achieve Sentinel-2 Level 2A bottom of atmosphere images where atmospheric effects are also corrected. The Sentinel 2 (atmospheric) Correction (Sen2cor) algorithm, developed by the ESA and implemented on GEE, performs several key functions for image processing. It carries out the radiometric calibration, the atmospheric correction, and the detection of the presence of clouds with three levels of probability: low, medium, and high.

From the perspective of Landsat 8 and 9 data, the raw thermal band 10 data is converted into radiance (power per unit area, per unit solid angle, per unit wavelength). This radiance can then be converted into BT using the inverse of Planck's law. Planck's law allows for determining the spectral radiance distribution of a body based on its temperature. Knowing the body's spectral radiance distribution (being captured by Landsat sensors), TOA BT can be found using the inverse of Planck's law as follows:

$$BT = \frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)} \tag{4.1}$$

where:

- BT is the brightness temperature in Kelvin
- K_1 and K_2 are Planck's law constants provided in the satellite metadata
- L_λ is the TOA spectral radiance in $(\frac{\text{watts}}{\text{m}^2 \times \text{srad} \times \mu\text{m}})$

Once the TOA BT is obtained, atmospheric and emissivity corrections are made. Atmospheric profiles of geopotential height, specific humidity, and air temperature are used to eventually compose a resampled Landsat product at 30 meters spatial resolution under the ST band 10 in order to obtain LST products [56].It is at this point that the thermal data used in the high-resolution indicator differ from the thermal data used by the FAO indicator, as the FAO uses the BT at the top of the atmosphere as a proxy for ST. In contrast, this high-resolution model uses actual ST data at bottom of the atmosphere.

4.2.4 Cloud masking

Once the data is pre-processed, it is necessary to apply scaling factors for Landsat data to adjust the radiance from the raw data (digital numbers)[57]. Next, the Sentinel-2 images were cloud-masked using the Sen2Cor algorithm (figure 13). Meanwhile, the Landsat images were cloud-masked using the QA PIXEL band. This band provides

information on the quality of the pixels in a satellite image by indicating the presence of clouds, cloud shadows, snow, and water, which are encoded in a binary manner.

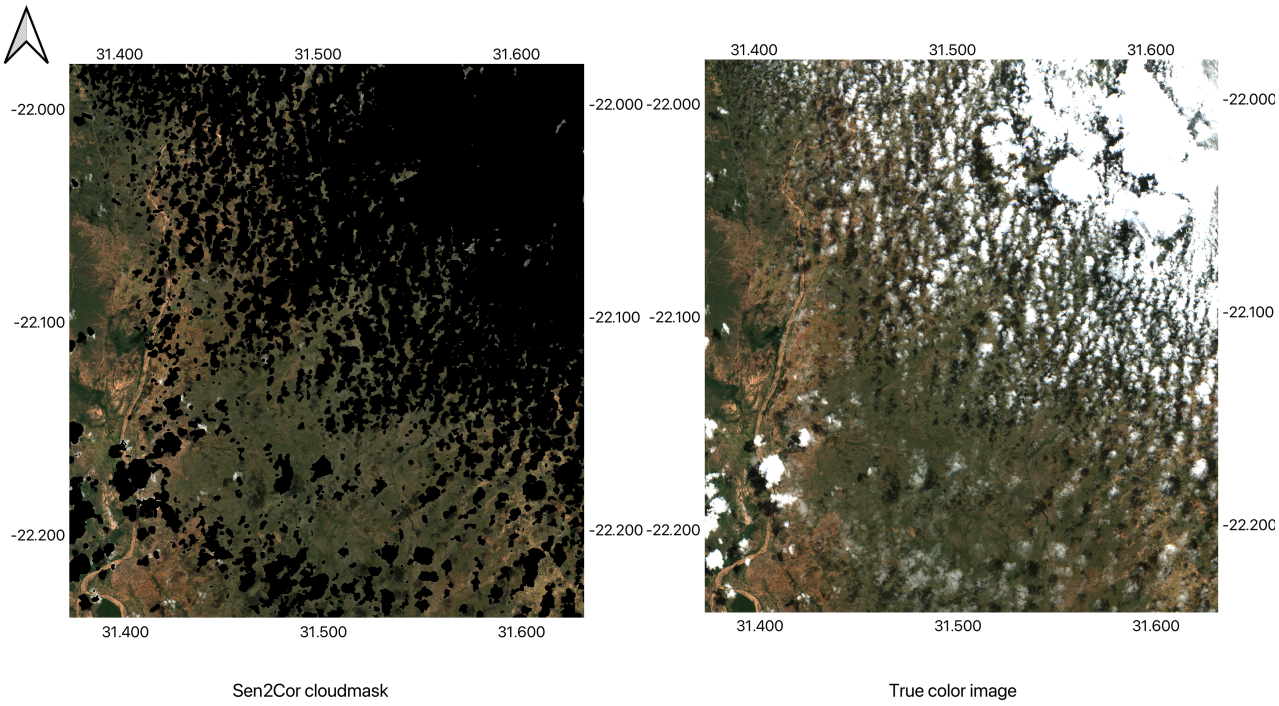


Figure 13: Comparison of a cloud-masked Sentinel-2 image processed with Sen2Cor and a true colour image captured on March 3, 2024

4.2.5 NDVI and LST time series, temporal filtering and smoothing

The first step involves ranging the NDVI from preprocessed images that have been cloud-masked (images ranging from 2018 to 2024). Next, once the NDVI and LST images are obtained, it is necessary to fill in the gaps where data are missing due to masked pixels, ensuring a continuous data series. This is achieved through temporal resampling at the pixel scale, creating an empty time series with regular intervals (every 5 days for Sentinel-2 and 16 days for Landsat images). These empty series provide a foundation to integrate with the NDVI and LST data. Linear interpolation is then employed to compute the values at these regular intervals using the closest available images in time before and after the gaps, thereby estimating values for the empty intervals and ensuring a complete and accurate dataset for environmental study.

Some pixel values may still exhibit noise due to sensor measurement errors or because some pixels escaped the various filters and corrections applied. To address this, the Savitzky-Golay filter smoothens the pixel values using a third-degree polynomial and

a 45-day moving window. Figure 14 illustrates the NDVI data of a crop pixel before and after smoothing with the Savitzky-Golay filter. For instance, the negative values observed in October 2022 are typical of noisy measurements and are corrected by this filter.

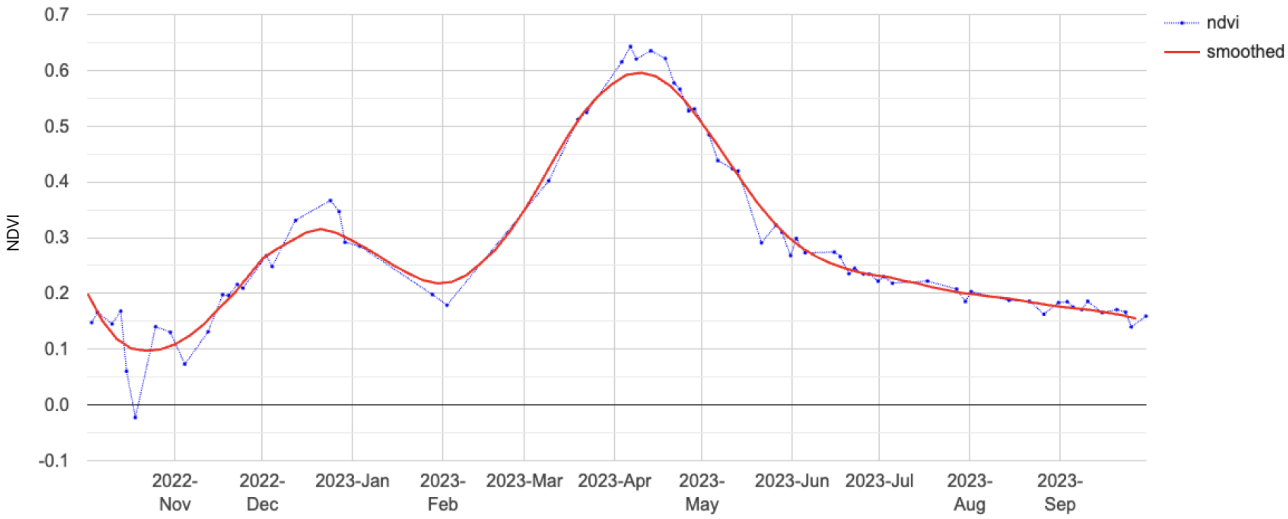


Figure 14: Annual (2022-2023) NDVI time series with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi

In terms of interpolation and smoothing of LST, data from Landsat 8 and 9 have been combined in order to mitigate the lower temporal resolution of these sensors (16 days). This integration allows for a higher frequency of LST data acquisition. Comparative analysis demonstrates that this integration increases the amount of data collected and maintains the same patterns as those observed using only Landsat 8 data 15.

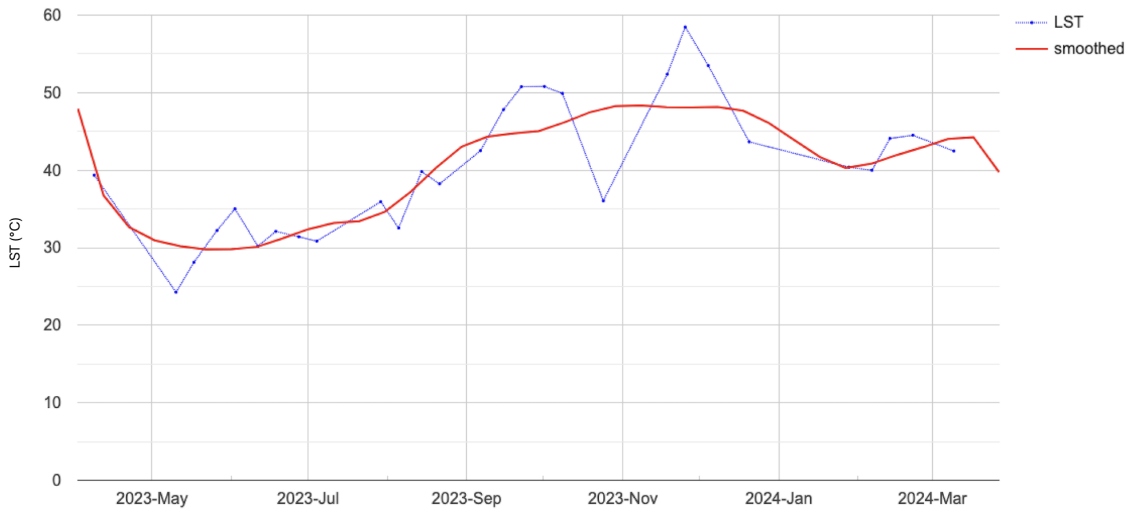


Figure 15: Annual (2023-2024) LST time series of only Landsat 8 data with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi

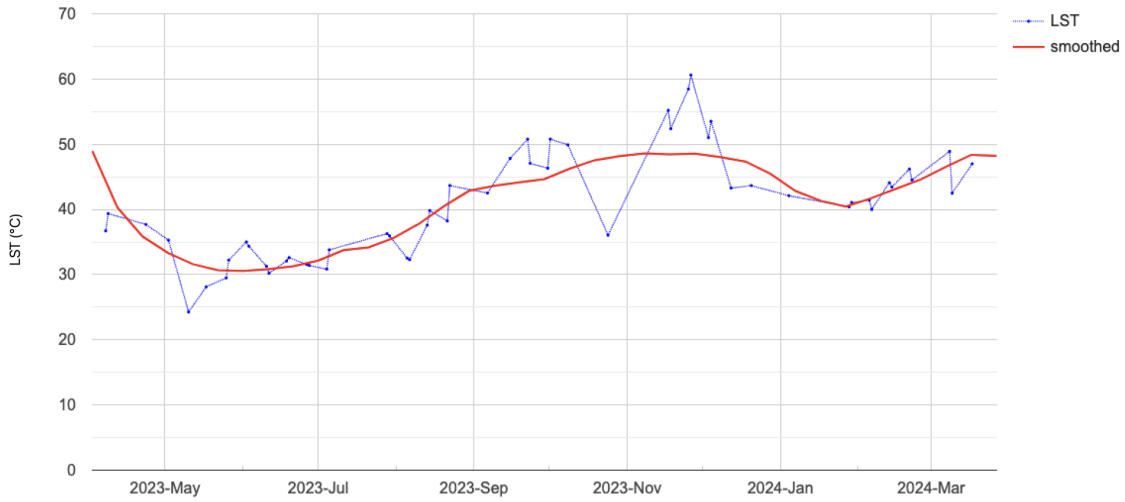


Figure 16: Annual (2023-2024) LST time series with L8 and L9 integrated data with Savitzky-Golay smoothing for a randomly selected crop pixel in Chiredzi

4.2.6 Calculation of VCI and TCI

Once continuous time series have been created from gap-filled and smoothed NDVI and LST data, it becomes possible to calculate the VCI and TCI indices, as detailed in equations 2.3 and 2.4 of chapter 2. These formulas will therefore not be further discussed in this section. It is, however, essential to specify the duration of the periods studied and the reference years used for comparisons. The study of the agricultural season spans from November 1, 2023, to April 29, 2024, and is divided into 11 segments, each lasting 16 days. This division aligns with the temporal frequency of the Landsat 8 and 9 satellites. For each 16-day period of the 2023-2024 season, a comparison is made with the corresponding periods from previous years. The historical data used for comparison is limited to 2018 onwards, which marks the beginning of Sentinel-2 data acquisition on the GEE platform.

Numerous similarities in VCI values can be observed between the low-resolution and high-resolution indices across the study area (figure 17). However, it is important to note that slight differences may arise due to variations in the satellite data collected—since 2018 for the high-resolution VCI and since 1984 for the FAO model. These differences can affect the recorded minimum and maximum NDVI values for each pixel in the VCI calculation, thereby altering the resulting VCI values. From a TCI perspective, a comparison is not feasible, as the FAO does not provide TCI data.

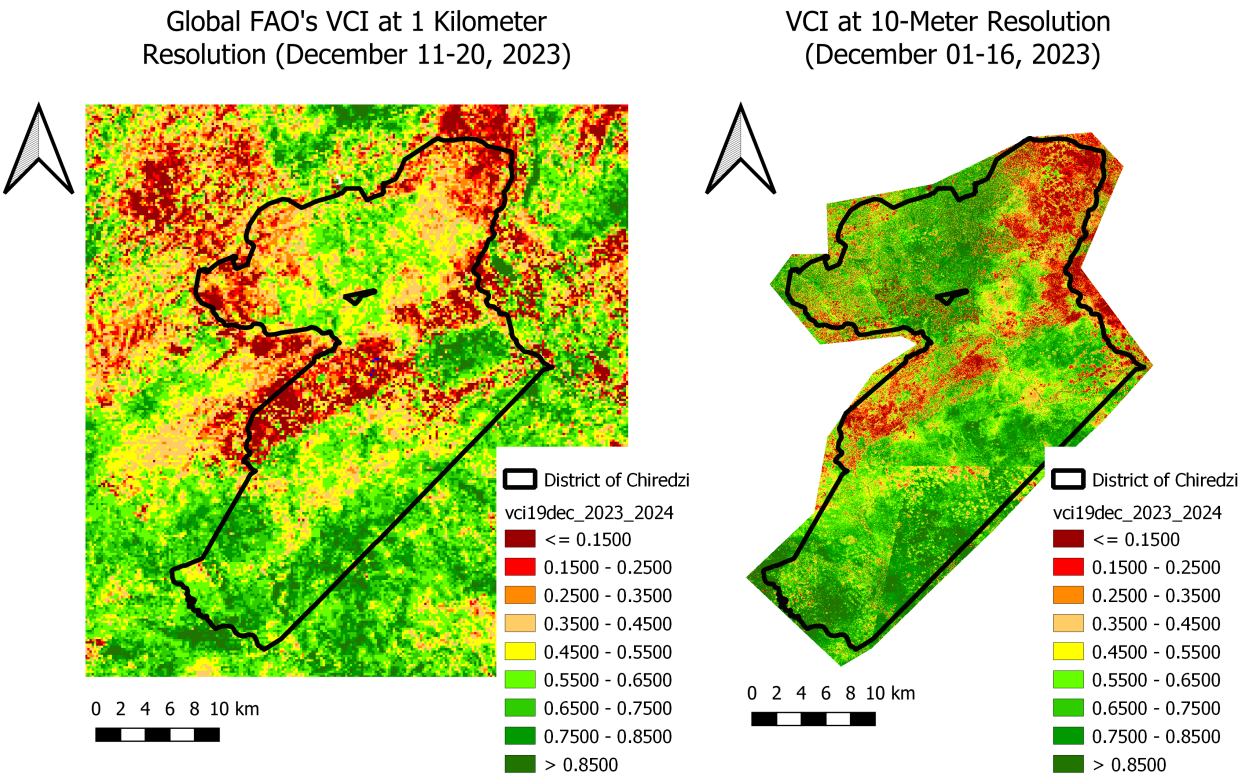


Figure 17: Comparison between global-resolution VCI (FAO) and high-spatial-resolution VCI in December 2023

4.2.7 Standardising the spatial resolution for both VCI and TCI indices

Before calculating the VHI, it is first necessary to standardise the spatial resolution between the VCI and TCI. Specifically, the VCI has a resolution of 10 meters, while the TCI has a resolution of 30 meters, leading to a difference in resolution that must be addressed.There are two scaling options to harmonise these resolutions.

The first option is to enhance the apparent resolution of the TCI to 10 meters. The bilinear interpolation enables to resize an image's pixel grid by calculating new pixel values. In this case, the goal is to scale up the pixel grid by a factor of 3, both in width and height. The new pixels are calculated based on the 4 nearest neighbouring pixels (those belonging to the original image before resizing) that are closest to the new pixel location. To perform the interpolation, a weight is assigned to the neighbouring pixels proportionally to the distance between them and the new pixel's location. Figure 18 illustrates the TCI at 10 meter resampled with bilinear interpolation (a) and the TCI at 30 meter spatial resolution (b).

