An assessment on field-scale spatial variability of sugarcane yield with satellite derived vegetation indices and evapotranspiration products

A case study on sugarcane fields in Mozambique

K.A. Beenen

An assessment on field-scale spatial variability of sugarcane yield with satellite derived vegetation indices and evapotranspiration products

A case study on sugarcane fields in Mozambique

by

K.A. Beenen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday March 30, 2020 at 11:00 AM.

Student number: 4139984

Thesis Committee:	Prof. Dr. ir. S.C. Steele-Dunne	TU Delft
	Prof. Dr. ir. P. van der Zaag	IHE Delft & TU Delft
	Dr. S. Lhermitte	TU Delft
	Dr. ir. A.M.J. Coenders-Gerrits	TU Delft
	MSc. N.I. den Besten	VanderSat & TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

VanderSat

UNESCO IHE DELFT

IWACA-TECH

Abstract

Due to increasing global population and increasing food demand, crop yield needs to be improved to forestall potential food shortages. To achieve optimal crop yield, irrigation is applied to replace water losses due to evapotranspiration (ET). Information on ET should be applied to optimize water allocations and water use. In this study, the correlation between ET and yield will be investigated on field-scale level to assess the potential of high resolution ET products as a tool to detect yield variation. The variability of crop yield and transpiration are caused by the variability in the topography, groundwater and soil properties. Agricultural practices and field-scale water management demand high resolution (in meters) and high temporal resolution (daily to sub-daily) remote sensing products. With the arrival of new satellite platforms, such as Sentinel-2, the aforementioned remote sensing data can be improved significantly in spatial and temporal resolution.

In order to compare the functionality of different remote sensing products, an assessment is executed for two satellite derived vegetation indices: normalized difference vegetation index (NDVI) and normalized difference water index (NDWI), and two satellite derived evaporation products: WaPOR (Water Productivity through Open access of Remotely sensed derived data) and a newly developed evaporation algorithm from VanderSat. Within this research, the focus lies on assessing which dataset is able to observe the spatial difference and temporal patterns on field-scale level. Using a large sugarcane plantation in Xinavane, Mozambique, as a case study, we demonstrate how the spatial variability of the remote sensing results are correlated to the sugarcane yield.

To assist irrigated agriculture we demonstrate that a high resolution evaporation product is needed to incorporate spatial variability in evaporation estimates. The analysis shows that the high resolution satellite derived vegetation indices are related to the spatial variability of yield. Our results indicate that NDWI has a strong positive correlation of 0.73 with yield, but NDVI has only 0.64. The actual evapotranspiration estimates have a moderately positive correlation with yield of 0.5 for WaPOR and 0.57 for VanderSat. Evaporation estimates should be related to yield to control irrigation properly. WaPOR and VanderSat use NDVI as a input for crop stress, these existing evaporation algorithms should incorporate high resolution spatial imagery as NDWI instead of NDVI to assist irrigation adequately. In order to use the satellite derived evaporation algorithms for agricultural practices and field-scale water management, future research should be focus on improving the relation between satellite derived evaporation algorithms and yield.

Keywords: evapotranspiration, irrigation, NDVI, NDWI, satellite remote sensing, spatial variability, sugarcane

Preface

This thesis is written as completion of my master degree in "Geoscience and Remote Sensing" at the Delft University of Technology. During my master track I finally found out what I am really passionate about. This thesis allowed me to combine my interest in remote sensing as well as agriculture. Through this project I learned many new things and meet interesting and exciting new people.

First of all, I want to thank Richard de Jeu and Nadja den Besten for given me the opportunity to do my master thesis at VanderSat. I am very happy to work as a graduate intern at a company that shares the same passion and values that I have. Nadja, I want to thank you for introducing this topic to me and giving me the opportunity to go on a fieldwork and to a conference in Johannesburg, South Africa. It was really nice working with you and I learned a lot from you. Thanks to Milton Paulao for helping me obtaining my field measurements in Xinavane, Mozambique. Unfortunately, we only worked together for a couple of weeks, but I got to know you as an enthusiastic and open person. Furthermore, I want to thank Nelson, Paulmo and Pitrosse, who I shared my accommodation with and who let me feel Mozambique as my second home.

I would like to thank Susan Steele-Dunne for her work as my daily supervisor and chair of my committee. Susan, you were an amazing supervisor. Thank you for being straightforward and keeping me on track and letting me focus on the parts which where relevant for my research. Thanks to the rest of the committee, Stef Lhermitte and Miriam Coenders, for sharing their ideas. Furthermore, I would like to thank Pieter van der Zaag. His input from a whole different perspective, namely integrated water resources management, was very welcoming.

Furthermore, this research was supported by the Lamminga Fund and VanderSat. With their funding it became financially possible for me to go to Mozambique and to present my results at the Waternet symposium in Johannesburg.

Last but not least, I want to thank my family and friends, who always support me. Especially I would like to thank my parents who always encourage me and give me the freedom to explore the most exotic places.

Kathelijne Beenen
kathelijnebeenen@gmail.com
Delft, March 2020

Contents

List of Figures

List of Tables

List of Abbreviations

AETI	Actual Evapotranspiration and Interception
ALEXI	Atmospheric-Land Exchange Inverse Model
C-GLOPS	Copernicus Global Land Cover
CGDD	Cumulative Growing Degree Days
CHIRPS	Climate Hazards Groups InfraRed Precipitation
CV	Coefficient of Variation
DEM	Digital Elevation Model
EC	Electrical Conductivity
E_i	Interception
E_s	Soil evaporation
E_t	Transpiration
ET	Evapotranspiration
EVI	Enhanced Vegetation Index
FAO	Food and Agricultural Organization
G	Soil heat flux
GEOS-5	An global atmospheric model called Goddard Earth Observing System Model, Version 5.
GLEAM	Global Land-surface Evaporation Amsterdam Model
H0	Null Hypothesis
HAND	Height Above Nearest Drainage
IIAM	Instituto de Investigação Agrárua de Moçambique
IPCC	Intergovernmental Panel on Climate Change
LAI	Leaf Area Index
LST	Land Surface Temperature
MSG	Meteosat Second Generation
NDSI	Normalized Difference Salinity Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
PROBA-V	Project for On-Board Autonomy-Vegetation
r_0	Surface albedo
R_n	Net radiation
SEBAL	Surface Energy Balance Algorithm for Land
STRM	Shuttle Radar Topography Mission

SWIR	Short-wave infrared
TC	Triple Collation
TCH	Tons cane per hectare
VNIR	Visible and near infrared
WDVI	Weighted Difference Vegetation Index
WP	Water productivity
WaPOR	FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data.
α	Priestley - Taylor coefficient
γ	Latent Heat Vaporization (2.45 MJ/kg)
λ	Psychometric constant
μ	Mean
σ	Standard deviation
τ	Transmissivity: amount of solar radiation that is propagated through the atmosphere.

1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) predicts that the global population increases to 9 billion people causing the food demand to rise by 60% by 2050 [IPCC, 2014]. The impact of population growth and increasing food demand will have an increased reliance on irrigation [IPCC, 2014]. To forestall potential food shortages, yields need to be improved.

Yield largely depends on availability of water, but with climate change temperature will rise and cause an increase in evaporation as a consequence [IPCC, 2014]. Climate change will have an influence on water resources, leading to a reduction of water allocation for downstream irrigated areas due to a higher evaporative demand [IPCC, 2014]. Therefore, water use efficiency of irrigation water needs to be improved.

Another reason to look into more effective irrigation management is that poor irrigation management can have a detrimental effect on yield, and can result in spatial variability within and between fields. Other contributors, which can have a detrimental effect on yield, are nutrients deficit and crop diseases [Sanches et al., 2020].

1.1. Problem statement

Most research focuses on the supply side of the water cycle such as precipitation, soil moisture and groundwater [Fisher et al., 2017]. However, evapotranspiration (ET) is important to understand the loss of water to the atmosphere, which can also be called the supply demand. Irrigation is applied to replace the water losses due to ET and achieve an optimal yield. For agricultural applications this can alleviate the vegetation stress for crops and irrigation is often applied in such a way that applied water approximately compensates the atmospheric demand for ET.

The current knowledge gap is how information on ET can be applied to optimize sustainable water allocations and agricultural water use. Hence meeting the water and food security goals in a changing climate [Fisher et al., 2017].

This study determines the relation between evapotranspiration and yield. Besides that the ET is linked to the water cycle, ET is also linked to the carbon and energy cycle [Fisher et al., 2017]. Therefore, a good ET product should be correlated with yield. In this study the correlation between evapotranspiration and yield will be investigated to assess the potential of high resolution evapotranspiration products as a tool to detect yield variations. Evapotranspiration varies both within and along days and spatially. Therefore, Fisher et al. (2017) state that ET products should have a high spatial resolution of 10-100 meter to detect spatially non-uniformity response of water at field-scale level. Furthermore, this study address that ET products should have a daily to sub-daily temporal resolution [Fisher et al., 2017]. Therefore, we used two newly developed products that met these crite-

ria: Vandersat evapotranspiration products and WaPOR [den Besten et al., nd, FAO, 2018]. Both ET products use the normalized difference vegetation index (NDVI) to detect crop stress [den Besten et al., nd, FAO, 2018]. Numerous studies demonstrate the correlation of NDVI as a proxy of vegetation growth [Pinheiro Lisboa et al., 2018, Lofton et al., 2012]. However, another study shows that the normalized difference vegetation index is not always the best indicator for yield stress [Bolton and Friedl, 2013]. If evapotranspiration estimates should be used as an indicator of crop yield, then the hypothesis is that NDVI estimates should also be correlated to crop yield. Therefore, this research determines the relation between yield and satellite derived vegetation indices such as NDVI and the normalized difference water index (NDWI). In this study, the spatial patterns of NDVI and NDWI are compared in relation to yield and their relationship are quantified.

This thesis is conducted in the context of the IWACA-TECH project, an abbreviation for "Improved Water efficiency Control based on remote sensing TECHnologies" (`http://www.iwacatech.com/`). The aim of the IWACA-TECH project is to improve water use efficiency at irrigation systems applying advanced Remote Sensing Technology [IWACA-TECH, 2020]. One of the partners of the project is the agriculture and agri-processing company Tongaat Hullett, which mainly focuses on producing and processing sugarcane and maize. One of their sugarcane plantation of Tongaat Hulett is located around the town Xinavane, which lies approximately 140 km northeast of Maputo, the capital of Mozambique. This location is used as the area of interest for this thesis.

One of the major problems at the Xinavane sugarcane plantation is the decrease in yield and the spatial variability of yield between fields. Since sugarcane requires a lot of water for its growth, irrigation is applied with a view to achieve the target yield. The yield is used as a measure of productivity. The yield of sugarcane is defined as the ratio of total cane mass to cropped area. The irrigation water is supplied by the Incomati river. To improve water use efficiency, spatial variability of yield within fields needs to be taken into consideration. Inefficient use of irrigation can cause problems such as waterlogging and salinity, which result in low yield.

1.2. Motivation of the research

Remote sensing can provide crop information with a high temporal and spatial resolution. To serve the needs of water management and agricultural applications on field level, high spatial (meter) and temporal resolution (daily to sub-daily) remote sensing and evaporation data are required [Martens et al., 2018]. Several studies show that remote sensing data can be used for vegetation applications [Steele-Dunne et al., 2017, Begue et al., 2008]. Other studies show that potential yield analysis can be done using satellite remote sensing [Lobell, 2013, Lobell et al., 2010]. Studies on sugarcane show that the spatial variability of sugarcane can be mapped by the normalized difference vegetation index (NDVI) [Morel et al., 2014, Bégué et al., 2010]. Recent studies investigate if sugarcane yield can be monitored by optical and radar satellite imagery [Molijn et al., 2019, Molijn et al., 2018]. When monitoring sugarcane Molijn et al. (2019) show that the signal transmitted by the C-band radar satellite saturates faster than optical sensors [Molijn et al., 2019].

With the launch of the Sentinel-2 satellites , the optical data is improved significantly in spatial and temporal resolution, although clouds still pose a problem for optical imagery as the signal cannot penetrate through clouds. Sentinel-2 has a revisit time of 5 days with 2 satellites (Sentinel-2A and Sentinel-2B) under cloud-free conditions. Using Sentinel-2 the data gaps can be resolved for vegetation indices using methods as linear interpolation, Harmonic Analysis of Time Series (HANTS) and machine learning.

By combining spectral bands satellite derived vegetation indices can be computed. Gao et al. (1996) introduce a vegetation index, which respond to changes in the vegetation liquid water content and structure, called normalized difference water index (NDWI)

[Mulianga et al., 2015, Gao, 1996]. Mulianga et al. (2015) have proved that NDWI can be used for identifying agricultural practices such as moment of harvest [Mulianga et al., 2015]. Multiple studies have been conducted on sugarcane yield assessment with NDVI [Mulianga et al., 2013, Lofton et al., 2012, Bégué et al., 2010]. However, to my knowledge, the relation between NDWI and sugarcane yield is not yet investigated and warrants investigation. Therefore, this study explores if the satellite derived vegetation indices capture the heterogeneity of the fields and quantifies the correlation between NDVI and NDWI estimates, and yield.

Table 1.1: Optical satellites and their spatial and temporal resolution

Optical satellites	Spatial resolution	Temporal resolution
Aster	15-90 m	16 day
Landsat-8	15-30m	16-day
MODIS	250-1000 m	1-2 day
PROBA-V	100-1000m	1-2 day
Sentinel-2	10-60 meter	5 day

Evapotranspiration is the flux of water from land (crop and soil) to the atmosphere [Margulis, 2017]. ET gives information about water deficit and it gives information on crop behaviour in relation to atmospheric conditions [Coenders-Gerrits et al., 2020]. Evapotranspiration cannot be measured directly using remote sensing satellites. To estimate evapotranspiration models are constructed using different satellite remote sensing variables as input [Martens et al., 2018]. For instance, Atmosphere-Land Exchannge Inverse model (ALEXI), Surface Energy Balance Algorithm for Land (SEBAL) and Global LAnd-Surface Evaporation Amsterdam Model (GLEAM) are examples of evapotranspiration model based on satellite remote sensing data.

However, ET models such as GLEAM have been hampered by its spatial resolution, resulting from microwave remote sensing data with coarse-scale resolution as inputs [Martens et al., 2018]. Other models such as disAlexi and SEBAL have a high spatial resolution (in meters), but are limited by data gaps caused by cloud coverage using optical imagery as inputs. Nowadays, high resolution satellites, see table 1.1, are available, which can be used as inputs for evaporation models. With the development of high resolution evaporation products such as WaPOR and the newly developed algorithm of VanderSat, evaporation estimates can meet the requirements for local-scale water management and agricultural applications such as irrigation planning. WaPOR version 2.1, released on 29/11/2019, provides evapotranspiration estimates with a dekadal temporal resolution and a spatial resolution of 100 meter for Mozambique [FAO, 2018]. The VanderSat algorithm provides evapotranspiration estimates with daily to sub-daily temporal resolution and a spatial resolution of 20 meter [den Besten et al., nd]. These newly developed high resolution evaporation products can therefore give better insight in spatio-temporal resolution in water use efficiency at irrigation systems, water deficit and identifying problematic areas within the fields. Yang et al. (2018) demonstrate that the ratio of actual to reference ET generated by high spatio-temporal satellite derived ET estimates can convey information for crop yield on field-scale level. In table 1.2 the above mentioned models are listed to provide an overview of their spatio-temporal resolutions for these models. Studies using WaPOR and VanderSat evaporation product are limited. Therefore, investigation is needed to know what the capabilities of these products are in providing spatial information on field-scale level. Since there is a lot of heterogeneity in the performance of the crop between and within fields at the study site, geostatical analysis can be used to quantify the spatial variation on field-scale level. This study focus on capturing the heterogeneity of the fields by these evaporation models, satellite derived vegetation index and on the relation between these parameters and yield.

Table 1.2: List of evapotranspiration models and their spatial and temporal resolution

ET models	Spatial resolution	Temporal resolution	Literature
ALEXI	5-10 km	daily	Anderson et al. (1997)
DisAlexi	30m	daily	Norman et al., (2003), Anderson et al. (2012)
GLEAM	25 km, in development 100m	daily	Martens et al. (2017), Martens et al. (2018)
SEBAL	30 m	16-day	Bastiaanssen et al. (1998, 2005)
VanderSat	20 m	daily - sub-daily	den Besten et al. (2019a)
WaPOR	Level 1 - Continental : 250 m	dekadel	FAO and IHE Delft (2019)
	Level 2 - Basin & Country : 100 m	dekadel	FAO and IHE Delft (2019)
	Level 3: Subbasin: 30 m	dekadel	FAO and IHE Delft (2019)

At the study site, yield is measured after harvest on field scale level by weighing trucks with and without sugarcane load. Visual inspection of the fields is done by field managers, making it labour intensive [Koenders et al., 2019]. Furthermore, visual inspections do not give an accurate yield estimation. At the study site, there is a high level of of heterogeneity in the performance of the crop between fields. Studies show that variation in crop yield can be caused by differences in topography, soil and groundwater properties [Kitchen et al., 2003, Kravchenko and Bullock, 2000]. Causes of this yield variation in relation to topography, soil and groundwater properties at the plantation are not yet investigated in detail. Therefore, this study looks at the relation between these field parameter and yield. For the plantation remote sensing data can be a tool for identifying non-heterogeneity between and within fields.

1.3. Research objectives

The aim of this study is to asses whether (1) the spatial variability of yield at the sugarcane plantation can be explained with topography, soil and groundwater properties and whether (2) the spatial patterns in the remote sensing data reflect the difference in crop performance. In particular, this thesis will determine if spatial patterns in yield variability are apparent in satellite derived vegetation indices and satellite derived evaporation products. In order to accomplish this, this research determines the correlation coefficient for sugarcane between yield and field properties, satellite derived vegetation indices and satellite derived evaporation products.

Based on the aim of this research the following sub-questions are formulated as:

1. What is the correlation between topography, soil and groundwater properties and yield?
   This is investigated by doing a correlation study between those different field properties and yield.
2. How are the spatial patterns in NDVI related to yield?
   This is investigated by computing the correlation between NDVI and yield throughout a growing season.
3. How are the spatial patterns in NDWI related to yield?
   This is investigated by computing the correlation between NDWI and yield throughout a growing season.
4. Can actual evapotranspiration from existing remote sensing evaporation products be used to detect yield variation on field-scale level?
   This is investigated by comparing the performance of WaPOR and VanderSat evaporation product in relation to yield.
5. Can evaporative stress indicators be used as yield indicators on field-scale level?
   This is investigated by computing the ratio of potential to actual evapotranspiration and the difference between accumulated potential ET and the accumulated actual ET, and by comparing these parameters to yield data.

With support of these sub-questions the main research question can be answered, which is:

How can the spatial variability of yield be explained and how is this reflected in existing Remote Sensing data?

1.4. Thesis outline

Chapter 2 provides the theoretical background and a literature review, which outlines the current state of understanding the topics discussed in this research. In Chapter 3 the project area is introduced and information about crop, agricultural practices and challenges in the area can be found. Chapter 4 gives a description of the different data and the tools which were used to analyse the data. Moreover, in chapter 4 an overview of the processing steps can be found, which were applied to convert the raw data into the actual results. In chapter 5 the results are presented and analyzed. Correlation coefficients between yield and the used field and remote sensing datasets are given. Chapter 6 provides the discussion centered around the main research question. The conclusions of this research are presented in Chapter 7, followed by an overview of improvements in the data and recommended future work in Chapter 8.

2

Background

This section is divided in two parts. The first section gives a literature review. The literature review outlines the current state of understanding the following topics of: (1) the relation between yield and field properties, (2) using satellite derived vegetation indices for yield assessment and (3) satellite based evaporation products. The second section describes the theoretical background. In this section a description is given of optical remote sensing data, evaporation and a product description of satellite derived evaporation products used in this study.

2.1. Literature review

In this section the literature review outlines the current state of art of understanding the following topics of (1) relation between topography, soil and groundwater properties on sugarcane yield, (2) assessing the satellite derived vegetation index in relation with yield and (3) an overview of state of the art satellite derived evaporation products. Throughout the years a lot of research has been done on these topics.

Yield affected by field properties

Field properties such as soil, topography and groundwater properties are important factors influencing the yield [Kravchenko and Bullock, 2000]. Understanding spatial soil variability related to topographic parameters such as elevation and convexity can improve soil management and can explain yield variability of the fields [Sanches et al., 2020]. Kravechenko et al. (2000) state that the amount of organic matter in the soil is an important factor that influences yield. Higher levels of organic matter are most found in concave areas [Sanches et al., 2020]. Therefore, Sanches et al. (2020) state that for sugarcane fields soil acidity management should be more rigorous in convex areas. Another method, Height Above Drainage Pumps (HAND) uses topography to compute the local relative heights between fields and drainage networks [Nobre et al., 2011]. Nobre et al. (2011) demonstrate that the HAND is highly correlated with the depth of the groundwater table. Areas with problems as salinization and waterlogging occur for 60%-70% in areas with a groundwater table close to the surface [Abbas et al., 2013]. Often these salinization problems are irrigation-induced and the areas are poorly drained [Abbas et al., 2013]. Hurst et al. (2004) show that there is an impact of shallow groundwater, which means that the groundwater table is close to the surface, on yield and evapotranspiration estimates. Furthermore, the paper implies when groundwater table is within 1.5 m of the soil surface, the groundwater can be used as a water resource for sugarcane and therefore water use efficiency in irrigation improves [Hurst et al., 2004].

Review of use of satellite derived vegetation indices on crop yield
Numerous studies are done using NDVI for crop condition assessment as well as yield forecast. To monitor sugarcane NDVI is the most common satellite derived vegetation indices to be used. Multiple studies have presented that there is a correlation between NDVI and sugarcane yield [Mulianga et al., 2013, Bégué et al., 2010]. Furthermore, several studies demonstrate that NDVI can characterize the spatio-temporal variability of sugarcane fields [Bégué et al., 2010] . NDWI has been applied in agricultural applications such as monitoring vegetation drought [Xu et al., 2020, Gu et al., 2008]. Bolton et al. (2013) investigated the correlation between multiple satellite derived vegetation indices and maize and soybean yield. The paper shows that for maize in non-semi-arid countries EVI2 (two band enhanced vegetation index) has the highest correlation with yield [Bolton and Friedl, 2013]. For maize in semi-arid irrigated areas NDWI has the highest correlation with yield [Bolton and Friedl, 2013]. For soybean NDVI and EVI2 correlate equally well with yield. Therefore, the paper demonstrates thus that performance of different satellite derived vegetation indices change per crop and per area. In addition, NDVI is used as input for numerous evaporation models as crop stress indicator, for example in WaPOR, Vandersat algorithm, SEBAL and disAlexi. However, NDVI does not have to be the best stress indicator for that crop and area.

Evapotranspiration
As mentioned in Section 3.1, evapotranspiration is the flux of water from land to the atmosphere. Evapotranspiration (ET) gives information about the water requirement of the crop. Spatio-temporal information on actual ET can be used for irrigation management purposes (water allocation), which can lead to improvement in water use efficiency and yield. Evapotranspiration varies highly in time and space. Instead of using in-situ measurements, which are point measurements, several remote sensed evaporation products are developed to give information on this flux in time and space. Numerous remote sensed models are constructed to estimate actual ET. WaPOR and VanderSat evaporation algorithm are relatively new evaporation models and not much research has be done yet using these models. Other models already have been used for numerous studies. Each ET model has there own methodology. Hence, some of these satellite derived evaporation models are briefly discussed below.

SEBAL (Surface Energy Balance Algorithm for Land) estimates the ET by computing the surface energy flux [Bastiaanssen et al., 2005, Bastiaanssen et al., 1998]. The spatial resolution of SEBAL is 30m. However, the temporal variability in SEBAL actual evapotranspiration depends on Landsat optical images. Landsat has a 16 day revisit time under cloud-free conditions. NDVI computed from Landsat is used as an input for the model.

GLEAM (Global Land-surface Amsterdam Model) computes ET using primarily remote sensing data from microwave sensors [Miralles, 2011]. By using microwave bands GLEAM provides ET data under all weather conditions. This is because microwave signals penetrate through clouds. At the moment ET data from GLEAM has a resolution of 0.25 degree, which means a resolution of about 27-28 km at the equator [Miralles, 2011]. However, attempts are made to downscale GLEAM to achieve daily values and a spatial resolution of 100m [Martens et al., 2018].

Alexi, Atmosphere-Land Exchange Inverse model, is based on the energy balance [Anderson et al., 2011]. It uses high resolution Land Surface Temperature (LST) from geostationary satellites and thermal infrared (TIR) remote sensing data to compute ET estimates for two sources: soil and canopy [Anderson et al., 2011]. The ET fluxes are mapped with a resolution of 5-10 km on a daily basis [Anderson et al., 2011].

A related algortihm to Alexi is *DisAlexi*, which disaggregates Alexi to get a higher resolution of 30 meter ET estimates using TIR data from polar orbiting satellites such as MODIS and Landsat [Anderson et al., 2010]. However, it is important to know that Alexi and DisALEXI are constrained by cloud coverage and therefore data gaps are occurring.

Currently, the potential of satellite observation of solar-induces chlorophyll fuorescence (SIF) normalized by the photosynthetically-active radiation (PAR) is examined to diagnose transpiration efficiency also formulated as the ratio of transpiration to potential evaporation [Pagán et al., 2019]. SIF can be detected in the spectral range of 650-800nm and SIF/PAR estimates can be computed using satellites as CERES and GOME-2 [Pagán et al., 2019]. Pagán

et al. (2019) demonstrate that the SIF/PAR estimates capture the effect of phenological changes and environmental stress.

Another new technique is proposed by Vanella et al. (2019). This research focuses on the usage of electrical resistivity tomography (ERT) to adjust the dual crop coefficient FAO-56 to improve evaporation estimates in irrigated conditions using Sentinel-2 and local meteorological data [Vanella et al., 2019]. The research explores if the ERT technique can be used to monitor the soil wetting distribution patterns during and after irrigation [Vanella et al., 2019].

McCabe et al. (2019) discuss the advances and issues of using terrestrial remote sensed evaporation data. For example the paper addresses the issues of scaling with remote sensing data. The paper emphasizes that improving ET estimates is not purely a function of increasing the resolution of satellite observations [McCabe et al., 2019]. Fisher et al. (2017) discuss requirements for solving the knowledge gap of ET in ecosystem functioning, carbon and climate feedbacks, agricultural management and water resources. To address the current ET questions in these sciences, ET must be mapped with very high fidelity [Fisher et al., 2017]. Fisher et al. (2017) discuss therefore the spectral and accuracy requirements and the spatial and temporal resolution needed to answer the ET question.

Last to mention is that Yang et al. (2018) explore if evaporation stress indicators such as the ratio of actual to potential ET can be used to improve yield and water productivity monitoring. The study shows the possibility of satellite derived evaporation estimates in conveying water stress information to yield prediction [Yang et al., 2018].

2.2. Theoretical background

In this section the techniques used in this research are described. First, optical remote sensing data is described. Secondly, the importance of evaporation in the hydrological cycle will be explained followed by a description of the used evaporation products: WaPOR and VanderSat evaporation algorithm.

Optical remote sensing data

This research will use reflectance data from passive optical sensors. This means: when solar radiation is incident on the surface, the radiation may be absorbed, transmitted or reflected. Optical satellite sensors measures the reflected radiation from the Earth. The spectral reflectance of the land surface depends on the land cover type, ground cover fraction, phenological stage and health of the vegetation. The spectral reflectance of a given surface also varies with wavelength. Differences in spectral reflectance signatures (i.e. reflectance as a function of wavelength) can be used to distinguish between cover types or to detect variations in the parameters listed above.

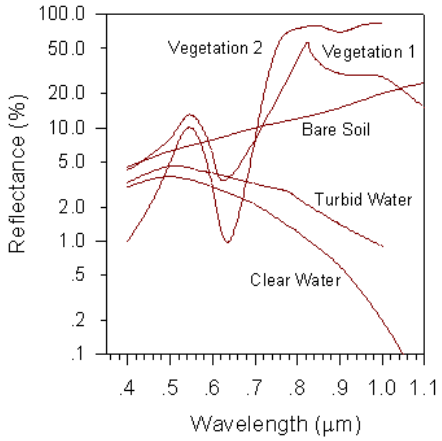


Figure 2.1: Spectral reflectance of different landcover types [Lieuw, 2001]

For this thesis the satellite derived vegetation indices are computed using Sentinel-2 data. Sentinel 2 consist of 13 spectral bands, which are presented in figure 2.2. More information about the acquired Sentinel-2 data can be found in chapter 4.

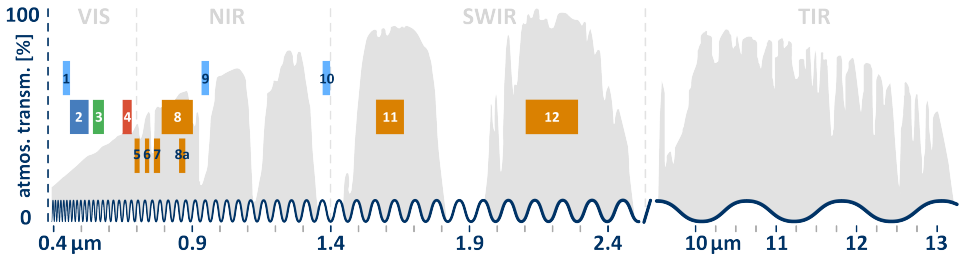


Figure 2.2: Spectral bands of Sentinel-2 [Berlin, 2020]

Normalized Difference Vegetation Index

The normalized difference vegetation index is used to quantify the chlorophyll content of the vegetation. NDVI quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs).

Rouse et al. (1974) define normalized difference vegetation index as:

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

For Sentinel-2 the near infrared (NIR) has a wavelength ranges from 785-899 nm and for the visible red band the band wavelength of ranges from 650 to 680 nm. The near InfraRed (NIR) reflects strongly at vegetation. The red light is absorbed by the vegetation.

To obtain NDVI images from Sentinel-2 reflectance data of band 4 (Red) and band 8a (Near InfraRed) are used to calculate the NDVI, see equation 2.2.

$$NDVI = \frac{Band8A - Band4}{Band8A + Band4} \tag{2.2}$$

The values of NDVI range from -1 to +1. The obtained spatial resolution of the Sentinel-2 NDVI images are 10 meter. Negative NDVI values correspond to water. Values close to zero correspond to bare soil.

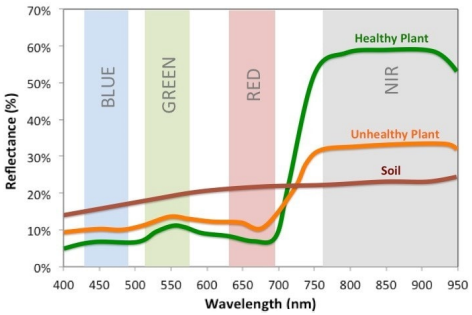


Figure 2.3: Spectral reflectance of vegetation [Shilo, 2018]

Several studies have shown that there is relationship between NDVI and sugarcane yield [Rao et al., 2002, Lofton et al., 2012, Pinheiro Lisboa et al., 2018]. For sugarcane, the NDVI increases in proportion to the density of crop. For example, at the planting date the sugarcane has a low NDVI value, because the density of the crop is low. The NDVI values increase gradually during the growing cycle until a certain density/height. Then the NDVI signal saturates [Molijn et al., 2016, Molijn et al., 2019]. This means that the sugarcane is still growing and thus increasing by density, but this no longer leads to an increase in the NDVI value.

Normalized Difference Water Index

The normalized difference water index (NDWI) reflects changes in the vegetation water content [Gao, 1996]. Gao at el. (1996) define the normalized difference water index as:

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} = \frac{860nm - 1240nm}{860nm + 1240nm} \tag{2.3}$$

The values of NDWI range from -1 to +1 and the NDWI value depends on the leaf water content but also on the vegetation cover. Table 2.1 gives an overview how to interpret the NDWI values. Looking at the table 2.1 the NDWI value increases as the crop grows, which indicates that NDWI is sensitive to crop phenological stage as well as the total liquid water amount in the crop [Gao, 1996].

Using the Sentinel-2 reflectance data the normalized difference water index (NDWI) is computed as follow:

$$NDWI = \frac{Band8A - Band11}{Band8A + Band11} \tag{2.4}$$

The obtained spatial resolution of the Sentinel-2 NDWI images are 20 meter.

NDWI can be used recognize the areas of the field with water stress problems. Within our knowledge no literature is found which quantifies the relation between NDWI and yield. However, studies show that there is a correlation between yield and NDWI for other crops types [Van Beek et al., 2015, Petersen, 2018].

Table 2.1: Interpretation NDWI values [Antognelli, 2020]

NDWI	INTERPRETATION
-1 – -0.8	Bare soil
-0.8 – -0.6	Almost absent canopy cover
-0.6 – -0.4	Very low canopy cover
-0.4 – -0.2	Low canopy cover, dry or very low canopy cover, wet
-0.2 – 0	Mid-low canopy cover, high water stress or low canopy cover, low water stress
0-0.2	Average canopy cover, high water stress or mid-low canopy cover, low water stress
0.2 – 0.4	Mid-high canopy cover, high water stress or average canopy cover, low water stress
0.4 – 0.6	High canopy cover, no water stress
0.6 – 0.8	Very high canopy cover, no water stress
0.8 – 1	Total canopy cover, no water stress or waterlogging

Water stress can be recognized with NDWI, when the crop tends to be at the same phenological stage within a field. Areas where the NDWI value is significantly lower than average normal can indicate vegetative development problems associated with water stress.

Normalized Difference Salinity Index

Salt affected areas can be identified using the Normalized Difference Salinity Index (NDSI) [Asfaw et al., 2018]. For soil salinity monitoring some papers suggest to use the bands within the SWIR spectrum and other suggest to use the VNIR bands [Bannari et al., 2018]. In this study the SWIR bands are used. However, studies for soil salinity monitoring using SWIR bands are mostly conducted with ASTER and Landsat [Al-Khaier, 2003]. Using the SWIR bands the normalized difference salinity index can be computed as follow:

$$NDSI = \frac{SWIR - SWIR}{SWIR + SWIR} = \frac{[1600 : 1700] - [2145 : 2185]}{[1600 : 1700] + [2145 : 2185]} \tag{2.5}$$

The first SWIR band should have a sensing range somewhere in 1600-1700 nm and the second SWIR band should have a sensing range of 2145-2185 nm. The values of NDSI range

from -1 to +1. For sentinel-2 reflectance data of band 11 (Shortwave infrared) and band 12 (Shortwave infrared) from Sentinel-2 are used to calculate the NDSI, which are in the spectral range mentioned in equation 2.5. For Sentinel-2 reflectance data the normalized difference salinity index (NDSI) is computed as follow:

$$NDSI = \frac{Band11 - Band12}{Band11 + Band12} \tag{2.6}$$

The obtained spatial resolution of the Sentinel-2 NDSI images are 20 meter.

A recent study from Qian et al. (2019) indicate that the wavelengths at 1358 nm and 2382 nm are the optimal wavebands for soil salinity monitoring [Qian et al., 2019]. Only these wavelengths are not sensed by Sentinel-2

Poor vegetation areas can be caused by salt-affected soils, thus the state of stressed vegetation can be an indirect sign of the presence of salts in the soils. Salt-affected soils mostly have a distinctive feature that salt crusts appear over the soil surface. The spectral response of the SWIR bands of the salt-affected soil is higher than that of normal soils. In this research pixel comparison within a field is used to map the salt-affected areas in the fields. The values of NDSI range from -1 to +1, where as from more than zero to +1 indicates salt-affected soils.

Studies varies on the salt tolerance threshold of sugarcane [Simões et al., 2016, Wahid and Ghazanfar, 2006]. Grieve et al. (2012) state that the salt tolerance for sugarcane is 1700 $\mu S/cm$ and this value is also used by the FAO [Grieve et al., 2012, Rhoades et al., 1992].

Evapotranspiration

The circulation of water among ocean, atmosphere and land is described as the hydrological cycle. Water is transferred as precipitation, evaporation and run-off. Evapotranspiration is the flux of water from land to the atmosphere [Margulis, 2017]. Evapotranspiration consists of transpiration, interception and soil evaporation and can be described as follow:

$$ET = E_s + E_t + E_i \tag{2.7}$$

where soil evaporation is E_s, transpiration is E_t and interception is E_i. Soil evaporation is the direct evaporation of water from the top layer of bare soils. Transpiration is described as water vapor loss through the stomata of the plant [Margulis, 2017]. Interception can be described as the rainfall intercepted by the plants and it is directly evaporated from the surface of the plants. Soil evaporation, transpiration and interception are limited by certain parameters as radiation, air temperature, wind speed and water availability.

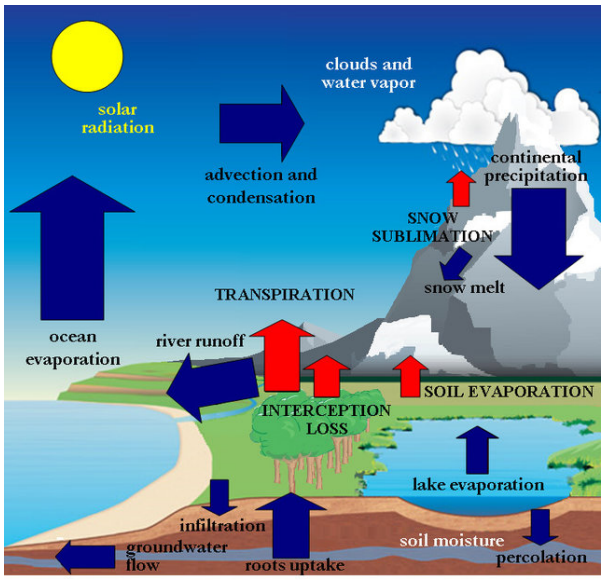


Figure 2.4: Global hydrological cycle [Miralles, 2011]. The red arrows are the different land-surface evaporation types.

Evapotranspiration can be classified as potential, actual and reference evapotranspiration [Hobbins and Huntington, 2016].

Potential Evapotranspiration is defined in Hobbins and Huntington (2016) as evapotranspiration that would take place from a well-watered surface under ambient atmospheric conditions.

Actual evapotranspiration is defined in Hobbins and Huntington (2016) as the actual transfer of moisture from the surface, which is the evaporation from bare soil, intercepted water on vegetation, and open water (and snow and ice) and transpiration from within vegetation, to the atmosphere under ambient conditions. Interception can be described as the part of rainfall that is intercepted by the earth's surface and directly evaporates [Gerrits, 2010].

Reference evapotranspiration is defined in Hobbins and Huntington (2016) as evapotranspiration that would take place from a well watered reference crop under specific surface moisture conditions but ambient atmospheric conditions.

Evaporative demand is defined in Hobbins and Huntington (2016) as the maximal rate of evapotranspiration given under ambient atmospheric condition but with an unlimited moisture supply.

The flux of water from land to atmosphere is one of the largest components of the hydrological cycle. However, evapotranspiration can not be directly measured from satellites [Martens et al., 2018]. Coenders-Gerrits et al. (2020) state that evapotranspiration is an important variable in the hydrological cycle and it gives information on crop behavior in relation to atmospheric conditions. Therefore, this thesis will look at the spatial variability of two new satellite derived evaporation products in relation to yield.

Evaporation products

In this part of the chapter the satellite derived evaporation products, VanderSat evaporation algorithm and WaPOR, are described. These products are used in this study because of their high spatial and temporal resolution. This section will focus on the methodology of the two evaporation products and especially on how they compute the actual ET estimates.

Product description VanderSat model

This section explained the VanderSat evaporation algorithm which is described in the following scientific paper:

den Besten, N., Kassing, R., Muchanga, B., Earnshaw, C., Muleya., M, de Jeu, R, Karmimi, P. and van der Zaag, P. (n.d.). Estimation of daily evaporation at field-scale for sustainable water management in a sugarcane plantation in Xinavane, Mozambique. Under review.

The newly created VanderSat evaporation model is based on the Priestley-Taylor (PT) equation [den Besten et al., nd]. In the VanderSat evaporation model the daily evaporation rate is computed using the Priestley-Taylor equation as:

$$E_{PT} = \frac{1}{\rho_w * \gamma} \frac{s(R_n - G)}{s + \lambda} * \alpha \tag{2.8}$$

Where the following symbols are:

E_{PT} potential reference evaporation flux [mm/day]

ρ_w density of water [kg/m^3]

γ Latent heat of evaporation [MJ/kg]

R_n net radiation [Wm^{-2}]

G soil heat flux [Wm^{-2}]

s slope of the saturation vapour pressure with temperature curve [$kPa \circ C^{-1}$]

λ psychometric constant [$kPa \circ C^{-1}$]

α Priestley-Taylor coefficient [-]

The model assumes that the soil heat flux can be neglected when the soil heat flux is averaged over a day [Price, 1982, den Besten et al., nd]. An important input of the model is the net radiation, which can be computed using the energy balance. The net radiation is described as:

$$R_n = (1 - r_0)R_g + L_{in} - L_{out} \tag{2.9}$$

In equation 2.9 the following symbols are:

R_n net radiation [Wm^{-2}]

r_0 the surface albedo [–]

R_g solar radiation [Wm^{-2}]

L_{in} long-wave radiation [Wm^{-2}] from the sky

L_{out} long-wave radiation [Wm^{-2}] from the Earth's surface

The long-wave radiation is computed using the Stefan-Boltzmann law, where outgoing and incoming long-wave radiations can be calculated using the surface temperature (T_s) and air temperature (T_a) respectively.

$$L_{in} = \epsilon_a[\sigma T_a]^4 \tag{2.10}$$

$$L_{out} = \epsilon_s[\sigma T_s]^4 \tag{2.11}$$

In equation 2.10 and 2.11 the ϵ indicates the emissivity for the surface and air and σ is the Stefan-Boltzmann constant ($5.67 * 10^{-8}\ W/(m^2K^4)$) [An et al., 2017].

The surface albedo in equation 2.9 can be described as the ratio of reflected to incoming shortwave radiation and depends on surface properties. More information about the computation of the surface albedo for this model can be found on the next page.

For the Priestley & Taylor coefficient two distinctions are made: between (1) a sugarcane pixel and (2) soil pixel [den Besten et al., nd]. This distinction is made by using NDVI (Normalized Difference Vegetation Index), where a pixel is allocated to a soil pixel when the NDVI values is below or equal to 0.3 [-] [den Besten et al., nd]. A pixel is allocated to a sugarcane pixel when the NDVI value is greater than 0.3 [-] [den Besten et al., nd]. As for the Priestley & Taylor coefficient (α) this model uses α=1.26 for sugar cane pixels α=1.08 for soil pixels [den Besten et al., nd]. However, it is important to know that α fluctuates [Priestley and Taylor, 1972]. Therefore, the α value can change when running this model for other study cases.

In the past the Priestley-Taylor equation has been proven that it can be used in calculating ET. The Global Land-Surface Evaporation Amsterdam Model has shown that the PT equation can be used to estimate evapotranspiration [Miralles, 2011, Martens et al., 2018].

To create actual evapotranspiration estimates, the VanderSat product uses NDVI observations as an indication for crop stress. For potential evapotranspiration estimates the model

assumes that the crop stress is zero. The relation between actual NDVI compared to the optimal NDVI (crop stress=0) was used to account for crop stress in the evapotranspiration estimates [den Besten et al., nd]. In general, the optimal NDVI ($NDVI_{opt}$) is assumed to be above the estimated NDVI curve [den Besten et al., nd]. This is because bare soil and stressed vegetation have a lower NDVI value than vegetation with no stress. Both observed NDVI and optimal NDVI are obtained as a function of cumulative growing degree days (CGDD) [den Besten et al., nd]. In this study the cumulative growing degree days is implemented as a measure of time. More information about the CGDD can be found in section 4.2.4.. The regression NDVI curve as function of CGDD for optimal is formulated in equation 2.12.

$$NDVI_{opt} = -2.779e^{-07} x CGDD^2 + 0.0008456 x CGDD + 0.3 \tag{2.12}$$

The crop stress factor is defined as:

$$F_s = \begin{cases} \frac{NDVI}{NDVI_{opt}}, & \text{for } NDVI \leqslant NDVI_{opt} \\ 1, & \text{for } NDVI > NDVI_{opt} \end{cases} \tag{2.13}$$

Albedo computation VanderSat model

For this research the VanderSat model computes the albedo by using different bands of Sentinel-2 and by using the spectral curves of clayey soil and sugarcane. The computed albedo of these wavelengths are then corrected by averaging and weighing the measurements by using equation 2.14.

$$C_f = \sum (\frac{\rho_{b,s}}{\rho_{b,rc}} * \omega_b) \tag{2.14}$$

In equation 2.14 $\rho_{b,s}$ is described as the observed reflectance of band b of Sentinel-2, $\rho_{b,rc}$ is the idealized reflectance of band b within the spectral reflectance curve of sugarcane and ω_b is a given weight of the band [den Besten et al., nd]. Finally, the MODIS BRDF-albedo model parameter are used to correct the computed albedo for overestimation caused limited observation bands of Sentinel-2 to get a daily mean albedo [den Besten et al., nd].

Product description WaPOR

As an evaporation product the WaPOR AETI dataset (Actual Evapotranspiration and Interception) is used in this research. The WaPOR AETI dataset computes the evaporation and transpiration using the ETlook model described in Bastiaanssen et al. (2012), which is based on the Penman-Monteith equation. The Penman-Monteith equation requires more input than the Priestley-Taylor equation, which is discussed above [Priestley and Taylor, 1972, den Besten et al., nd].

Penman-Monteith equation

The Penman-Monteith equation can be used to compute the evapotranspiration from a sugarcane crop [McGlinchey and Inman-Bamber, 1996]. The Penman-Monteith equation is described in equation 2.15.

$$\lambda ET = \frac{\Delta(R_N - G) + \rho c_p((e_s - e_d)/r_a)}{\Delta + \gamma(1 + r_s/r_a)} \tag{2.15}$$

In which the following symbols are:

- λ Latent heat of evaporation [Jkg^{-1}]
- E evaporation [$kg\ m^{-2}\ s^{-1}$]
- T transpiration [$kg\ m^{-2}\ s^{-1}$]
- Rn net radiation [Wm^{-2}]
- G soil heat flux [Wm^{-2}]
- ρ density of air [$kg\ m^{-3}$]
- c_p specific heat of air [$J\ kg^{-1}\ K$]
- e_d measured vapour pressure of the air [Pa]
- e_s saturated vapour pressure of the air [Pa]
- Δ slope of the saturation vapour pressure with temperature curve [$Pa\ K^{-1}$]
- γ psychometric constant [$Pa\ K^{-1}$]
- r_a average aerodynamic resistance [$s\ m^{-1}$]
- r_s bulk surface surface [$s\ m^{-1}$]

For the ETlook model an adaption of the Penman-Monteith (P-M) equation is used to predict the evaporation and transpiration [Bastiaanssen et al., 2012, FAO, 2018]. It divides the Penman-Monteith equation in two components, namely the soil evaporation and transpiration by the plant, which are formulated in 2.16 and 2.17 [Bastiaanssen et al., 2012, FAO, 2018]. The equation distinguish a plant and a soil class for the net radiation, aerodynamic resistance and the surface resistance.

$$\lambda E = \frac{\Delta(R_{N,soil} - G) + \rho c_p((e_s - e_d)/r_{a,soil})}{\Delta + \gamma(1 + r_{s,soil}/r_{a,soil})} \tag{2.16}$$

$$\lambda T = \frac{\Delta(R_{N,plant}) + \rho c_p((e_s - e_d)/r_{a,plant})}{\Delta + \gamma(1 + r_{s,plant}/r_{a,plant})} \tag{2.17}$$

Net radiation

Net radiation is an important parameter in the Penman-Monteith equation. For the AETI product the net radiation is computed as follows:

$$R_N = (1 - r_0)R_s - L - I_{energy} \tag{2.18}$$

Where the following symbols are:

- R_N net radiation [Wm^{-2}]
- R_s incoming solar radiation [Wm^{-2}]
- r_0 surface albedo [-]
- L net long wave radiation [Wm^{-2}]
- I energy dissipation due to interception losses [Wm^{-2}]

In WaPOR the net radiation is derived differently for soil and vegetation. This is because an increase in vegetation, in the form of the leaf area index, results in an exponential decrease in the fraction of radiation available for the soil. By using Beer's law, which described the attenuation of light through a material, the net radiation for the soil and vegetation computed as in formula 2.19 and 2.20. The parameter a ([-]) is the light extinction factor for net radiation.

$$R_{N,soil} = R_N * exp(-aI_{lai}) \tag{2.19}$$

$$R_{N,plant} = R_N * (1 - exp(-aI_{lai})) \tag{2.20}$$

To compute the dissipated energy due to interception losses the normalized difference vegetation index is used [eLEAF, 2020]. NDVI is used to compute the vegetation cover (c_{veg}), which is then converted into leaf area index (LAI) [eLEAF, 2020]. Equation 2.21 and 2.22 describe the conversion of NDVI index to leaf area index [eLEAF, 2020]. The first step is to convert the NDVI index to vegetation cover.

$$\begin{cases} c_{veg} = 0, & I_{ndvi} \leq 0.125 \\ c_{veg} = 1 - (\frac{0.8 - I_{ndvi}}{0.8 - 0.125})^{0.7} & 0.125 < I_{ndvi} < 0.8 \\ c_{veg} = 1 & I_{ndvi} \geq 0.8 \end{cases} \tag{2.21}$$

The second step is to convert the vegetation cover to leaf area index.

$$\begin{cases} I_{lai} = 0, & I_{ndvi} \leq 0.125 \\ I_{lai} = \frac{ln(-(c_{veg}-1))}{-0.45} & 0.125 < I_{ndvi} < 0.795 \\ I_{lai} = 7.63 & I_{ndvi} \geq 0.8 \end{cases} \tag{2.22}$$

eLEAF has derived this conversion to LAI based on literature of Carlson and Ripley (1997) and Duchemin et al (2006).

Interception

Interception data component of WaPOR can be downloaded from the database and it is used in the computation of the WaPOR AETI dataset. How much water evaporates by interception is dependent on precipitation, vegetation cover and leaf area index [Von Hoyningrn-Huene, 1983, eLEAF, 2020]. WaPOR only takes into account the precipitation and not the amount water fed to irrigated crops.

$$I[mmday^{-1}] = 0.2 * I_{lai}(1 - \frac{1}{1 + \frac{c_{veg} * P}{0.2 * I_{lai}}}) \tag{2.23}$$

Interception is related to precipitation, where a small amount of precipitation gives a high interception ratio and high amount of precipitation gives a low interception ratio. The upper limit of total intercepted evaporation is 15% [Gerrits, 2010]. The energy needed for the interception can computed with equation 2.24.

$$I_{energy} = I * \frac{\lambda}{86.400} \tag{2.24}$$

3

Project area

3.1. Mozambique

Despite of economic growth Mozambique is still one of the poorest countries in the world [IFAD, 2016]. However, Mozambique is recognized as a country with a high potential in the agricultural sector [G4AW, 2014]. In total 62% of the total area of the country is agricultural land [FAO, 2016] and in 2019 73% of the working population was employed in the agriculture sector [UN, 2019]. Mozambique faces many problems like decline in natural soil fertility and increased salinization and natural hazards such as drought, cyclones and floods [FAO, 2016].

3.2. Study site

This thesis is conducted in the context of the IWACA-TECH project, therefore the sugarcane plantation in Xinavane is used as area of interest. The Köppen climate classification system classified the project area partly as Tropical savanna wet climate (AW) and warm semi-arid climate (BSh) [Kottek et al., 2017]. The last few years, the plantations experienced seasons of floods (2012 until 2014) and droughts (2015 and 2016) [den Besten et al., nd]. The floods have affected the operations within the plantation [den Besten et al., nd]. At the plantation two meteorological stations are available. These stations record short wave incoming solar radiation ($MJ/hr/m^2$), temperature ($\circ C$), relative humidity (%), rainfall (mm/hr), and wind speed (m/s) [den Besten et al., nd]. Data from the meteorological station is used by the plantation to compute the FAO-56 Penman-Monteith evaporation [den Besten et al., nd]. Looking at the weather data the summers (December-January-February) are relatively wet and the winters (June-July-August) dry.

This thesis focuses only at a specific part of the Xinavane sugarcane plantation, which is called Aguiar. The time of interest of this performed research is the growing season of 2017-2018. The Aguiar area is located west of the village Xinavane and it is located along the Incomati river at the North. In this area irrigation water is provided by pumping out water of the Incomati river, which is part of the Incomati river basin [Khalili, 2007]. This basin is shared by the countries of South Africa and Eswatini [Van der Zaag and Carmo Vaz, 2003]. Located along the south of Aguiar is a lagoon. This still standing waterbody is used by the plantation as a reservoir. When the Incomanti river has a high water table, water is pumped to the lagoon. On the other hand when the watertable of the incomanti river is too low, the plantation is forbidden to pump the water out of the river.

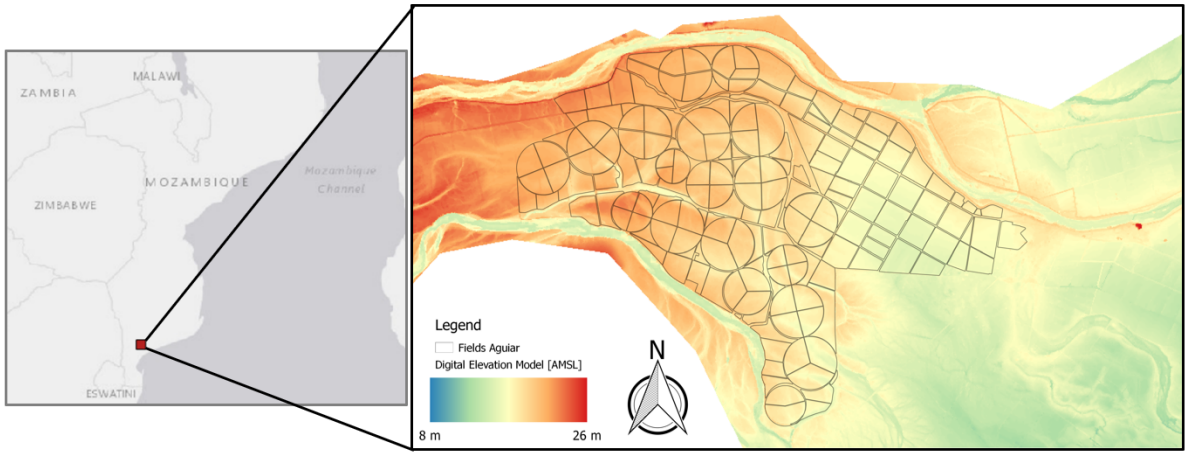


Figure 3.1: Map of location and a digital elevation model of the study site.

At the plantation the precipitation is recorded by using manual rain gauges. At the study site Aguiar 5 manual rain gauges are located. The agricultural department of the plantation keeps track off the data. In the area there are some small differences in receiving rainfall per month. For this research the time of interest ranges from May 2017 until December 2019.

3.3. Crop characteristics

Sugarcane belongs to the group of grasses. At the plantation the sugarcane is harvested yearly. However, the growing period of the fields differs. When sugarcane is harvested, a portion of the stem is left underground. After a while new ratoons emerge from the same root system. Over time the yield decreases when using the same root system.

Sugarcane is divided into four crop growth stages. The first stage is the establishment phase of the crop. This stage occurs after planting or harvesting the crop and it is characterized by the germination. The second stage is the tillering. At this stage the canopy is developed. The grand growth phase is characterized by the elongation of the stem [den Besten et al., nd]. The last stage is called the ripening phase. At the ripening phase sucrose is stored in the stem. Therefore, the fields are not irrigated in the period just before harvest. In this way the glucose will be forced in sucrose [Clements, 1962].

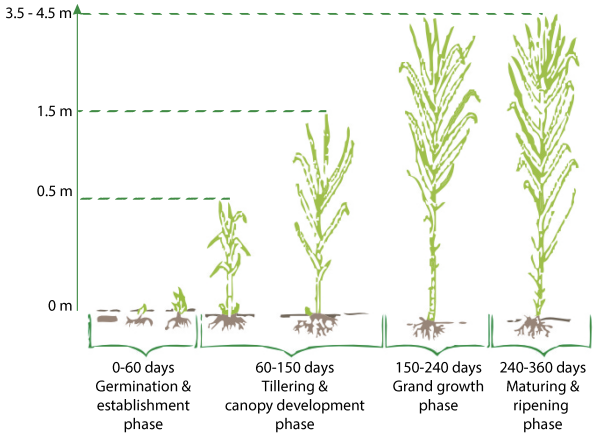


Figure 3.2: Phenological phases of sugarcane [Molijn et al., 2019]

3.4. Agricultural practices

The duration of the harvest seasons depend on the process capabilities of the local sugarcane mill. The harvest season starts generally from April until December. In Aguiar different irrigation systems as furrow, pivot and sprinklers are used. The different irrigation systems are shown in figure 3.3 and they can be described as followed:

furrow irrigation, also known as flood irrigation, is an irrigation technique which uses small parallel channels to carry water in order to irrigate the sugar cane. Usually the crop is cultivated on the ridges between the furrows.

sprinklers irrigation is an irrigation method which uses a distribution systems of pipes (sprinklers) to irrigate the land.

pivot irrigation is also called circle irrigation. Sprinklers are rotating around a pivot and applying water to a circular area.

As can be seen in figure 3.4, the area of Aguiar is divided in three subblocks called Aguiar 1,2 and 3. Aguiar 1 and Aguiar 2 consist of sprinklers and pivots irrigation systems. In Aguiar 3 furrow irrigation is the most common irrigation form but also some fields with sprinkler and pivot irrigation systems are present. Aguiar 3 is also known as an area with salinity and drainage problems.

Each field has their own field name. The first three letters of the field name indicate the subblocks Aguiar 1, 2 or 3 by using the terms AG1 , AG2 and AG3. After that the field number is given. For pivots the field name ends with a P. For sprinklers the field name ends with a S. For furrow fields the fields will not end with a letter. For example field name AG2A03P means that the pivot is located in Aguiar 2.