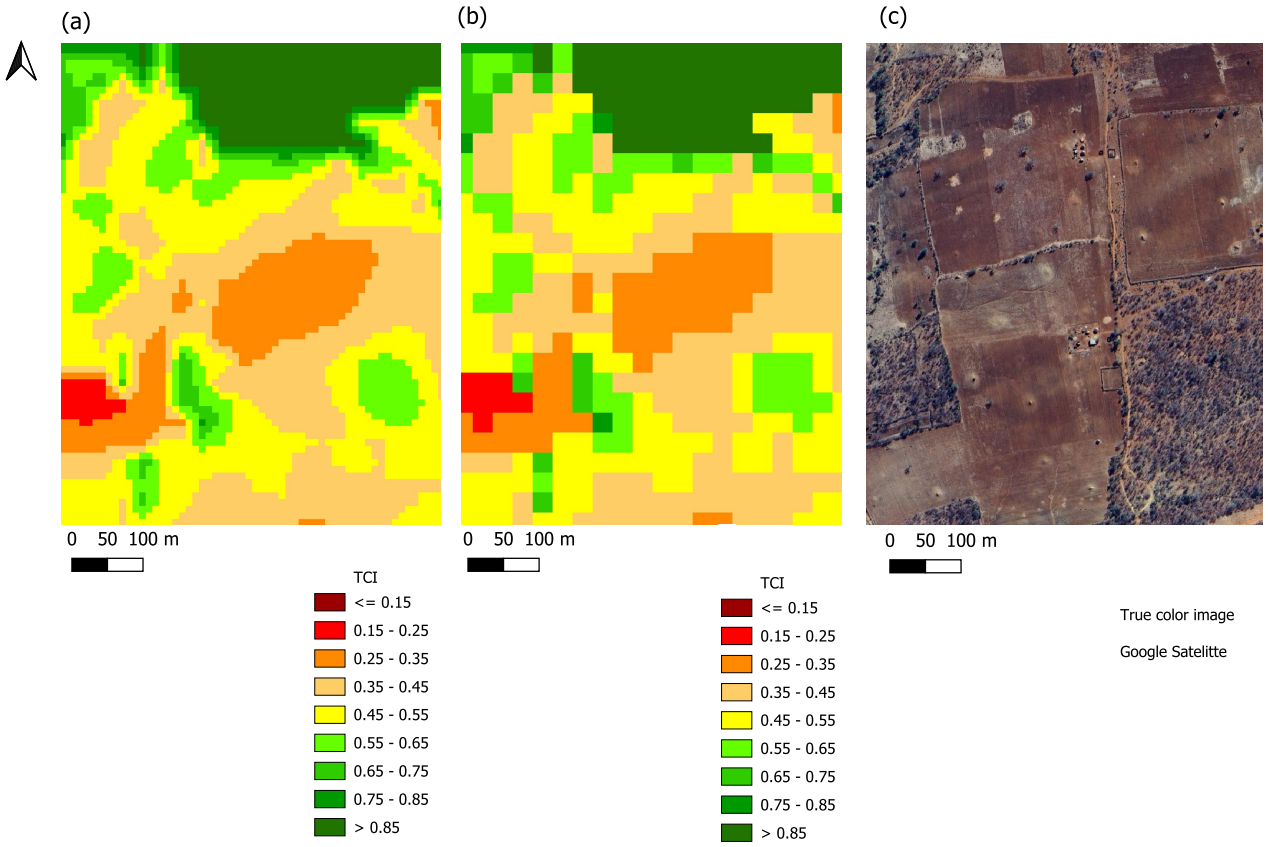


Figure 18: Comparison between the resampled TCI from January 20 to February 5, 2024, at 10-meter spatial resolution with bilinear interpolation method (a), and the original TCI at 10-meter spatial resolution (b), alongside a representation of the study area with a true-colour image from Google Satellite (c)

The second option is to resize the VCI resolution to 30 meters. To better understand how this resizing should have been performed, it is essential to comprehend how spatial quality in optical remote sensing is determined by two key elements: Ground Sample Distance (GSD) and the Point Spread Function (PSF) [58].

Firstly, GSD refers to the distance between the centers of adjacent ground pixels. It is the smallest distance that allows two objects on the ground to be distinguished and thus corresponds to the theoretical spatial resolution.

The PSF represents a measure of how a point of light is spread on the ground once captured by the remote sensing system. Various elements can affect the sharpness of the image, including the optics of the instrument, the electronics of the sensor, and the atmosphere. These elements are considered in the PSF. The PSF thus represents the dispersion of the light signal, i.e., how a point source of light is captured and spread across multiple pixels instead of being confined to a single one. The PSF therefore enables the measurement of the actual spatial resolution of the instrument [59]. Wenny et al. (2015) successfully estimated the PSF of the TIRS on Landsat 8. Assuming isotropy, meaning that the PSF is the same in all directions, and after

normalising this function to have a peak amplitude of 1 and a means of 0, it can be calculated as follows:

$$f(x, y) = e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}} \tag{4.2}$$

With:

- σ, the standard deviation, valued at 2.65

The VCI can now be calculated at the same spatial resolution as the TCI (30 meters), following the determination of the PSF for the TIRS sensors. The figure 19 illustrates the VCI signal at its original spatial resolution of 10 meters (a), as well as resampled to 30 meters taking into account the PSF of the TIRS sensors (b), and finally, the equivalent of the outlined region in true colour (c). However, the figure shows that analysing at a field level using a spatial resolution of 30 meters appears too coarse and is no longer sufficiently interpretable.Consequently, the decision to standardise the spatial resolution between the VCI and TCI was made at 10 meters, rather than 30 meters, and the initial method of bilinear interpolation was then employed.

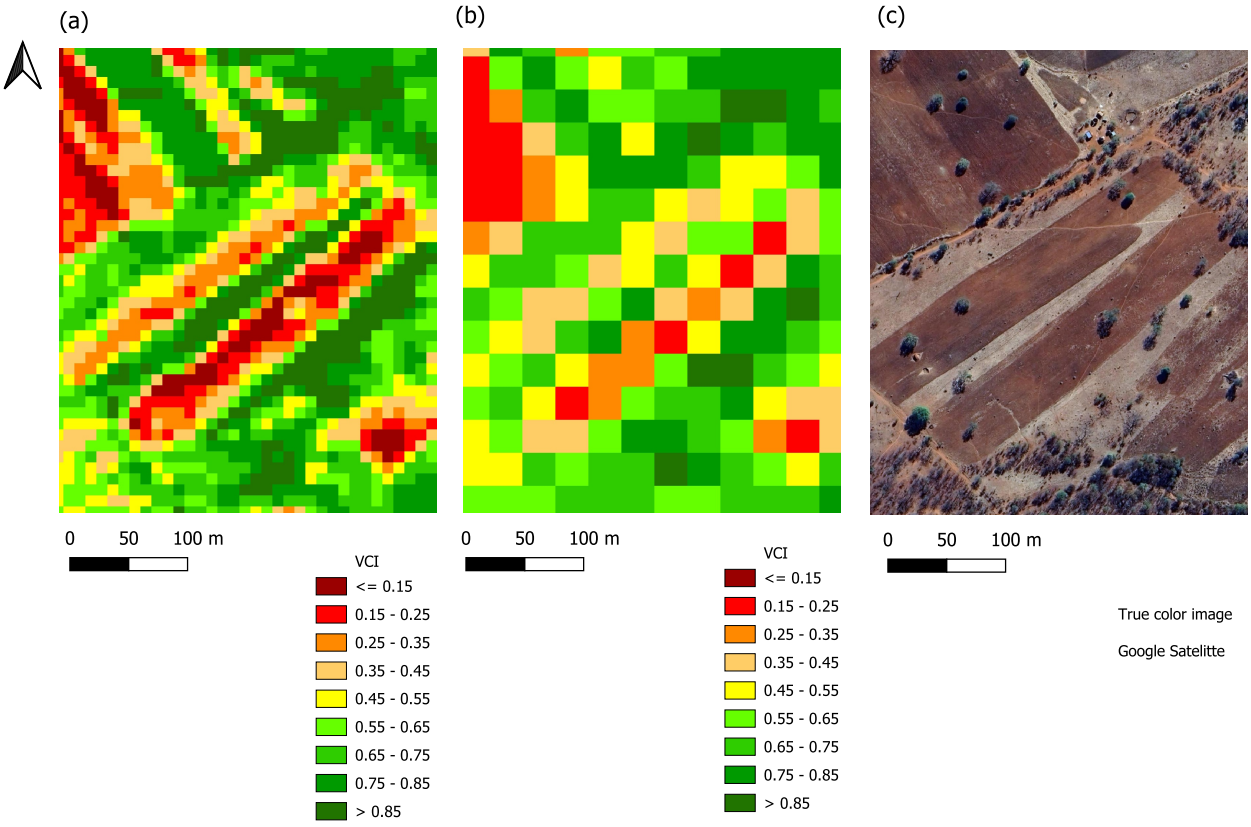


Figure 19: Comparison between the original VCI from January 20 to February 5, 2024, at 10-meter spatial resolution (a), and the resampled VCI at 30-meter spatial resolution, considering the PSF of Landsat 8 TIRS sensors (b), alongside a representation of the study area with a true-colour image from Google Satellite (c)

4.2.8 Calculation of the VHI

Now that the VCI and TCI indices have been standardised, it is finally possible to calculate the VHI. The VHI is calculated as described in equation 2.5 of chapter 2.

4.2.9 Adjustment of VHI according to the agricultural calendar

The current methods for evaluating the VHI do not take into account seasonal shifts. The weather or climatic conditions can cause a crop's growing period to start earlier or later in a given year. This shift in the plant's developmental cycle from one year to another is currently not accounted for in the VHI. Indeed, the VHI assumes that the development periods in the plant growth cycle, such as germination, flowering, or maturation, always occur at the same time each year.

Figure 20 shows by observing the NDVI profiles of a crop that the development periods in the plant's growth cycle are not in phases.

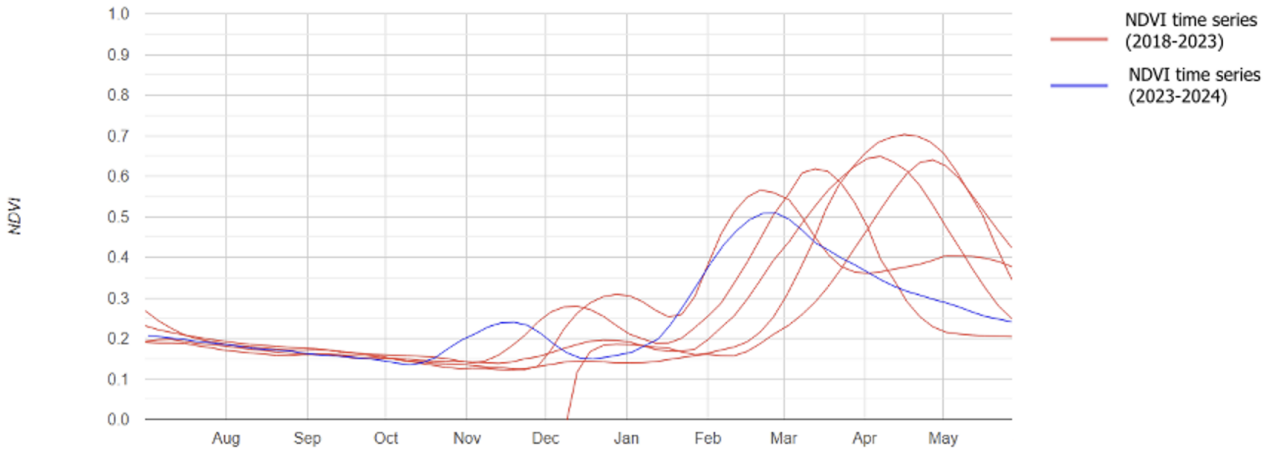


Figure 20: NDVI profiles of the 2023-2024 agricultural season and historical NDVI profiles since 2018 of a crop located in Chiredzi

The impact of seasonal shifts on VHI varies with the scale of analysis. Coarser spatial resolution VHIs, such as the FAO's VHI, have a too low resolution to capture details at the individual field level. Instead, the smallest unit analysed encompasses several fields, and individual variations in seasonal shifts within one field may be masked by the general trends of other fields. Conversely, with a high spatial resolution of 10 meters, seasonal shifts specific to a single field are not diluted by the stability of surrounding fields. Consequently, each shift directly influences the VHI of that particular field and must be taken into account.

To address the issue of seasonal shifts, it would be more appropriate to normalise the NDVI according to seasonal variations, ensuring that each measurement is compared to a similar growth period from previous years, thus ensuring a more consistent and

precise comparison. Therefore, one should identify common phases in crop growth cycles through NDVI profiles.

The period marking the start of the growth season, designated as "start of season" (SOS), is determined using a method based on an average: the 50% threshold [60]. This threshold is calculated as the average between the maximum NDVI and the preceding minimum NDVI, which must be below 0.2. In fact, this minimum (0.2) represents the highest level of NDVI observed during periods when there is no active agricultural activity, such as during fallow. This choice makes sure that the value used to mark the beginning of the season really shows a meaningful increase in plant growth. This calculation approach allows for the calibration of different NDVI profiles as obtained on figure 21. The development periods in the plant growth cycle are now normalised, allowing for a more robust establishment of the VHI.

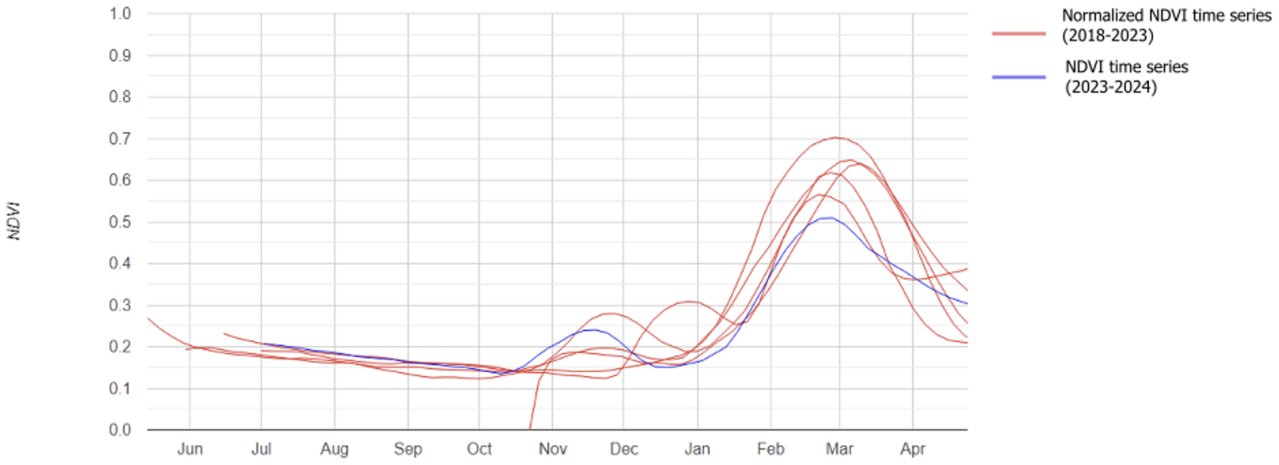


Figure 21: NDVI profiles of the 2023-2024 agricultural season and historical NDVI profiles since 2018 after normalisation of the seasonal shift of a crop located in Chiredzi

4.3 Crop Type Classification

With the high spatial resolution VHI completed, it is now crucial to conceive a crop type map to precisely assess the impact of drought on agricultural crops. Without the crop type map, targeting the affected agricultural areas wouldn't be possible.

In this section, the key steps of crop type classification are examined. This process includes handling in situ data, which involves the preparation, calibration, and validation of the data. Additionally, it encompasses the extraction of specific spectral signatures for the different land categories within the in situ data. To obtain these spectral signatures, Sentinel-2 data acquisition is required, followed by feature extraction using Sentinel-2 bands and spectral indices. Once the in situ data is processed and the specific spectral signatures for the different classes are obtained, it is possible to generate a crop type classification map using RF classification and the application

of majority filtering.

4.3.1 In situ data preparation

The FAO, with financial support from the African Development Bank, conducted a survey on Zimbabwe's 2024 summer crops to enhance their spatial distribution. Data collection was performed using two distinct methods. Initially, a statistical survey involved trained teams carrying out field measurements on 10m x 10m secondary sampling units (SSU), which were randomly selected within 2km x 2km primary sampling units across Zimbabwe. The primary goal of this survey was to estimate acreage statistics for key crops. Additionally, a windshield survey was conducted, where teams collected numerous data points along roadways. This approach aimed to gather diverse class samples to improve the training of the classification algorithm. In total, 4,601 points were sampled across Zimbabwe. In the study area of this master's thesis, 326 samples were collected, with 89 coming from the statistical survey and 237 from the windshield survey.

The distribution of the number of samples across different land categories in the study area is provided in Table 1. It is important to note that the in situ sample dataset is unbalanced. A minimum threshold of 5 samples was established for a class to be considered as represented in the collected dataset. Therefore, classes like cowpea, groundnut, and fallow are underrepresented and will not be considered in the subsequent analysis. In contrast, classes such as sugarcane, woodland, sorghum, and shrubland are heavily represented.

Table 1: Distribution of samples by land category in the study Area, based on FAO's windshield and statistical survey

	Statistical Survey	**Windshield Survey**	**Total samples**
Urban	6	5	11
Shrubland	29	3	32
Fallow	3	/	3
Bare Land	9	10	19
Agriculture	11	/	11
Forest	7	2	9
Water	1	6	7
Woodland	23	24	47
Maize	/	7	7
Sunflower	/	2	2
Sugarcane	/	132	132
Sesame	/	6	6
Cowpea	/	1	1
Pearl Millet	/	6	6
Groundnut	/	1	1
Sorghum	/	32	32

Polygons are manually created using QGIS software, based on images from Google Satellite and Sentinel-2 (figure 22). While Google Satellite provides very high-resolution images, they are not always up-to-date, so a more recent composite image from Sentinel-2 (February 2024) is used to verify any changes in land cover. Once the polygons are drawn, a 10-meter buffer is applied to avoid edge effects that could bias the analysis. The polygons are also designed to balance the dataset across different land use classes. To achieve this, polygons were not established for each individual sample in the most frequently sampled classes. Despite this, some classes still lack sufficient samples and are therefore underrepresented in the dataset. These classes include bare soil, millet, maize, and sesame (figure 23). The approach to address this issue will be explained in the section on calibration and validation of the data. Additionally, it is important to ensure that the proximity of polygons does not result in similar spectral signatures. To avoid this proximity bias, samples are selected from across the entire study area to ensure better representation of each category.