The size of the fields also differs. The area of the sprinkler fields ranges from 9 ha to 62. The area of the pivots ranges from 26 to 65 ha. The area of the furrow fields ranges from 3.6 to 30.6 ha.

(a) Furrow irrigation

(b) sprinkler irrigation

(c) Pivot irrigation

Figure 3.3: Irrigation systems at Aguiar

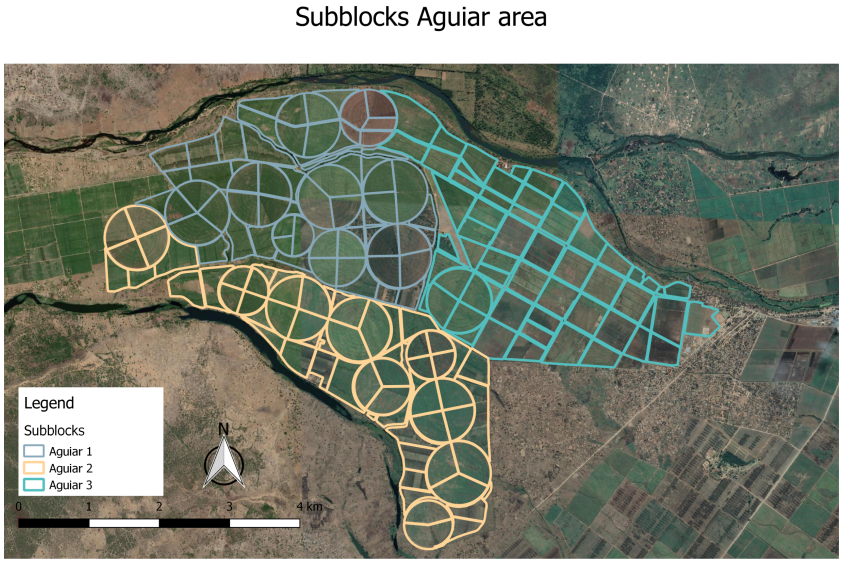


Figure 3.4: The study site is divided in three subblocks called Aguiar 1,2 and 3.

3.5. Challenges in irrigation management

Yield varies spatially in the area of Aguiar. Part of this thesis is to identify these problematic fields and mapping these field-scale spatial variability. The problematic fields with low yield can be recognized by several issues. The issues happening in the field are mainly caused by poor irrigation management. The poor irrigation management results in poor drainage and salinization issues. The main drainage canal in Aguiar is overgrown with vegetation. Due to poor drainage, still standing waterbodies occur in the areas which are linked to increasing salinization issues and shallow groundwater tables (groundwater table close to the surface) [Dinka and Ndambuki, 2014]. In Aguiar salinization issues occur mostly in areas with furrow irrigation. In furrow irrigation systems salt tends to accumulate on the top of the ridge as a saline crust, which is visualized in figure 3.5a [Pereira et al., 2002]. The saline crust reduces the infiltration of irrigation water and thus can cause higher run-off losses. The reducing infiltration of irrigation water can cause crop stress which can lead to inducing yield losses [Pereira et al., 2002].

Some fields have parts where no sugar cane grows. These fields can be recognized by bare soil or by areas where weeds have the upper hand. Especially weeds like cynodon dactlyon and Ipomoea quatica can be found in Aguiar. These crops are known to be grown in moist soils [Koenders et al., 2019]. Cynodon is known as a plant that grows under in very acid, alkaline and saline conditions [Koenders et al., 2019]. Ipomoea quatica needs significantly more water than sugarcane [Koenders et al., 2019]. The salinization problems, which occur in the fields, are an ideal condition for these weeds to grow. The weeds grow at the expense of sugarcane, because the weeds compete with sugarcane crops for the same water and nutrients in the soil. Examples of the problems discussed in this section are shown in figure 3.5.

(a) Salinity issues

(b) weeds issues

(c) Drainage canal with aquatic weeds

(d) Not growing spots

Figure 3.5: Examples of current challenges in the research area Aguiar. These pictures are made during fieldwork (July 2019).

4

Data & Methods

This chapter is divided in two section. First, data sets used in this thesis are described. Second, the data processing steps and the major data analysis methods are described.

4.1. Data

4.1.1. Data provided by the plantation and fieldwork data

This section describes the data provided by the plantation and the data collected during fieldwork. The plantation provided yield data, meteorological data and contour lines of altitude for the selected study site. Furthermore, in 2011 a soil study was executed by Instituto de Investigação Agrária de Moçambique (IIAM), the Institute of agricultural research, which is part of the Ministry for Agriculture in Mozambique. The outcome of this study is used as soil data input. During fieldwork auger measurements and root-zone measurements, groundwater measurements were collected.

Yield data

At the plantation sugarcane yield data is measured per field after harvest in tons cane per hectare (TCH). A weighing system for loaders is used by measuring the trucks with and without their sugarcane load. For this thesis it is assumed that the yield data is reliable and that measurements have a high precision. For this research yield data is collected for the years 2013 up to and including 2018.

In this study a field is classified as a good performing field, when the yield is above average. A field performs poor, when the yield is below average. Furthermore, a field is called stable, when the interannual yield variation is low. A field is called unstable when the field has a high interannual variation of yield. The thresholds are arbitrarily defined.

Meteorological data

Two meteorological stations, named Chibandza and Timanguene, are located at the plantation. The meteorological station closest to the study site is Chibandza. The meteorological data of Chibandza is used as input for the VanderSat evaporation product. At Chibandza station the radiation, precipitation, humidity, minimum and maximum temperature are measured on a daily basis. Linear interpolation is used to fill the gaps between time-adjacent measurements.

Elevation data

Elevation data is provided as contour lines obtained from LIDAR (LIght Detection And Ranging of Laser Imaging Detection And Ranging) measurements. For this study the contour lines are converted into a Digital Elevation Model (DEM). In this research the contour lines are used to create a digital elevation model (DEM).

Soil data

In 2011 the Instituto de Investigação Agrárua de Moçambique (IIAM), performed a soil research at the plantation, where the different soil types were mapped for the entire plantation [Vilanculos and Mafalacusser, 2012]. The data provided for this research were verified by auger measurements and then used as input for the field analysis. The different soil types are ranked based on their clay content, which is presented in table 4.1. Here, the FG means that soils are clayey soils and FS indicates alluvial soils of stratified texture
[Vilanculos and Mafalacusser, 2012]. The FG group is ranked on their clay content at the top part of the soil. The FS group has been divided in four groups due to their difference in physical characteristics [Vilanculos and Mafalacusser, 2012]. Figure 4.1 visualizes the categories based on clay content. The small letters **g**, **h**, ***z*** characterize a certain phase [Vilanculos and Mafalacusser, 2012]. The small letter **g** characterizes a mottling phase. This means that a layer within a 100 cm depth and thicker than 20 cm consist of 15% mottles [Vilanculos and Mafalacusser, 2012]. The small letter **h** refers to soil with a groundwater table within a depth of 50 cm for part of the year 2011 or most of the year
[Vilanculos and Mafalacusser, 2012]. The small letter ***z*** stands for a salic phase
[Vilanculos and Mafalacusser, 2012]. IIam identifies a soil as a salic phase if the soils are characterise by a EC of more than 4000 $\mu S/cm$ withinn some layer at depth of 100 cm from the surface [Vilanculos and Mafalacusser, 2012]. The different legend symbols are described in more detail in appendix A.

Table 4.1: Ranking of different soil types based on the clay content. The soil types are based on IIAM report (2011).

Ranking based on clay content	Soil type
1	FG/FGg
2	FGg/FG
3	FGg
4	FGg/FGgh + FGgz/FGgh
5	FS1
6	FS1g
7	FS1/FS2g
8	FS1/FS3
9	FS2
10	FS2g
11	FS2/FS3
12	FS3
13	FS3/FS4
14	FS4

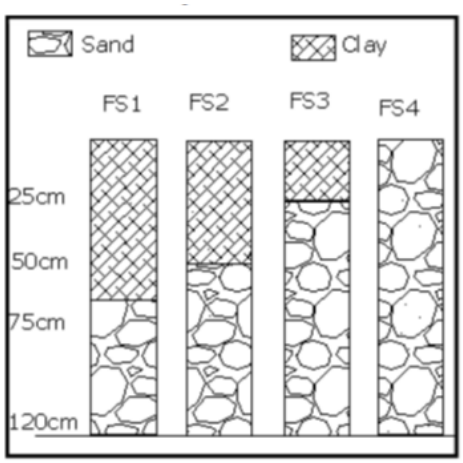


Figure 4.1: Categories of the alluvial soils of stratified texture. The alluvial soils of stratified are divided in four groups due to their variable physic characteristics. Diagram is taken from the Ilam soil report [Vilanculos and Mafalacusser, 2012].

Auger measurements

Auger measurements are executed on selected fields using an Edelman auger. The Edelman auger has a depth of 1.5 meters. While doing the auger measurements soils texture are estimated by hand. At fieldwork soils are identified as clay, clay loam, loam, sandy loam, loamy sand or sand.

The soil texture is estimated in the field as follow:

1. Take a soil sample from the field.
2. Add water to the sample and kneed the sample into a bolus/ball. Identify how the soil feels in your hand. If a bolus cannot be formed, then the soil is sand. Identify how the soil feels when kneading it: gritty (sandy), silky (silty) or plastic/sticky (clay).
3. A soil ribbon is being formed. Classify soil texture using figure 4.2.

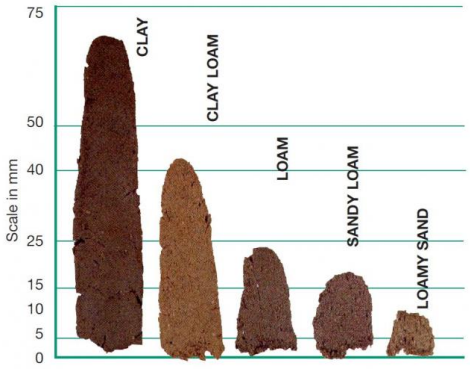


Figure 4.2: Soil ribbons classification to soil texture [WAAA, 2020].

Root-zone measurements

Using contour lines fields with a low elevation and water logging issues are selected to conduct root-zone measurements. At these location the depth of the roots are measured relatively to the soil surface with a tapeline. When the roots reach the ground water depth, it can give an indication of the water availability in the ground. The contribution of groundwater and irrigation to the sugarcane can lead over-irrigation which reduces yield [den Besten et al., 2019]. Poor subsurface drainage at irrigated fields can cause a shallow groundwater table [den Besten et al., 2019]. However, it is important to keep in mind that shallow groundwater tables can occur naturally.

Electrical conductivity & groundwater depth measurements

When conducting the auger measurements, electrical conductivity and groundwater depth measurements are taken when the groundwater table is reached. The depth of the groundwater table is measured with a tapeline and the electrical conductivity (EC) of the groundwater is measured with EC meter. The electrical conductivity determines the total amount of dissolved salt in the groundwater. The analysis of the electrical conductivity in groundwater is done with the EC meter Greisinger GMH 3400 Series. The aim of these measurements are to see if there is a relationship between poor yield fields and the occurrence of saline groundwater.

4.1.2. Sentinel-2 data

The satellite derived vegetation indices are derived using Sentinel-2 data. At the area of interest the revisit time of Sentinel-2 is 5 days under cloud-free conditions and the spatial resolution of the bands varies from 10, 20 and 60 meter. Therefore, in this study an assessment is done if the spatial variability of vegetation indices derived with Sentinel-2 data is in line with the yield.

Sentinel-2 is an optical remote sensing satellite and therefore affected by clouds. For this study Sentinel-2 images with cloud coverage ≤ 10% are used. The time of interest is May 2017 up to and including December 2018. For the area of interest in total 40 Sentinel-2 images are found with a cloud coverage ≤ 10%. The downloaded images are orthoimage Bottom of Atmosphere Level 2A products. This means that the images are corrected for atmospheric interference by making use of the Sen2Cor processor [ESA, 2020a]. The Level-2A product provides Bottom Of Atmosphere (BOA) reflectance images derived from the associated Level-1C products [ESA, 2015]. The images are acquired over Tile 36JVT . At the area of interest the overpass time of Sentinel 2 is 07:28 AM. Different bands of the sentinel-2 are used to get information about specific characteristics of the fields. Therefore the normalized difference vegetation index, normalized difference water index and the normalized difference salinity index are computed using the Sentinel-2 data. For all 40 cloud-free images the NDVI and NDWI are computed. For NDSI only the fields are selected, which have a high measured electrical conductivity level during the fieldwork. Since NDSI can be used to map salt-affected areas in fields, the first acquired cloud-free Sentinel-2 image after harvest is chosen. An overview of salt-affected fields with their corresponding harvest dates and the date of first cloud-free acquired Sentinel-2 image can be found in table 4.3.

Table 4.2: Overview of acquisition dates of cloud-free Sentinel-2 images

Month	Year 2017	Year 2018
January	-	24
February	-	13
March	-	15, 30
April	-	14, 24
May	19	9, 24, 29
June	8, 28	8, 13, 18, 28
July	3, 13, 23	13
August	2, 12, 17, 22	7, 17
September	11, 16	1, 6, 26
October	11	1, 6 , 26
November	20	5, 15
December	25	30

Table 4.3: Chosen Sentinel-2 image to compute NDSI

Fields	Date harvest	Closest S2 image
AG1D07P	14 sept 2018	26 sept 2018
AG3C32	6 jul 2018	13 jul 2018
AG2A02P	31 may 2018	8 june 2018
AG2A01P	22 nov 2018	30 dec 2018
AG2BO2S	25 jun 2018	28 jun 2018

4.1.3. Satellite-based evaporation models

In this thesis the newly satellite-based evaporation models VanderSat evaporation algorithm and WaPOR are used.

VanderSat evaporation product

Actual and potential ET estimates are computed using the VanderSat evaporation product. The evapotranspiration estimates have a spatial resolution of 20 meter and a daily to sub-daily temporal resolution. Local weather data and satellite data are used as inputs for the algorithm. An overview of the datasets used in the algorithm can be found in figure 4.3. The VanderSat algorithm uses four types of datasets namely the surface albedo, weather data, NDVI and the land surface temperature (LST). The NDVI is computed using Sentinel-2 images. As weather data the solar radiation , minimum and maximum temperature are used in the algorithm. Data gaps of weather data are filled in with linear interpolation. For weather measurements the Chibanza meteorological station is used. After 18-12-2918 no meteorological data is avalaible for Chibanza meteological station. Therefore, instead of using the Chibandza station the second meteorological station of the plantation, Timanguene, is used as input for the remaining time of interest. The surface albedo is calculated using Sentinel-2 and MODIS (Moderate-resolution Imaging Spectroradiometer) satellite sensor. MODIS is a optical satellite sensor just as Sentinel-2. The land surface temperatures are also obtained by MODIS. More information about the MODIS albedo and MODIS land surface temperature can be found here below.

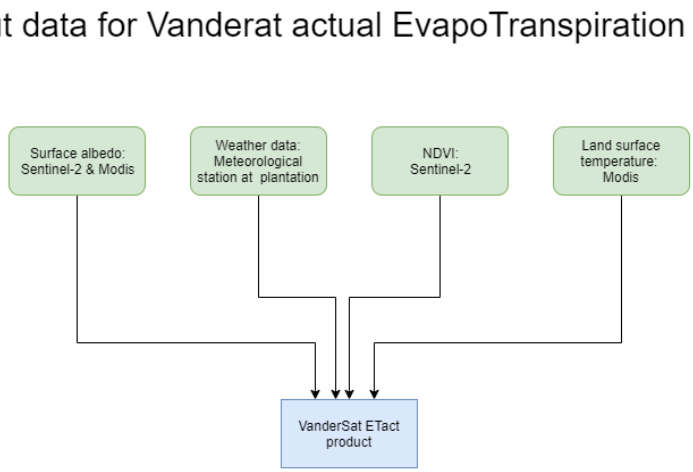


Figure 4.3: Input data components for ETact product of the VanderSat evaporation model with their sources.

For the albedo data the MODIS MCD43A3.006 product is used and it is downloaded from the AppEAARS software. The BRDF Albedo Mandatory Quality for shortwave broadband, the Black-Sky Albedo for shortwave broadband ($Albedo\_BSA\_shortwave$) and White-Sky Albedo for shortwave broadband ($Albedo\_WSA\_shortwave$) MCD43A3.006 are downloaded. The acquisition dates of the MODIS albedo data is the same as Sentinel-2. MODIS MCD43A3.006 has a temporal resolution of one day and a spatial resolution of 500 meters [USGS, 2020a]. The MODIS BRDF-albedo model parameter is used to correct the albedo for overestimation caused by limited observation bands of Sentinel-2 to get a daily mean [den Besten et al., nd]. More information on the albedo computation for this product can be found in chapter 2.

For the land surface temperature MODIS MOD11A1.006 is used and downloaded from the AppEEARS software. The land surface temperature of the day (LST_Day_1km) and land surface temperature of the night (LST_Night_1km) of MOD11A1.006 are downloaded as a point sample. The temporal resolution of the MODIS LST is one day and the spatial resolution is 1 kilometer [USGS, 2020b]. The data gaps of LST day and LST night are filled in with linear interpolation. The average LST is used as input for the VanderSat algorithm, which is formulated as:

$$LST = \frac{LST\_Day\_1km + LST\_Night\_1km}{2} \tag{4.1}$$

WaPOR

Water Productivity through Open access of Remotely sensed derived data , also called (WaPOR), of the Food and Agricultural Organization (FAO) is a database which provides near-real-time data for 21 parameters such as Actual Evapotranspiration and Interception (AETI) data, phenology data, land cover classification [FAO and IHE DELFT, 2019]. At the moment of writing the newest version WaPOR version 2.1, which was released 29th of November 2019, is used for this thesis. For Mozambique the dataset Actual Evapotranspiration and Interception (AETI) is available with a resolution of 100 by 100 meter (0.000992 degree) [FAO, 2018]. The AETI dataset used for this thesis has a dekadal temporal resolution. So for every tenth day, twentieth and last day of the month an image is available, which supposed to provide an average daily value for AETI in a given dekad [FAO and IHE DELFT, 2019].

WaPOR provides AETI data with a dekadal temporal resolution and 100 meter spatial resolution for January 2009 until present. For this thesis, data from the period May 2017 to December 2018 were obtained from the data portal. This portal can be accessed using the following link: `https://wapor.apps.fao.org/home/WAPOR_2/1`. A conversion factor of 0.1 needs to be applied on the dekadal AETI dataset. The data uses a spatial reference system of EPSG:4326-WGS84, which is, in this case study, a different coordinate system then the data provided by the plantation. Except for AETI WaPOR provides also reference ET data and precipitation data. In table 4.4 the resolution of the dekadal dataset of AETI, the reference ET and percipitation can be found.

Table 4.4: Relevant dekadal data from WaPOR model with its resolution for this study

Dataset v2.1 WaPOR	Spatial resolution
AETI	0.000992 degree (±100m)
ETref	0.17 degree (±20 km)
Percipitation (CHIRPS)	0.05 degree (±5 km)

For the WaPOR AETI dataset the evaporation and transpiration are computed using the ETlook model described in chapter 3. The actual ET product of WaPOR uses different kind of data as input, which is visualized in a flowchart in figure 4.4. The blue colored boxes indicate WaPOR data products, which can be downloaded from the WaPOR portal. A small description of each input can be found below. Furthermore, to understand the WaPOR data it is important to know that only the optical data is available at a 100 m meter resolution. The other input data all have a lower resolution. Since the spatial variability of these input data is limited by the coarse resolution, this can affect the spatially variability of the AETI product.

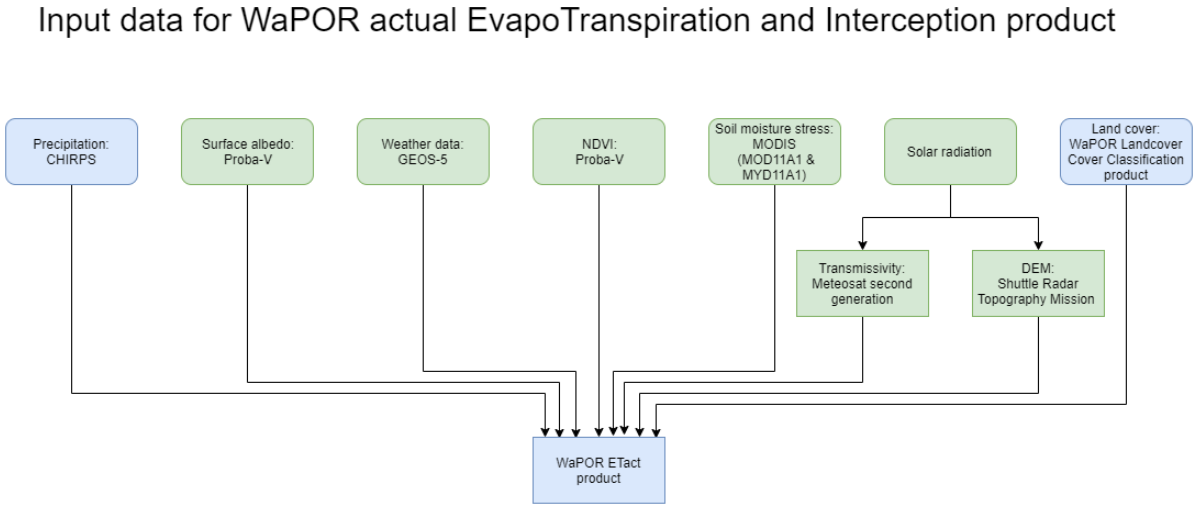


Figure 4.4: Input data components for AETI product of WaPOR with their sources. Blue boxes indicate WaPOR data products, which can be downloaded freely from the WaPOR portal

Precipitation: CHIRPS

CHIRPS (Climate Hazards Group InfraRed Precipitation with Station) are rainfall estimates obtained from rain gauge and satellite observations. In the WaPOR portal total daily rainfall estimates are available with a spatial resolution of approximately 5 km (0.05 degree) [FAO, 2018]. When data gaps occur in the CHIRPS data, the data gaps will be filled with other data sources to get daily precipitation data in the WaPOR portal [FAO, 2018]. The WaPOR portal also provides precipitation with a dekadal, monthly and annual resolution. This data is obtained by taking the daily precipitation and summing the days of a dekad, month or year to get the total amount of precipitation, and dividing the total by the number of days of that period [FAO, 2018].

Surface albedo and NDVI: PROBA-V

For both the albedo and NDVI the PROBA-V satellite is used. PROBA-V (Project for On-Board Autonomy - Vegatation) is a satellite developed as part of the earth observation mission [ESA, 2020b, Francois et al., 2014]. The PROBA-V has a daily temporal resolution for the latitudes 35-75°N and 35-56°S, and a two day temporal resolution between the latitudes of 35°N and 35°S [ESA, 2020b, Francois et al., 2014]. PROBA-V measures in the visible and near infrared (VNIR) and short-wave infrared (SWIR) region of the spectrum with four spectral bands, namely blue, red, NIR and one band in the SWIR spectral range [ESA, 2020b, Francois et al., 2014]. PROBA-V has a spatial resolution of 350 m and 660 m at the swath extremes for the VNIR and SWIR channels, respectively [Wolters et al., 2018]. It can however, be up to 100 m by 100 m at nadir [Wolters et al., 2018, Francois et al., 2014]. The SWIR band has a spatial resolution of 600 meter [ESA, 2020b]. Final PROBA-V products are disseminated at 100 m, 300 m and, 1 km resolution [Wolters et al., 2018]. For the WaPOR portal only the PROBA-V dataset with a resolution of 100 m is used [FAO, 2018]. For the surface albedo a dekadal temporal resolution is used as input [FAO, 2018].

Weather data: GEOS-5

GEOS-5 (Goddard Earth Observing System Model) is a global atmospheric model. WaPOR uses this model to acquire meteorological parameters such as temperature, wind speed and relative humidity [FAO, 2018]. These parameters are also available at most meteorological stations. However, since WaPOR provides data for the entire African continent and Arabian peninsula, the areas covered by these meteorological stations are small [FAO, 2018]. Due to lack of meteorological stations the GEOS-5 global atmospheric model is used [FAO, 2018]. GEOS-5 runs globally between a resolution of 3.5 km up to a resolution of 28 km [Putman and Suarez, 2011]. The spatial resolution of the input weather data for WaPOR is approximately 25 km (0.25 degrees) and it has a daily temporal resolution [FAO, 2018]. In order to match the resolution of the WaPOR product level, the GEOS-5 data will be resampled using a bilinear method [FAO, 2018]. Since temperature is affected by elevation, the temperature data will additionally be resampled using the digital elevation model [FAO, 2018], where WaPOR uses a temperature lapse rate of $6°C/km$ [FAO, 2018].

Soil moisture stress - Modis

For soil moisture stress the daily land surface temperature (LST) from Modis MOD11A1 and MYD11A1 are used as input. The daily Modis LST has a spatial resolution of 1 km. The soil moisture stress in WaPOR has a dekadal temporal resolution and is processed at a resolution of 250 m, 100m and 30m, which are the resolutions of the different level products of WaPOR. The calculation of the soil moisture stress is based on the triangle method described by Yang et al. (2015) and described in FAO (2018), the WaPOR methodology report.

Transmissivity: Meteosat second generation

The daily transmissivity (τ), which is a measure of the amount of solar radiation that propagated through the atmosphere, is derived from Meteosat second generation (MSG) [FAO, 2018]. The atmospheric transmissivity has a spatial resolution of 4 km [FAO, 2018]. Solar radiation, which is computed using DEM data and transmissivity, is computed at all three product levels with a resolution of 250m, 100m an 30m by resampling the input data.

DEM: Shuttle radar topography mission
The Shuttle Radar Topography Mission (SRTM) provides land elevation data. The SRTM data, used in WaPOR, is at a 90 meter spatial resolution [Rabus et al., 2003, FAO, 2018]. The elevation data is acquired by C-Band interferometric radar and a X-band radar on board of the space shuttle endeavour [Rabus et al., 2003]. For WaPOR the DEM of the SRTM is resampled to 25km (0.25 degrees), for the reason that it has the same pixel resolution as the temperature of the global atmospheric model [FAO, 2018].

WaPOR land cover classification product
The Level 2 WaPOR land cover classification product provides land cover classification with the classes: rainfed cropland, irrigated cropland, tree cover, shrubland, grassland, wetland, artificial land, bare soil, permanent water bodies and seasonal waterbodies [FAO, 2018]. The product will be updated yearly and the main goal of the product is identifying cultivated land and, more specifically, to make a distinguish between irrigated and rainfed areas [FAO, 2018]. It has, just as the other Level 2 products, a spatial resolution of 100 meter (0.000992 degree). To obtain a landcover classification product WaPOR uses NDVI time series, seasonal phenology information, EVI, multispectral imagery and as reference data the Copernicus Global Land Cover maps (C-GLOPS) at a 100 meter resolution [FAO, 2018].

4.2. Methods

In this section the previously mentioned data are combined to assess whether the spatial variability of yield at the sugarcane plantation can be explained with topography, soil and groundwater properties and whether the spatial patterns in satellite derived vegetation indices and satellite derived evaporation products reflect the difference in crop performance. The data-processing steps are explained in this section and divided in three subsection: (1) data processing for field analysis, (2) data processing for satellite derived vegetation indices and (3) data processing for satellite derived evaporation products.

4.2.1. Data processing: field analysis

A digital elevation model is created in python and QGIS using the contour lines. Figure 4.5 shows the processing steps made to create the digital elevation model in a diagram.

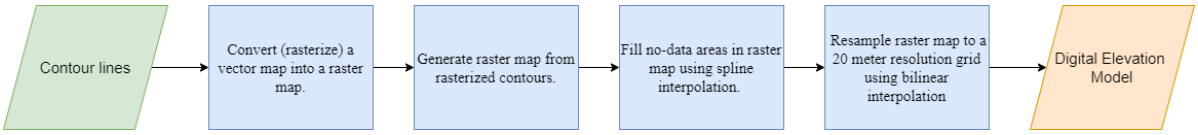


Figure 4.5: Diagram of the processing steps to create a digital elevation model.

The first step is to convert the vector map of the contour lines into a raster map with contour lines in QGIS. The second step is to create a elevation map from the rasterized contour lines. By using the GRASS function r.surf.contour, a linear interpolation of the rasterized contour lines is applied to compute the elevation for a certain pixel size. The rasterized elevation has some data gaps. These no-data areas are filled using regularized spline with tension interpolation. Literature shows that this method can be applied for terrain modeling [Mitášová and Hofierka, 1993]. For filling the no-data areas an spline tension parameter of 40 and spline smoothing parameter of 0.1 is used. At last, the created digital elevation map is resampled to a map with 20 meter resolution using bilinear interpolation. Bilinear interpolation uses the 4 nearest pixel cells to generate a new map.

The digital elevation map with a resolution of 20 meter is used as input to execute different terrain analysis maps, such as a convexity, a HAND map and a slope map. The processing steps to create these maps are visualized in a diagram in figure 5.2. Moreover, figure 5.2 shows how the data is connected to execute a correlation study between yield and topographic properties. In the correlation study the relationship between the following parameters are compared to each other:

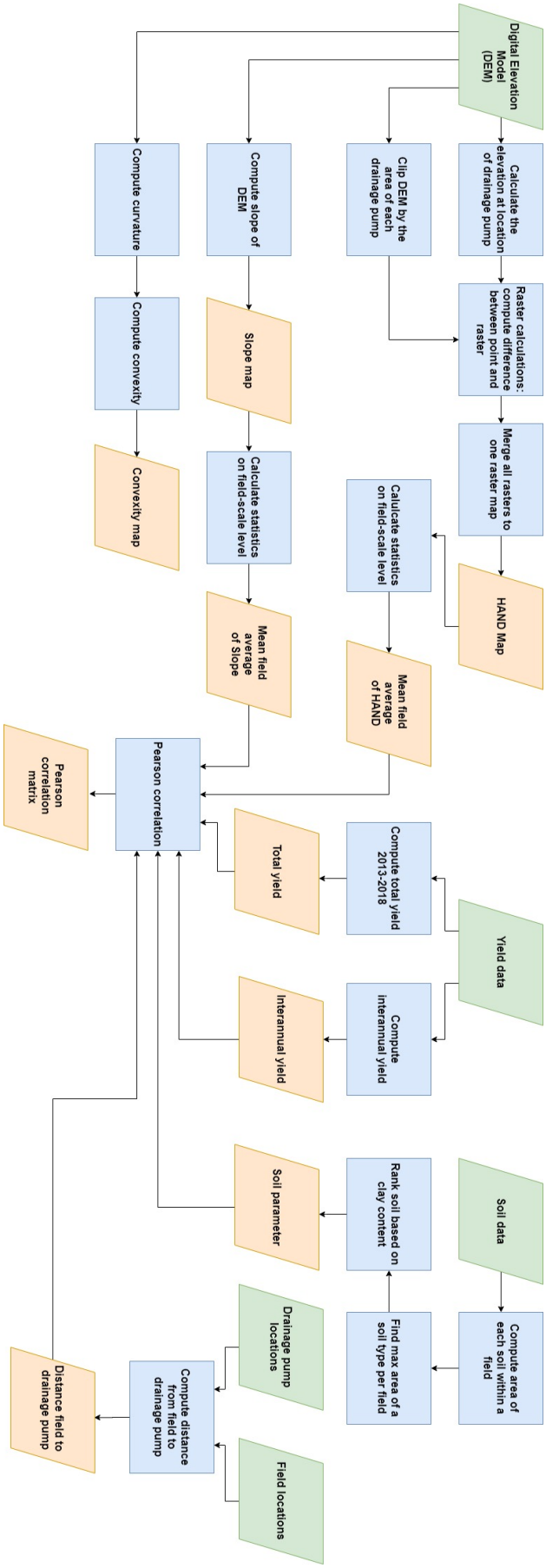


Figure 4.6: Flowchart of correlation study

1. Total yield
The total yield is the sum of yield data for the year 2013 up to an including the year 2018.

2. Interannuel yield
The interannual yield gives information about the stability of the performance of the fields. When the interannual yield variation is low, the field has a stable performance. When the field has a high interannual variation of yield, the field has an unstable performance.

3. Slope
The slope represents the rate of change of elevation. A slope map is created with a spatial resolution of 20 meter. For the correlation study the average/mean slope is extracted on field-scale level using a shapefile with the fields.

4. HAND
A map with the height above nearest drainage (HAND) map is produced. In this study the HAND is computed using the height above the nearest drainage pump. There are three drainage pumps in the study area. Each field is connected to a drainage pump by a network of drainage canals. The HAND is produced on a pixel basis as follow:

$$HAND\ pixel = DEM\ pixel\ - height\ above\ nearest\ drainage\ pump \tag{4.2}$$

The HAND normalizes topography according to the local relative heights found at the drainage pump [Nobre et al., 2011]. The HAND presents the topology of the relative soil gravitational potentials [Nobre et al., 2011].

5. Convexity
The convexity map can give an indication where potential waterlogging areas are located. In concave areas water is likely to pool and therefore evaporates. At convex areas runoff of irrigation can take place. Concave areas are areas which have a negative curvature, whereas convex areas are areas with a positive curvature.

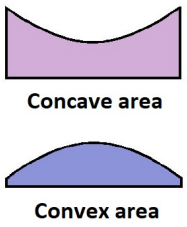


Figure 4.7: Convex and concave shape

6. Soil parameter
As discussed in section 4.1.1 the soil types are ranked based on their clay content. Using the shapefile with the fields the area of each soil type is computed within the field. For the correlation study all the data is on field-level. Therefore, the soil type representing the largest surface area within that field is selected as the soil type for that field.

7. Distance from field to drainage pump
The flow path length, which follows the drainage canals, of each field to the outlet (drainage pump) is computed. The assumption is made that the outlets are the drainage pumps. The flow path length is manually computed in QGIS by creating a shapefile as a line from the field, where the closest corner of the field to the drainage pump is chosen, to drainage pump. The drainage system removes excess water from the irrigated land. This excess water may be e.g. waste water from irrigation or surface runoff from rainfall. How longer the flow path length, how more water will be leaked away from the drainage system to the other fields. In literature it is stated that the water flow is negatively correlated with the flow path length [Zhangzhong et al., 2016].

4.2.2. Data processing: satellite derived vegetation indices

For the satellite derived vegetation indices the Sentinel-2 data is processed to get the vegetation indices NDVI and NDWI. The main processing steps of analyzing satellite derived vegetation indices are visualized in a flowchart in figure 4.8. The disadvantage of optical images is the limited availability of data with a consistent temporal resolution due to cloud coverage. In this study the optical images of Sentinel-2 are filtered on cloud coverage. In this study only images are used which have a cloud coverage less or equal than 10 %. The vegetation indices NDVI and NDWI are computed from the Sentinel-2 data. The images are clipped to the size of the study area and linear interpolation is applied between the time-adjacent images to get daily values. Linear interpolation can be applied, since NDVI and NDWI also contained information about the vegetation canopy [Gao, 1996]. Using a shapefile containing the fields, polygon statistical parameters such as mean (μ), standard deviation (σ) and coefficient of variation (CV) are computed daily for the time series. Per field the sugarcane is planted and harvested on different days in the year. Therefore, as a measure of time the cumulative growing degree day (CGDD) is used. When using the cumulative growing degree days the same time index can be implemented for all fields. The NDVI and NDWI data is corrected for the effects of temperature on phenology by using the cumulative growing degree day [Lofton et al., 2012, den Besten et al., nd]. More information on how to compute the cumulative growing degree day can be found in section 4.2.4. To analyse the satellite derived vegetation indices different data analysis tools are applied on the data. For example the correlation between yield and the vegetation indices are computed throughout the growing season. The spatial distribution of the slopes has been used for time series analysis and vegetation performance analysis.

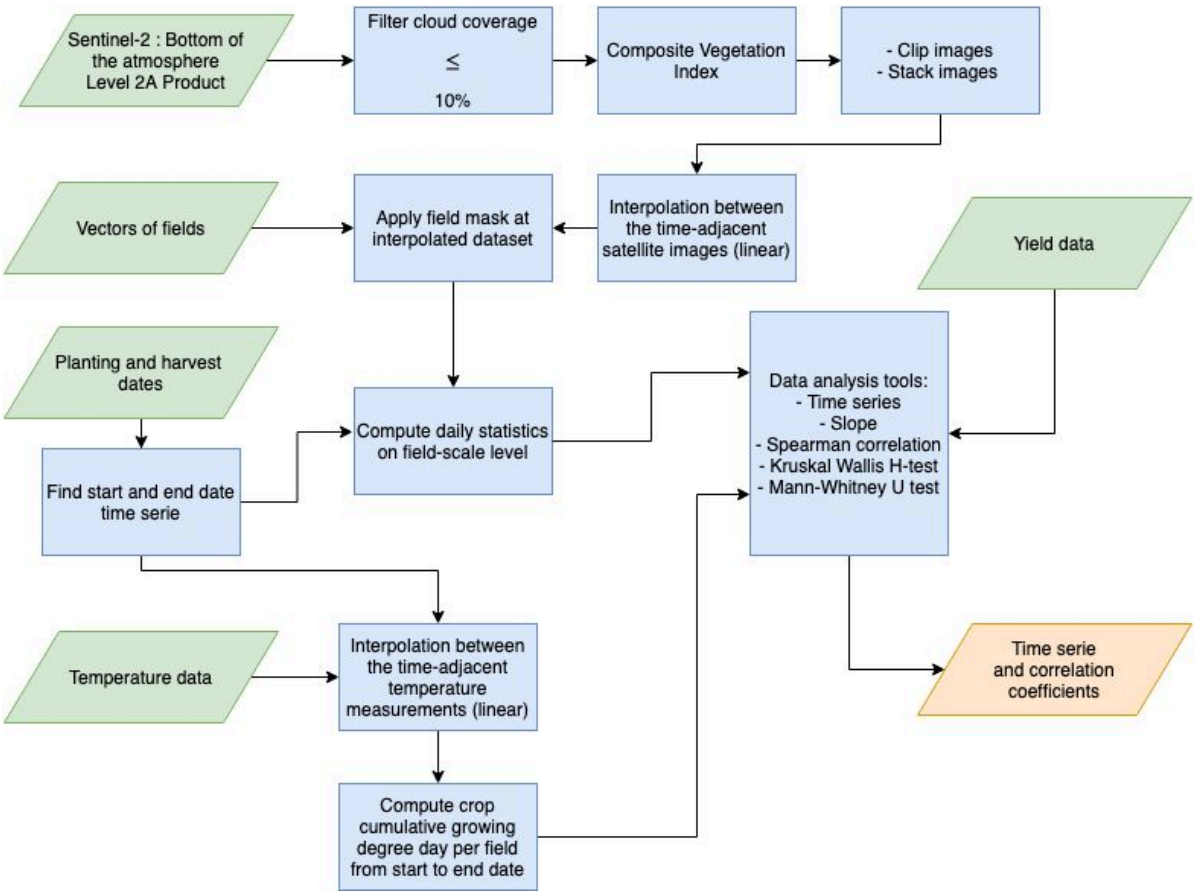


Figure 4.8: Flowchart diagram of Sentinel-2 data processing. The green boxes represents the input data. The orange box represents the output data. The blue boxes represents the processing steps.

4.2.3. Data processing satellite derived evaporation indices

This section focuses on the data processing steps of the ET data to the data presented in the results. There are some small changes between the processing steps of the ET data from VanderSat and WaPOR. Flowcharts are presented in figures 4.9 and 4.10, which show the processing steps for VanderSat and WaPOR, respectively. The Vandersat ET estimates have a resolution of 20 meter. For WaPOR it is important to know that a conversion factor of 0.1 needs to be applied before using the AETI data. Furthermore, the AETI data of WaPOR has a resolution of 100 meter. These images are resampled to a resolution of 20 meter, so that WaPOR and VanderSat ET datasets have the same resolution. For WaPOR every tenth day, twentieth and last day of the month an AETI image is available, which supposed to provide an average daily value for AETI in a given dekad. Hence, linear interpolation is applied between the images to get daily ET estimates. Moreover, for VanderSat actual and potential ET data, linear interpolation is applied to fill in the no-data gaps. As with the satellite derived vegetation indices CGDD is used as a measure of time. To analyse the satellite derived evaporation products, different data analysis tools are applied on the data. For example the water productivity as well as the correlation between yield and the actual ET are computed throughout the growing season. VanderSat has provided potential and actual ET data. Using the VanderSat datasets the difference between potential and actual ET and the ratio of potential and actual ET are computed. A large difference between potential and actual ET indicates more stress. The same applies to the ratio of potential and actual ET. The lower the ratio value, the larger the crop stress.

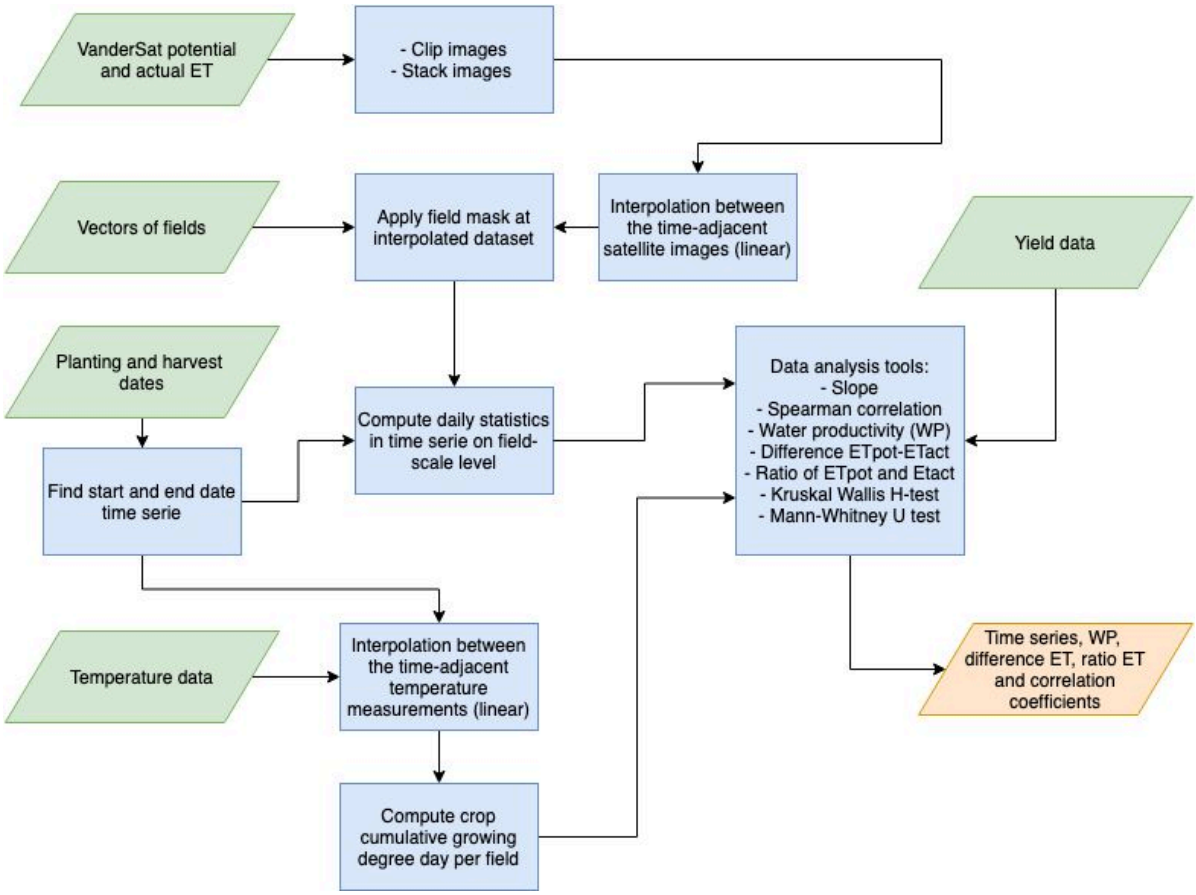


Figure 4.9: Flowchart diagram of VanderSat evaporation algorithm data processing. The green boxes represents the input data. The orange box represents the output data. The blue boxes represents the processing steps.

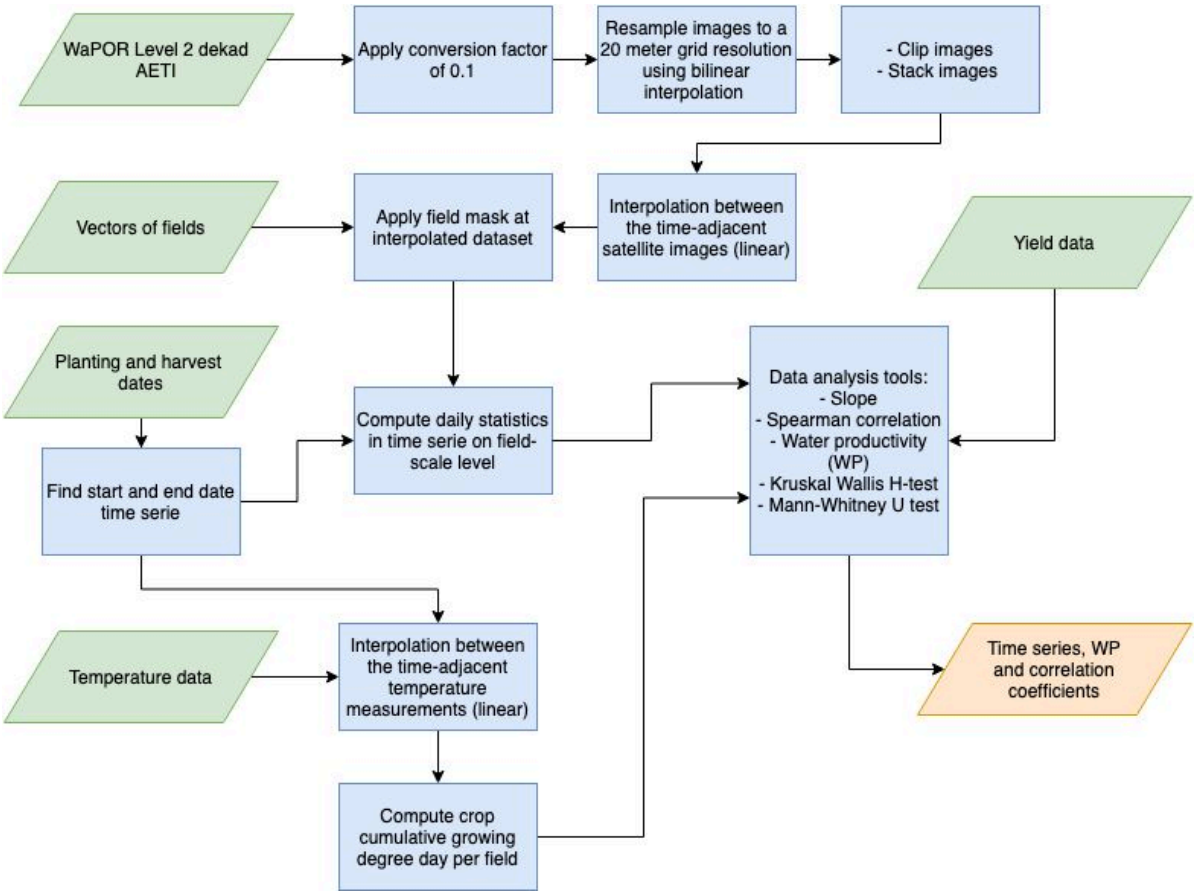


Figure 4.10: Flowchart diagram of WaPOR evaporation data processing. The green boxes represents the input data. The orange box represents the output data. The blue boxes represents the processing steps.

4.2.4. Data analysis tools

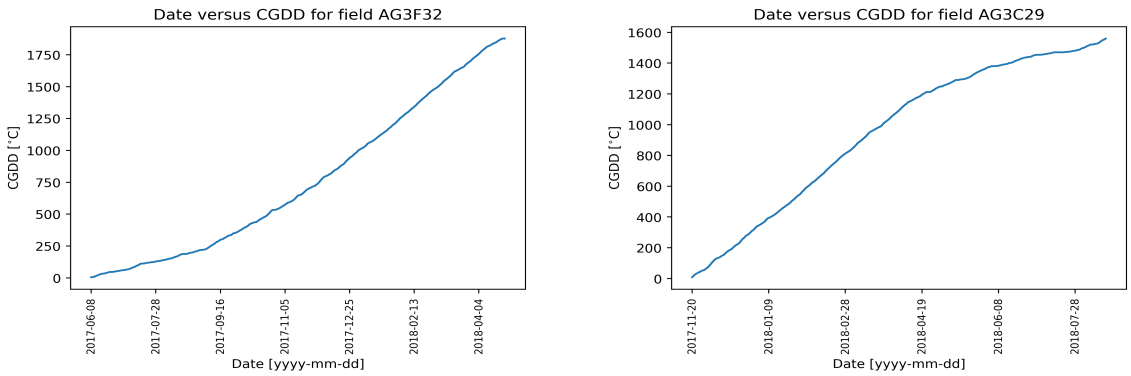
The different analysis tool used in this research are described in this section.

Cumulative Growing Degree Days (CGDD)

Since sugarcane is planted and harvested at the plantation at different times in the year, cumulative growing degree days is used to implemented as a measure of time for all the fields. Using the CGDD the data is corrected for the effects of temperature on phenology [Lofton et al., 2012, den Besten et al., nd]. The cumulative growing degree days are the total sum of all growing degree days when added together over time. A growing degree day is computed as follow:

$$Growing\ Degree\ Day\ [°C] = \frac{(Temp_{max} - Temp_{min})}{2} - Base\ Temperature \tag{4.3}$$

where:
$Temp_{max}$ = maximum daily temperature;
$Temp_{min}$ = minimum daily temperature;
Lofton et al. (2012) state that the base temperature is 18 °C for sugarcane production. The base temperature can be used as an indication of a minimum temperature for the crop to grow and to develop [Steduto et al., 2012]. When the average temperature of a day is below 18 °C, then the growing degree day is set as 0. Figure 4.11 visualizes the date versus the CGDD for two fields, which have different planting and harvest dates.



(a) (b)

Figure 4.11: Date versus Cumulative Growing Degree Days for the fields (a) AG3F32 and (b) AG3C29

Cumulative mean

The mean is the average outcome of all values for a field. In the thesis the cumulative mean is computed per field for a specific parameter, for example actual ET. The cumulative mean indicates the total sum of all mean values for a field when added up to a certain point in time, which is for this thesis given in CGDD.

Cumulative standard deviation

The standard deviation gives information about how the values are spread out from the mean. A low standard deviation means that most of the values are close to the average. A high standard deviation means that the values are more spread out. Cumulative standard deviation is the sum of all standard deviation values up to a point in time, which is for this thesis expressed in CGDD.

Coefficient of Variation

In this thesis the coefficient of variation is used a measure of relative variability within a field. The coefficient of variation (CV) is the ratio of the standard deviation to the mean, which is formulated in equation 4.4. In this thesis the CV is used as mathematical expression for relative variability of a specific component within a field. In the thesis the cumulative CV is computed, which means that the cumulative coefficient of variation is the total sum of all coefficients of variation when added up to a point in time.

$$Coefficient\ of\ Variation = \frac{standard\ deviation}{mean} \tag{4.4}$$

Correlation coefficients

In this thesis the field data in relation to yield is computed using Pearson correlation coefficient. Moreover, the remote sensing data for specific days in relation to yield is computed using the Pearson correlation coefficient. The interpretation of the correlation coefficient used for this thesis is presented in table 4.5.

Table 4.5: Meaning of correlation coefficient value

Correlation coefficient value	Interpretation correlation
-1	Perfectly negative
-0.8	Strongly negative
-0.5	Moderately negative
-0.2	Weakly negative
0	No correlation
0.2	Weakly positive
0.5	Moderately positive
0.8	Strongly positive
1	Perfectly positive

Pearson correlation coefficient

The Pearson correlation coefficient determines the degree to which a relationship is linear [Laerd Statistics, 2018a]. The Pearson correlation coefficient ranges from -1 to 1, where 1 means a positive correlation, -1 means a negative correlation and 0 means no correlation. For using the pearson correlation coefficient variables must be normally distributed and there should be a linear relationship between the two datasets.
Pearson correlation coefficient is computed as follow:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m^y)^2}} \tag{4.5}$$

Looking at equation 4.5 the m_x and m_y represent the mean of variable x and the mean of variable y, respectively.

Spearman rank correlation coefficient

The Spearman rank correlation computes the nonparametric measure of a monotonic relation between two datasets [Laerd Statistics, 2018b].

A monotonic relationship can be explained as a relationship between two datasets, which does the following things:

- if one variable increases, so does the other variable [Laerd Statistics, 2018b];
- if one variable increases, the other variable decreases [Laerd Statistics, 2018b].

The Spearman correlation does not assume that both datasets are normally distributed, where the Pearson correlation assumes that they are both normally distributed [Laerd Statistics, 2018b].

In this thesis the Spearman rank correlation is computed as equation 4.6.

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{4.6}$$

For this thesis the Spearman rank correlation is used to determine correlation between different field parameters. Furthermore, it is used to determine the relationship between remote sensing data, such as NDVI and NDWI, and yield data. Moreover, the Spearman rank correlation is computed to determine the relationship between the satellite-based evaporation algorithm and the yield data. The Spearman rank correlation is computed in python using the scipy package for the satellite data and pandas.Dataframa.corr() for the field data. Using the pandas dataframe correlation function the correlation of columns are compute pairwise and excludes the not a number and/or null values.

Kruskal-Wallis H-test

In this thesis the Kruskal-Wallis H-test is used to compare different datasets. The Kruskal-Wallis H-test is a nonparametric version of the one-way analysis of variance test [Brownlee, 2019]. The null hypothesis (H0) for the Kruskal-Wallis H-test is that all data samples were drawn from the same distribution [Brownlee, 2019]. A rejection of the H0 indicates that one of the datasets is significant different than the other datasets. When the test fails to reject H0 it indicates that the datasets are not significant different than the other datasets. It is important to know that the Kurskal Wallis H-test does not show where and what is different [Brownlee, 2019]. The Kruskal-Wallis H-test is calculated in python using the scipy package. It is used to compare the datasets of the different irrigation systems for a similar parameter. A disadvantage of the Kruskal-Wallis H-test is that it only reject the H0 hypothesis if the majority of the distributions of the datasets are significant from each other and not all datasets. To go more in depth, which distributions of the datasets are significant different from each other, the Mann-Whitney U test is executed.

Mann-Whitney U test

The Mann-Whitney U test is a nonparametric test, which determines if two independent datasets have the same distribution [Brownlee, 2019]. The null hypothesis (H0) is that the datasets have the same distribution [Brownlee, 2019]. A rejection of the H0 indicates that the two datasets are significantly different from each other. When the test fails to reject H0 it means that the datasets are not significantly different compared to the other datasets. The Mann-Whitney U test is computed in python using mannwhitneyu() scipy function. The function is used to compare if the distribution of the datasets for different irrigation system are equal.

Water productivity

In addition to yield data, the water productivity is computed to express the production of the crop per unit of water consumed [Molden et al., 2010]. In this study water productivity (WP) is measured using equation 4.7.

$$WP\ (\frac{kg}{m^3}) = \frac{Yield\ (ton/ha)}{\sum ET_{act}\ (mm)} * 0.1 \tag{4.7}$$

By multiplying the ratio of yield to ET_{act} with 0.1 the water productivity can be expressed in units of kg/m^3, because 1 ton/ha is 0.1 kg/m^2. Water productivity gives an indication whether the water use efficiency is improved and gives the difference between water use efficiency between fields [Molden et al., 2010]. Molten et al. (2010) state that the methods to increase the water productivity are:

- increasing the crop productivity per unit of ET at field-scale.
- reducing non-productive water consumption, which will have a effect on reducing salinization occurrence.
- improving irrigation management and irrigation equipment.
- re-allocating and co-managing water.

5

Results & Discussion

In this chapter the results, which are relevant to the research questions, are shown. The results of this research are subdivided into five parts: analysis of yield data (section 5.1), the field data (section 5.2), normalized difference salinity index (section 5.3), the satellite derived vegetation index (section 5.4) and the satellite-based evaporation products (section 5.5).

5.1. Yield analysis

At the plantation sugarcane yield is measured on field-scale basis after the harvest in tons cane per hectare (TCH). To understand the spatial variability and consistency of the yield performance, yield data for the years 2013 up to and including 2018 are analysed. Looking at figure 5.1 the yield is decreasing over years from 2014 until 2017. The boxplot shows that for the year 2018 a slight increase in yield occurs. Note that when plotting the yield data as maps on field-scale level for the years 2013-2018, this slight increase in the yield data for the year 2018 can be explained by not growing sugarcane on fields which have a poor yield the year before. These yield maps for the years 2013-2018 can be found in figure 5.2. Figure 5.1 shows that for the year 2013 the yield median is lower than for the years 2014 and 2015. Additionally, as can be seen in table 5.1, the median of 2013 is higher than that of the years 2016, 2017 and 2018. Since sugarcane is a semi-perennial crop, the crop can be harvested without replanting [Lofton et al., 2012]. It is known that the yield decreases over time when using the same ratoon [Molijn et al., 2019]. The effect of using the same ratoon is not investigated in the data and it should be wise to take into account when doing yield analysis for sugarcane. Furthermore, weather events such as drought and floods could have an effect on the yield performance of the crop. The plantation experienced floodings in 2012,2013 and 2014 [den Besten et al., nd]. For the years 2015 and 2016 a drought occurred in the area [den Besten et al., nd]. Yield data needs to be collected over a longer time period (>10years) in order to estimate the effects of droughts and floodings. The yield data only spans from 2013 up to and including 2018 and therefore, the influence of droughts and floodings cannot be estimated and eliminated from the yield data. This needs to be taken into consideration for the yield analysis.