Figure 22: Geospatial polygons created based on a composite cloud-masked Sentinel-2 image from February 2024 and a Google Satellite image, using in situ samples from Chiredzi District, Zimbabwe

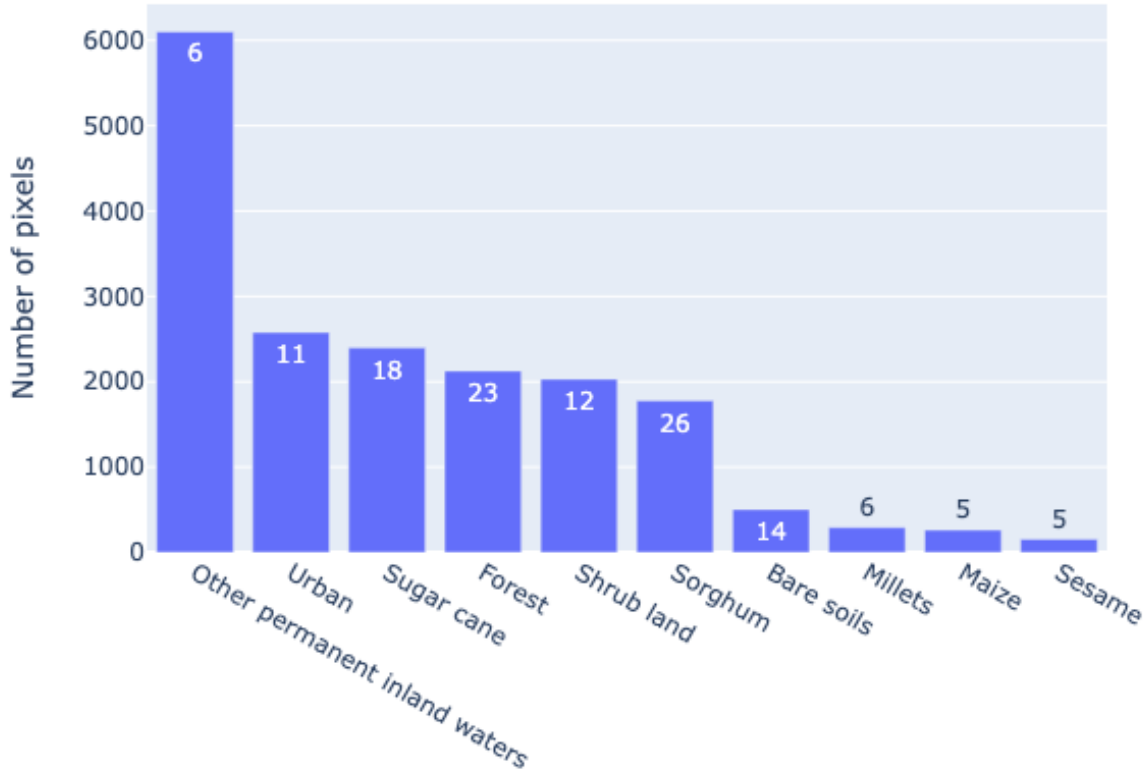


Figure 23: Number of polygons and number of 10-meter by 10-meter pixels for the different classes selected after data preparation

4.3.2 Calibration data and validation data

Once the polygons have been created, they need to be divided into two sets, each with a distinct role in the classification algorithm that will produce a crop type map. The first set, known as the calibration or training set, is used to train the classification algorithm. The second set, referred to as the validation set, is used to evaluate the algorithm's performance after training. This validation set allows for comparing the classification results with known data to assess the accuracy of the algorithm.

The imbalance in data (see section 4.3.1) can bias the algorithm in favour of classes with more pixels, which is problematic for image classification.

To mitigate this bias, the proportion of data used for calibration is adjusted according to the frequency of each class within the dataset. The distribution is as follows:

- **Major classes (more than 10% of the total data)**: 55% of the available data is used for calibration.
- **Intermediate classes (between 3% and 10% of the total data)**: 70% of the available data is used for calibration.

- **Minor classes (less than 3% of the total data)**: 85% of the available data is used for calibration.

This approach ensures that less frequent classes are not underrepresented in the calibration set, giving them proportionally more importance based on their rarity. In other words, the rarer a class is, the larger the proportion of its data that is used for calibration, improving the algorithm's performance on these underrepresented classes.

Although the classification operates on a per-pixel basis, it is essential to perform the split at the polygon level to ensure that the validation is independent and to avoid spatial autocorrelation between training and validation pixels.

4.3.3 Sentinel-2 Data Acquisition

After properly processing the samples, it is necessary to assign them distinct spectral signatures so that the classification algorithm can learn to associate each signature with a specific class. These spectral signatures are obtained using Sentinel-2 data. Sentinel-2 images could not be processed on GEE because the development of the classification model requires combining these satellite images with in situ data. This in situ data cannot be shared on the cloud due to confidentiality reasons, which prevents their use on GEE. Therefore, 11 images from a Sentinel-2 tile with a cloud coverage of less than 30% were manually selected during the crop season from the Copernicus Dataspace (figure 34). These images were downloaded as Level 2A (georeferenced and orthorectified), as explained in section 4.2.3. To ensure that all the Sentinel-2 bands selected for classification have the same spatial resolution, the 20-meter bands were resampled to a 10-meter resolution through bilinear interpolation. Subsequently, these were re-masked using the Sen2Cor algorithm.

4.3.4 Metrics and time series of spectral indices for feature extraction

To accurately extract features from Sentinel-2 images, it is useful to calculate time series of various spectral indices and their corresponding metrics.Initially, various spectral indices were calculated for each of the 11 cloud-free Sentinel-2 images, and time series were derived from these. The spectral indices used include NDVI, NDWI, and Brightness. For more details on these indices and their selection for classification, refer to section 2.2.4 and 2.3.1.

To illustrate the advantage of time series, the curves of average NDVI (figure 24) highlight the typical variations in NDVI of crops. These curves start with lower values at the beginning of the season, which gradually increase as vegetation develops during the growing season. This differentiation helps distinguish crops from urban areas and bare soils, which exhibit a more constant NDVI close to 0. Furthermore,

a significant difference in NDVI helps distinguish water bodies from crops. Lastly, forests and sugarcane plantations, characterized by much denser vegetation, fall within different NDVI ranges.Thus, the use of NDVI time series proves particularly effective in identifying and differentiating cultivated from non-cultivated areas.

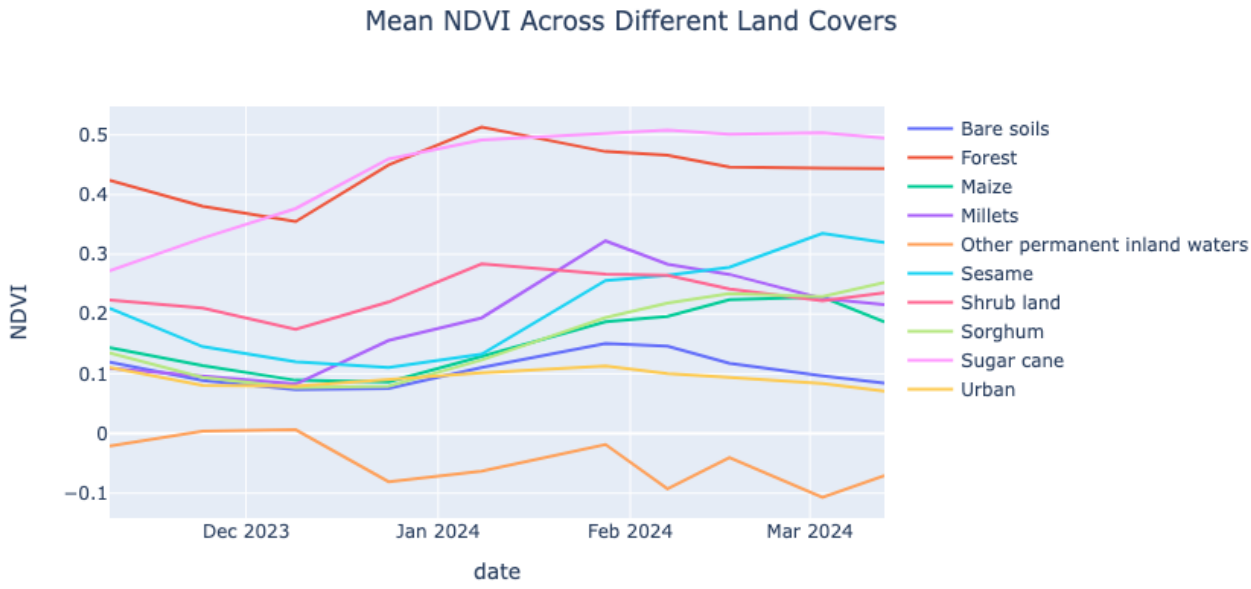


Figure 24: Mean NDVI time series of different land cover classes

To differentiate among crop types, more meticulous analysis is required, and specific trends unique to each crop must be highlighted. By observing the time series of all sampled crops 25, it is obvious that sorghum, which has many more sampled data points, shows a time series similar to corn. Millet and sesame can be slightly distinguished from their NDVI profiles, but some fields appear anomalously elevated, which could be defined by a different cultivation type (irrigation, intensive farming).

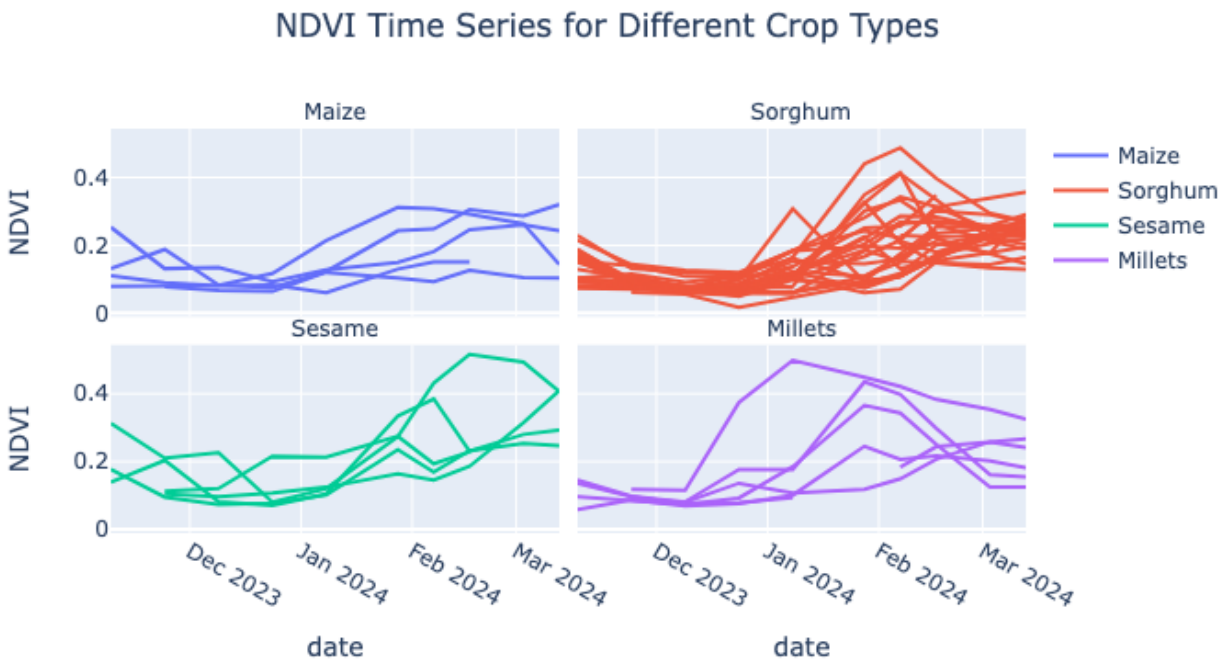


Figure 25: NDVI time series of different crop type classes

To better utilize time series, various metrics are calculated from the spectral indices. For NDVI, the maximum, minimum, date of maximum, and maximum and minimum differences within a temporal neighbourhood of size 3 are considered. For NDWI and Brightness, basic statistics such as mean, minimum, maximum, and standard deviation are taken into account [51].
The Sentinel-2 spectral bands can themselves be used for discriminating different types of land cover. Therefore, all Sentinel-2 bands, except for atmospheric bands 1, 9, and 10, were considered at various dates during the vegetation growth period, specifically on February 5, February 22, March 10, and March 26.

4.3.5 Random Forest classification

Principles and main advantages

Among the many classification methods available, Random Forest (RF) was selected. This model is more robust and accurate than classical parametric models and DTs [61]. The robustness of RF relies on several principles. The principle of bagging involves training each decision tree on a subset of the available data and then combining the decisions of these trees to obtain a more accurate overall prediction. The "out of bag" concept allows evaluating the predictions of DTs using data that was not included in their training, thus ensuring a more reliable evaluation. Moreover, RF allows evaluating the importance of different input variables by observing the impact of altering one variable at a time on the accuracy of the tree's decisions. This ability to assign different weights to input variables is crucial in this case, where many variables

(spectral bands, temporal spectral indices, and their metrics) are included in the algorithm, thereby reducing the risk of bias in the predictions [61]. One advantage of the RF model is that it requires few parameters to be set, including:

- **n_estimator= 100**: The number of DTs used.
- **max_features**: The maximum number of features to consider for the best split. The square root of the total number of features was used for each split.
- **max_depth=None**: This means that the trees will be built until each leaf is pure or contains fewer than min_samples_split samples.
- **min_samples_split=2**: At least 2 samples are required to split a node.
- **min_samples_leaf=1**: At least 1 sample is required in a leaf node.

Hierarchical Classification

Once all parameters were established, two approaches were tested. The first approach, referred to as the classical approach, involved directly classifying each class with available training samples. In this approach, the model attempted to distinguish all available classes, such as different crop types and other land covers, in a single step without prior distinction between crop and non-crop areas.

The second approach, known as the hierarchical approach, employed a two-step classification process. In the first step, a distinction was made between crop and non-crop classes, such as forests, urban areas, and other land covers. Once the crop areas were identified, the second step involved classifying specific crop types, such as maize, wheat, and sorghum, under the crop mask.

A key aspect of the hierarchical approach is the management of sample distribution, particularly during the second classification step. Since certain crops, like sorghum, were more represented in the training samples, there was a risk of these crops being overrepresented in the classification results. To address this potential bias, a sample rebalancing method was implemented. Specifically, an equal number of calibration samples (100 samples per crop class) were allocated to each crop type. This rebalancing ensured that no single crop class would dominate the predictions simply due to its higher representation in the training data.

4.3.6 Majority Filtering

The majority filtering method is applied once the crop type map has been obtained. It involves applying a sliding window over each pixel of the image and reassigning each pixel to the most frequent class among the pixels within this window. A defined-size

window (3x3) is centered on each pixel of the image. To ensure that the window can be applied to all pixels of the image, even those on the edges, the image is extended by padding with the values of the pixels on the edges. For each pixel, the sliding window selects all neighbouring pixels within the window. The most frequent class among these pixels is determined (i.e., the majority class). The central pixel of the window is then reassigned to the majority class determined in the previous step. This process is repeated for each pixel of the image, resulting in a new image where each pixel corresponds to the majority class of its window. This allows for a smoothing of transitions and a reduction of noise 26.

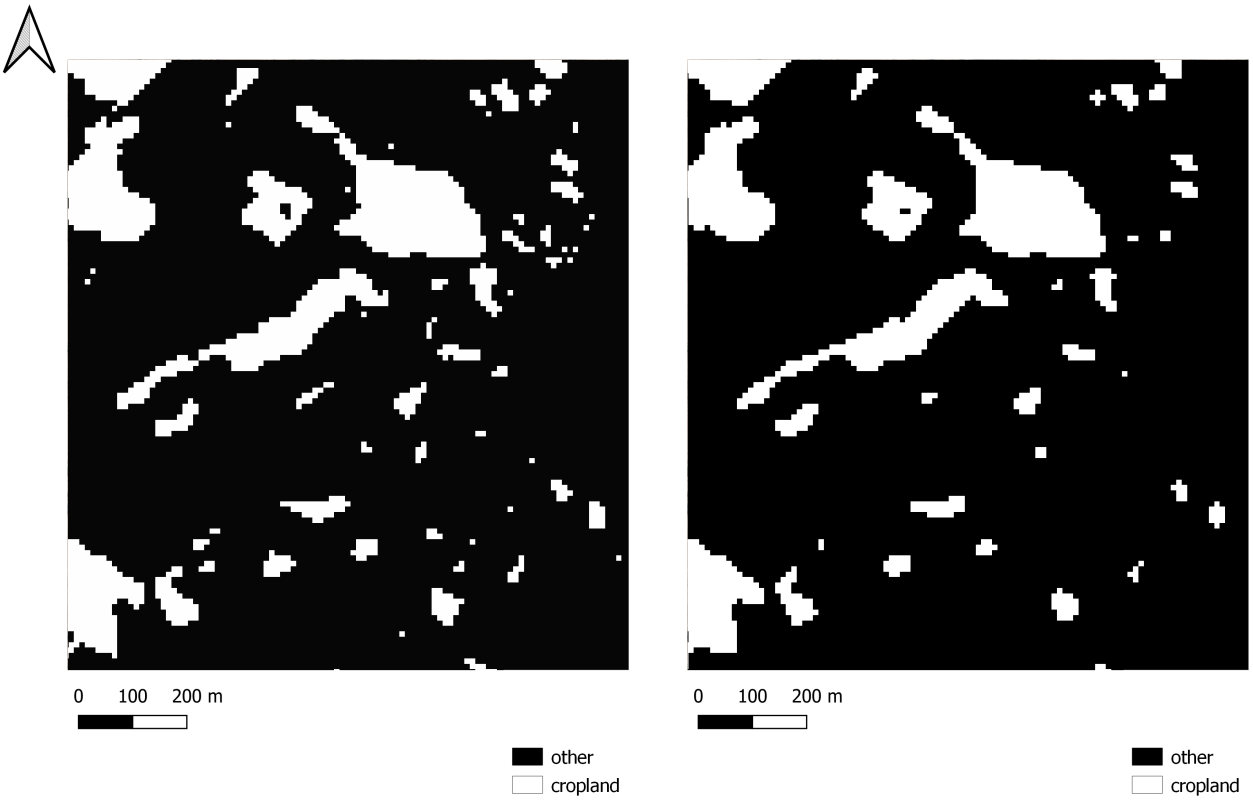


Figure 26: Majority filtering application

Chapter 5

Results and discussion

This chapter presents the key results from the study. We begin with an analysis of the VHI at the district level, comparing the VHI products from the FAO with those generated in this study. The focus then shifts to the benefits of high spatial resolution data for monitoring drought impacts on crops, highlighting specific areas where this improved resolution provides significant insights. The chapter also evaluates the accuracy of crop classification through various metrics and a confusion matrix, discussing its implications for crop type differentiation. Finally, the integration of VHI time series with the crop type map is explored, with detailed analyses of sugarcane and sorghum fields. By the end of this chapter, the reader will gain a clear understanding of vegetation health dynamics and the value of high-resolution monitoring in drought-prone areas.

5.1 VHI Results

This section presents the evolution of the VHI at the district level. It then allows for a visual comparison between the VHI products from the FAO and those obtained in this study. Finally, it shifts to a finer scale to illustrate the benefits of high spatial resolution in monitoring crops water stress.