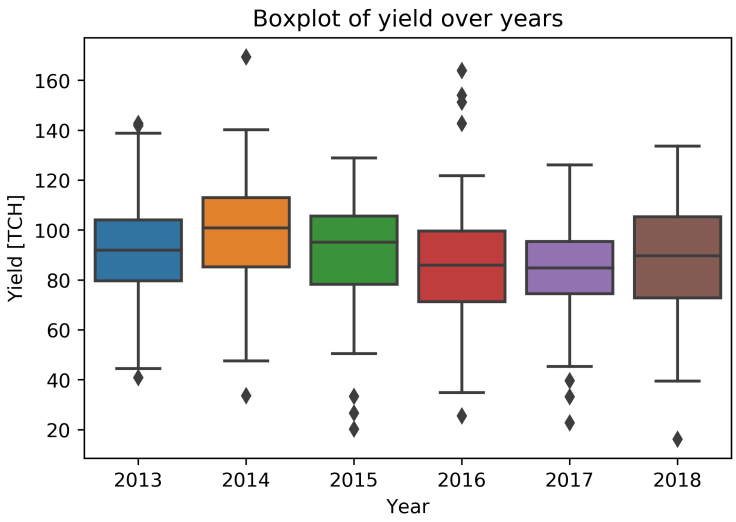
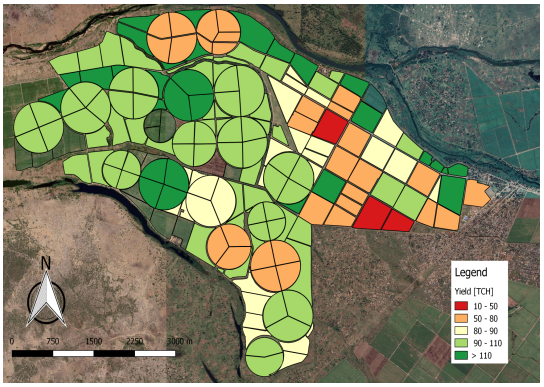


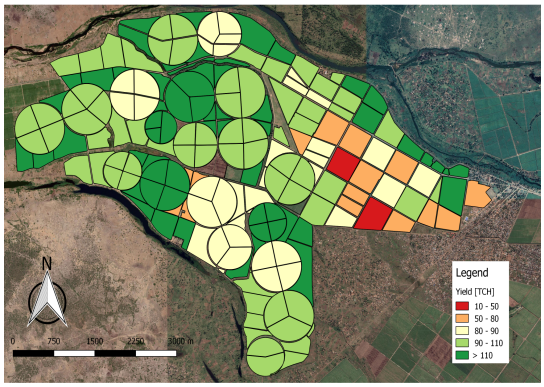
Figure 5.1: Boxplot of yield for the 2013 - 2018 for the entire Aguiar area. At the study site floods occurred in the years 2013 and 2014 and for the years 2015 and 2016 a drought occurred.

Table 5.1: Yield statistics of the study site for the years 2013-2018

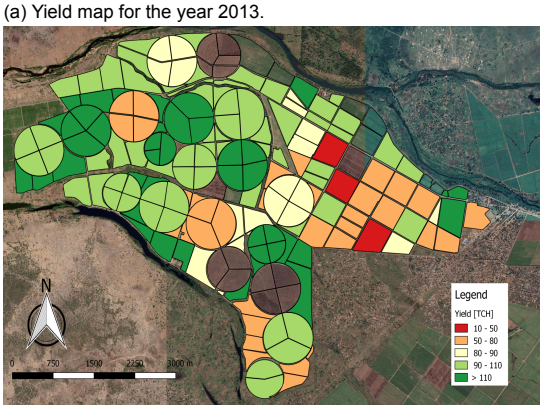
Yield per year	Median [TCH]	Std [TCH]	Var [TCH]	Mean [TCH]
2013	91.9	22.1	486.6	91.8
2014	100.9	22.4	494.6	98.9
2015	95.2	22.9	523.3	90.2
2016	86.0	26.6	706.1	87.2
2017	84.8	19.6	385.3	84.1
2018	89.8	22.5	504.9	87.7



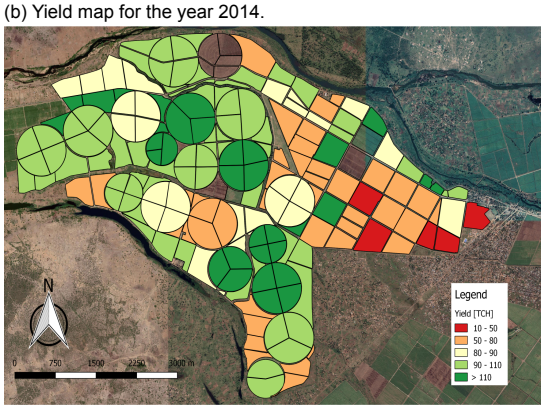
(a) Yield map for the year 2013.



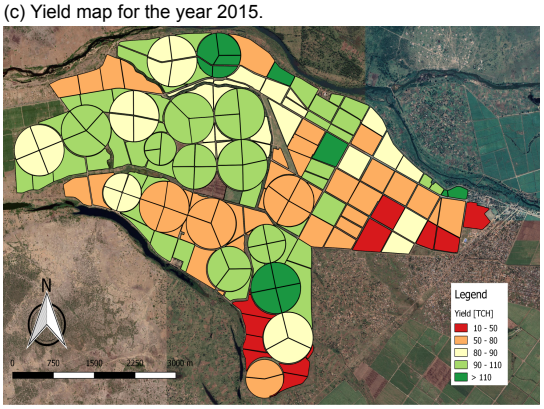
(b) Yield map for the year 2014.



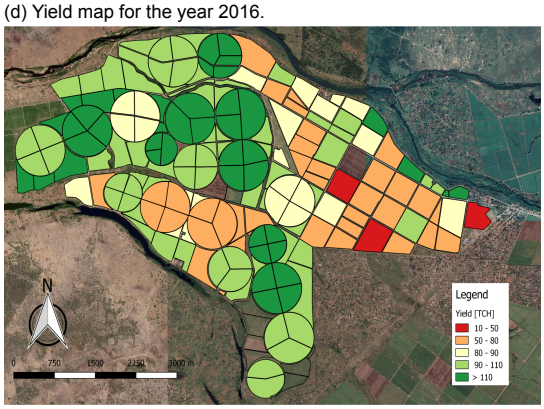
(c) Yield map for the year 2015.



(d) Yield map for the year 2016.



(e) Yield map for the year 2017.



(f) Yield map for the year 2018.

Figure 5.2: Yield map for the years 2013-2018. Color indicates the yield value between a specific range. No color indicates that the field was not planted and harvested that specific crop season.

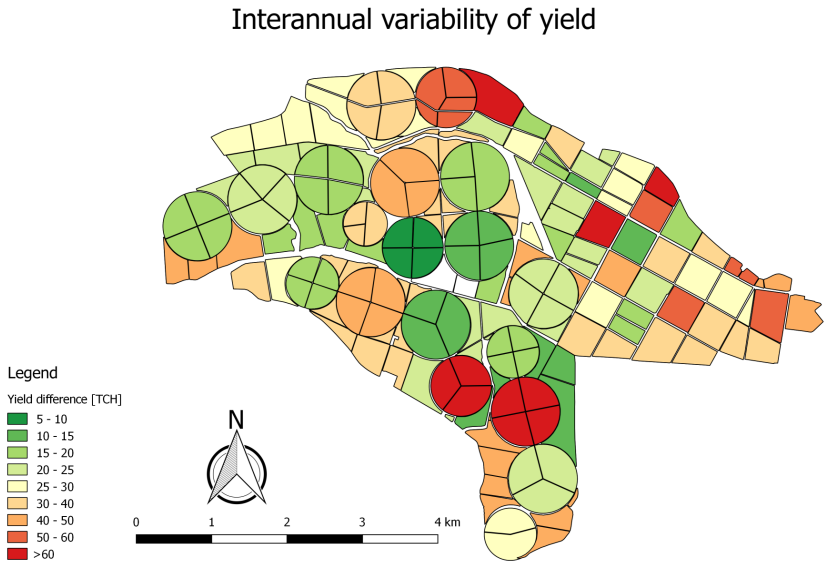


Figure 5.3: Map of interannual variability of yield by using the yield data for the years 2013-2018. Colors indicate the yield difference, which is defined as the difference of the maximum yield and minimum yield, see equation 5.1.

Yield maps can shown in figure 5.3 and 5.2 to visualize spatial and temporal trend. The colors of the fields mapped in figure 5.3 and 5.2 indicate a specific range of yield. The RdYlGn colorramp is used to classify the yield data. Looking at the yield data for the years 2015-2018 the mean yield is between 80 and 90 TCH. Therefore, in figure 5.2 the average yield performance is chosen as 80-90 TCH and colored as yellow in the maps. For figure 5.2 green indicates good performing fields based on that their yield is above average. Orange and red indicates fields which are performing below average. If a field is not colored, then it represents no growing areas, which means that sugarcane was not growing on that field for that specific growing season.

Figure 5.3 shows the interannual variability of yield by computing it as:

$$Interannual\,variability = Yield_{max} - Yield_{min} \tag{5.1}$$

To obtain the maximum and minimum value for the yield data, the yield data on field-scale from 2013 up to and including 2018 are taken. Figure 5.3 shows that the interannual variability of yield differs from 5 to more than 60 TCH. The interannual variability gives an indication how stable or unstable a field performed throughout the years. In figure 5.3 the colors indicate a specific value of interannual variability of yield, which is explained in the legend. Green colors indicate a lower interannual variability than red colors. Combining the interannual variability with the yield maps can give information if a field is performing consistently below the mean yield.

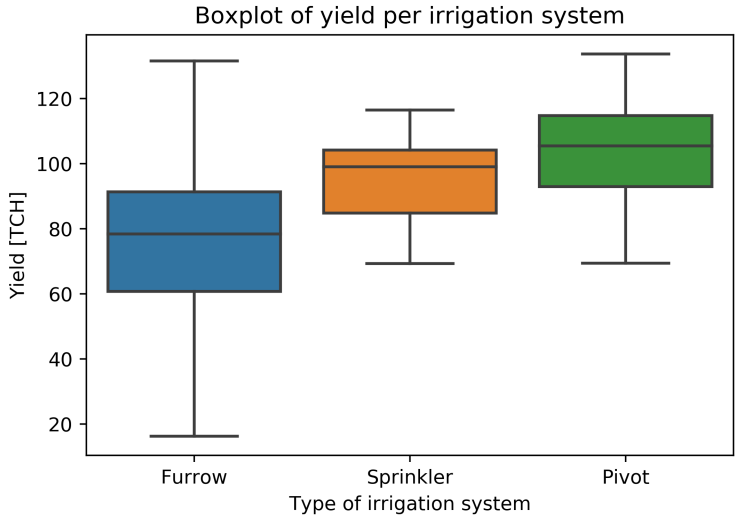


Figure 5.4: Boxplot of yield data for the year 2018 per irrigation system.

In this study an assessment of spatial variability on sugarcane yield with remote sensing data is executed for the growing season of 2018. The yield data per irrigation system is plotted in figure 5.4. As can be seen in the figure, differences in the spatial variability of the irrigation systems are clearly visible. For furrow the distribution of the yield is much larger than that of sprinkler and pivot. Moreover, for furrow irrigation the median of the yield is much lower than that of pivot and sprinkler irrigation. In the next sections the yield data for the year 2018 is used as a validation dataset to investigate if the spatial patterns in field and remote sensing data can explain the variation in yield and can detect yield variation on field-scale level.

5.2. Field analysis

To understand the yield variability between fields, a field study is executed. Several field parameters are compared with respect to yield data. In this section yield data at field-scale level is compared to soil texture, height above drainage pumps, distance from field to drainage pump, slope, groundwater depth, and the electric conductivity of groundwater. Hence, a correlation study between these parameters is executed.

Yield compared to soil texture

A soil map with major soil types at the study site is visualized in figure 5.5. The soil map is based on the soil study of IIAM. The soil types can be subdivided into two groups: clayey alluvial soils and alluvial soils of stratified texture. Every soil type has a symbol consisting of letter and/or numbers, which can be be found in the legend of figure 5.5. The meaning of the symbols are explained in section 4.1.1. and more information about these symbols can be found in the appendix A. As can be seen in figure 5.5 the lines of the parcels are colored by their yield value. When looking at figure 5.5 fields mostly consist of more than one soil type. Auger measurements are taken to validated the soil study of IIAM. The locations of the auger measurements are visualized in figure 5.6. Furthermore, the boundaries of this soil types are not straight. Using the same irrigation approach for one entire field can cause some problems. Irrigation water infiltration rate depends among other things on soil type. Therefore, irrigation management should consider applying nonuniform irrigation depending on soil variability within fields.

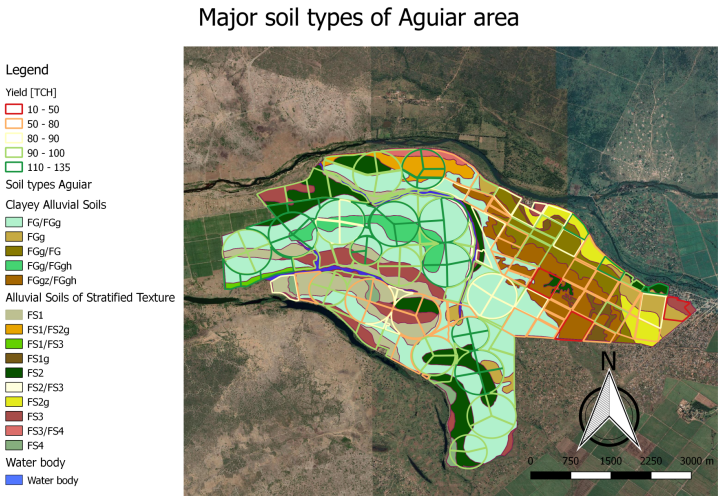


Figure 5.5: Yield distribution per irrigation system for the year 2018. The fields are divided between the two soil types: clayey alluvial soils and alluvial soils of stratified texture. The boundaries of the parcels are colored by the yield value for the year 2018.

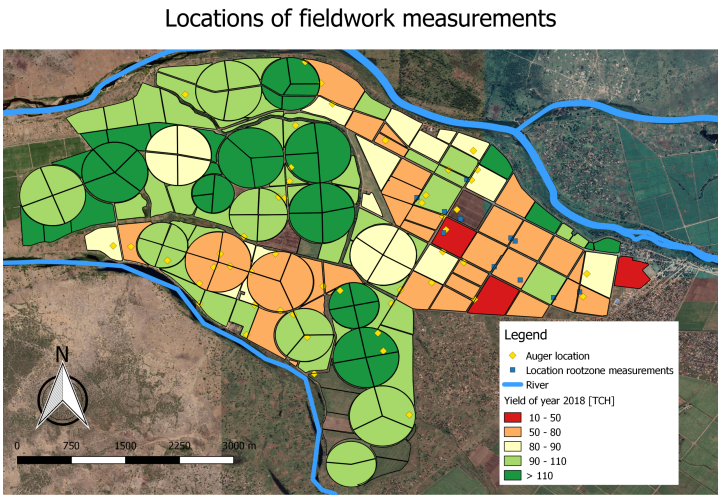


Figure 5.6: Locations of fieldwork measurements. Blue squares presents the locations of the rootzone measurements. The yellow rhombus presents the location of auger measurements.

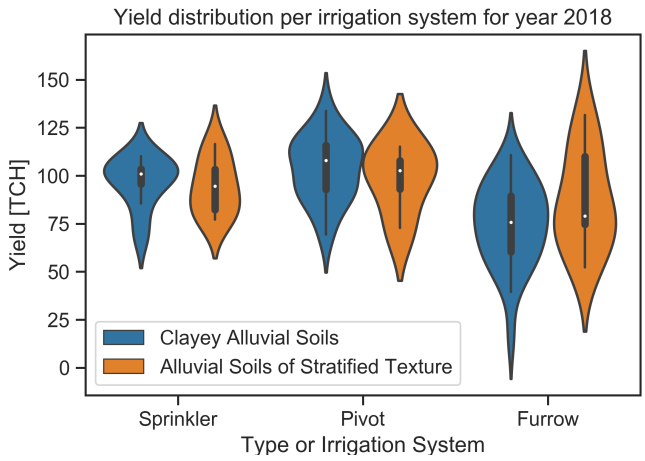


Figure 5.7: Violin plot where the yield is compared to soil texture. The soil types are based on the IIAM research (2012).

To explain the yield variability at the study site, it is interesting to look if there is a difference in yield distribution per irrigation system and soil type. Figure 5.7 visualizes the yield distribution per irrigation system for the two major soils types. A field is categorized to one of these two groups by computing the area of the different subs-types within the field. A soil type with the largest area in the field is selected as main sub-type. This sub-type is linked to one of these major soil types. When looking at the yield distribution per main soil type for each irrigation system, it can be seen that clayey alluvial soils perform better for sprinkler and pivot fields than furrow fields. For pivot and sprinkler the yield median of is higher for clayey alluvial soils than alluvial soils of stratified texture. Fields with furrow irrigation show a higher variability in yield for both major soil types compared to the two other irrigation systems. For furrow irrigation the yield median for fields with alluvial soils of stratified texture are higher than that of fields with clayey alluvial soils. However, the yield difference per irrigation type could also be explained by field parameters other than soil type. This is because the furrow is located at the lower elevated part of the selected study site. Furthermore, pivot and furrow fields are mixed in the same area and furrow irrigated fields not.

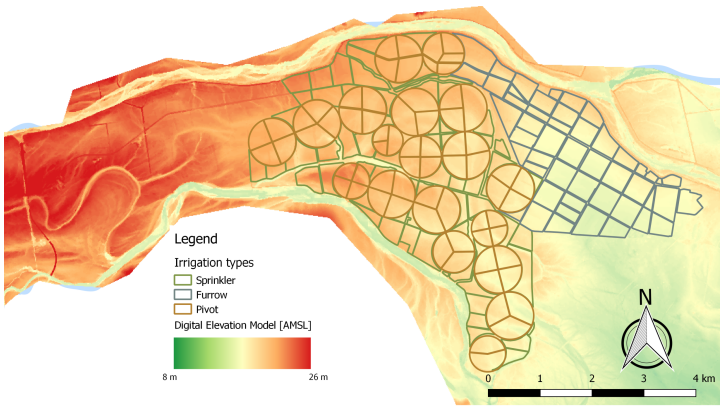


Figure 5.8: Fields colored by type of irrigation systems and plotted on top of the digital elevation model. The brown color presents pivot irrigated fields. Green indicates sprinkler irrigation and the grey/blue color presents furrow irrigation.

Yield compared to terrain

Using the digital elevation model a new terrain model named HAND (Head Above Near Drainage) is created. In this research the HAND model is based on the elevation difference between the fields and the drainage pumps on pixel-scale level. The drainage pumps are connected to

the fields through drainage canals. Three drainage pumps are located in the selected study site. Table 5.2 gives the elevation of these drainage pumps. The HAND map is created with the same resolution as the DEM, namely with a pixel size resolution of 20 meter. Figure 5.9 visualizes the HAND map. Here, it can be seen that a significant portion of the fields have a negative hand. A negative HAND means that the elevation of the fields is lower than the height of the nearest drainage pump. The local relative heights in the study site represents the local draining potentials based on the topology of the relative soil gravitational potentials [Nobre et al., 2011]. As can be seen in figure 5.9, the eastern part of the study site has the highest negative elevation differences, which is also the area where the furrow fields are located. The areas with the highest negative elevation differences are also the areas with poor yields. The hand map is also used in the correlation analysis.

Table 5.2: Drainage pump with elevation height from DEM model

Drainage pump	Elevation [meter above mean sea level]
DPD7	19.43 m
DPSAN	16.94 m
DPTSA	18.10 m

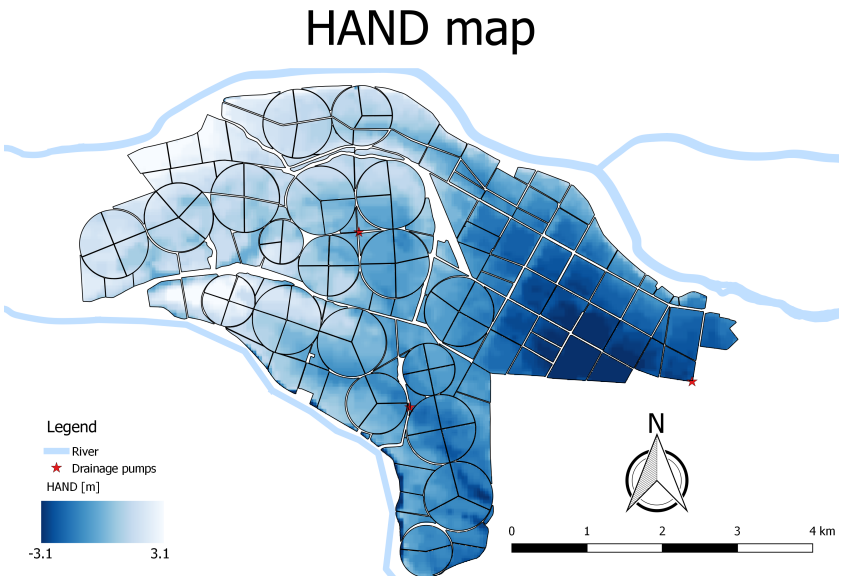


Figure 5.9: Hand Above Near Drainage (HAND) map with the locations of the drainage pumps at the study site. The pixel resolution of the HAND map is 20 meter.

Correlation analysis

As part of the field analysis a correlation study is executed for several parameters. Field parameter such as slope, HAND, the distance from field to drainage pump and soil parameter are compared to the total yield and interannual yield. All parameters are average values on field-scale level. The slope and HAND are first computed with a pixel-size resolution of 20 meters. After that, the mean value per field is calculated. The distance from field to drainage pump is manually computed in QGIS by measuring the distance from the drainage pump to the closest corner with the lowest elevation of each field by following the drainage canals and its corresponding height. Here, an assumption is made that the water flows in the drainage canal from higher elevated locations to lower elevated places. For the soil parameter, the different soil types are ranked based on their clay content. For the total yield, the yield data from 2013 up to and including 2018 is added up.

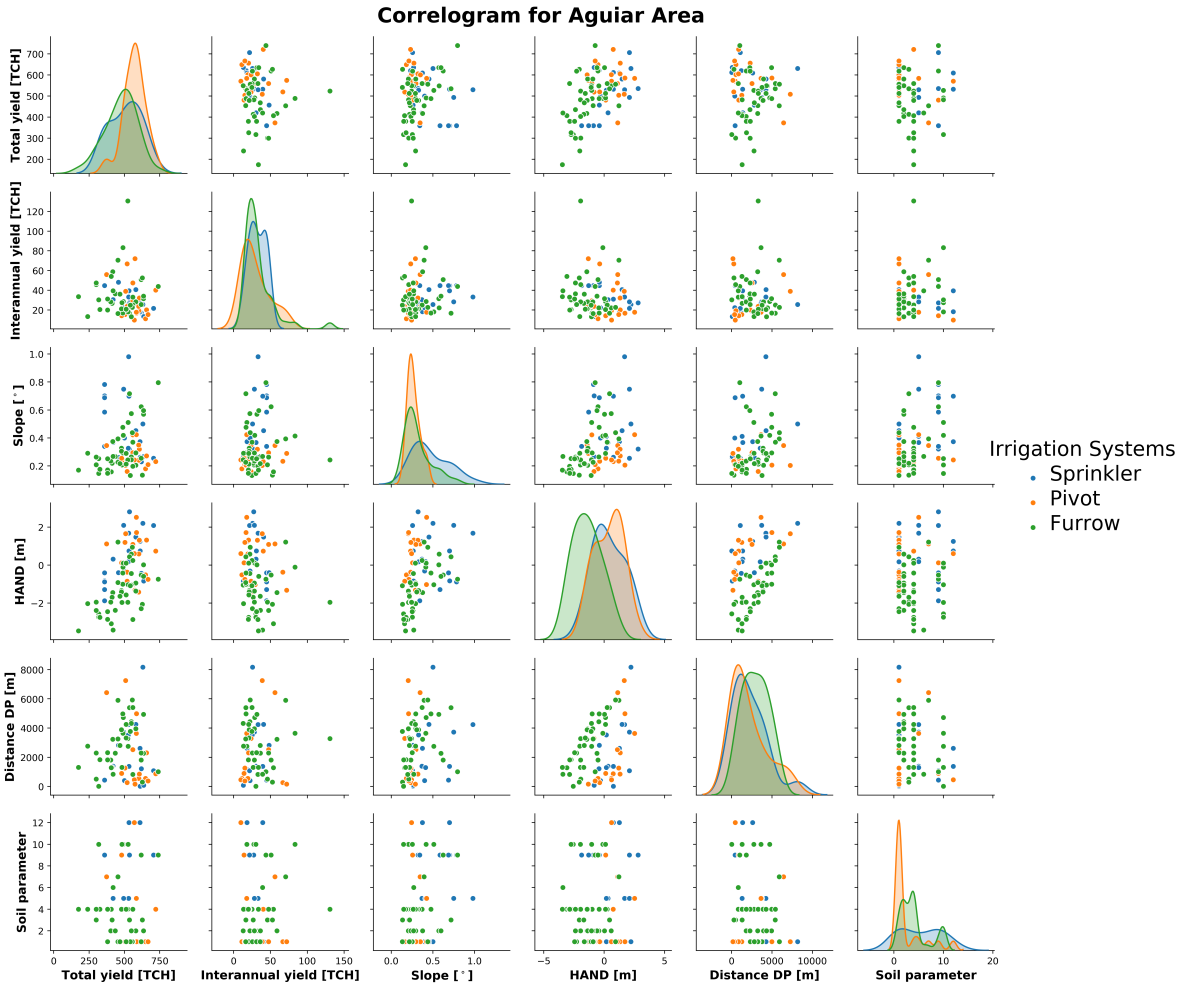


Figure 5.10: Correlogram between total and interannual yield and field parameters slope, HAND, distance to drainage pump (DP) and soil rank using Pearson correlation. The colored dots indicates the different irrigation systems, whereas blue is sprinkler irrigation, orange is pivot irrigation and green is furrow irrigated fields.

Table 5.3: Pearson correlation coefficients between total and interannual yield and field parameters slope, HAND, distance to drainage pump and soil rank

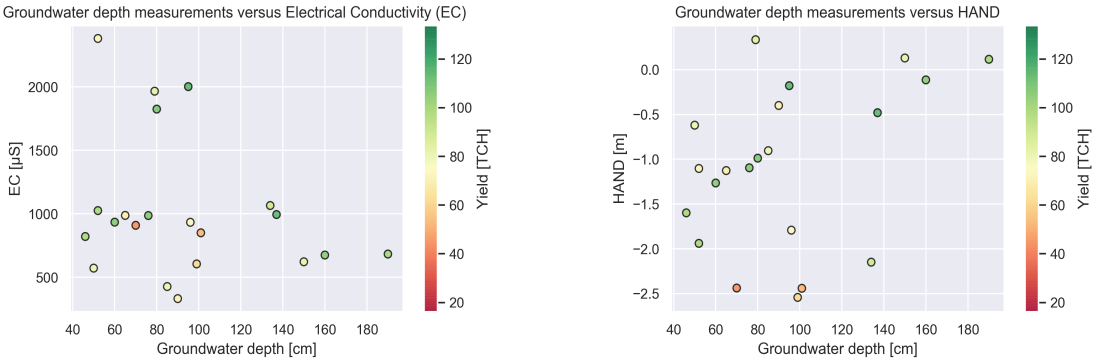
	Total yield	Interannual yield	Slope	HAND	Distance to DP	Soil parameter
Total yield	1	-0.14	0.10	0.52	0.01	-0.23
Interannual yield	-0.14	1	0.11	-0.13	-0.01564	0.14
Slope	0.10	0.11	1	0.38	0.25	0.35
HAND	0.52	-0.13	0.38	1	0.37	0.03
Distance to DP	0.01	-0.02	0.25	0.37	1	-0.12
Soil parameter	-0.23	0.14	0.35	0.03	-0.12	1

The method for determining the interanual yield is explained in section 5.1. Since the assumption is that field parameters in the correlogram are consistent over time, these two yield parameters are chosen to understand the relation with the field parameters. Another assumption is that the relationship between yield and field parameters is linear. Therefore, the correlation study is executed using the Pearson correlation coefficients. The results of this study are visualized in form of a correlogram in figure 5.10 and in the form of correlation coefficients in table 5.3.

The graphs on the diagonal of the correlogram in figure 5.10 show the distribution of the parameters. All other graphs are scatterplots, where the dots are colored by type or irrigation systems. The Pearson correlation coefficients of these parameters can be found in table 5.3, where the highest correlation occurs between the HAND and total yield. This is with a Pearson correlation coefficient of 0.52 and can be labelled as moderately positive correlated. Slope and the soil parameter have a weak to no correlation with total yield. The distance between fields and the drainage pump have no correlation to the total yield. The interannual yield shows no to weak correlation for all field parameters. We need to keep in mind that the slope and the HAND are computed using a digital elevation model. Digital elevation models are an interpolation of the actual surface. The DEM is useful for visualizing the area. However, it should not be considered ground-truthed. For the variables considered and for this study the HAND is the most important field property to explain yield variation between fields. However, more field properties should influence the yield variability. Research show that other contributors such as nutrients and planting dates can have effect on yield [Sanches et al., 2020].

Groundwater depth

Groundwater measurements are taken in the field in the summer of 2019. One of the measurements taken are the depth of the groundwater table and its electrical conductivity. From all fields measurements, groundwater was reached 23 times, where for 21 locations the yield data is known. Therefore, the plots show only the measurements for the fields where the yield data is available. Figures 5.11a and 5.11b visualizes the depth of the groundwater in relation to EC and HAND. In figure 5.11a there is no relation visible between the depth of the groundwater and the EC measurements of the groundwater at that location. For the HAND observations in figure 5.11b only the pixel value in the HAND map of that specific location is taken for the groundwater depth measurements. The paper of Nobre et al. (2011) demonstrates that there is a high correlation between the HAND and the groundwater depth [Nobre et al., 2011]. Looking at the results 18 of the 21 groundwater measurements are taken at a location where there a negative HAND. However, an increasing depth in groundwater does not go together with an increasing negative HAND. Furthermore, it is visible that the most negative HAND values are located at the fields with the poorest yield values. However, looking at figure 5.11b no clear relation is visible between the groundwater depth and the HAND. The fields with their groundwater closest to the surface do not have the most negative HAND values. Furthermore, there is no clear relation visible between groundwater depth and yield. Moreover, during fieldwork root zone measurements were taken. Measurements show that sugarcane roots reach an average depth of around 50 cm below the surface. This indicates that for some fields sugarcane can reached the groundwater.

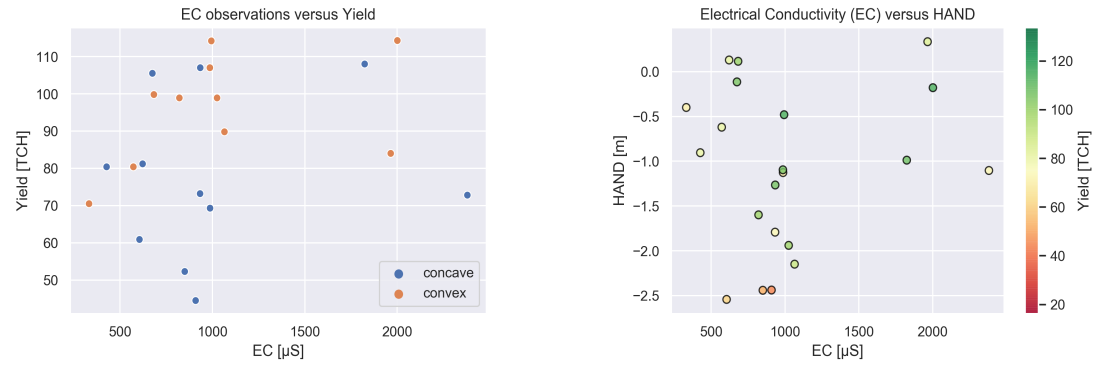


(a) (b)

Figure 5.11: Groundwater measurements compared to EC and HAND. (a) Groundwater depth measurements compared to EC measurements in the field (b) Groundwater depth measurements compared to pixel value of the HAND map

Electrical conductivity

Electrical conductivity measurements are taken at groundwater level. Figures 5.12a and 5.12b visualize EC measurements compared to the yield data and the HAND. Figure 5.12a shows EC in relation to yield data. The colors of the dots refer to the convexity. When the measurements took place in a concave area, the dots are colored blue. For convex areas, the color is orange. The convex and concave areas are derived from the convexity map using the pixel value at the location, where the EC measurement is taken. It can be found in literature that groundwater with salinity levels results in crop yield losses [Talebnejad and Sepaskhah, 2015]. However, in this study the high saline levels in groundwater are not related to significant yield losses. For example, the yield of field AG1D07P for the year 2018 is 114.3 TCH and the measured electrical conductivity is 2001 $\mu S/cm$. Another example is the field AG2A02P, where the yield of the field is 108 TCH is and the measured EC is 1824 $\mu S/cm$. Furthermore, the salinity levels in groundwater are very localized. For the year 2018 the field AG3C32 was not used to grow sugarcane and for that reason these two measurements are not visualized in the plots. For Ag3C32 two times groundwater was reached during field measurements. At one location the groundwater was reached at a depth of 76 m and the EC was recorded at 1547 $\mu S/cm$. For the other location in the field the groundwater was reached at a depth of 50 cm with an EC of 517 $\mu S/cm$. So within a field the EC can vary significantly. Lastly, figure 5.12b shows that EC is not related to the HAND. The highest EC values do not occur at location with the most negative HAND.



(a) (b)

Figure 5.12: EC measurements compared to yield and HAND. (a) EC measurements compared to the yield data. The blue color means that the EC measurements are taken in a concave area. Orange means that the EC measurements are taken in a convex area. The convex and concave areas are derived from the convexity map. (b) EC measurements compared to pixel value of the HAND map. The colors specify the yield data for the corresponding fields where these EC measurements are taken from.

5.3. Analysis of normalized difference salinity index

The normalized difference salinity index (NDSI) can be computed by combining spectral bands from Sentinel-2 imagery. Using NDSI salt affected areas can be identified and therefore, NDSI can help with identifying locations where sugarcane poorly grows. This section shows the results of identifying salinized areas using the normalized difference salinity index (NDSI). NDSI maps of the study site are presented in figure 5.13 and 5.14. The values at the study site range from 0 to 0.4, which is not high enough to indicate salt effected soils. One of the reasons might be that the crusts formed on top of the soil have not effected a large enough area that it has an influence on the signal reflectance. The NDSI maps have a pixel resolution of 20 meter. During fieldwork we notice that the salinization occurs locally. The salinized areas are noticed by salt cursts on soil, which is shown in figure 3.5, and by no growing spots within fields. Another reason that salt effected areas might not be recognized is that the EC measured in the fields are below 2500 $\mu S/cm$. Researchs, which uses the NDSI, are in study sites, where soils are more saline. For example Asfaw et al. (2018) categorize soil salinity levels as follow:

- Non-saline soils have a EC below 2000 $\mu S/cm$.
- Slightly saline soils have a EC between of 2000-4000 $\mu S/cm$.
- Moderately saline soils have a EC between 4000-8000 $\mu S/cm$.

This means that most areas are non-saline soils and the other areas are slightly saline. Furthermore, we need to keep in mind that the fieldwork measurements are acquired in July 2019 and then acquired satellites images are from August and September 2018. Situations could have change in a year, which can explain that the saline areas are not visible in these maps.

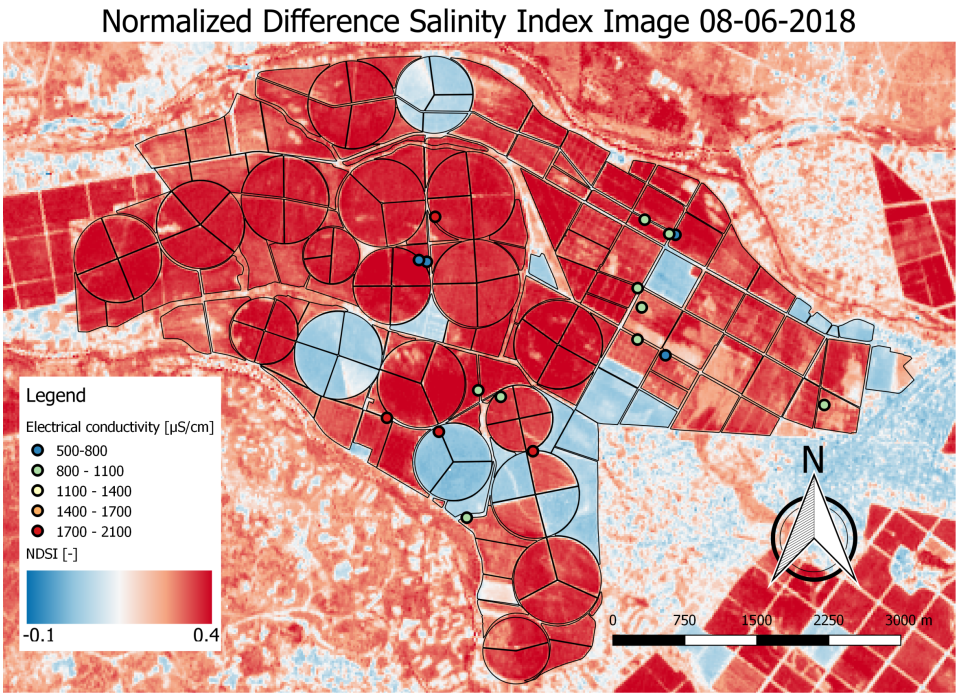


Figure 5.13: Normalized Difference Salinity Index Image of 08-06-2018. The colorramp ranges from -1 to 1. The colored dots indicate the locations of the field measurements where groundwater is reached. The color in the dots indicate the measured electrical conductivity of groundwater. Field measurement were taken in July 2019.

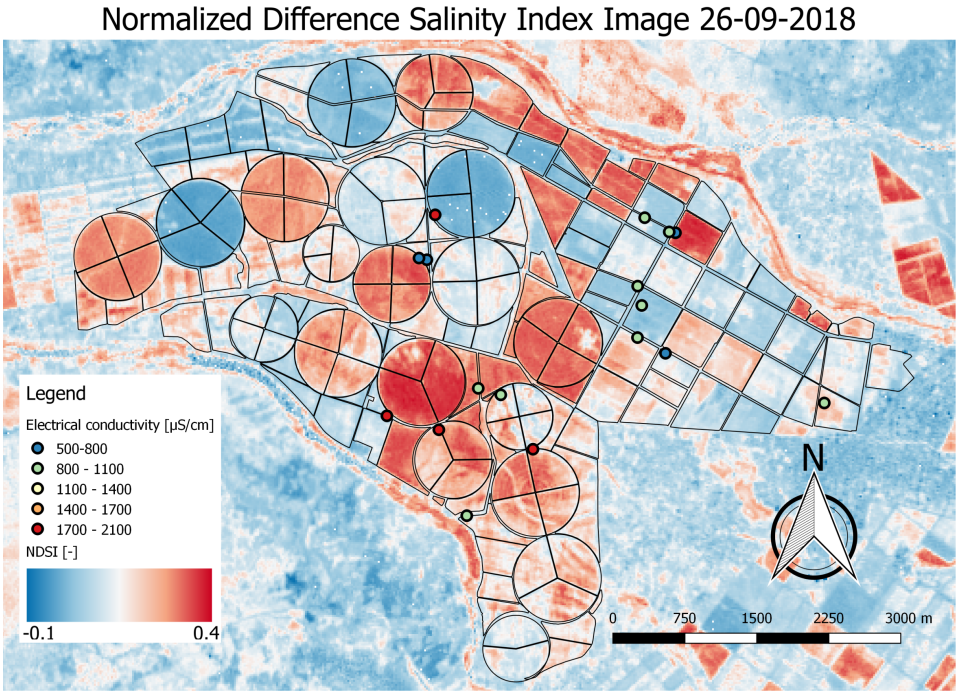
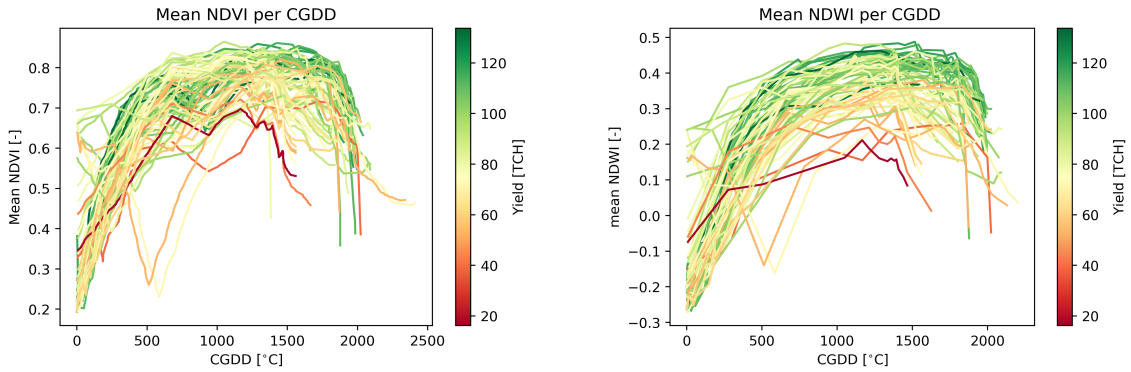


Figure 5.14: Normalized Difference Salinity Index of 29-09-2018.The colorramp ranges from -1 to +1. The colored dots indicate the locations of the field measurements where groundwater is reached. The color in the dots indicate the measured electrical conductivity of groundwater. Field measurement were taken in July 2019.

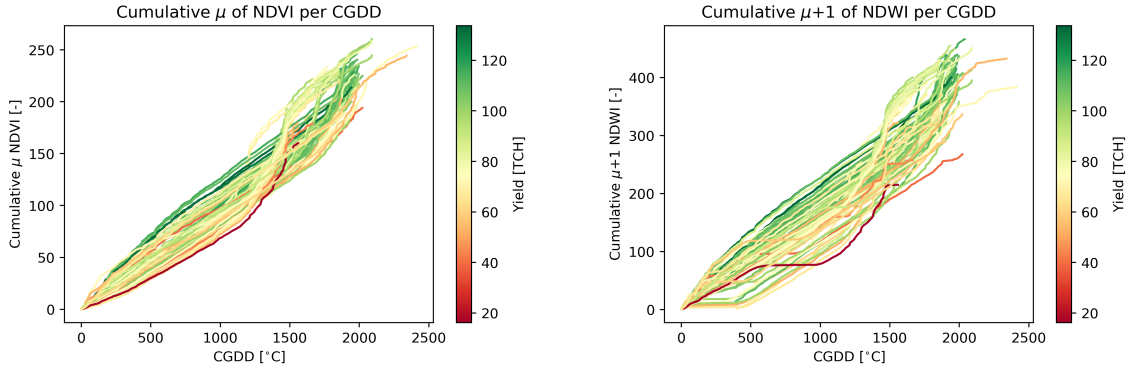
5.4. Analysis of satellite derived vegetation indices

Studies on sugarcane show that the spatial variability of sugarcane yield can be mapped by NDVI [Morel et al., 2014, Bégué et al., 2010]. Bolton and Friedl (2013) have shown that for maize and soybean NDVI is not always the best correlated satellite derived vegetation index with yield. Therefore, this section explores whether NDVI and NDWI capture the heterogeneity of the fields and whether it is correlated with yield. A map of NDVI and NDWI estimates is presented for a certain day in the growing season 2017-2018 in appendix B. The hypothesis is that fields with poor yield have a lot of non-uniformity within a field compared to fields with good yield. This non-uniformity should be visible in high resolution spectral indices as NDVI and NDWI. To test that this non-uniformity is present, the standard deviation and coefficient of variation within the performing fields is computed throughout one growing season. All statistical values are computed on field-scale level. For the mean the hypothesis is that fields with a high yield should have a higher mean values compared to fields with poor yields. Figure 5.15a and 5.15b show a time serie of mean NDVI and NDWI values on field-scale level. The time is expressed in crop growing degree days (CGDD). Knowing the planting and harvest date of each field and the temperature for each day, the time is converted from days to cumulative growing degree days. The reason to convert the days into CGDD is done to correct the influence of temperature on the crop development over the growing season. This study selects the planting date as the first cloud free sentinel-2 overpass after the provided planting date of the plantation. In this study the used harvest is the first cloud free sentinel-2 overpass before the actual harvest date. For all NDVI and NDWI figures the fields are plotted in the same order. Zooming in at figures 5.15a and 5.15b, a decrease in NDVI and NDWI is visible for two fields at around 500-700 CGDD. This dip can be explained by the fact that the provided data about the planting dates does not correspond with the actual planting date. Comparing both mean NDVI and mean NDWI figures, the transition from good to poor yield and the extremes are better represented in the mean NDWI.



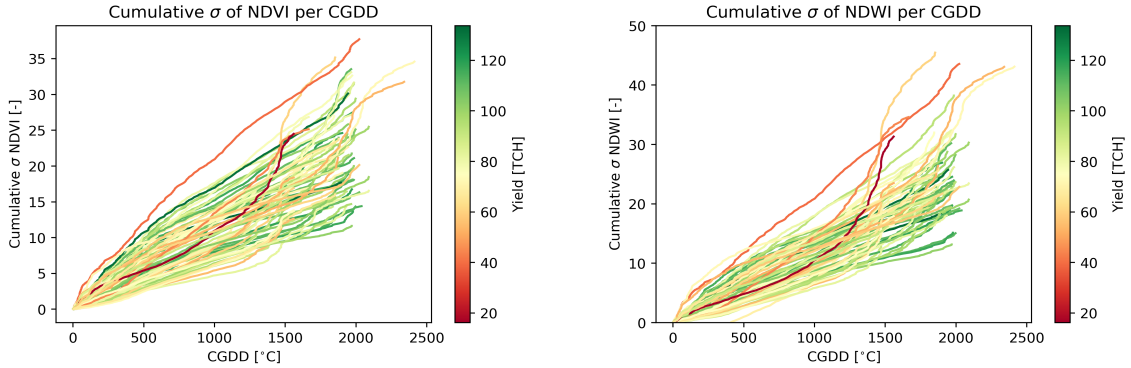
(a) (b)

Figure 5.15: Time series of field-scale satellite derived vegetation indices observations in relation to CGDD. Colors are specified based on the yield of the fields. (a) NDVI observations per field in relation to CGDD. (b) NDWI observations per field in relation to CGDD.



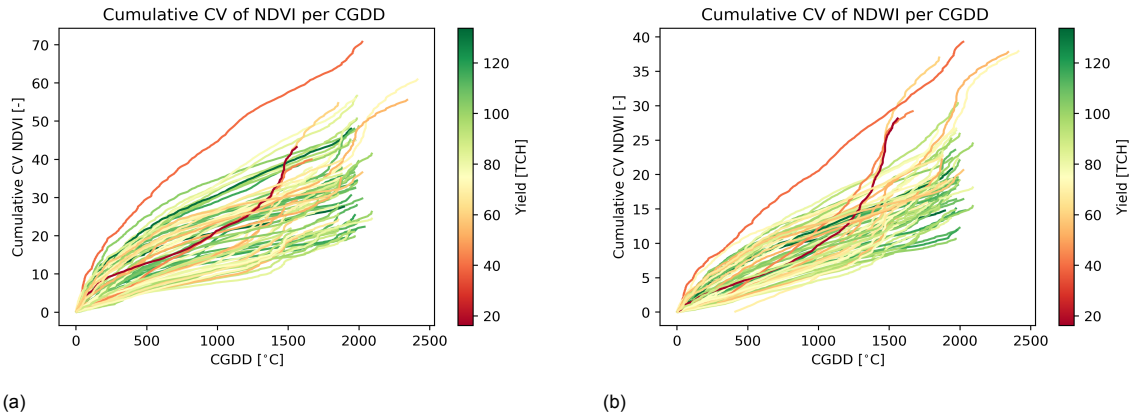
(a) (b)

Figure 5.16: Cumulative mean satellite derived vegetation indices observations in relation to CGDD. Colors are specified based on the yield of the fields. (a) Cumulative mean NDVI observations per field in relation to CGDD. (b) Cumulative mean NDWI observations per field in relation to CGDD.



(a) (b)

Figure 5.17: Cumulative standard deviation of satellite derived vegetation indices observations in relation to CGDD. Colors are specified based on the yield of the fields. (a) Cumulative standard deviation of NDVI observations per field in relation to CGDD. (b) Cumulative standard deviation of NDWI observations per field in relation to CGDD.



(a) (b)

Figure 5.18: Cumulative coefficient of variation (CV) of satellite derived vegetation indices in relation to CGDD. Colors are specified based on the yield of the fields. (a) Cumulative coefficient of variation of NDVI observations per field in relation to CGDD. (b) Cumulative coefficient of variation of NDWI observations per field in relation to CGDD.

Figures 5.16, 5.17 and 5.18 show the cumulative mean, standard deviation and coefficient of variation of NDVI and NDWI in relation to CGDD. The cumulative standard deviation and coefficient of variation (figures 5.17 and 5.18) give information about the non-uniformity of the crop growth. At the beginning of the crop season, looking at the figures, there is not a clear distinction visible between good and poor performing fields. Looking at the end of the crop season, the poor performing fields have clearly a higher cumulative non-uniformity than good performing fields. In figures 5.16a and 5.16b, a strong increase is visible around 1300-1400 CGDD for the satellite derived vegetation indices. The reason for the strong increase can be explained by a strong increase in crop growth. For that reason, a correlation study is executed until 1350 CGDD. The results of this correlation study are put in table 5.4. The red colored values point out the maximum correlation values. For the cumulative mean, the maximum correlation occurs at 1000 CGDD. For the average values the highest correlation with yield occurs at 1350. The mean NDWI shows a stronger correlation to yield than NDVI. However, it needs to be taken into account that the correlation study is executed only for the Aguiar area thus for a limited number of fields and not for the whole plantation. Currently , most studies use NDVI to estimate sugarcane yield [Lofton et al., 2012, Mulianga et al., 2013, Bégué et al., 2010] . Besides that, most correlation studies between different crops compare NDVI with other vegetation indices rather than NDWI [Simões et al., 2005, Panda et al., 2010].

Huang et al. (2009) also prove that the NDWI has a superior correlation compared to NDVI [Huang et al., 2009]. However, other studies show that NDVI has a higher correlation in crop yield rather than NDWI [Bolton and Friedl, 2013]. Bolten (2012) concludes that the enhanced vegetation index (EVI) is more effective at forecasting crop yield than both NDVI and NDWI. EVI is another satellite derived vegetation index. This index is not investigated in this study.

Table 5.4: Spearman correlation between NDVI & NDWI parameters and yield

CGDD	Cumulative mean NDWI	Mean NDWI	Cumulative mean NDVI	Mean NDVI
200	0.31	-0.14	0.3	0.29
300	0.34	0.17	0.34	0.51
400	0.35	0.36	0.39	0.52
500	0.37	0.45	0.41	0.4
600	0.43	0.54	0.47	0.31
750	0.49	0.56	0.51	0.13
1000	0.55	0.59	0.52	0.16
1150	0.54	0.64	0.48	0.41
1250	0.5	0.66	0.41	0.57
1350	0.41	0.73	0.28	0.64

The Kruskal Wallis H-test is used in agricultural applications to test the yield difference between different groups. In this study, we explore if the yield significantly differs between irrigation system using the NDVI and NDWI observations instead of yield data. Plotting the yield data per irrigation system, see section 5.1, it can be seen that the distribution between irrigation systems differs significantly. The Kruskal Wallis H-test is performed at 1000 CGDD and it ueses a significance level (α) of 0.05. The results of the test are presented in table 5.5. The test rejects the null hypothesis for the mean, standard deviation and CV observations of NDVI. This means that for one or more groups the distributions are significant different. Thus the distribution of the NDVI measurements differs significantly per irrigation system. This test is also performed using NDWI measurements. However, looking at the Mann-Whitney U test and testing it on the mean NDVI values, the null hypothesis is not rejected for the distribution of the mean NDWI values for 1000 CGDD between furrow and sprinklers. This means that the mean NDVI values for 1000 CGDD are not significantly different when looking at the distribution of furrow and sprinklers. However, looking at figure 5.4 the yield distribution of furrow and sprinklers looks very different. In table 5.6 the results of the Kruskal Wallis H-test are presented for 1000 CGDD. The test rejects the null hypothesis for NDWI observations of mean, standard deviation and cumulative mean for 1000 CGDD. Thus, the distribution of the NDWI measurements differs significantly when comparing the irrigation systems. However, when computing the Mann Whitney U-test for the cumulative mean and mean of NDWI observation, it shows that the distribution of sprinklers and pivot does not differ significantly. The H0 is not rejected. Computing the Mann Whitney U-test for NDWI standard deviation on 1000 CGDD, the result shows that the H0 is rejected for comparing the distribution of sprinklers and pivot. However, the H0 is not rejected when comparing the distribution of sprinklers and furrow.

Table 5.5: Kruskal Wallis H-test on NDVI measurements for 1000 CGDD and α =0.05 to test the distribution of furrow, sprinkler and pivot irrigation systems.

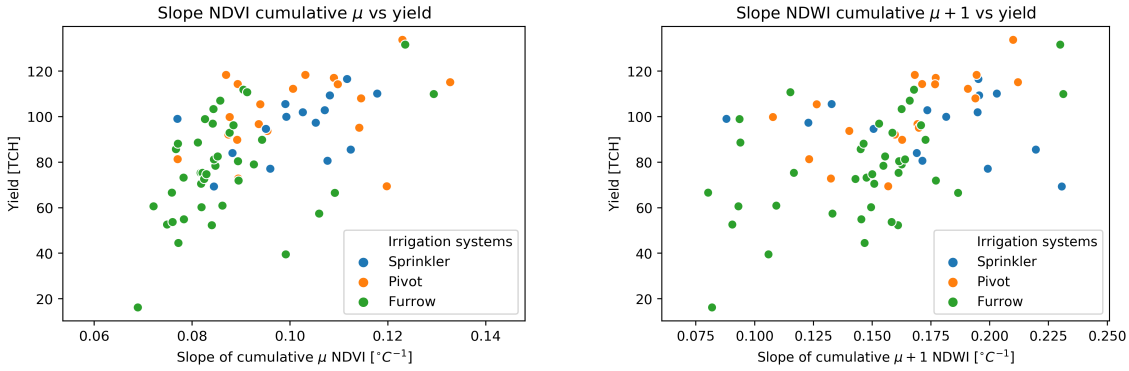
NDVI	Test statistics	p-value
Mean	6.819	0.033
Standard deviation	6.137	0.046
Coefficient of variation	6.879	0.032

Table 5.6: Kruskal Wallis H-test on NDWI measurements for 1000 CGDD and α =0.05 to test the distribution of furrow, sprinkler and pivot irrigation systems.

NDWI	Test statistics	p-value
Cumulative mean	11.760	0.003
Mean	29.248	0.00
Standard deviation	11.133	0.004

Slope NDVI & NDWI

In this thesis we want to know if the spatial patterns in yield variation are apparent in satellite derived vegetation indices. Therefore, we want to know if the slope of the NDVI and NDWI time series is related to the yield. Since the correlation between the mean values and the yield are the highest at 1000 CGDD, the slope of the cumulative mean is computed by using the first 1000 CGDD. The slope gives information about the rate of increase of the satellite derived vegetation indices per CGDD. The hypothesis is that fields with high yields have a steeper slope than fields with poor yields. The Spearman correlation coefficient is used to quantify the relation between the satellite derived vegetation indices and yield. For NDVI and NDWI the slope of the cumulative mean have a Spearman correlation coefficient of 0.562 and 0.487, respectively. This means that both NDVI and NDWI are moderately correlated to yield. The slope of the cumulative mean for NDVI and NDWI are plotted in relation to yield in figure 5.19. The fields with the lowest and highest yield are having the lowest and highest rate of increase, respectively. Looking at the mid range values of the yield data, no clusters are observed when plotting the slope in respect of yield.

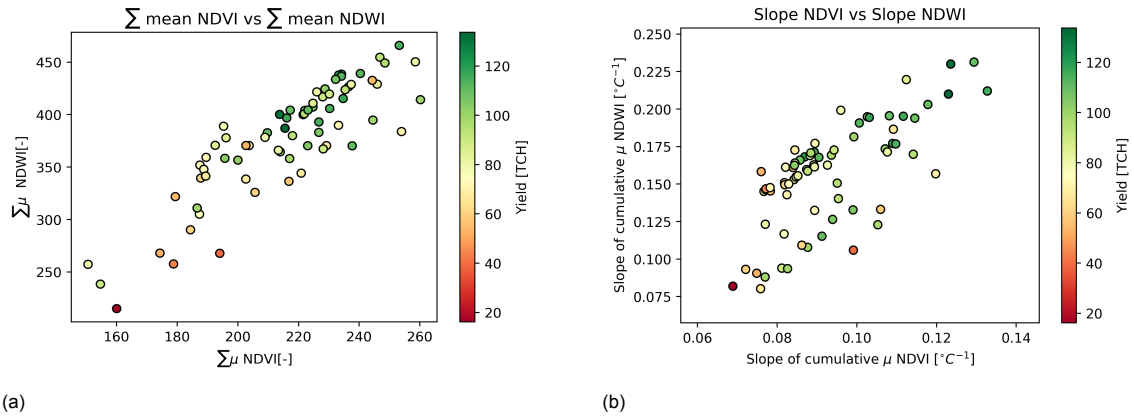


(a) (b)

Figure 5.19: The slope, also called as rate of increase, is computed on field-scale basis using the cumulative mean (μ) for the first 1000 CGDD. The colors indicates the irrigation system of the fields. (a) Rate of increase for NDVI cumulative mean. (b) Rate of increase for NDWI cumulative mean.