5.1.1 Evolution of the VHI at the district level

With the VHI calculation completed, it is possible to assess the evolution of vegetation health across the entire Chiredzi District during the 2023-2024 agricultural season (figure 27). Until the end of January 2024, the drought appears to have affected only certain areas of the district. However, starting in early February, vegetation health began to decline. This deterioration is also detected by NASA Earth Observatory, which reports that Zimbabwe experienced prolonged drought periods from February onward [62]. The persistent water deficit led to an intensification of the drought in March and April. This is reflected in the increasingly lower VHI values and their wider distribution across the Chiredzi District.

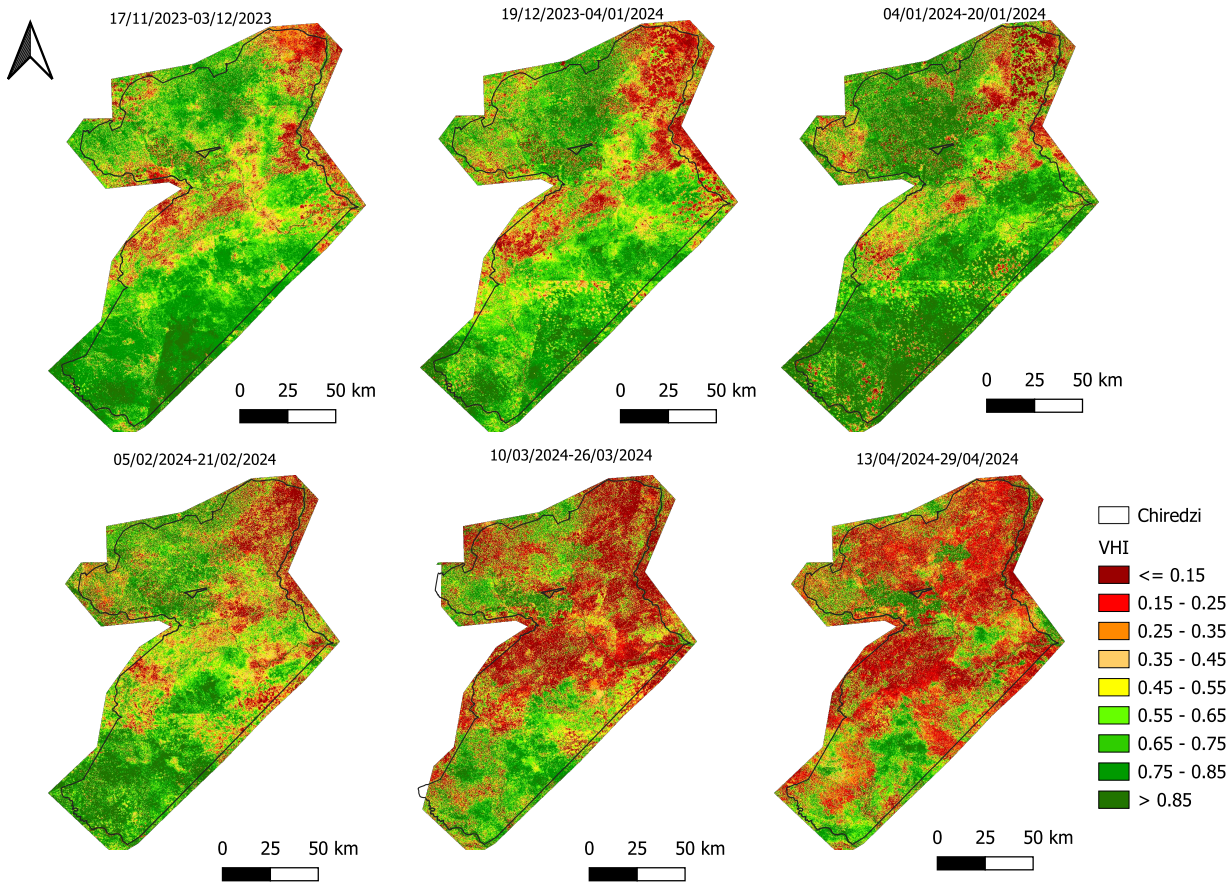


Figure 27: Evolution of vegetation health throughout the 2023-2024 agricultural season in Chiredzi District, Zimbabwe

5.1.2 Comparison between FAO's VHI at the district level

The results obtained at the district level for two different models of the VHI show similarities, particularly in January, where regions affected by water stress are located in the same areas in both models. However, in March, the results diverge more significantly. The high-resolution VHI model indicates a more intense drought compared to the FAO model, with values tending more toward extreme lows in the VHI, indicating poorer vegetation health. This can be explained by the fundamental difference between these two models, which lies in the historical data on which they are based. The FAO model uses data from 1984 onwards, whereas the VHI model used in this master's thesis relies on data starting from 2018. This difference in the reference periods may account for why the March values are more extreme in the high-resolution model. The FAO model, by using a longer comparison period, incorporates years when vegetation health may have been very poor, thus making the current values less extreme by comparison.

Nonetheless, it is also noteworthy that VHI trends are similar in March, with regions showing higher VHI values largely corresponding between the two models.

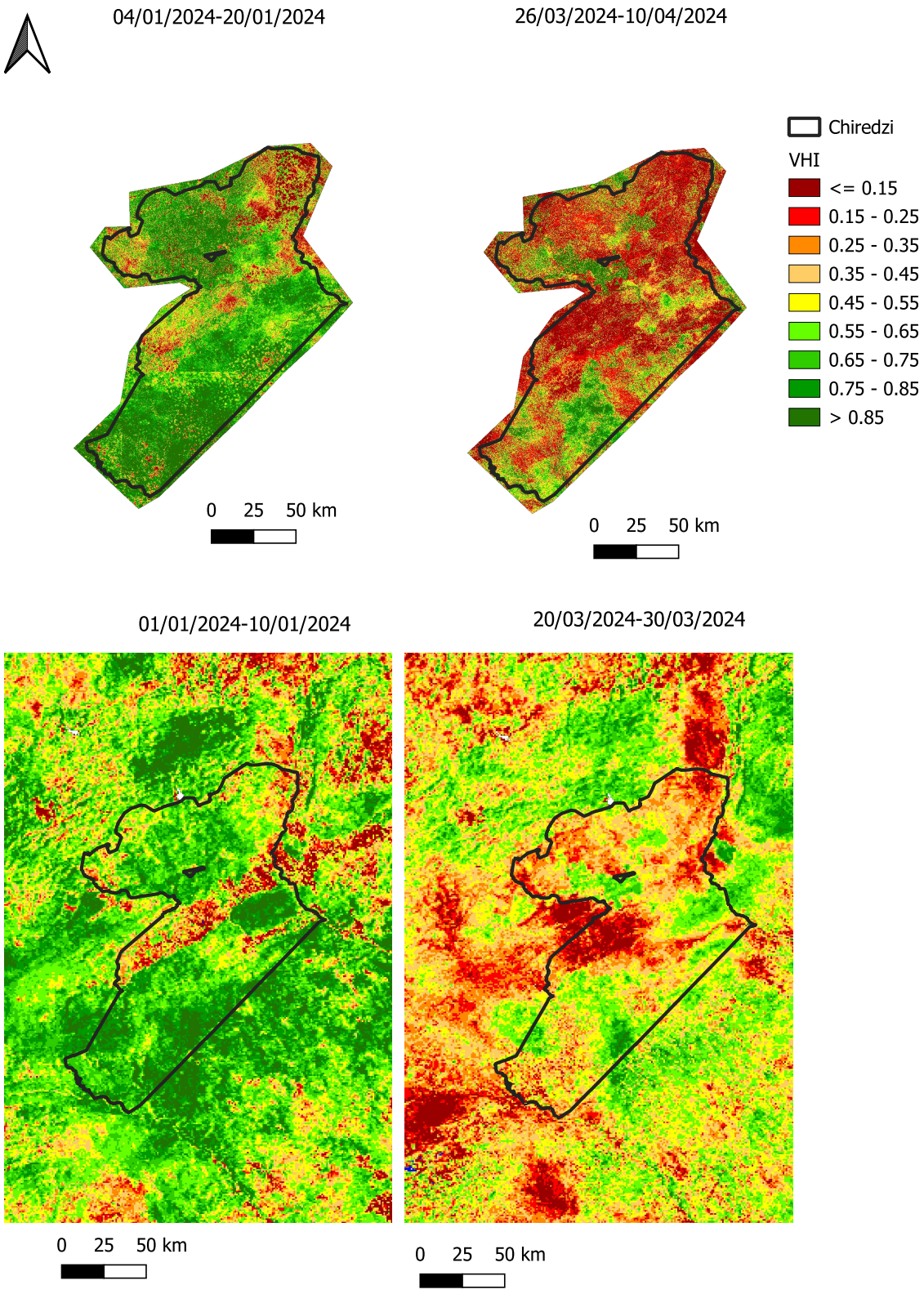


Figure 28: Comparison between the two VHIs (FAO's VHI and the high spatial resolution VHI) in January 2024, corresponding to the agricultural crop growth period, and in March 2024, which is close to the harvest period, in the Chiredzi District

Now that the consistency and reliability of the high-resolution VHI products have been assessed, the contribution of the high resolution can be confirmed. A quantitative analysis, which would have allowed for the calculation of correlations between the

FAO images and those generated in this thesis, was considered. However, the FAO VHI images are only available for visualisation, and not for manipulation or analysis. Consequently, the correlations between the two VHI datasets could not be carried out.

To make comparisons and correlations between VHI data of different resolutions, it would have been necessary to resample the VHI data from 10 meters to 1 km resolution. This process would have involved resampling the VCI to 30 meters, taking into account the PSF of the Landsat 8 sensor. Subsequently, the resolution of the resampled 30-meter VCI would have to be harmonised to a spatial resolution of 1 km, considering the specific characteristics of the AVHRR.

The complexity of resampling arises because the AVHRR, when not in nadir position (directly above the observed point), experiences a coarser resolution due to the sensor's inclination. As the sensor observes the Earth with a certain tilt relative to its central axis, it captures an inclined surface, leading to a coarser spatial resolution. The tilt limit is set at 30°, and at this inclination angle, the average resolution between the along-track and across-track spatial resolution becomes 1.4 km. With this effective resolution in mind, it would have been possible to correctly resample the 30-meter VHI to 1 km, enabling the correlation operations between the FAO products and the resampled 1 km VHI ([63]).

5.1.3 High spatial resolution for a better monitoring of drought impact on agricultural crops

Figure 29 (c) shows an area with both natural vegetation and agricultural crops, comparing two vegetation indices from January 2024: a FAO product with low spatial resolution (figure 29 (a)) and a high-resolution VHI (figure 29 (b)). The FAO product, due to its low resolution, combines natural vegetation and agricultural crops within a single pixel, making it difficult to accurately distinguish the condition of the crops. In contrast, the high-resolution VHI clearly differentiates crops from natural vegetation, which is crucial for accurately assessing the impact of drought on food security. Crops, especially in the early growth stages, have less developed roots to draw water compared to natural vegetation, making them more vulnerable to water stress. This distinction is essential for better evaluating the impact of droughts on food production, highlighting the importance of improved spatial resolution in capturing the actual condition of crops.

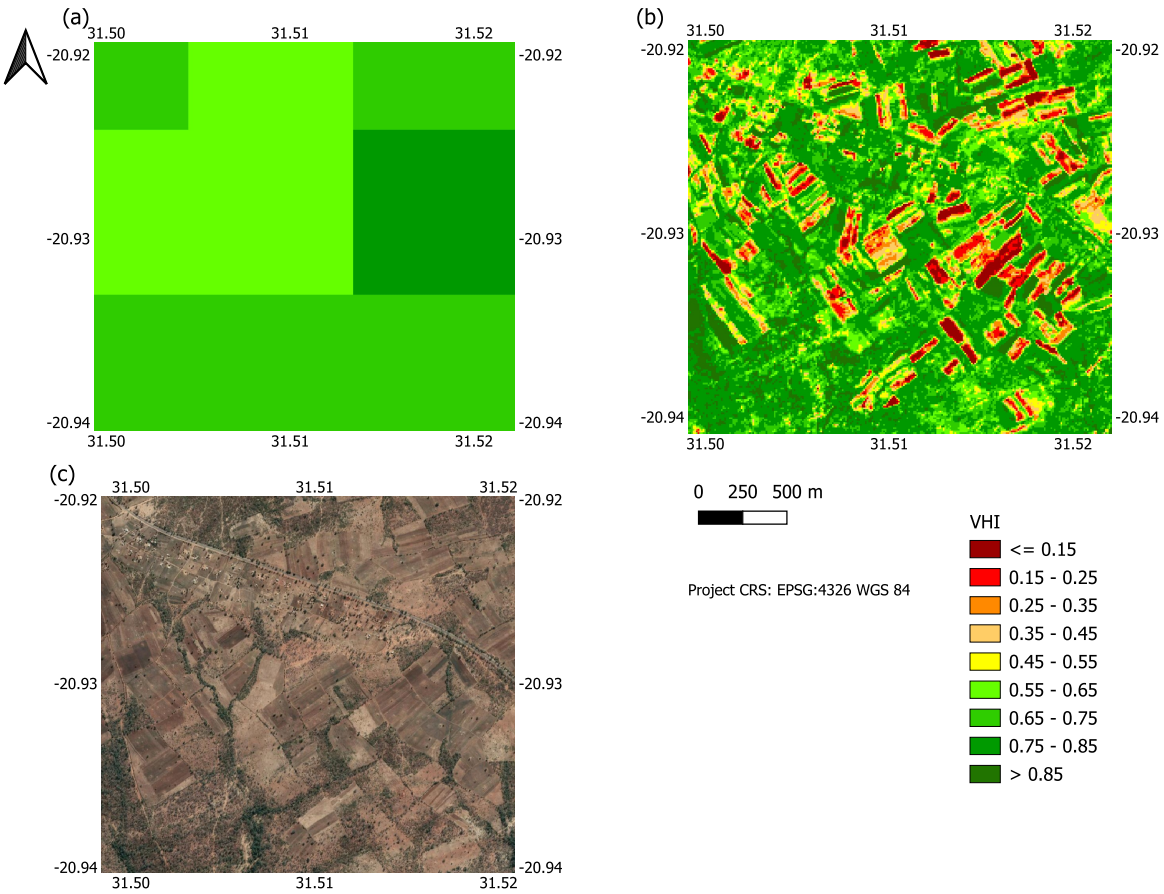


Figure 29: Comparison between the FAO's VHI and the high-resolution VHI developed in this master's thesis: (a) the FAO's VHI at a spatial resolution of 1 km from 01/01/2024 to 10/01/2024, (b) the VHI at a spatial resolution of 10 meters from 04/01/2024 to 20/01/2024, (c) a true-colour image of the area distinguishing natural vegetation from agricultural zones

5.2 Classification Results

Figure 30 (A) and (B) illustrate the accuracy of the field classification, allowing for a comparison with the corresponding true-colour image. The study area initially reveals a large green zone, corresponding to different sugarcane crops. Upon closer inspection, the algorithm's ability to distinguish between natural vegetation, roads, sugarcane, and other crop types with remarkable precision is relevant. The crop type map effectively reflects the landscape and the various types of land cover. It is worth noting that natural vegetation and roads are excluded from the cropland mask, as they are not the focus of the subsequent analysis. Furthermore, there is a clear distinction between sugarcane areas and other types of crops.

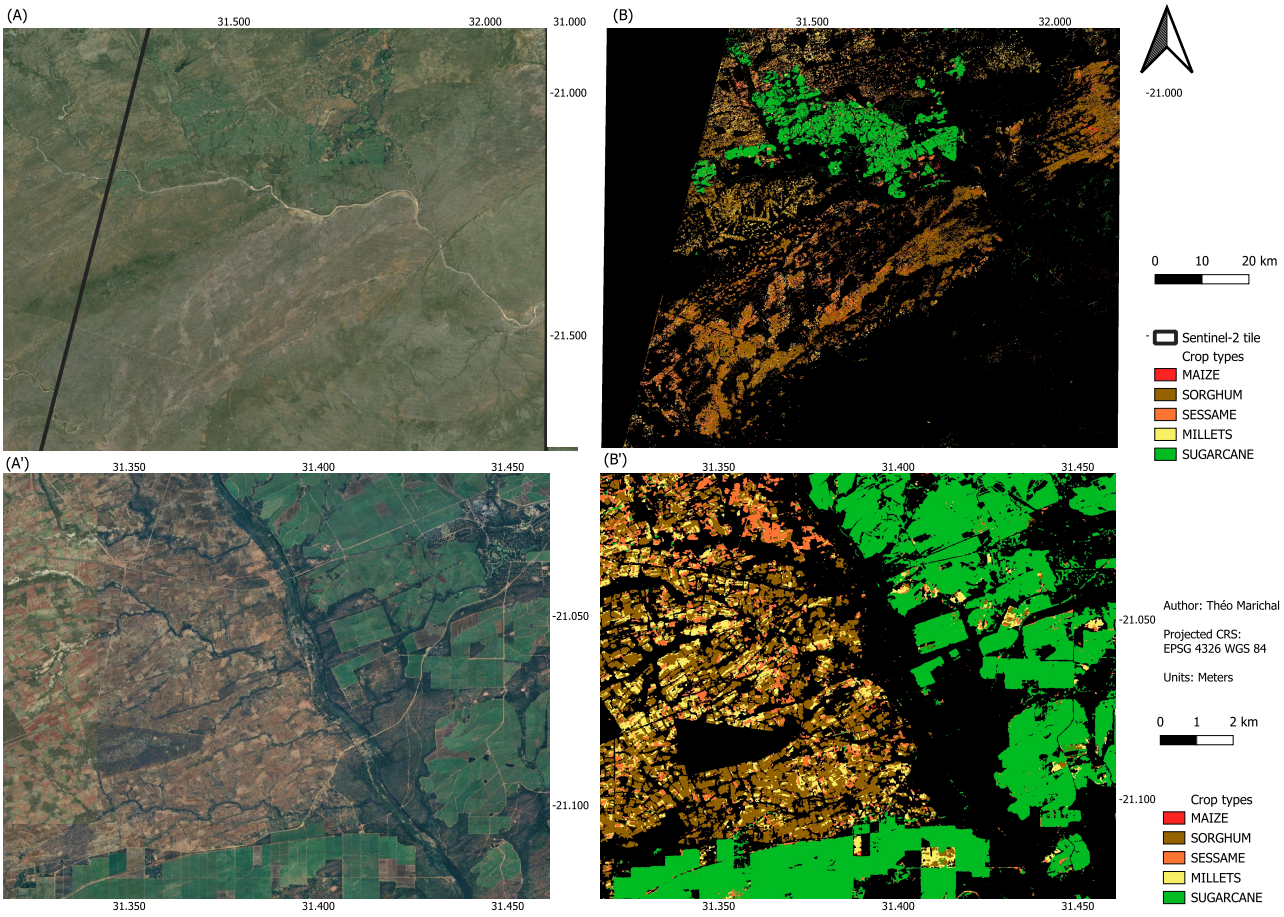


Figure 30: Crop type map of the study area at the district level for the 2023-2024 agricultural season: displayed in true color (A) and with a higher-resolution (A'), as well as with crop classification (B) and its higher resolution counterpart (B')

5.2.1 Confusion matrix and accuracy metrics

The classification performed allows for highly accurate differentiation between agricultural (crops) and non-agricultural (non-crops) surfaces. The confusion matrix (figure 31), developed using in situ validation data, demonstrates precise identification of water bodies, built-up areas, and forests. While some surfaces, such as shrubland and bare soil, may be confused with other non-agricultural classes, this does not present a significant issue since these confusions occur only among non-crop classes (figure 31). Consequently, a highly accurate map of cultivated lands can be produced.