Correlation between the satellite based vegetation indices

The sum of mean NDVI values and the sum of mean NDWI values are plotted against each other to see if the two satellite derived vegetation indices are correlated with each other. Figure 5.20 presents the sum of mean NDVI values plotted against the sum of mean NDWI values and the slope for the first 1000 CGDD of NDVI plotted against the slope of the first 1000 CGDD of NDWI. As can be seen in figure 5.20a NDVI and NDWI are highly correlated. The Spearman rank correlation coefficient of figure 5.20a is 0.85, which means the sum of mean NDVI and NDWI values are very strongly positive correlated. The Spearman rank correlation coefficient of figure 5.20b is 0.70, which means that also the slope of the cumulative mean of NDVI and NDWI are moderately to strongly positive correlated.



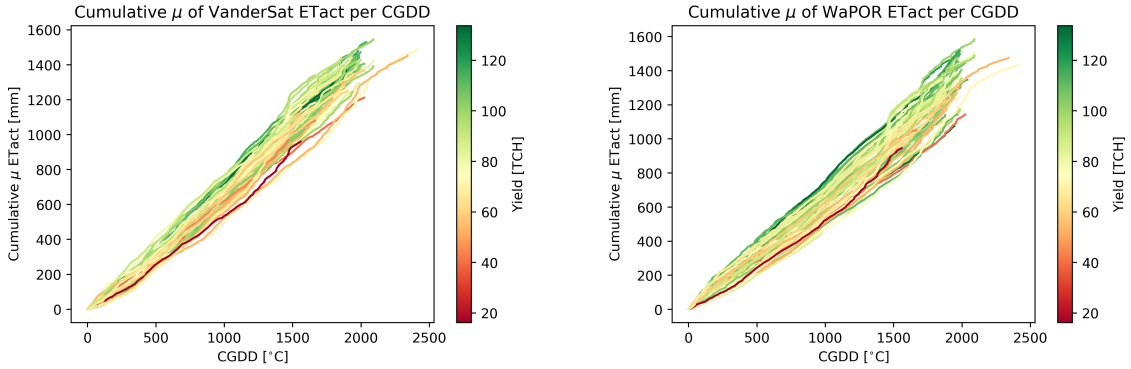
(a) (b)

Figure 5.20: Scatterplots of spectral index pairs. (a): cumulative mean of NDVI - cumulative mean of NDWI. (b): slope of cumulative mean NDVI - slope of cumulative mean NDWI. The slope of line is computed for 0 - 1000 CGDD.

5.5. Analysis of satellite-based evaporation products

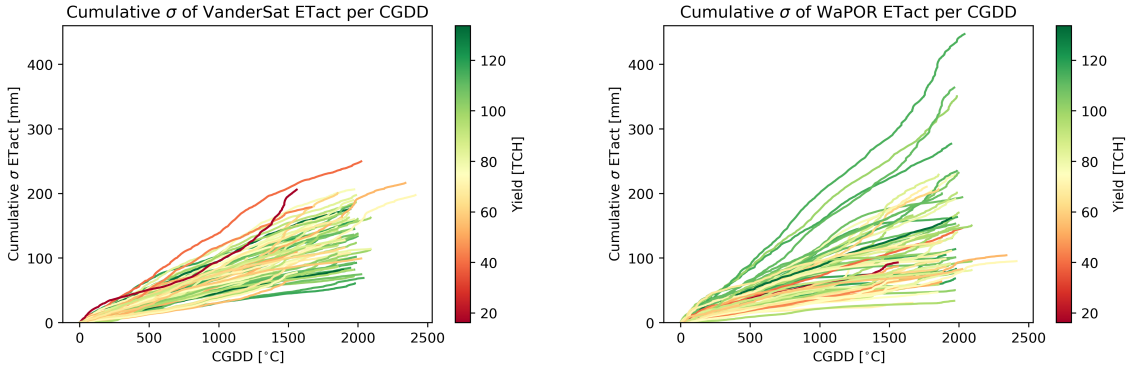
ET is linked to the water, carbon and energy cycle [Fisher et al., 2017]. Therefore, a good ET product should be correlated with yield. This study determines the relation between ET and yield.

In this section the satellite-based evaporation products of VanderSat and WaPOR are compared in relation to yield. A map of actual evapotranspiration estimates of these products is presented for a certain day in the growing season 2017-2018 in appendix D. Figure 5.21 shows the cumulative mean of field-scale actual ET estimates using Wapor and the VanderSat product. The expected actual ET response is that fields with high yields have a higher actual ET, also symbolized as ETact, than fields with low yields. In figure 5.21 the extremes are clearly visible. Fields with the highest yields have a higher ETact estimate than fields with the lowest yield values. However, no difference is observed when a field has a better or poorer performance if the two fields have yields in the mid range. When looking at the results per field, it seems that both figures give very similar actual ET estimates. WaPOR and VanderSat show different outcomes when visualizing the spatial variability within the field. To test the non-uniformity within a field, the cumulative standard deviation and cumulative coefficient of variation are visualized in figure 5.22 and 5.23. The expected signal should be that fields with poorer yields have a larger standard deviation and a larger coefficient of variation and thus also larger values for their respective cumulatives. The results of the cumulative standard deviation and cumulative CV of VanderSat correspond with the expected signal. WaPOR gives an opposite result. WaPOR shows that fields with good fields have a higher cumulative standard deviation and higher cumulative CV compared to fields with low fields. This means that fields with a high cumulative CV and standard deviation have a higher non-uniformity. Moreover, keep in mind that the actual ET values do not give any information about the irrigation and soil management.



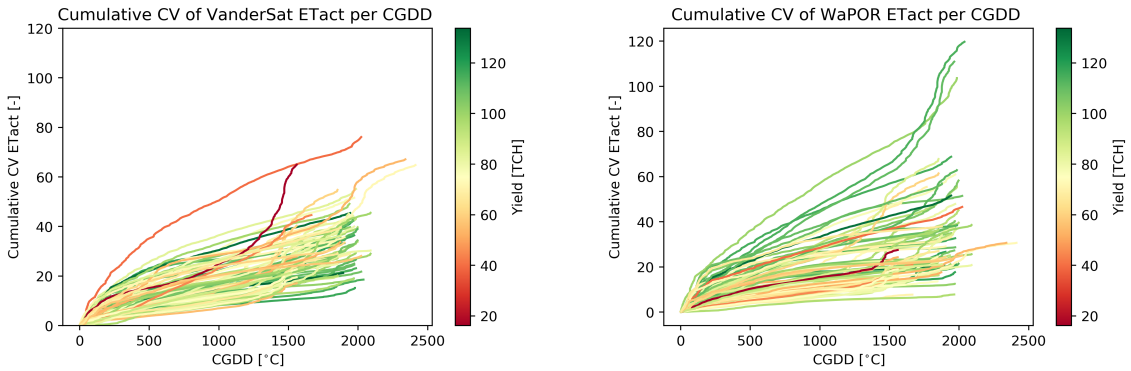
(a) (b)

Figure 5.21: Cumulative mean (μ) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Cumulative mean observations are averaged per field. The color of the lines indicates the yield.



(a) . (b)

Figure 5.22: Cumulative standard deviation (σ) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Cumulative mean observations are averaged per field. The color of the lines indicates the yield.



(a) (b)

Figure 5.23: Cumulative coefficient of variation (CV) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Cumulative mean observations are averaged per field. The color of the lines indicates the yield data.

Table 5.7: Spearman correlation between VDS evaporation & WaPOR parameters and yield

CGDD	Cumulative mean Wapor	Mean Wapor	Cumulative mean VDS	mean VDS
200	0.32	-0.076	0.15	-0.24
300	0.34	-0.059	0.16	-0.21
400	0.38	-0.026	0.20	0.058
500	0.4	0.036	0.21	0.046
600	0.44	0.11	0.33	0.39
750	0.47	0.38	0.39	0.12
1000	0.5	0.36	0.54	0.2
1150	0.48	0.35	0.57	0.46
1250	0.4	0.5	0.51	0.31
1350	0.36	0.44	0.45	0.11

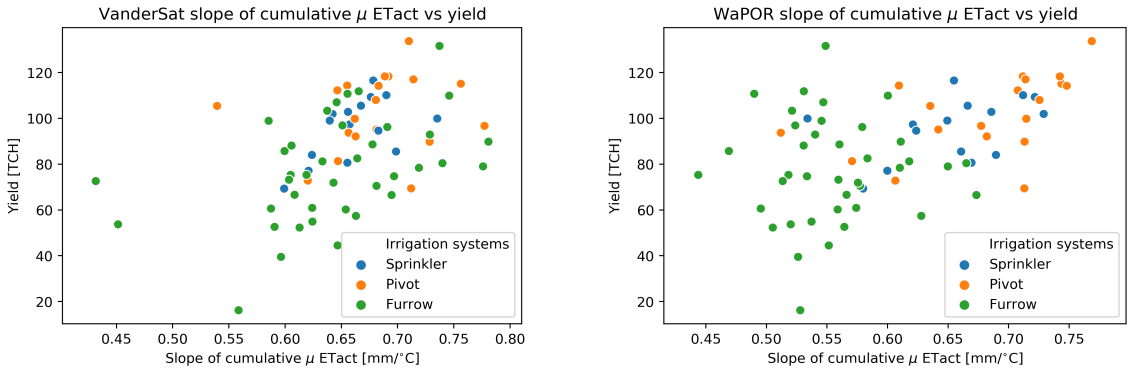
Just like the satellite derived vegetation indices, the correlation between yield and the evaporation products of VanderSat and WaPOR is quantified using the Spearman rank correlation coefficient, which are found in table 5.7. The red colored values point out the maximum correlation values found. For both the mean and cumulative mean of WaPOR, the maximum correlation is found at 0.4, which indicates a moderately positive correlation. For VanderSat the cumulative mean of ETact estimates at 1150 CGDD have the best correlation with yield, which is a Spearman correlation coefficient of 0.57. For the VanderSat ET product the field-scale mean ETact values show a maximum correlation coefficient 0.46. Mladenova et al. (2017) show that the strength of correlation between yield depends on the interannual variability in yield. Correlation results on corn and soybean indicate ET can provide better information than satellite based vegetation indices for yields [Mladenova et al., 2017]. This case study shows that satellite based vegetation indices show significantly more correlation for yield than ETact.

The Kruskal Wallis H-test is applied on ET measurements to look if the ET observations significantly differ per irrigation system. Section 5.1 shows by plotting the yield data per irrigation system that the distribution of yield data between irrigation system significantly differs. The Kruskal Wallis H-test is performed at 1000 CGDD and using a significance level (α) of 0.05. The results of the test are presented in table 5.8. The test reject the null hypothesis for VanderSat actual ET measurements for cumulative mean and the standard deviation. When the null hypothesis (H0) is rejected, this means that for one or more groups the distributions are significantly different. However, looking at the Mann-Whitney U test and testing it on the cumulative mean and mean actual ET values for VanderSat, the null hypothesis is not rejected when looking at the distribution for pivot and sprinklers on 1000 CGDD. Furthermore, for the mean actual ET observations of the algorithm of VanderSat the null hypothesis is not rejected for furrow and pivot as well. This means that the observations for 1000 CGDD are not significantly different comparing the distribution of furrow and pivot. However, looking at figure 5.4 the yield distribution of furrow and pivot looks very different. For the standard deviation observations the test does not reject the null hypothesis for the distribution of observations for pivot and furrow. Due to the inconclusive results of the actual ET observations the Kruskal Wallis H-test and the Mann Whitney U test is not performed for the WaPOR dataset. These inconclusive results might occur due to moderately correlation of WaPOR and VanderSat data with yield.

Table 5.8: Kruskal Wallis H-test on VanderSat actual ET measurements for 1000 CGDD and α =0.05 to test the distribution of furrow, sprinkler and pivot irrigation systems.

VanderSat actual ET	Test statistics	p-value
Cumulative mean	13.194	0.001
Mean	4.523	0.104
Standard deviation	6.595	0.037

As well for the ETact products, the slope is calculated using the first 1000 CGDD. The assumption is that fields with high yields have a steeper slope than fields with poor yields. The Spearman rank correlation coefficient is used to quantify the relation between actual evapotranspiration and yield. The actual evapotranspiration of the VanderSat product gives a Spearman rank correlation coefficient of 0.46. For WaPOR the correlation coefficient is 0.46. This means that both products have a moderately positive correlation with yield. A correlation of 0.46 is not enough to give crop yield performance estimates on a field basis. Improvements in the algorithm should be made if ET products are being used to assist in spatial variability of sugarcane yield. Figure 5.24 visualizes the slope of the ETact per field in relation to yield. Looking at the WaPOR data the furrow have generally a less steep slope than most pivot and sprinkler fields. Looking at the VanderSat data, there is no cluster visible between slope and irrigation system.



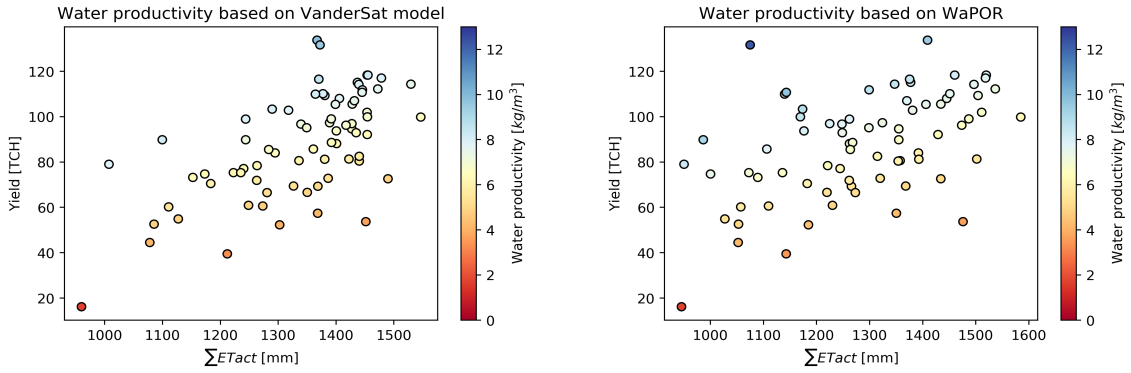
(a) (b)

Figure 5.24: The slope, also called as rate of increase, is computed on field-scale basis by the cumulative mean (μ) using the first 1000 CGDD. The colors indicates the type of irrigation system of the fields. (a) Rate of increase of the cumulative mean VanderSat ETact. (b)Rate of increase of the cumulative mean WaPOR ETact.

Water Productivity

An agricultural performance indicator as water productivity can be used to monitor and evaluate water use efficiency [Blatchford et al., 2019]. The water productivity should be used in a relative way to compare the water productivity between fields or over time for the same user and not as a absolute value [Blatchford et al., 2019]. By knowing the ETact and the yield the water productivity can be computed. The water productivity per field is plotted in figure 5.25. For WaPOR the maximum value for water productivity is 12.24 kg/m^3 and the minimum value is 1.71 kg/m^3. For Vandersat the maximum value for water productivity is 9.78 kg/m^3 and the minimum value is 1.69 kg/m^3. The maximum water productivity value for WaPOR is for the field AG3B28. VanderSat shows a significant lower value for field AG3B28, namely 9.78 kg/m^3. The water productivity based on VanderSat and WaPOR show both a high variability in field performance. In literature water productivity values for sugarcane in Brazil can be found ranging from 8.13 to 16.53 kg/m^3 [Leal et al., 2017]. Studies in South Africa show water productivity values for sugarcane of 4.8 to 12.1 kg/m^3 [Robertson and Muchow, 1994]. These water productivity values of South Africa are in the same range as the water productivity value at this study site. In this study case some fields

have water productivity values lower than 4.8 kg/m^3. The assumption can be made that the water productivity of these fields can be significantly improved. In this study the difference between sugarcane species are not taken into account. It is not known if clusters exists in water productivity. The correlation coefficient is not computed for the water productivity since the yield is used to compute this parameter.



(a) (b)

Figure 5.25: Water productivity on field-scale level for the study site. On the x-axis the accumulation of actual evapotranspiration can be found and the y-axis the yield (a) Water productivity on field-scale level based on VanderSat ETact. (b) Water productivity on field-scale level based on WaPOR ETact.

Since potential evapotranspiration is only available for the Vandersat evapotranspiration product, the ratio of actual to potential evapotranspiration can only be computed for this product. The ratio of actual to potential evapotranspiration measures crop water supply in relation to the crop water demand [Yao, 1974]. By computing the ratio of actual to potential evapotranspiration the ET observations can be linked to crop stress. The ratio of actual to potential evapotranspiration of VanderSat is presented in figure 5.26. Using Spearman rank correlation the ratio of actual to potential evapotranspiration shows a strong positive correlation of 0.71 with yield. The values of the ratio of actual to potential evapotranspiration ranges from 0.72 to 0.91. Studies have shown that a ratio of 0.85 - 0.9 is the optimum evapotranspiration [Yao, 1974]. Ratios of actual to potential evapotranspiration of 0.60 or lower are considered as requiring irrigation for crop growth [Yao, 1974]. The fields in this study site are all irrigated, which can be the reason that the minimum value is 0.72. Studies assume that optimum evapotranspiration have a close relationship with potential yield [Yao, 1974]. Therefore, in this study it is assumed that for fields with the ratio of actual to potential evapotranspiration below 0.85, crop stress and thus yield stress is occurring.

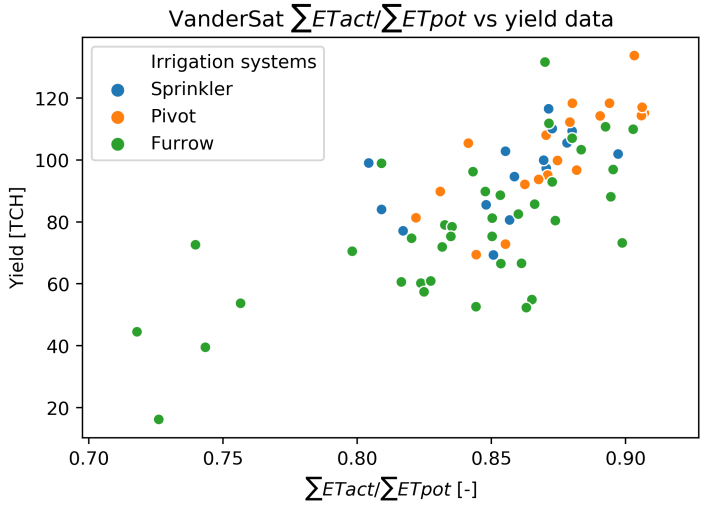


Figure 5.26: $\sum ETact/\sum ETpot$ based on VanderSat ET data and compared to yield data.

Figure 5.27 visualizes the difference between accumulated potential ET and accumulated actual ET on a field level over a crop season. Fields with high yields show smaller differences between accumulated potential and accumulated actual evapotranspiration than fields with poor yields. In figure 5.27 a cluster of 5 fields are visible. This cluster is marked by a black oval circle in figure 5.27. In table 5.9 the difference between accumulated potential evaporation and accumulated actual evaporation and their associated yield and field name can be found. The three poorest performing yield fields are found in that table. There is a clear relation visible between yield and $\sum ETpot$-$\sum ETact$. The difference between accumulated potential and accumulated actual evapotranspiration show a negative moderate to strong correlation with a Spearman rank correlation coefficient of -0.59. No clusters between type of irrigation systems are visible in figure 5.27. The spatial information provided in 5.27 is used to identify problematic fields. The potential evapotranspiration is provided on field level. However, when potential evapotranspiration is provided on a pixel-based level, information about in-field spatial variability and thus the non-uniformity can be provided.

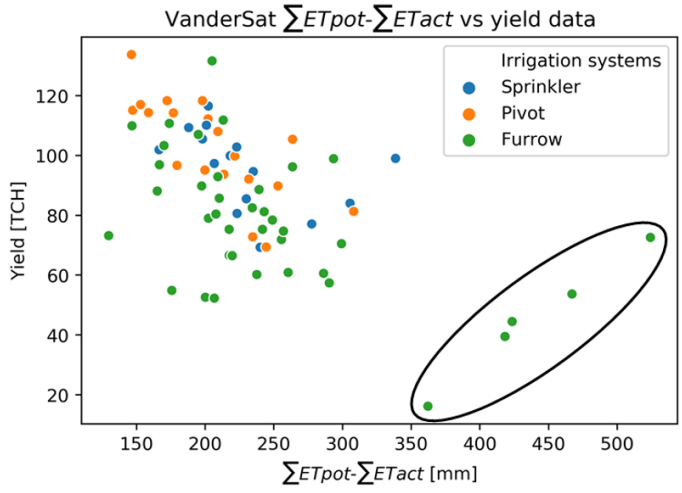


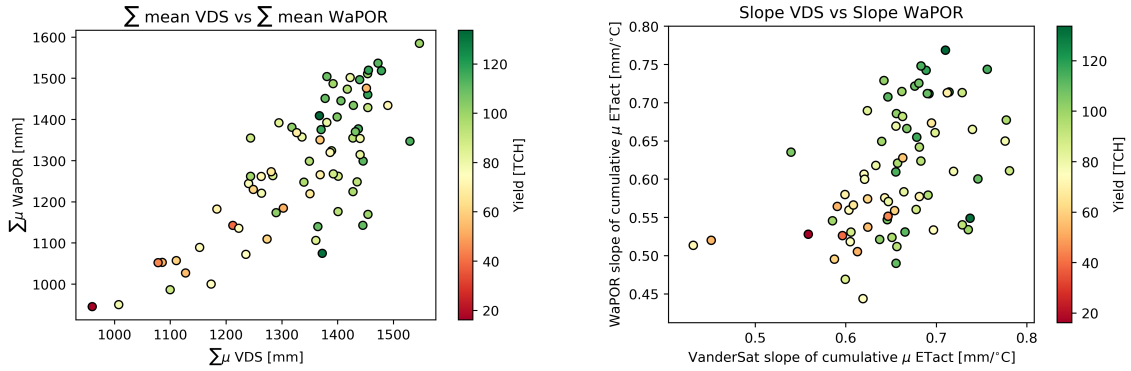
Figure 5.27: $\sum ETpot$-$\sum ETact$ based on VanderSat ET data and compared to yield data on field level. The results are colored by their type of irrigation system. The dots in the black oval circle indicate the outliers and the corresponding field names and yield can be found in table 5.9.

Table 5.9: highest $\sum ETpot/\sum ETact$ values and their corresponding fields

Field name	$\sum ETpot$-$\sum ETact$ [mm]	Yield [TCH]
AG3D30	524.20	72.6
AG3D31	467.11	53.7
AG3D32	423.53	44.5
AG3C27	418.31	39.5
AG3E30	362.19	16.2

Correlation between satellite derived evaporation product

The sum of mean ETact of WaPOR values and the sum of mean ETact values of VanderSat are plotted against each other to see if the two satellite derived ET products are correlated with each other. As can be seen in figure 5.28a the sum of mean ETact for both ET products have a strong positive correlation with each other. The Spearman rank correlation coefficient of the sum of mean of ETact between these products is 0.71 [-]. The strong correlation between these products can be explained by the similarities of the two approaches the ET products use. WaPOR uses the Penman-Monteith equation to compute the actual evapotranspiration. VanderSat ET product uses the Priestley-Taylor equation. Den Besten et al. (nd) show that the Priestly-Taylor and the Penman-Monteith equation have a strong positive correlation of 0.82. Tukimat et al. (2012) shows a Pearson correlation of 0.90 between the Priestly-Taylor and the Penman-Monteith equation [Tukimat et al., 2012]. By comparing the slope of both ET products for the first 1000 CGDD the correlation is 0.52 [-], which means that the slope of the products are moderately positive correlated with each other.



(a) (b)

Figure 5.28: Scatterplots between VanderSat and WaPOR. (a): Sum of mean of VDS - Sum of mean of WaPOR. (b): Slope of cumulative mean VanderSat - slope of cumulative mean WaPOR. The slope of line is computed for 0 - 1000.

6

General discussion

In the introduction the main research question is presented. This chapter provides the discussion around the main question:

How can the spatial variability of yield be explained and how is this reflected in existing Remote Sensing data?

First, this chapter discusses how the spatial variability of yield can be explained. Second, this chapter discusses how yield data is reflected in existing remote sensing data. The reason for this approach is that some of the limitations in determining the relationship between yield and remote sensing data occur for both the results of satellite derived vegetation indices as well as for the results of satellite derived evaporation product. Last, the limitations of the satellite derived vegetation indices and satellite derived evaporation products are discussed.

Correlation between topography, soil and groundwater properties

How can the spatial variability of yield be explained?

To explain the yield variability for sugarcane the relationship between yield and different field parameters such as elevation, slope and distance from the field to drainage pump are analyzed.

The obtained results show that the HAND has a moderately positive Pearson correlation of 0.52 with total yield. Thereby providing a possible explanation for the spatial variability of yield in this case study. However, the variables considered in this research have an influence on the obtained results. Therefore, it is important to evaluate the different data sets used in this thesis and their limitations.

The correlation study is executed by accumulating the yield data of the years 2013-2018 and computing the interannual variability of yield. These two yield variables are used to identify which fields perform consistently poor or good and determine the yield variability of the fields through time. The total yield and interannual variability are used because field parameters such as slope, HAND and soils also do not change over time. Alternatively, the yield of the year 2018 could be compared with the field parameters, possibly giving different results.

Since the yield data is provided on field-scale level, the other variables in the correlation study are aggregated to field level. This is done to compare the variables on the same scale-level. However, by taking only the mean elevation for a field or mean HAND for a field information within the field is not taken into considerations. To do research within the field a pixel-based analysis is preferred. Ideally, yield should also be recorded at a finer resolution

than field-scale. However, this is impractical to achieve since a field is harvested all at once.

There are more variables to be considered, such as nutrient deficits and crop diseases, which have an effect on yield. Moreover, it is important to understand that the results provide evidence to explain yield variation in this case. This does not mean that these same variables considered in this thesis have the same effect on yield at other locations in the world.

Furthermore, it is known that meteorological conditions and ratoon age of the crop have a significant effect on the yield. After the first harvest, another crop is regenerated from the cut stalks, where cut stalks are left to stick out just above the ground. Thus the yield decreases over time with increasing ratoon age. These parameters were not taken into account in this research.

By using the Pearson correlation coefficient, a fundamental assumption is introduced, namely that the variables are normally distributed and that there is a linear relationship between yield and the considered variables.

Spatial patterns in remote sensing data

How is the yield data reflected in existing Remote Sensing data?

In this study the spatial patterns of remote sensing data are compared with yield data and the variation of remote sensing data in one specific field is quantified by using the cumulative coefficient of variation as a statistical analysis.

There are examples of other methods to quantify the spatial variation in an area. An example of this is variograms [Inamura et al., 2004]. The variogram express the way that a parameter varies over a specific distance. For this study a variogram can be computed per field per obtained image. However, in this study the areas of the fields vary a lot. The smallest field is around 3 ha and the biggest field is approximately 60 ha. Therefore, the distance is constraint for the use of variograms. This is because the fields are demarcated.

Another method to quantify the non-uniformity in the field is the Shannon-Wiener diversity index (SWDI) [Diker et al., 2004]. Diker et al. (2004) clip the yield provided by the yield monitor in smaller areas such as it is corresponding with the spatial resolution of the NDVI images.

However, we need to keep in mind that for this research only field-scale yield data is provided. Therefore, by using the coefficient of variation the spatial variation within the field is expressed in the coefficient of variation. The coefficient of variation is computed by the standard deviation divided by the mean, where the standard deviation is a measure of the amount of variation in a field. A low standard deviation indicates that the values of a parameter tend to be close to the mean, which means the non-uniformity is low. A higher standard deviation indicates that the values of the parameter are more spread out and have a larger deviation of the mean, which means the non-uniformity is high.

The obtained correlation values in this study should be interpreted with caution, because the data analyzed is only from one growing season. To obtain more reliable information multiple years of data should be analyzed.

The Kruskal Wallis H-Test and the Mann-Whitney U test are used in agricultural applications to test if the yield distribution differs between groups. In this case the groups are the different irrigation management systems (furrow, sprinkler, pivot). The type of irrigation systems used in the field are affecting the yield. To understand how big the influence on yield is for these different irrigation systems, a performance comparison study should be executed. In this research the influence of the irrigation systems are treated secondary. The results of the Kruskal Wallis H-test show that the distribution of the satellite derived vegetation indices and satellite derived ET estimates for these irrigation management systems differs significantly with each other.

Spatial patters in satellite derived vegetation indices

The satellite derived vegetation indices are obtained by optical imagery. A reason to interpret the optical results with caution is the fact that, because of cloud coverage, huge data gaps occur between acquired optical images. For some months only one optical satellite image is acquired. This occurs for example for the months October, November and December in 2017. To obtain more reliable results multiple images in a month should be acquired. This thesis resolved the data gaps using linear interpolation. However, closing the gap with more images where possible can improve the accuracy of the results.

Spatial patterns in satellite derived evaporation products

Actual evapotranspiration estimates are plotted as time series for the cumulative mean, cumulative standard deviation and the cumulative coefficient of variation. The results are colored with the corresponding yield data for that specific field. Looking at the ET time series of the cumulative mean the results of WaPOR and VanderSat have the shame shape. In general, fields with high yield have a higher cumulative mean ET than fields with poor yields. However, the cumulative CV of these two products look very different. The results shows that for WaPOR fields with high yield have a high cumulative CV and fields with poor yields have a lower cumulative CV. The results of VanderSat show the opposite. For VanderSat fields with high yield show a low cumulative CV and fields with poor yield show a higher cumulative CV. A possible cause for the high standard deviation and high CV values of the WaPOR results is that the WaPOR images are resampled to a 20 meter grid resolution using bilinear interpolation. The VanderSat ET estimates are not resampled.