Regarding the discrimination between different crop types, certain crops, such as sugarcane, were classified with high accuracy due to their specific spectral signatures, which facilitate their distinction by the RF algorithm. Sorghum was classified with the accuracy of 79.2%. Millet achieved the accuracy of 69%, with an excellent recall, indicating the absence of false negatives for this class. However, it is important to interpret these results cautiously due to the limited number of validation pixels. Maize, on the other hand, was often confused with sorghum, likely because of their similar spectral characteristics, and there were very few validation pixels available for maize. Lastly, sesame (categorised under "other oilseed crops") was poorly classified, with

the accuracy of only 9%.

The crop type map achieved an overall accuracy of 89.4%, largely due to the accurate classification of the best-sampled classes.



Figure 31: Classification confusion matrix



Figure 32: Classification confusion matrix

Hierarchical classification confusion matrix

Hierarchical classification did not succeed in better differentiating the crop types from each other. This can be explained by the selection of a limited number of calibration pixels (100), which was not sufficiently representative for each type of crop (figure 33). Therefore, classical classification is used for the remainder of the analysis.

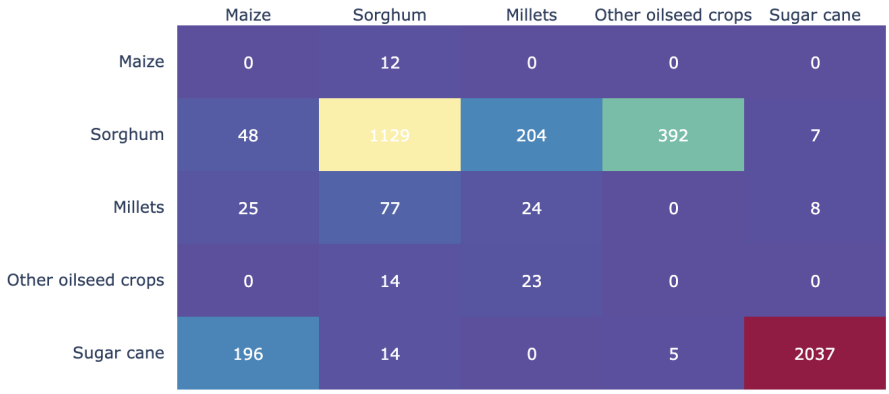


Figure 33: Hierarchical classification confusion matrix

5.2.2 Gini Feature Importance

The Gini feature importance highlights that the different spectral bands of Sentinel-2 had a greater impact on the algorithm's ability to differentiate the classes compared to the spectral indices and their metrics. Among the indices, NDWI had the most significant influence on the classification, followed by NDVI and brightness. Considering the NDVI time series (figure 24), the differences between the NDVI profiles of crops were too subtle to be captured by the algorithm or were not sufficiently highlighted.

Several insights can be derived from these results. Firstly, to enhance the utilisation of metrics from various spectral indices, it would be beneficial to consider smoothing and extrapolating the data, similar to the approach used for calculating the VHI with the Savistky-Golay filter in order to reduce the noise in the data, which can negatively affect the metrics like maximum, minimum and standard values. Secondly, the agricultural season was impacted by drought at critical moments of crop development, preventing optimal growth and good differentiation between crop types. This can affect the ability of spectral indices like NDVI, which are normally essential for classification, to effectively discriminate between crops.

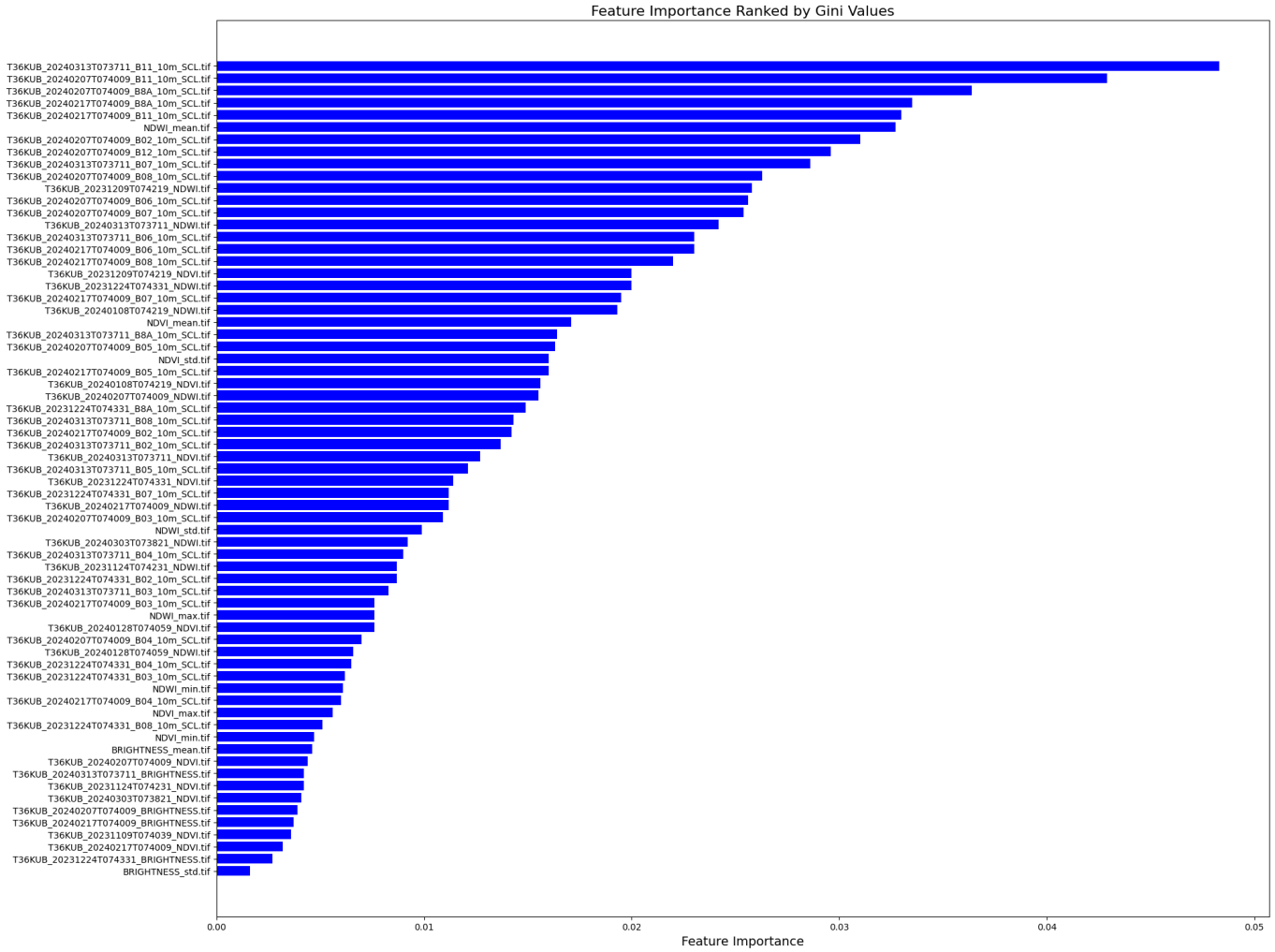


Figure 34: Gini index

5.3 Integration of the VHI time series and the crop type map

The vegetation health of all crop fields in Chiredzi District is first assessed. Following this, the application of the VHI to specific fields is examined in detail. For this purpose, fields of sorghum and sugarcane have been randomly selected for focused analysis.

5.3.1 Assessing the VHI by crop type

To evaluate the health distribution of various crop types across the territory, a VHI analysis is conducted throughout the season. In this analysis, we will assess the data at different levels. We will start by examining the in-situ data points collected by the FAO. Next, we will analyse polygons that were manually digitised as part of this thesis, based on the provided in-situ points (section 4.3.1). Finally, we will evaluate the results derived from the classification data (section 5.2).

Analysis of sugarcane fields

The analysis first focused on sugarcane fields, which are particularly well classified. Initially, in situ data were analysed (see figure 35). These data indicate that sugarcane is in relatively good health, with a gradual improvement until December, followed by a steady decline. Overall, the median VHI fluctuates between 0.6 and 0.8 throughout the season.

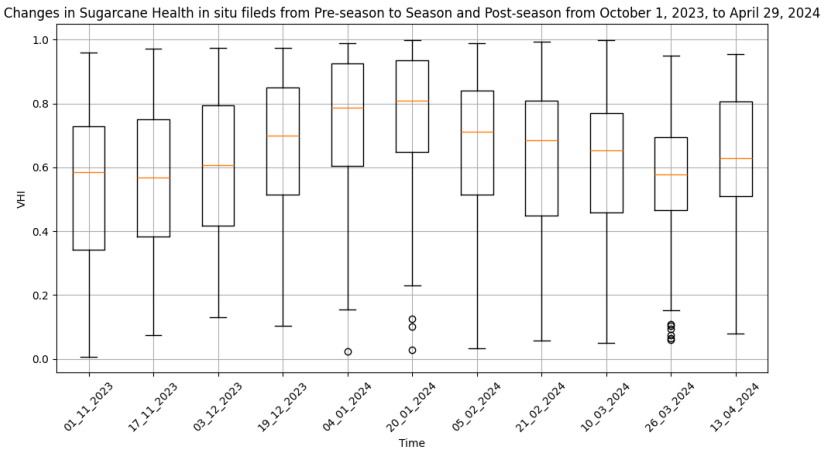


Figure 35: VHI time series of in situ points of Chiredzi

When doing the same analysis using in-situ data polygons instead of points, the boxplot highlights, the boxplot (figure 36) highlights a similar trend in VHI data over time. However, a notable difference lies in the increase in the variation of the VHI data distribution. This is logically explained by a larger quantity of analysed data, leading to greater diversity in environmental conditions and increased natural variability.

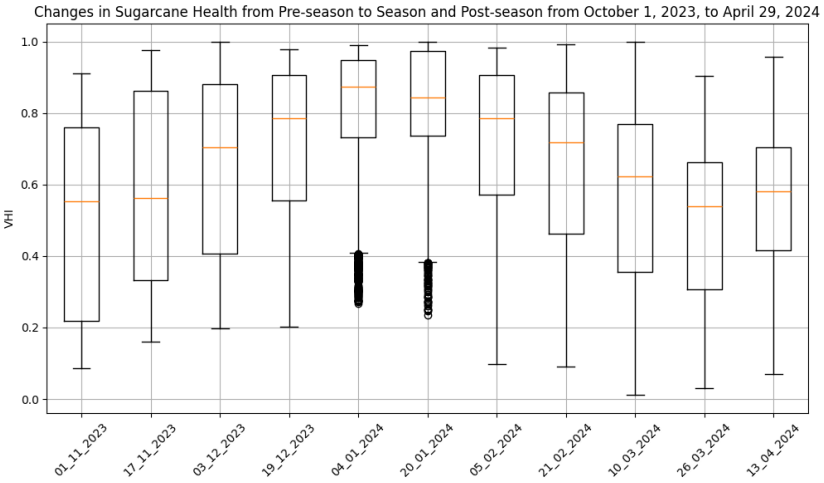


Figure 36: VHI time series of in situ polygons of Chiredzi

Finally, a similar analysis was performed on the entire set of reclassified fields. A comparable trend in median VHI values was observed 37. The gap between the 75th and 25th percentiles becomes more regular and shorter, indicating a better representation of the majority of the sugarcane fields. However, the variability, beyond these

percentiles, increases for the same reasons as explained above.

In general, sugarcane is not particularly affected by drought. A reason could be that these crops are grown intensively and irrigated to mitigate the effects of drought. This will be further analysed in section 5.3.2.

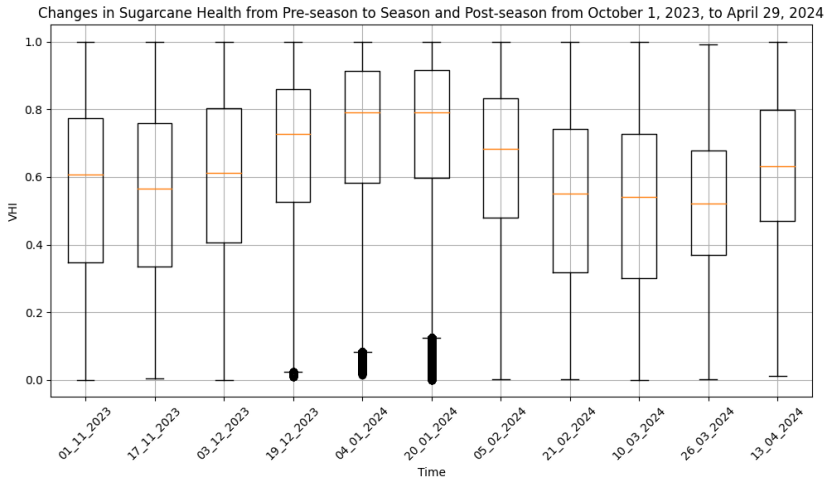


Figure 37: VHI time series of sugarcane classified fields in Chiredzi

Analysis of sorghum Fields

A similar analysis was performed over sorghum. In this case, we can draw the same conclusions regarding the increase in VHI variability when transitioning from in situ data to reclassified data (see figures 38, 39, 40). Additionally, a similarity is observed in the median trend of sorghum health condition through the transition from in situ fields to classification fields. However, this median indicates that the sorghum fields have been severely affected by drought. From the end of January, there is a noticeable decline in the condition of the sorghum fields. This poor condition occurs at crucial times for crop development, signalling a significant risk of impact on sorghum field production. This is further exacerbated by the ongoing drought, preventing the crops from recovering.

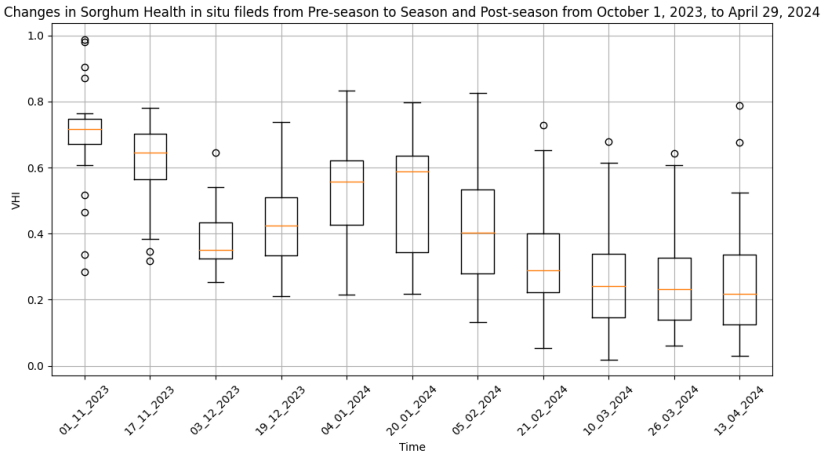


Figure 38: VHI time series of in situ points of Chiredzi

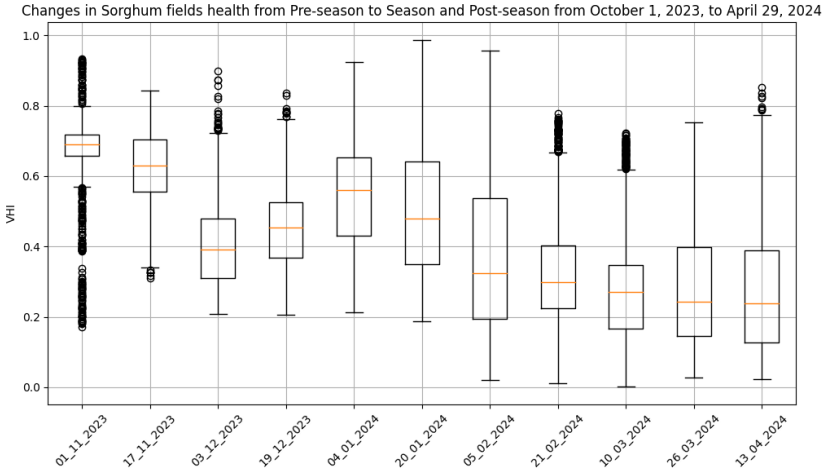


Figure 39: VHI time series of in situ polygons of Chiredzi

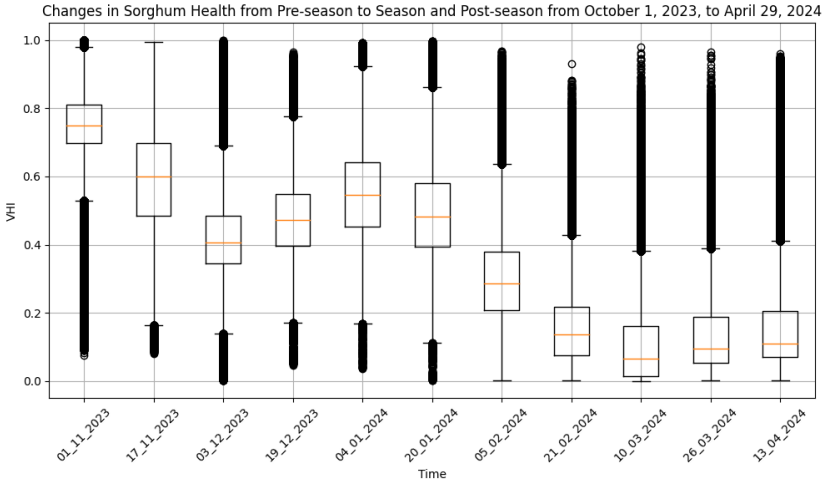


Figure 40: VHI time series of in situ reclassified sorghum fields in Chiredzi

Analysis of other crop types fields

The same analysis was conducted for the other crop types, but only based on in situ data because of the low quality of the classification for these classes. Thus, it is observable that the crop least affected by the drought is sesame (see figure 41), followed by millets (figure 42), and finally maize, which is the most affected by the drought (figure 43).

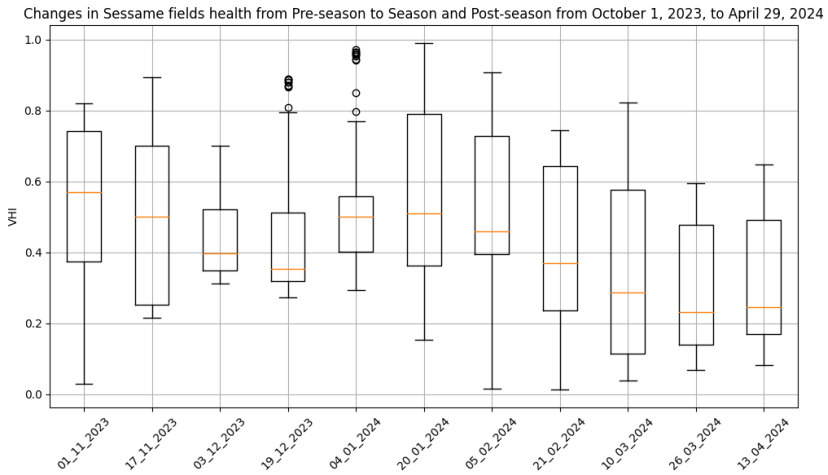


Figure 41: Evolution of vegetation health of sesame fields across the 2023-2024 agricultural season

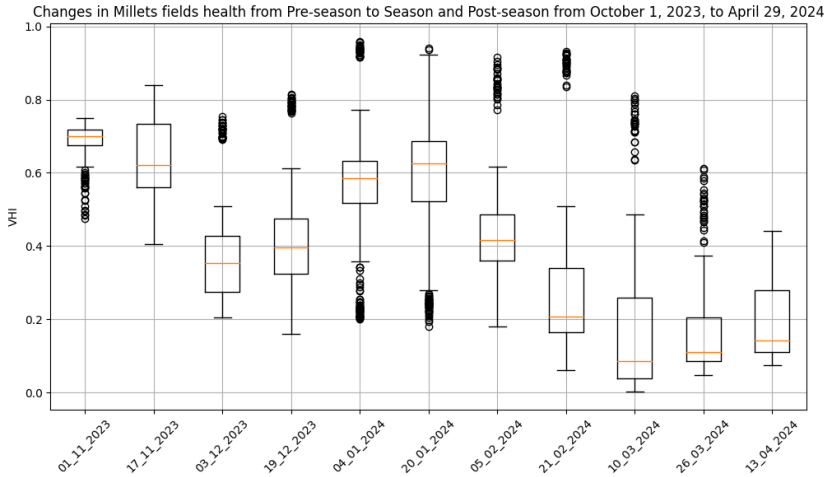


Figure 42: Evolution of vegetation health of millets fields across the 2023-2024 agricultural season

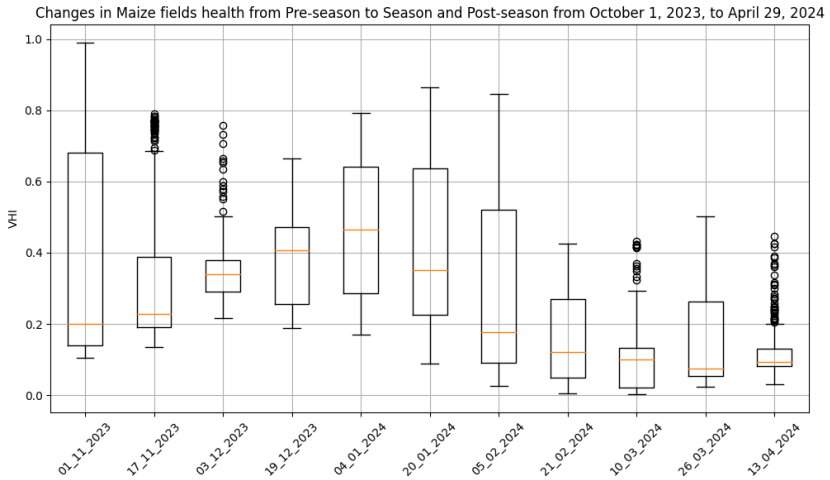


Figure 43: Evolution of vegetation health of maize fields across the 2023-2024 agricultural season

5.3.2 Assessing the health of sugarcane at the field scale on targeted parcels

The selected sugarcane fields are illustrated in figure 44.

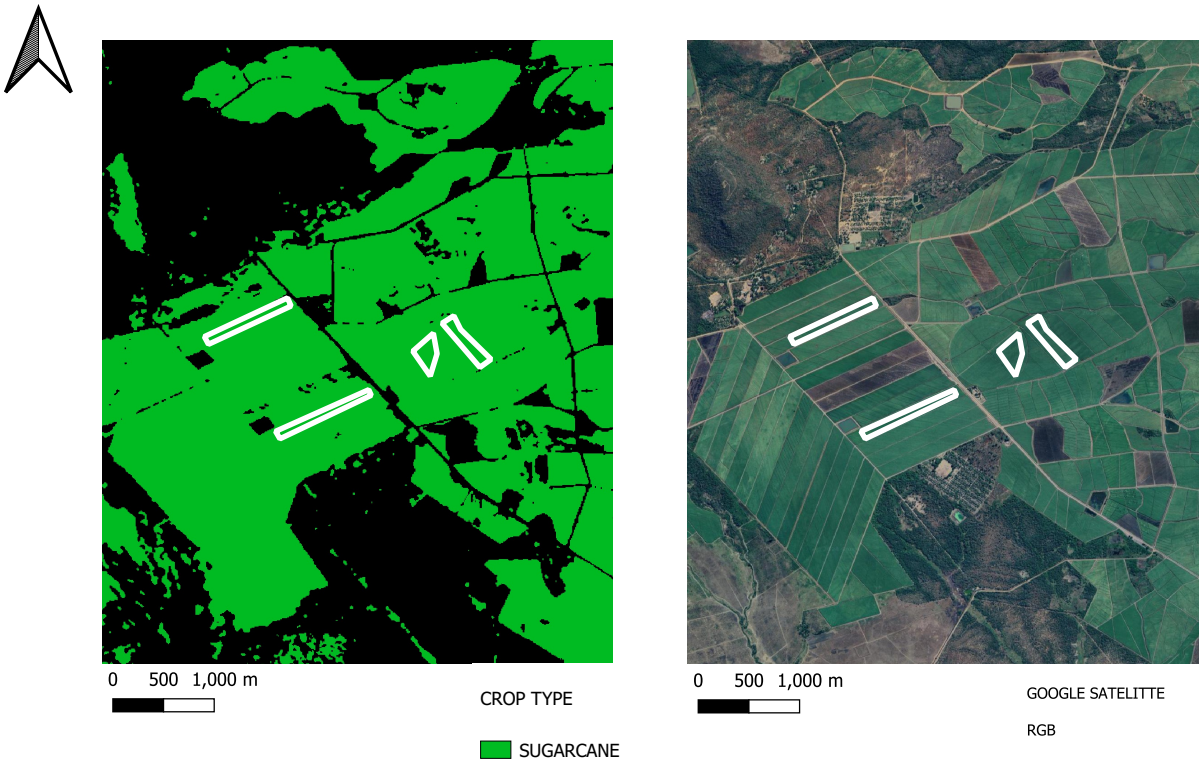


Figure 44: Randomly selected sugarcane fields

The analysis of the VHI indicates that the sugarcane fields are in good health (figure 45). Moreover, the similarity in behaviour among these fields indicates relatively uniform agricultural practices. The studied fields are also representative of the general trend observed for sugarcane across the Chiredzi territory, implying that the results can be generalised to other fields in the region, as established in the previous section 5.3.1.

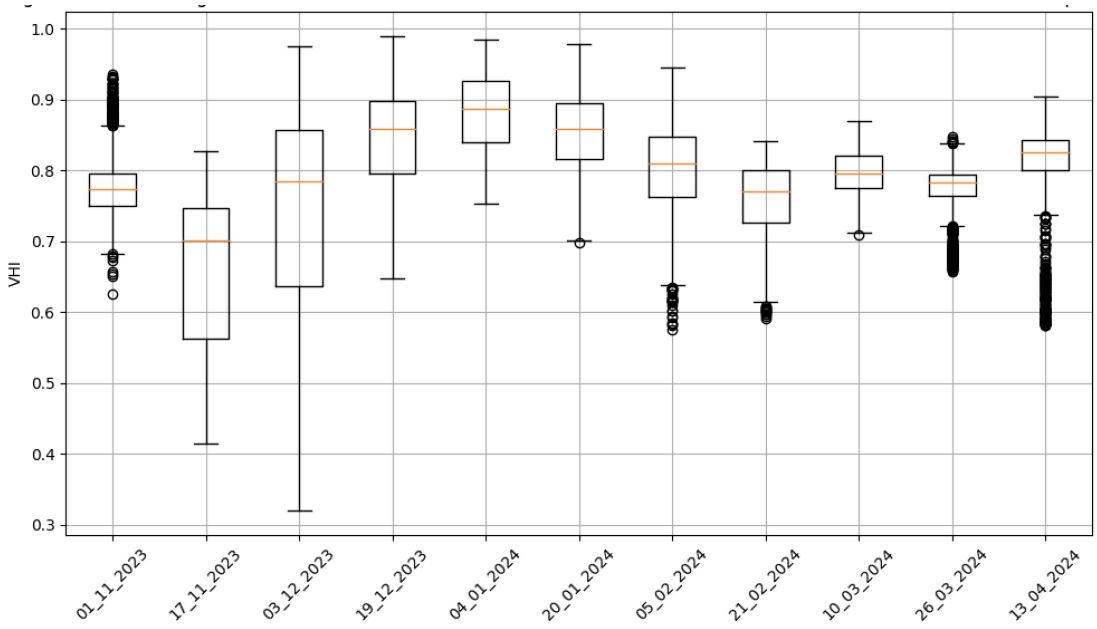


Figure 45: Evolution of the VHI of intensive sugarcane fields from October 1, 2023, to April 29, 2024.

The added value of this study lies in the analysis of NDVI profiles, which were generated using the average NDVI of the four delineated fields shown in Figure 44. These profiles exhibit remarkable similarity, with a consistent pattern over time: a harvest in August results in a significant drop in NDVI, followed by a gradual recovery, reaching values between 0.75 and 0.85. The 2023-2024 season does not appear to be impacted by drought (Figure 46). This resilience to drought, with no observable dips in the NDVI profile, is likely due to the implementation of intensive and irrigated agriculture. This hypothesis is supported by the structure of the sugarcane plots (Figure 44), which consist of large rectangular and pivot-shaped fields, optimised for mechanised, intensive cultivation with efficient pivot irrigation systems.

Furthermore, by examining the monthly precipitation time series (figure 47), it can be observed that precipitation levels drop significantly starting in January, reaching as low as 18.544 mm in February. However, this decrease in precipitation is not reflected in the NDVI profile (figure 47), which further supports the likelihood of irrigation being used in these sugarcane fields.

Therefore, it is the agricultural practices, rather than drought conditions, that influence the NDVI profiles of these sugarcane fields. Since these fields are irrigated, this aspect is not central to our study on drought monitoring and will therefore not be further analysed.

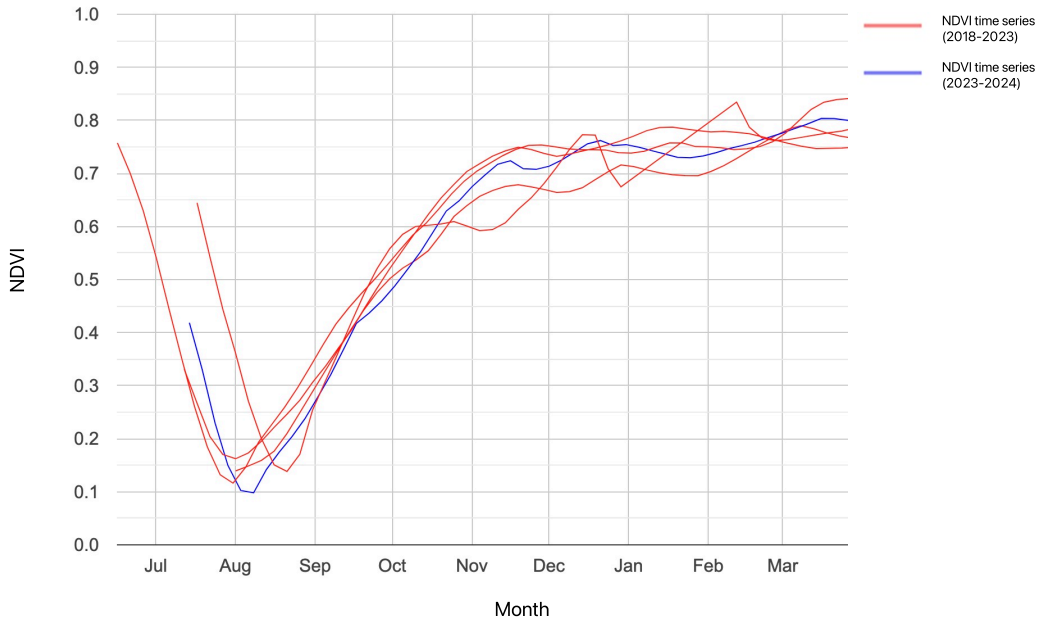


Figure 46: Comparative study of the 2023-2024 season NDVI time series with the time series from previous seasons for randomly selected sugarcane fields.

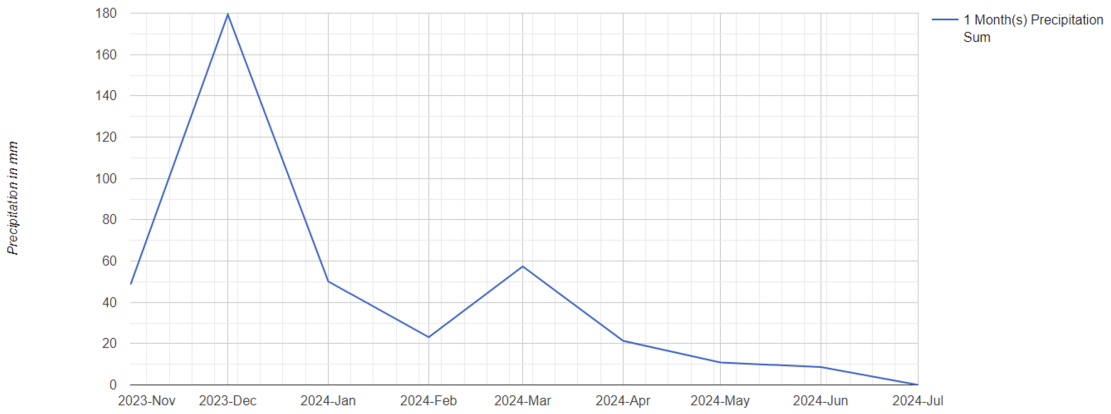


Figure 47: Monthly precipitation time series for the 2023-2024 agricultural season, based on CHIRPS data, covering a 5 km^2 area that includes randomly selected sugarcane fields (figure 45)

5.3.3 Assessing the health of sorghum at the field scale on randomly selected fields

To better understand the impact of droughts on rain-dependent crops, which are more vulnerable to extreme weather, sorghum fields were selected in a region suitable for extensive agriculture. This selection focused on areas where fields are randomly dispersed across a varied landscape, including different types of soil cover such as forests and shrublands (figure 48).

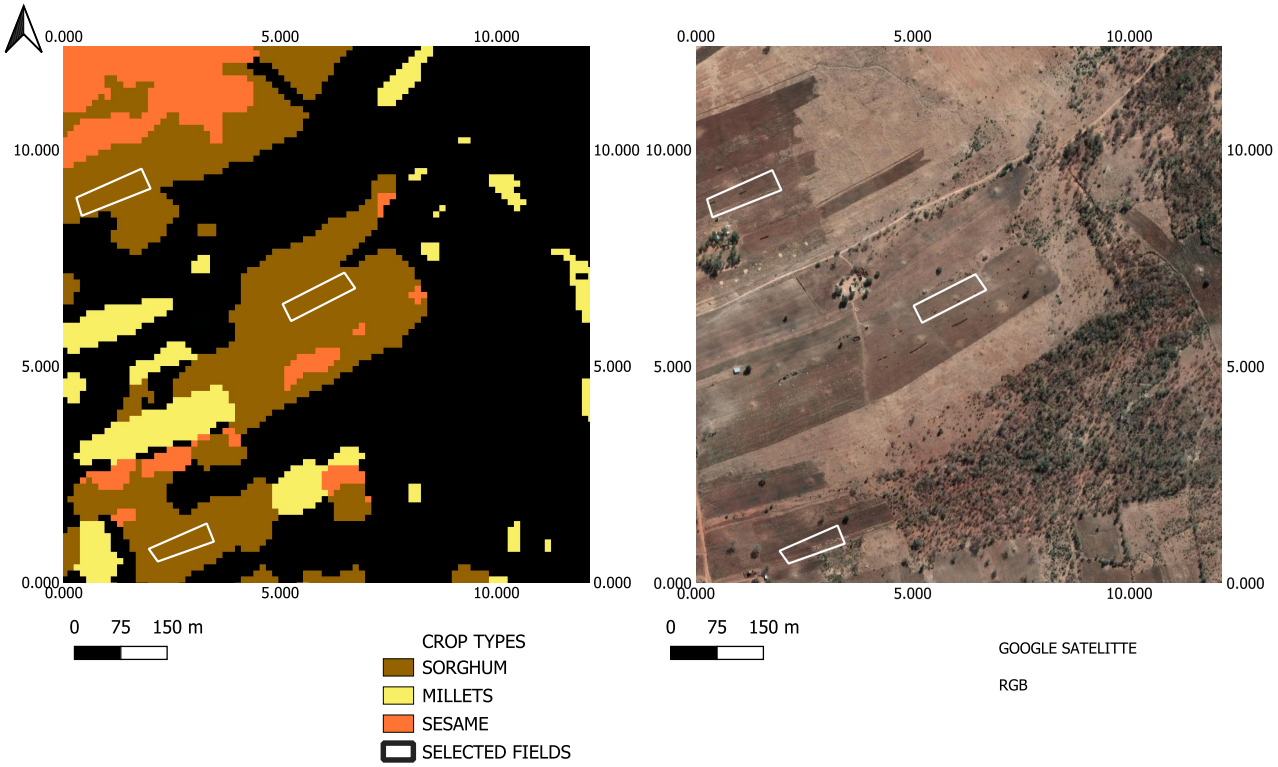


Figure 48: Randomly selected extensive sorghum fields in Chiredzi

Monitoring changes in VHI throughout the season across the selected sorghum fields

The figure (49) illustrates the changes in the VHI over time for selected fields. High VHI values (in green) indicate healthy vegetation, while low values (in yellow to red) suggest poorer vegetation health. Significant variations in VHI are observed throughout the agricultural season. For instance, on December 19, 2023, some fields display high VHI values while others show lower values. By March 10, 2024, the VHI values become more extreme, partly due to a prolonged drought that affected Zimbabwe from late January to mid-March. This drought occurred during a critical period for crop growth, likely impacting vegetation health as reflected by the VHI.

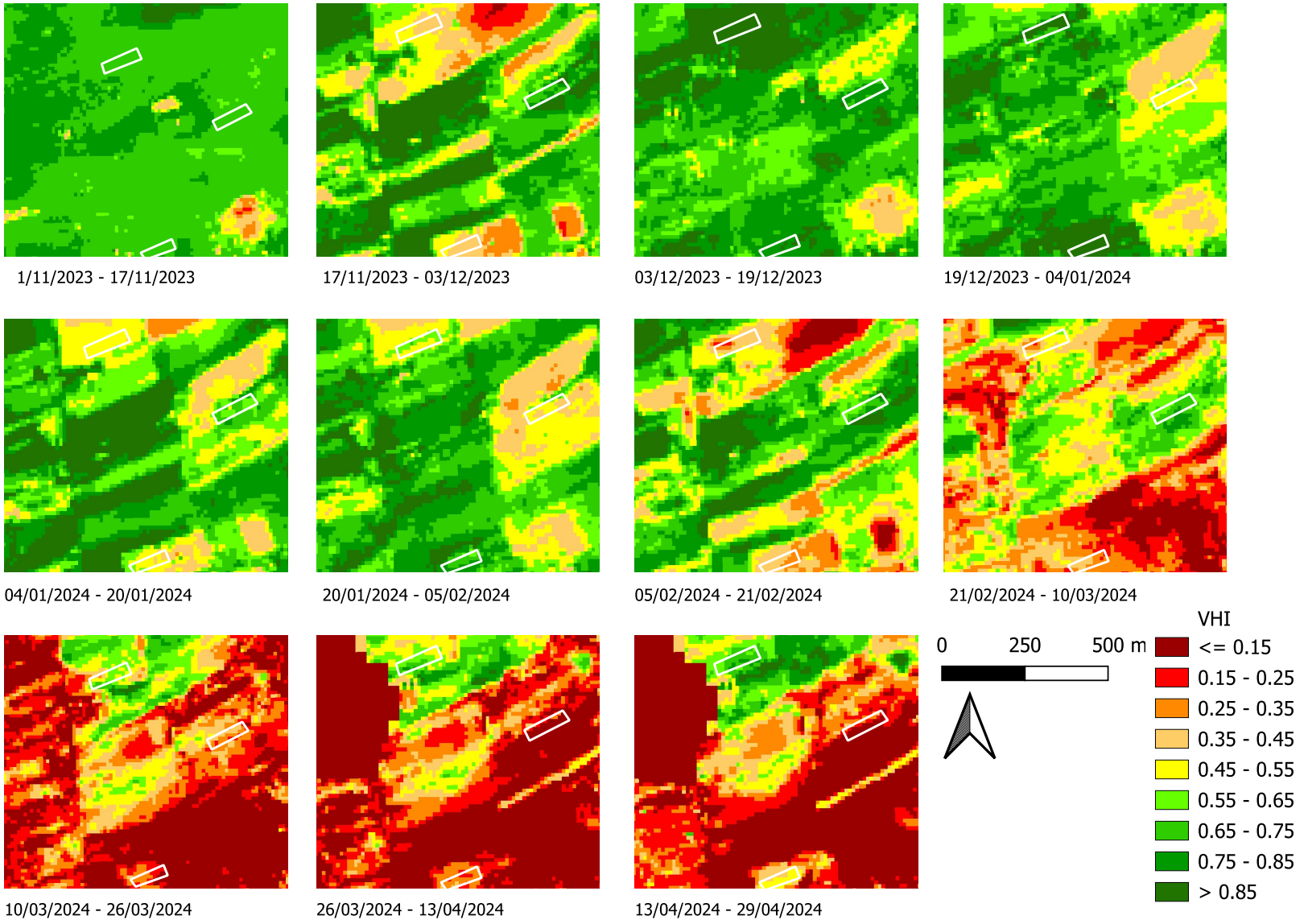


Figure 49: Evolution of VHI of randomly selected sorghum fields throughout the 2023-2024 agricultural season

Tracking the evolution of VHI for a specific sorghum field during the 2023-2024 agricultural season using NDVI and LST time series data

The evolution of VHI values throughout the season in sorghum fields shows significant variation due to different responses to drought among the fields (figure 49). Consequently, it is essential to focus on a single field to analyse the progression of VHI throughout the 2023-2024 season in order to better understand the VHI calculation mechanism. A detailed analysis is conducted by comparing the NDVI and LST time series of the 2023-2024 season with those from previous seasons (available since 2018) for a specific sorghum field, as illustrated in figure 50 (a) and (b). This comparison between the current season's profiles and those of previous seasons allows for a better understanding of the underlying mechanisms of the VCI and TCI.

For figure 50 (a), the higher the NDVI curve compared to the curves from previous years, the denser and healthier the vegetation is relative to other years, resulting in a higher VCI. The VCI is obtained by calculating the average NDVI for each 16-day period and scaling it between the maximum and minimum values for that 16-day period. It is important to note that the maximum and minimum values of the time series are no more likely to be biased due to the smoothing and gap-filling techniques applied to these data.

In contrast to the NDVI time series, higher LST values during the agricultural season indicate less favourable conditions for vegetation, as shown in figure 50 (b). Consequently, the higher the LST during the 2023-2024 agricultural season compared to other years, the higher the TCI will be. The methodology for scaling between the maximum and minimum values of the TCI is the same as for the VCI.

With an understanding of how these NDVI and LST profiles contribute to the derivation of the TCI and VCI indices, the VHI can be calculated by performing a weighted linear combination of these two indices. The evolution of the VHI for 11 periods of 16 days during the agricultural season in the sorghum field is illustrated in figure 50 (c). This analysis allows for a better understanding of the seasonal dynamics of vegetation health in the context of climatic variability.

Examining the VHI images alongside the NDVI and LST time series shows key patterns for the 2023-2024 season (figure 50). In November, the NDVI jump, indicating dense and healthy vegetation. The average LST from November 1st to 17th is relatively low, creating favourable conditions. This results in high VHI values. In December, the NDVI declines, signalling weaker vegetation health. This could have reduced VHI values significantly, but the low LST helps to mitigate the impact. In January, the NDVI increases rapidly, showing a quick improvement in vegetation health. The LST remains low, so VHI values stay high. From March onwards, VHI values decrease. This drop is due to a higher LST compared to previous years and a lower

NDVI. These factors suggest that climatic conditions are becoming less favourable for vegetation growth.
However, it is important to note that the NDVI profiles are not synchronised in their growth cycles, and seasonal shifts are present. At high spatial resolution, this must be taken into account to achieve a more robust VHI (figure 50 (a)).

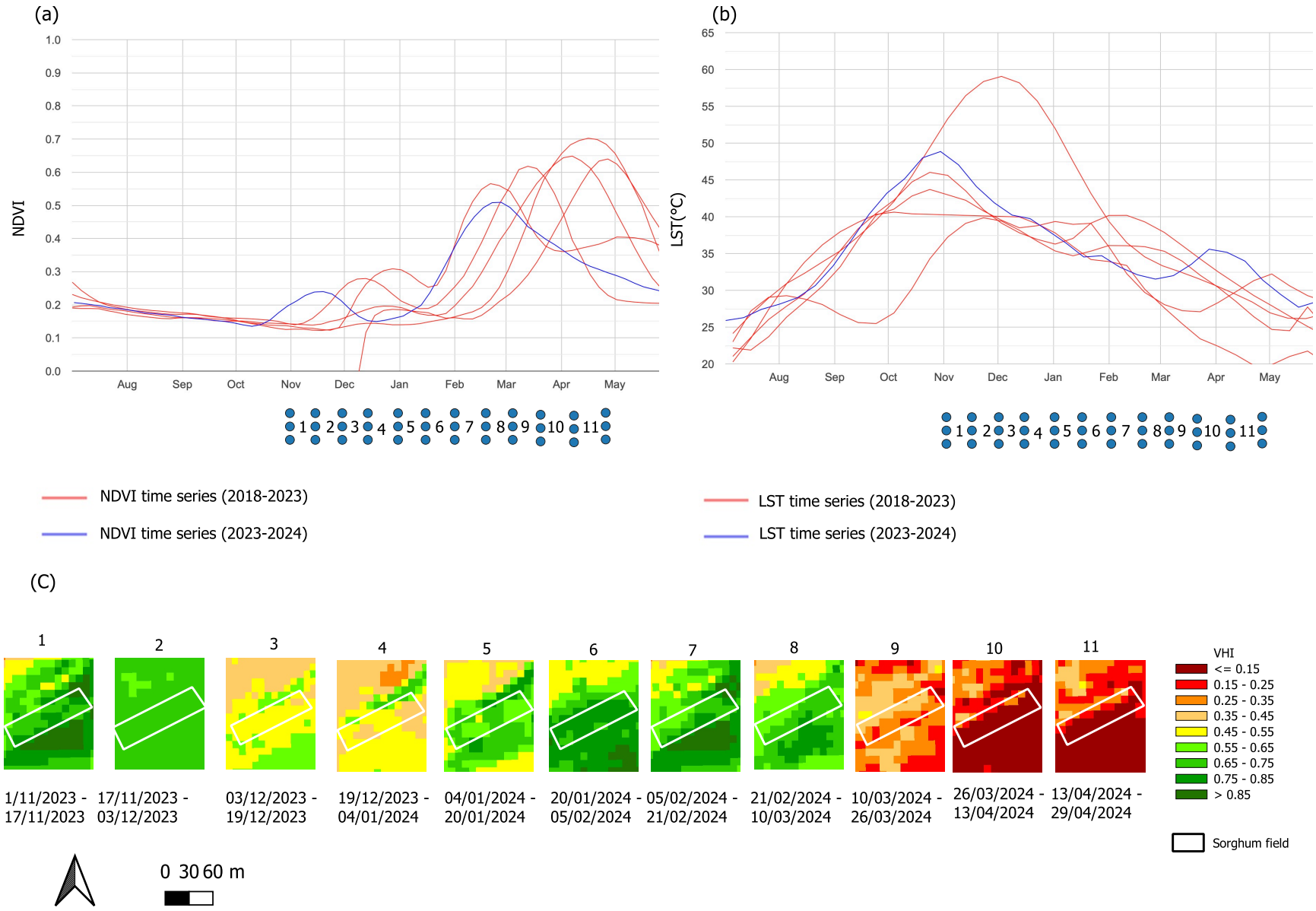


Figure 50: Analysis and explanation of the evolution of VHI values (distributed over 11 periods of 16 days) for a sorghum field during the 2023-2024 agricultural season, using a comparative study of the 2023-2024 season NDVI and LST time series with the time series from previous seasons

Establishing a robust, high-resolution VHI for a specific sorghum field, taking into account seasonal variations

The method explained in Section 4.2.9 successfully eliminated the seasonal shifts in the data, facilitating the alignment of crop development phases across multiple years and thereby allowing for more accurate comparisons.

Figure 51(a) illustrates that the NDVI curves are now synchronized with each other, which was not the case before these adjustments. As a result, crop development phases are now comparable from one year to the next.

The analysis of NDVI and LST time series for the 2023-2024 season reveals significant differences between the VHI corrected for seasonal shifts and the uncorrected VHI. For instance, in November, although the NDVI shows an increase, the corrected VHI remains high but is lower than the uncalibrated VHI. This demonstrates that adjusting for seasonal shifts allows for more accurate data interpretation. From December onwards, the corrected VHI exhibits very low values throughout the season. This phenomenon is explained by a lower NDVI and higher LST compared to other years, indicating less favorable climatic conditions. These variations would have been difficult to discern without the seasonal shift correction.

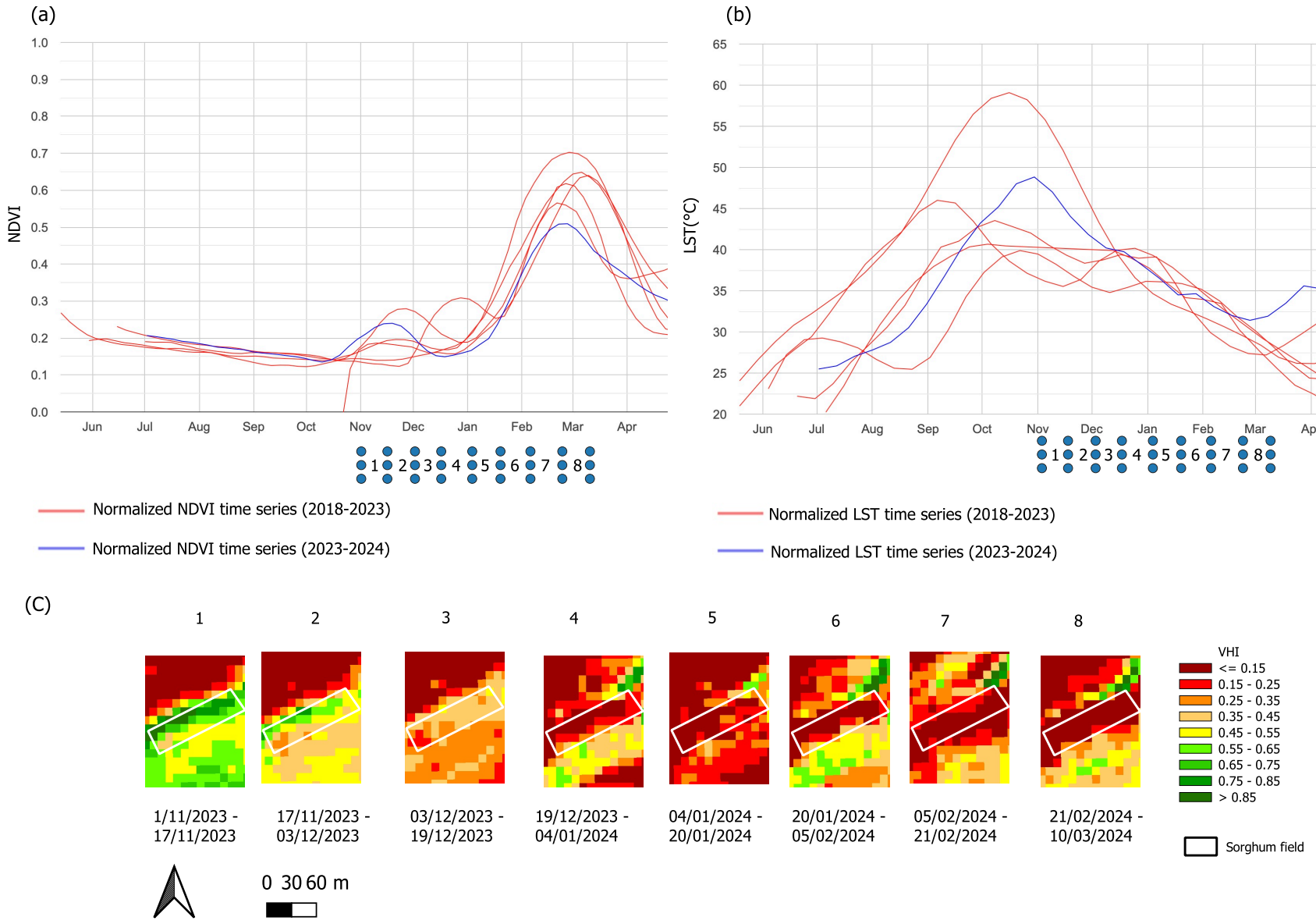


Figure 51: Analysis and explanation of the evolution of VHI values for a sorghum field during the 2023-2024 agricultural season, using a comparative study of the 2023-2024 season NDVI and LST time series with the time series from previous seasons.

5.3.4 General discussion

This study has demonstrated the significant potential of high-resolution spatial data in enhancing drought monitoring and response strategies, particularly in agriculturally vulnerable regions like Chiredzi, Zimbabwe. By utilizing the VHI at a finer spatial resolution, we were able to achieve a more precise assessment of the impacts of drought on various crop types at the field level and down to a detail of 10 meters. This precision is crucial for guiding policymakers in developing targeted interventions that can enhance the resilience of different crop types to drought.

One of the key contributions of this study is the demonstration that existing EWS can be effectively upgraded with high-resolution data. The technology allows for a combined approach using Sentinel-2 and Landsat data to provide detailed insights necessary for distinguishing between different crop types and even different agricultural practices, such as intensive irrigated versus extensive non-irrigated farming.

The high-resolution VHI tool developed in this study provides a valuable mechanism for identifying the location, extent, intensity, and duration of drought periods with unprecedented detail. By assigning weights to the VHI based on the crop growth cycle and specific crop coefficients, the tool allows for a more accurate reflection of water stress impacts. For example, during critical phenological phases such as flowering or grain filling in cereals, the VHI anomalies can be weighted more heavily, offering a more realistic assessment of drought's impact on crop yields.

However, the study also highlighted some limitations that need to be addressed in future work. One of the main challenges is the necessity to account for crop rotation and seasonal shifts at the pixel level, which can complicate data storage and processing. These factors, if not carefully managed, could introduce biases or inaccuracies in the VHI calculations.

The study also underscores the critical importance of having sufficient quality and quantity of in situ data for classification purposes. In this study, the lack of adequate in situ data for certain crop classes limited the accuracy of classification. This highlights the need for more comprehensive and systematic data collection efforts to support the full potential of high-resolution drought monitoring tools.

Chapter 6

Conclusion

In conclusion, this study highlights the critical importance of high-resolution spatial data for assessing the impacts of drought on agricultural crops, particularly in vulnerable regions like Chiredzi, Zimbabwe. By using indicators such as the VHI with fine spatial resolution, it becomes possible to precisely distinguish cultivated areas and better understand the effects of drought at the field level, down to a detail of 10 meters.

Integrating crop type maps with vegetation health indices is essential for accurate drought monitoring. This combination allows for the localization of different crops and analysis of their specific responses to drought conditions, ensuring that high-resolution data is accurately interpreted.

The findings demonstrate that while global drought monitoring models are useful, they have limitations at the local level, especially in areas with high spatial variability. Without sufficient resolution to distinguish different types of land cover, these models can become biased. In contrast, high-resolution indices capture fine-scale dynamics, revealing patterns of water stress that would often be missed with coarser resolutions.

Adjusting the VHI to the agricultural calendar helps better align crop development phases over the years, enabling more precise comparisons of vegetation health. The VHI also aids in anticipating critical periods of water stress and identifying when crops are most vulnerable.

The study also underscores the influence of agricultural practices, such as irrigation, which can be detected by unexpectedly high VHI values, particularly when cross-referenced with precipitation data.

The implications of this research are broad, offering a methodology that can be adapted to other regions for VHI use and contributing to the improvement of early warning systems for food security.

Bibliography

[1] Elisa Savelli et al. "Drought and society: Scientific progress, blind spots, and future prospects". In: *Wiley Interdisciplinary Reviews. Climate Change* 13.3 (2022), e761. ISSN: 1757-7780. DOI: 10.1002/wcc.761. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9286479/.

[2] Anne F. Van Loon et al. "Drought in the Anthropocene". In: *Nature Geoscience* 9.2 (Feb. 2016), pp. 89–91. ISSN: 1752-0894, 1752-0908. DOI: 10.1038/ngeo2646. URL: https://www.nature.com/articles/ngeo2646.

[3] Michael Hagenlocher et al. "Drought vulnerability and risk assessments: state of the art, persistent gaps, and research agenda". In: *Environmental Research Letters* 14.8 (Aug. 1, 2019), p. 083002. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab225d. URL: https://iopscience.iop.org/article/10.1088/1748-9326/ab225d.

[4] Ashok K. Mishra, Vijay P. Singh. "A review of drought concepts". In: *Journal of Hydrology* 391.1 (Sept. 2010), pp. 202–216. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2010.07.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022169410004257.

[5] Fernando Maliti Chivangulula, Malik Amraoui, Mário Gonzalez Pereira. "The Drought Regime in Southern Africa: A Systematic Review". In: *Climate* 11.7 (July 2023). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 147. ISSN: 2225-1154. DOI: 10.3390/cli11070147. URL: https://www.mdpi.com/2225-1154/11/7/147.

[6] NASA. *El Niño: Pacific Wind and Current Changes Bring Warm, Wild Weather*. Publisher: NASA Earth Observatory. Feb. 14, 2017. URL: https://earthobservatory.nasa.gov/features/ElNino.

[7] Hugo Carrão, Gustavo Naumann, Paulo Barbosa. "Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability". In: *Global Environmental Change* 39 (July 2016), pp. 108–124. ISSN: 09593780. DOI: 10.1016/j.gloenvcha.2016.04.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959378016300565.

[8] Janna Frischen et al. "Drought Risk to Agricultural Systems in Zimbabwe: A Spatial Analysis of Hazard, Exposure, and Vulnerability". In: *Sustainability* 12.3 (Jan. 2020). Number: 3 Publisher: Multidisciplinary Digital Publishing

Institute, p. 752. ISSN: 2071-1050. DOI: 10.3390/su12030752. URL: https://www.mdpi.com/2071-1050/12/3/752.

[9] Johannes Belle, Moyo Sithabile, Abiodun A. Ogundeji. "Assessing communal farmers'preparedness to drought in the Umguza District, Zimbabwe". In: *International Journal of Disaster Risk Reduction* 22 (June 2017), pp. 194–203. ISSN: 22124209. DOI: 10.1016/j.ijdrr.2017.03.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S2212420916306689.

[10] NASA. *Severe Drought in Southern Africa*. Publisher: NASA Earth Observatory. Apr. 22, 2024. URL: https://earthobservatory.nasa.gov/images/152711/severe-drought-in-southern-africa.

[11] Harry West, Nevil Quinn, Michael Horswell. "Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities". In: *Remote Sensing of Environment* 232 (Oct. 2019), p. 111291. ISSN: 00344257. DOI: 10.1016/j.rse.2019.111291. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425719303104.

[12] Ademola Braimoh et al. *Assessment of Food Security Early Warning Systems for East and Southern Africa*. World Bank, Washington, DC, Jan. 2018. DOI: 10.1596/29269. URL: http://hdl.handle.net/10986/29269.

[13] David Hunt. "Introduction to Remote Sensing". In: ().

[14] NASA. *ARSET - Remote Sensing of Drought | NASA Applied Sciences*. July 12, 2017. URL: http://appliedsciences.nasa.gov/get-involved/training/english/arset-remote-sensing-drought.

[15] Godfrey Mutowo, David Chikodzi. "Remote sensing based drought monitoring in Zimbabwe". In: *Disaster Prevention and Management* 23.5 (Oct. 28, 2014), pp. 649–659. ISSN: 0965-3562. DOI: 10.1108/DPM-10-2013-0181. URL: https://www.emerald.com/insight/content/doi/10.1108/DPM-10-2013-0181/full/html.

[16] Adam L. Mahood et al. "Ten simple rules for working with high resolution remote sensing data". In: *Peer Community Journal* 3 (2023). ISSN: 2804-3871. DOI: 10.24072/pcjournal.223. URL: https://peercommunityjournal.org/articles/10.24072/pcjournal.223/.

[17] Meteosat. *Meteosat series | EUMETSAT*. Apr. 15, 2020. URL: https://www.eumetsat.int/our-satellites/meteosat-series.

[18] Soroosh Sorooshian et al. "Advanced Concepts on Remote Sensing of Precipitation at Multiple Scales". In: *Bulletin of the American Meteorological Society* 92.10 (2011). Publisher: American Meteorological Society, pp. 1353–1357. ISSN: 0003-0007. URL: https://www.jstor.org/stable/26218594.

[19] NASA. *GPM - NASA Science*. URL: https://science.nasa.gov/mission/gpm/.

[20] CHIRPS. *CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara*. URL: https://www.chc.ucsb.edu/data/chirps.

[21] Thomas B McKee, Nolan J Doesken, John Kleist. "THE RELATIONSHIP OF DROUGHT FREQUENCY AND DURATION TO TIME SCALES". In: (1993).

[22] A. AghaKouchak et al. "Remote sensing of drought: Progress, challenges and opportunities". In: *Reviews of Geophysics* 53.2 (2015). _eprint: https://onlinelibrary.wiley.com pp. 452–480. ISSN: 1944-9208. DOI: 10.1002/2014RG000456. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000456.

[23] Ikhlas Ghiat, Hamish R. Mackey, Tareq Al-Ansari. "A Review of Evapotranspiration Measurement Models, Techniques and Methods for Open and Closed Agricultural Field Applications". In: *Water* 13.18 (Jan. 2021). Number: 18 Publisher: Multidisciplinary Digital Publishing Institute, p. 2523. ISSN: 2073-4441. DOI: 10.3390/w13182523. URL: https://www.mdpi.com/2073-4441/13/18/2523.

[24] NASA. *MODIS Web*. URL: https://modis.gsfc.nasa.gov/.

[25] NASA. *Suomi National Polar-orbiting Partnership | NASA's Earth Observing System*. URL: https://eospso.nasa.gov/missions/suomi-national-polar-orbiting-partnership.

[26] Julia S. Stoyanova, Christo G. Georgiev, Plamen N. Neytchev. "Drought Monitoring in Terms of Evapotranspiration Based on Satellite Data from Meteosat in Areas of Strong Land–Atmosphere Coupling". In: *Land* 12.1 (Jan. 2023). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 240. ISSN: 2073-445X. DOI: 10.3390/land12010240. URL: https://www.mdpi.com/2073-445X/12/1/240.

[27] Daniel Blank et al. "A global analysis of water storage variations from remotely sensed soil moisture and daily satellite gravimetry". In: *Hydrology and Earth System Sciences* 27.13 (July 4, 2023). Publisher: Copernicus GmbH, pp. 2413–2435. ISSN: 1027-5606. DOI: 10.5194/hess-27-2413-2023. URL: https://hess.copernicus.org/articles/27/2413/2023/.

[28] R. Albertz et al. "Investigation and mapping of forest damage using remote sensing methods. Final report. Digitale geometrische Aufbereitung multisensoraler und multitemporaler Fernerkundungsdaten fuer die Waldschadenforschung. Abschlussbericht". In: (Mar. 1, 1991). URL: https://www.osti.gov/etdeweb/biblio/5372089.

[29] NASA. *SMAP*. URL: https://smap.jpl.nasa.gov/.

[30] Oscar Cárdenas Hernández. "LAS REDES AVANZADAS DE CÓMPUTO EN LA PERCEPCIÓN REMOTA Y LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICA". Oct. 24, 2017. DOI: 10.13140/RG.2.2.25155.71208.

[31] J. W. Rouse et al. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*. E75-10354. NTRS Author Affiliations: Texas A&M Univ. NTRS Document ID: 19750020419 NTRS Research Center: Headquarters (HQ). Nov. 1, 1974. URL: https://ntrs.nasa.gov/citations/19750020419.

[32] Assefa M. MelesseUSGS Agenagnew A. Gessesse. *Normalized Difference Vegetation Index - an overview | ScienceDirect Topics*. URL: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/normalized-difference-vegetation-index.

[33] Samantha Lavender, Andrew Lavender. *Practical Handbook of Remote Sensing*. 2nd ed. Boca Raton: CRC Press, Apr. 20, 2023. 319 pp. ISBN: 978-1-00-327227-4. DOI: 10.1201/9781003272274.

[34] Bo-cai Gao. "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space". In: *Remote Sensing of Environment* 58.3 (Dec. 1, 1996), pp. 257–266. ISSN: 0034-4257. DOI: 10.1016/S0034-4257(96)00067-3. URL: https://www.sciencedirect.com/science/article/pii/S0034425796000673.

[35] Felix N. Kogan. "Droughts of the Late 1980s in the United States as Derived from NOAA Polar-Orbiting Satellite Data". In: (May 1, 1995). Section: Bulletin of the American Meteorological Society. ISSN: 1520-0477. URL: https://journals.ametsoc.org/view/journals/bams/76/5/1520-0477_1995_076_0655_dotlit_2_0_co_2.xml.

[36] F. N. Kogan. "Application of vegetation index and brightness temperature for drought detection". In: *Advances in Space Research*. Natural Hazards: Monitoring and Assessment Using Remote Sensing Technique 15.11 (Jan. 1, 1995), pp. 91–100. ISSN: 0273-1177. DOI: 10.1016/0273-1177(95)00079-T. URL: https://www.sciencedirect.com/science/article/pii/027311779500079T.

[37] Virgílio A. Bento et al. "A climatological assessment of drought impact on vegetation health index". In: *Agricultural and Forest Meteorology* 259 (Sept. 15, 2018), pp. 286–295. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2018.05.014. URL: https://www.sciencedirect.com/science/article/pii/S0168192318301667.

[38] Jahangir Khan et al. "Mapping MODIS LST NDVI Imagery for Drought Monitoring in Punjab Pakistan". In: *IEEE Access* 6 (2018). Conference Name: IEEE Access, pp. 19898–19911. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2821717. URL: https://ieeexplore.ieee.org/abstract/document/8329424.

[39] Virgílio A. Bento et al. "The roles of NDVI and Land Surface Temperature when using the Vegetation Health Index over dry regions". In: *Global and Planetary Change* 190 (July 1, 2020), p. 103198. ISSN: 0921-8181. DOI: 10.1016/j.

gloplacha.2020.103198. URL: https://www.sciencedirect.com/science/article/pii/S0921818120300898.

[40] Felix N. Kogan. "Global Drought Watch from Space". In: (Apr. 1, 1997). Section: Bulletin of the American Meteorological Society. ISSN: 1520-0477. URL: https://journals.ametsoc.org/view/journals/bams/78/4/1520-0477_1997_078_0621_gdwfs_2_0_co_2.xml.

[41] NASA Earth Science Data Systems. *AVHRR | Earthdata*. Publisher: Earth Science Data Systems, NASA. June 6, 2024. URL: https://www.earthdata.nasa.gov/sensors/avhrr.

[42] ESA. *Sentinel-2*. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2.

[43] GEOGLAM. *GEOGLAM*. URL: https://old.earthobservations.org/geoglam.php.

[44] Inbal Becker-Reshef et al. "Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning". In: *Remote Sensing of Environment* 237 (Feb. 1, 2020), p. 111553. ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.111553. URL: https://www.sciencedirect.com/science/article/pii/S0034425719305735.

[45] ESA. *The Sentinel missions*. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions.

[46] Paweł Hawryło, Piotr Wezyk. "Predicting Growing Stock Volume of Scots Pine Stands Using Sentinel-2 Satellite Imagery and Airborne Image-Derived Point Clouds". In: *Forests* 9 (May 17, 2018), p. 274. DOI: 10.3390/f9050274.

[47] Jordi Inglada et al. "Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series". In: *Remote Sensing* 8.5 (May 2016). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 362. ISSN: 2072-4292. DOI: 10.3390/rs8050362. URL: https://www.mdpi.com/2072-4292/8/5/362.

[48] Markus Immitzer, Francesco Vuolo, Clement Atzberger. "First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe". In: *Remote Sensing* 8.3 (Mar. 2016). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 166. ISSN: 2072-4292. DOI: 10.3390/rs8030166. URL: https://www.mdpi.com/2072-4292/8/3/166.

[49] Mmamokoma Grace Maponya, Adriaan van Niekerk, Zama Eric Mashimbye. "Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning". In: *Computers and Electronics in Agriculture* 169 (Feb. 1, 2020), p. 105164. ISSN: 0168-1699. DOI: 10.1016/j.compag.2019.105164. URL: https://www.sciencedirect.com/science/article/pii/S0168169919307586.

[50] Pranay Panjala, Murali Krishna Gumma, Pardhasaradhi Teluguntla. "Machine Learning Approaches and Sentinel-2 Data in Crop Type Mapping". In: *Data Science in Agriculture and Natural Resource Management*. Ed. by G. P. Obi Reddy et al. Singapore: Springer, 2022, pp. 161–180. ISBN: 9789811658471. DOI: 10.1007/978-981-16-5847-1_8. URL: https://doi.org/10.1007/978-981-16-5847-1_8.

[51] Silvia Valero et al. "Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions". In: *Remote Sensing* 8.1 (Jan. 2016). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 55. ISSN: 2072-4292. DOI: 10.3390/rs8010055. URL: https://www.mdpi.com/2072-4292/8/1/55.

[52] D. Swets et al. "A weighted least-squares approach to temporal NDVI smoothing". In: From image to information: 1999 ASPRS Annual Conference. 1999. URL: https://pubs.usgs.gov/publication/70201050.

[53] Roel Van Hoolst et al. "FAO's AVHRR-based Agricultural Stress Index System (ASIS) for global drought monitoring". In: *International Journal of Remote Sensing* 37.2 (Jan. 17, 2016), pp. 418–439. ISSN: 0143-1161, 1366-5901. DOI: 10.1080/01431161.2015.1126378. URL: http://www.tandfonline.com/doi/full/10.1080/01431161.2015.1126378.

[54] S. Kandasamy et al. *A comparison of methods for smoothing and gap filling time series of remote sensing observations: application to MODIS LAI products*. Dec. 4, 2012. DOI: 10.5194/bgd-9-17053-2012. URL: https://bg.copernicus.org/preprints/9/17053/2012/bgd-9-17053-2012.pdf.

[55] Jacques Delince et al. *Handbook on remote sensing for agricultural statistics*. Aug. 10, 2017. DOI: 10.13140/RG.2.2.13259.69920.

[56] USGS. *Landsat 8-9 Collection 2 Level 2 Science Product Guide | U.S. Geological Survey*. URL: https://www.usgs.gov/media/files/landsat-8-9-collection-2-level-2-science-product-guide.

[57] USGS. *How do I use a scale factor with Landsat Level-2 science products? | U.S. Geological Survey*. URL: https://www.usgs.gov/faqs/how-do-i-use-a-scale-factor-landsat-level-2-science-products.

[58] Alvaro Valenzuela, Karin Reinke, Simon Jones. "Assessing the spatial resolution distance of satellite images: SuperDove versus Landsat 8". In: *International Journal of Remote Sensing* 45.12 (June 17, 2024). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431161.2024.2357839, pp. 4120–4159. ISSN: 0143-1161. DOI: 10.1080/01431161.2024.2357839. URL: https://doi.org/10.1080/01431161.2024.2357839.

[59] Brian N. Wenny et al. “Pre- and Post-Launch Spatial Quality of the Landsat 8 Thermal Infrared Sensor”. In: *Remote Sensing* 7.2 (Feb. 2015). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, pp. 1962–1980. ISSN: 2072-4292. DOI: 10.3390/rs70201962. URL: https://www.mdpi.com/2072-4292/7/2/1962.

[60] Michael A. White, Peter E. Thornton, Steven W. Running. “A continental phenology model for monitoring vegetation responses to interannual climatic variability”. In: *Global Biogeochemical Cycles* 11.2 (1997). _eprint: https://onlinelibrary.wiley.c pp. 217–234. ISSN: 1944-9224. DOI: 10.1029/97GB00330. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/97GB00330.

[61] Victor Rodriguez-Galiano, Mario Chica. “Land cover change analysis of a Mediterranean area in Spain using different sources of data: Multi-seasonal Landsat images, land surface temperature, digital terrain models and texture”. In: *Applied Geography* 35 (Nov. 1, 2012), pp. 208–218. DOI: 10.1016/j.apgeog.2012.06.014.

[62] NASA. *Severe Drought in Southern Africa*. URL: https://earthobservatory.nasa.gov/images/152711/severe-drought-in-southern-africa.

[63] Oscar Karl Huh. “Limitations and capabilities of the NOAA satellite advanced very high resolution radiometer (AVHRR) for remote sensing of the Earth’s surface”. In: *Preventive Veterinary Medicine* 11.3 (Dec. 1, 1991), pp. 167–183. ISSN: 0167-5877. DOI: 10.1016/S0167-5877(05)80003-0. URL: https://www.sciencedirect.com/science/article/pii/S0167587705800030.

Towards a crop-specific drought impact assessment tool in Chiredzi District, Zimbabwe, using Sentinel-2 and Landsat-8/9 imagery

Theo Marichal

The increasing frequency and severity of droughts due to climate change pose significant threats to agricultural production, particularly in regions like Southern Africa. Accurate monitoring and assessment of drought impacts at the crop-specific level are crucial for effective agricultural management and food security. This study proposes a valuable tool, a high-resolution Vegetation Health Index (VHI) tailored for monitoring drought effects on various crop types within the Chiredzi District of Zimbabwe. Utilizing Sentinel-2 satellite imagery, NDVI and LST time series were calculated to produce a VHI with a spatial resolution of 10 meters per pixel. The VHI was further refined by correcting for seasonal shifts, ensuring that crop development phases were accurately aligned across different years. This high-resolution index was integrated with a crop type classification map, derived using Random Forest (RF) classification, enabling precise assessment of drought impacts on individual crops at this fine spatial scale.The results reveal significant variability in drought resilience among different crops, with rain-fed crops like sorghum being particularly vulnerable. The study underscores the importance of high-resolution data and seasonal adjustment in enhancing the accuracy of drought monitoring systems, offering a valuable tool for improving agricultural decision-making in drought-prone regions. The methodology developed here can be adapted to other regions, contributing to the refinement of early warning systems for food security.

Université catholique de Louvain

Faculté des bioingénieurs

Croix du Sud, 2bte L7.05.01, 1348 Louvain-La-Neuve, Belgique | www.uclouvain.be/agro