When determining the relation between ET and yield, the results show that the ratio of actual to potential evapotranspiration and the difference between potential and actual evapotranspiration have a higher correlation with yield than solely the correlation between yield and actual evapotranspiration. The limitations of only using the actual evapotranspiration is that there is no information indicating that the water requirements is met. Important information is missing regarding the ET estimates under ideal circumstances.In other words, we want to know the specific water requirements of the sugarcane crop. The potential ET gives information about what the ET should be under well-watered surface and ambient atmospheric conditions. Ideally, potential and actual ET should be the same. Then the water requirements are met. However, actual ET gives information to what extent the water requirements are being met under prevailing conditions where constraints in radiation, air temperature, wind speed and water availability occur. The ratio of actual to potential evapotranspiration and the difference between potential and actual evapotranspiration may reveal areas/fields where water shortages and/or waterlogging occur. Hence the ratio of actual to potential evapotranspiration and the difference between potential and actual evapotranspiration gives information about the ET gap. These results can lead in improving irrigation or drainage in these areas and this can eventually lead to increase crop yields.

Water productivity can be computed to monitor and evaluate water use efficiency. However, to make a conclusion about increasing or decreasing water productivity, more than one year of data should be analyzed. It is important to take into account that increasing water productivity does not necessarily mean that water is saved. An increased water productivity can be caused by an increase of yield or a decrease in ET, while the same amount of water is irrigated on land. Other aspects, such as fertility management, can have an influence on the increase of yield. Furthermore, a high water productivity does not necessarily mean higher yields. High water productivity can occur at fields with low yields.

Preliminary results of this thesis have been shared by the scientists who are developing the VanderSat evaporation product and WaPOR. The interest of these scientists in how these products are correlated with yield emphasize how important it is to have a good ET product, which is correlated with yield, so it can be applied to optimize the sustainable water allocations and agricultural water use.

7

Conclusion

To forestall potential food shortages, crop yield and water use efficiency of irrigation water need to be improved. The current knowledge gap is how information on ET can be applied to optimize sustainable water allocations and agricultural water use. Hence meeting the water and food security goals in a changing climate. ET is linked to the water, carbon and energy cycle [Fisher et al., 2017]. Therefore, a good ET product should be correlated with yield. The aim of this research is to explain the spatial variability at the sugarcane plantation and to investigate whether the spatial patterns in the remote sensing data reflect difference in crop performance. This chapter will first answer the mean research question before answering the sub-questions.

In this research, the main research question to answer is:

How can the spatial variability of yield be explained and how is the spatial variability of yield reflected in existing remote sensing data?

In this case and for the variables considered in this research HAND is the most important field property to explain yield variation between fields. However, it is suspected that more field properties influence the yield variability. These factors did not come to light yet. The results presented here demonstrate that satellite derived vegetation indices as NDWI and NDVI can provide valuable insights in the spatial variability of yield. ***A relationship has been shown between yield and NDVI and NDWI by conducting a correlation study.*** The results show that NDWI shows a higher correlation with yield than NDVI. This study demonstrates that actual evapotranspiration satellite derived evaporation products show a moderately strong positive correlation with yield. ***Should actual evapotranspiration be used to analyse yield variation within and inter fields, the algorithm needs to be improved.*** By combining actual and potential evapotranspiration the difference between potential and actual evapotranspiration can be made as well as the ratio of actual to potential evapotranspiration. ***The difference between potential and actual evapotranspiration and the ratio of actual to potential evapotranspiration show a clear relation in yield and these two parameters can assist in detecting yield variability on a field-scale level.***

From here on the sub-questions are answered.

What is the relation between yield and topography, soil and groundwater properties?
In this study the HAND (height above near drainage) is the most important field property to explain yield variation between fields. Yield analysis shows a decrease in yield over years. For explaining yield variability relationships between yield, topography, groundwater and soil are analyzed on a field-scale level. A correlation study has been conducted between yield, topography and soil properties. The results show that the HAND has a moderately positive Pearson correlation of 0.52 with the total yield. Other topographical features

such as elevation, slope, curvature/convexity are analyzed and show no to low correlation with total as well as interannual yield. Furthermore, there is a low correlation between yield parameters and soil rank based on clay content. Comparing yield to soil type and type of irrigation system, furrow irrigation shows a higher spatial variability of yield between fields for all soil types. The clayey alluvial soils show a higher median yield for both sprinkler and pivot irrigation systems than alluvial soils of stratified texture. Groundwater properties such as groundwater depth and electrical conductivity are analyzed and show no relation with yield.

How are the spatial patterns in NDVI and NDWI related to yield?

In this study NDWI shows a stronger relation with yield than NDVI. This study analyzes the satellite derived vegetation indices NDVI and NDWI on field-scale level to investigate if their spatial patterns reflect the differences in crop performance by calculating the correlation between NDVI and NDWI in relation to yield. By computing the cumulative coefficient of variation, non-uniformity within the performing fields can be investigated. The hypothesis is that fields with yield above average show lower non-uniformity than fields with yield below average. The results show that this signal is reflected in the cumulative coefficient of variation at the end of the growth season. Since planting and harvest dates differ per field, cumulative growing degree days (CGDD) is used to implement a similar index for all fields. Using the CGDD the data is corrected for the effect of phenology and temperature. The results show that for 1350 cumulative growing degree days, mean NDWI field-scale values show a strong positive correlation with yield illustrated by a Spearman rank correlation coefficient of 0.73, whereas for NDVI the Spearman rank correlation coefficient is only 0.64. Therefore, this study demonstrates that the NDWI is better correlated with yield than NDVI.

Can actual evapotranspiration from existing remote sensing evaporation products be used to detect yield variation on field-scale level?

Improvements in the algorithm should be made if these actual ET products are to be used in analyzing spatial variability of sugarcane yield. The best correlation with yield is found for the cumulative mean of the VanderSat evaporation product. The WaPOR and VanderSat evaporation products are satellite derived evaporation products and these are analyzed on field-scale level to investigate if the spatial pattern for these products reflect difference in crop performance by calculating the correlation of these products with yield. WaPOR is based on the ET look model which uses the Penman-Monteith equation and VanderSat uses the Priestley-Taylor equation. Both evaporation products use NDVI as input for crop stress. Cumulative coefficient of variation is used to investigate the non-uniformity within the performing fields. VanderSat actual evapotranspiration (ET) estimates show a higher non-uniformity for poor performing fields (yield below average) than for good performing fields (yield above average). WaPOR actual ET estimates show the contrary. The actual ET estimates of WaPOR show for good performing fields a higher cumulative standard deviation and a higher cumulative coefficient of variation than for poor performing fields. Both evaporation products show a moderately positive correlation with yield. For the VanderSat evaporation product the highest correlation with yield is found with a Spearman rank correlation coefficient of 0.57 for the cumulative mean actual ET estimates for 1150 CGDD. For WaPOR the highest correlation with yield is found with a Spearman rank correlation coefficient of 0.50 for the cumulative mean actual ET estimates for 1000 CGDD and for the mean actual ET at 1250 CGDD. Therefore, for both products improvements in the algorithm should be made.

Can evaporative stress indicators be used to detect yield indicators on field-scale level? ***Satellite derived evaporation products can assist in spatial variability of yield when combining the potential and actual ET estimates.*** As evaporative stress indicators the difference between potential and actual evapotranspiration and the ratio of actual to potential evapotranspiration are computed. For this study only the potential evapotranspiration for VanderSat is available. Using Spearman rank correlation the ratio of actual to potential evapotranspiration shows a strong positive correlation of 0.71 with yield. The difference between potential and actual evapotranspiration shows a negative moderately to strong correlation with a Spearman rank correlation coefficient of -0.59. Since the results show that the parameters are strongly correlated, we can conclude that evaporative stress indicator can be used in analyzing the spatial variability of sugarcane yield on field-scale level.

8

Recommendations

The recommendations of this research are described in this chapter. Recommendations are based on the conclusions and purpose of this study

In this study only a moderate correlation has been found between the HAND and total yield. In order to understand what other factors cause the yield variability in the field, it would be interesting to know how much the influence of irrigation management systems was on the yield. This can be done by a performance comparison study between the three types of irrigation management systems: sprinklers, pivot and furrow. Improving irrigation management can result in an improvement of the yield and water use efficiency.

This study investigates if the spatial patterns in satellite derived vegetation indices NDVI and NDWI are in relation to the yield data. It would be wise to also investigate what the correlation is for yield with other satellite derived vegetation indices like EVI (Enhanced Vegetation Index) and WDVI (Weighted Difference Vegetation Index). Considering that vegetation indices can be used to estimate the performance of the field throughout the crop season. Besides that, future studies should also focus on how many cloud free satellite images are needed to be acquired in order to give an accurate estimation of crop yield performance and crop yield estimation.

The results show that, currently, there is no strong correlation between yield and the used satellite derived evaporation products. Therefore, future work on satellite derived evaporation products should focus on getting a stronger correlation with the yield.
The results show that NDWI has a stronger correlation to yield than NDVI. Both WaPOR and the VanderSat evaporation product use NDVI as their crop stress parameter. However, it might be wise to use NDWI instead of NDVI as a crop stress parameter in WaPOR and the VanderSat evaporation product. More research is needed to determine the usefulness of NDWI in satellite derived evaporation products and to investigate if this results in a higher correlation on sugarcane yield variability for these evaporation products.

Considering the approach that the evaporation products can be used for evaporation-based irrigation scheduling, it would be ideal to get daily actual evaporation estimates. Currently, optical imagery is used as a crop stress indicator. Optical data depends on cloud free conditions. This condition limits the quantity of optical data, which can have an effect on the accuracy of the evaporation estimates. To improve the evaporation estimates and thus the correlation with the yield throughout a crop season it is interesting to look at the possibility to implement radar to assess the optical data as stress indicator.

In this thesis the relationship between yield and the satellite derived evaporation products, WaPOR and VanderSat, are computed. This thesis does not investigate the biases and errors of these products, which are important to understand how reliable the relation between satellite-based evaporation products and yield data is. Triple collocation (TC), introduced by Stoffelen (1998), is a method to determine the systematic biases and random errors in products [Stoffelen, 1998, Vogelzang and Stoffelen, 2012]. Recent studies of TC show that it can be used on soil moisture datasets retrieved by satellites [Scipal et al., 2010, Miralles et al., 2010, Gruber et al., 2016, Dong and Crow, 2017]. The triple collocation provides the errors of the measurements without having access to the true value [Nearing et al., 2017]. It must be taken into account that three independent measurement systems are needed. These three independent systems need to deliver simultaneous collocated measurements. Due to the high spatial resolution of the evaporation products used in this thesis, triple collocation can be a good application for errors analysis. Nonetheless, a third independent measurement system is still needed for this application.

Lastly, due to limited time this study uses mean, standard deviation and coefficient of variation of variables on a field-scale level to analyze the spatial variability of sugarcane yield with remote sensing data. An interesting next step would be, to do an analysis of spatial variability on sugarcane yield on a pixel based level per field. Then, on a pixel-scale level it can be mapped if a certain pixel performs consistently poor throughout one crop season or even over multiple years. The standard deviation and the coefficient of variation on a field-scale level can identify fields with a high non-uniformity of yield, but these results on pixel level can assist to map the areas where over-irrigation or under-irrigation is happening.

Appendices

A

Information about soil legend symbols

In this appendix additional information about the soil symbols are visualized in figure A.1.

Simplified soil serie legend for Chibanza, Taninga, East block (Mepambe, Madala, Xinavane and Chiannissano) and Aguiar blocks

Map Symbol	Soil Unit	Geomorphology and geology	Landscape	Soil Texture Topsoil/subsoil	Drainage	FAO Classification	Main Limitations
FG	Fine to medium textured alluvial soils developed from alluvial sediments	Holocene Alluvial (fluvio) deposits	Almost flat (slopes < 2%), sometimes with micro undulation	Sandy clay to clay - Silt clay to clay	Well to Moderately drained	Haplic Fluvisoil	Sometimes drainage
FGg	Fine textured alluvial soils with Mottling phase			Sandy clay to clay - Silt clay to clay	Imperfectly Well drained	Haplic / Gleyic Fluvisoil	Drainage
FGgh	Fine to medium textured alluvial soils (hidromorphic phase)			Sandy clay to clay - Silt clay to clay	Imperfectly to poorly drained	Gleyic Fluvisoil	Drainage
FGgz	Fine to medium textured alluvial soils (mottling and salic phase)			Sandy clay to clay - Silt clay to clay	Imperfectly to poorly drained	Salic Luvisols	Drainage, salinity and sometimes sodicity
FS1	Stratified alluvial soil with > 50 cm of medium to fine texture layer over coarse texture layers		Gently slopping to almost flat levee with coarse to medium stratified alluvium texture	Clay loam to clay - Sandy to clay-loam	Well drained	Haplic Fluvisols	Sometimes drainage
FS1g	Stratified alluvial soil with > 50 cm of medium to fine texture layer over coarse texture layers (mottling phase)			Sand to sandy loam - Sandy loam to sandy-clay-loam	Moderately to imperfectly well drained	Haplic / Gleyic Fluvisoil	Sometimes imperfectly drained
FS2	Stratified alluvial soil with 25 - 50 cm of medium to fine texture layer over coarse texture layers			Sandy loam to clay - sandy to clay	Moderately to well drained	Haplic Fluvisoil	Sometimes drainage
FS2g	Stratified alluvial soil with 25 - 50 cm of medium to fine texture layer over coarse texture layers (mottling phase)			Sandy loam to clay - sandy to clay	Moderately to imperfectly well drained	Haplic / Gleyic Fluvisoil	Sometimes imperfectly drained
FS3	Medium textured soils with < 25 cm of top coarse textured layer			Clay loam to clay - Sandy to clay	Well drained	Haplic Fluvisoil	Sometimes low water holding capacity
FS4	Stratified alluvial soil with > 50 cm of coarse texture layer			Sandy loam - Sandy loam	Well to excessively drained	Haplic Fluvisoil	Low natural fertility Low water holding capacity
FGM1	Fine to medium textured alluvial soils developed from alluvial sediments over Mananga sediments	Alluvial (fluvio) deposits under Pleistocene indurate sodic deposits material denominated "Mananga"	lower slopes areas, seasonally inundated	Silt loam to clay - Silt clay to clay	Moderately well to imperfectly drained	Haplic Fluvisoil / Glayic Luvisols	Poor drainage, poor workability, salinity and sodicity
CM	Clay soils derived from Mananga sediments	Colluviums derived from Mananga	Flat alluvial plain with micro-undulations	Sandy clay loam to clay - Sandy clay loam to clay loam	Moderately well to imperfectly drained	Haplic Glaysoil	Poor drainage, poor workability, salinity and sodicity

Figure A.1: Copy of the simplified soil legend used in the report Soil Survey Report of Açucareira de Xinavane, SA [Vilanculos and Mafalacusser, 2012].

B

Map of NDVI and NDWI

Using Sentinel-2 NDVI and NDWI are computed. The NDVI and NDWI data can be visualized as maps for the study site to provided spatial information within the fields. The maps shows a certain day (dd-mm-yyyy), which is within the time period of interest for this study.

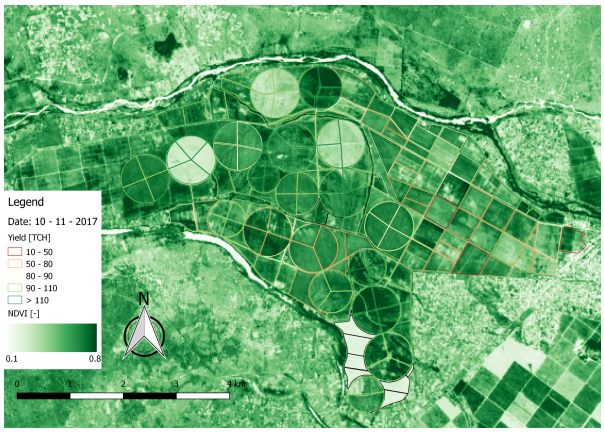


Figure B.1: NDVI map for 10-11-2017. The boundaries of the parcels are colored by the yield value for the year 2018. Black boundaries means that no sugarcane crops are grown in those fields.

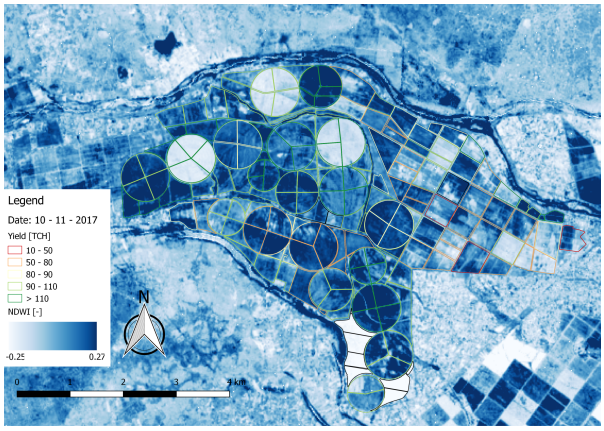
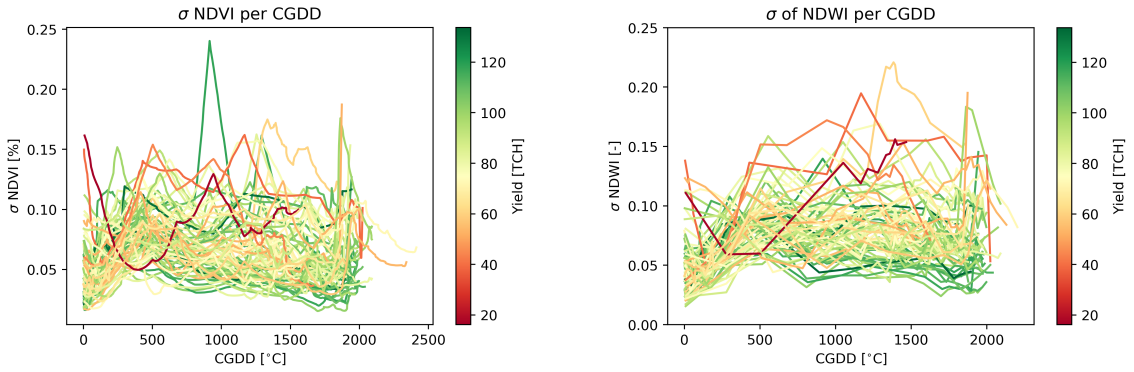


Figure B.2: NDWI map for 10-11-2017. The boundaries of the parcels are colored by the yield value for the year 2018. Black boundaries means that no sugarcane crops are grown in those fields.

C

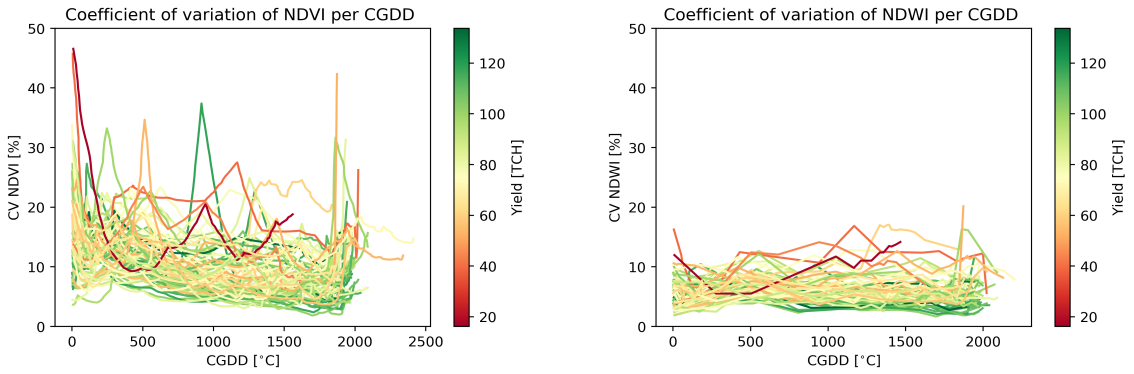
NDVI & NDWI in relation to CGDD

In this appendix the standard deviation and coefficient of varation of NDVI and NDWI observations on field-scale level in relation of CGDD are presented. Both the standard deviation plot as well as the coefficient of variation plots show that the variation of the spectral response signal in the fields varies through time.



(a) (b)

Figure C.1: Standard deviation of NDVI and NDWI observations per field in relation to CGDD. The color of the lines indicates the yield data.(a) Standard deviation of NDVI observations per field in relation to CGDD. (b) Standard deviation of NDWI observations per field in relation to CGDD.



(a) (b)

Figure C.2: NDVI and NDWI observations per field are visualized in relation to CGDD. The color of the lines indicates the yield data. (a) Coefficient of variation of NDVI observations per field in relation to CGDD. (b) Coefficient of variation of NDWI observations per field in relation to CGDD.

D

Map of actual evapotranspiration

WaPOR and VanderSat actual evapotranspiration estimates are provided as tiff files. Figures D.1 and D.2 visualize a tiff file for a certain day (dd-mm-yyyy), which is within the time period of interest for this study. The evapotranspiration data can be visualized as maps for the study site to provided spatial information within the fields.

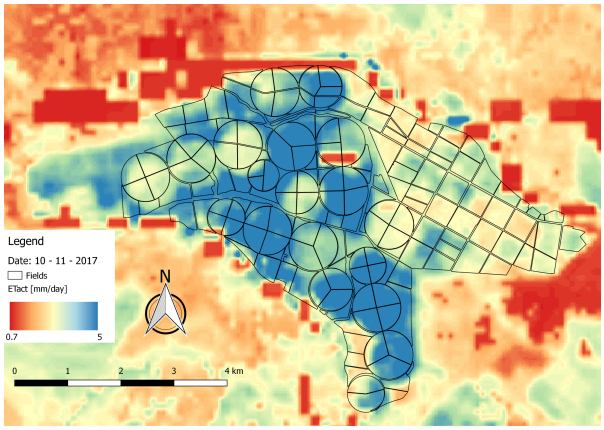


Figure D.1: Actual ET map of WaPOR for 10-11-2017. The colors indicates the actual evapotraspiration.The pixel resolution of the map is 20 meter.

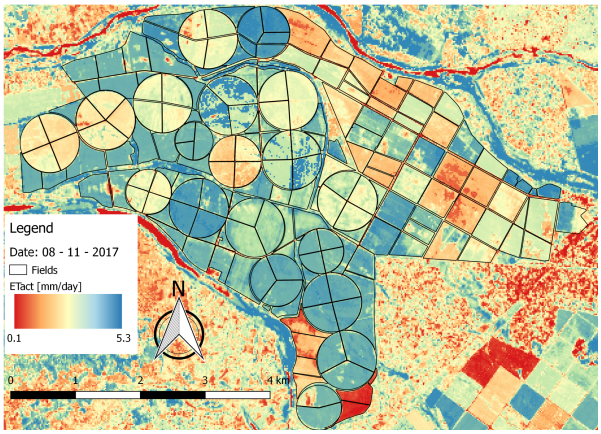
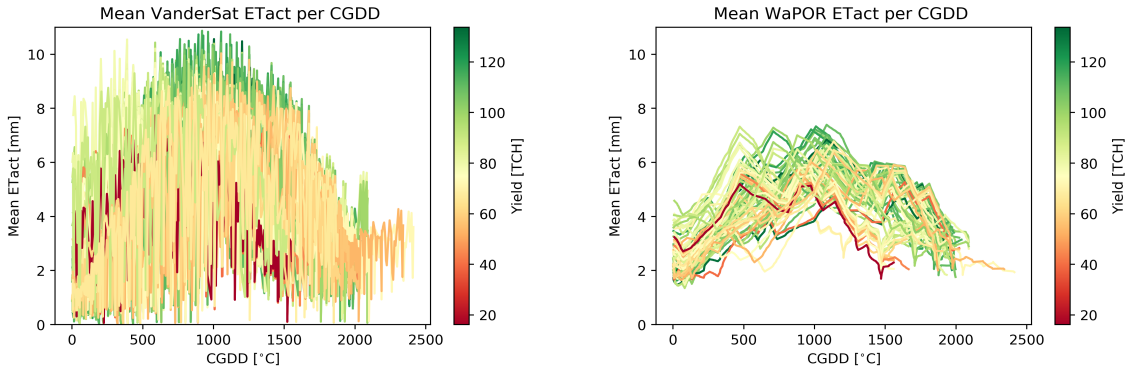


Figure D.2: Actual ET map of VanderSat evaporation product for 08-11-2017. The colors indicates the actual evapotraspiration. The pixel resolution of the map is 20 meter.

E

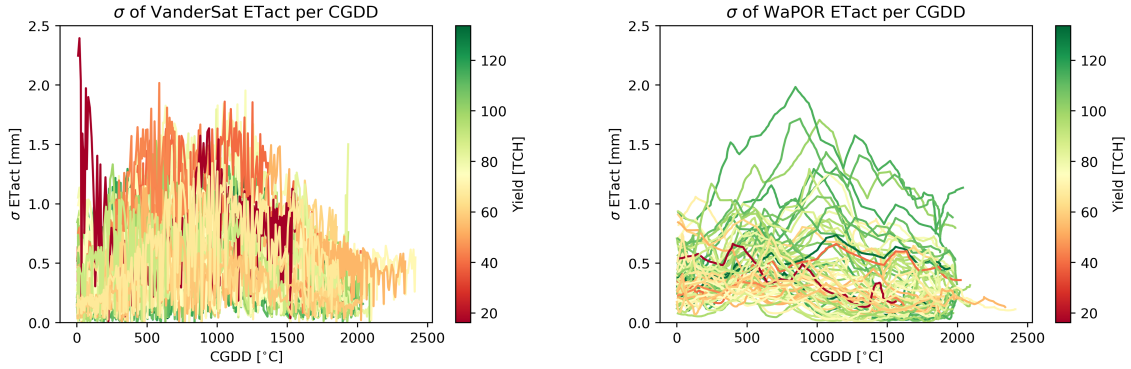
VanderSat & WaPOR ET data in relation to CGDD

In this appendix the mean, standard deviation and coefficient of variation of evapotransipration data on field-scale levl are presented in relation of CGDD. The mean ETact data show that the Priestley-Taylor equation show more variability than by using the Penman-Monteith equation.



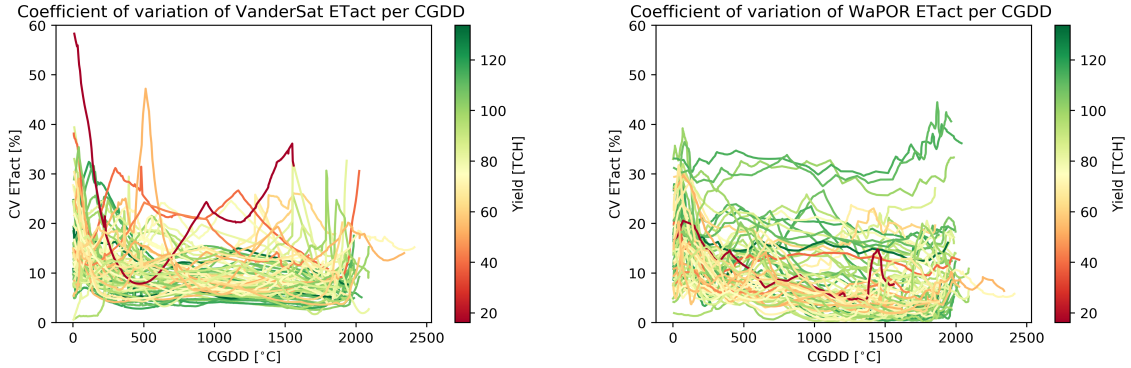
(a) (b)

Figure E.1: Mean (μ) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Mean observations are averaged per field. The color of the lines indicates the yield.



(a) (b)

Figure E.2: Standard deviation (σ) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Standard deviation observations are averaged per field. The color of the lines indicates the yield.



(a) (b)

Figure E.3: Coefficient of variation (CV) observations of actual evapotranspiration ($ETact$) in relation to CGDD are visualized for VanderSat in (a) and for WaPOR in (b). Coefficient of variation observations are averaged per field. The color of the lines indicates the yield.

Bibliography

[Abbas et al., 2013] Abbas, A., Khan, S., Hussain, N., Hanjra, M. A., and Akbar, S. (2013). Characterizing soil salinity in irrigated agriculture using a remote sensing approach. *Physics and Chemistry of the Earth, Parts A/B/C*, 55:43–52.

[Al-Khaier, 2003] Al-Khaier, F. (2003). Soil salinity detection using satellite remote sensing.

[An et al., 2017] An, N., Hemmati, S., and Cui, Y.-J. (2017). Assessment of the methods for determining net radiation at different time-scales of meteorological variables. *Journal of Rock Mechanics and Geotechnical Engineering*, 9(2):239–246.

[Anderson et al., 2010] Anderson, M., Kustas, W., Norman, J., Hain, C., Mecikalski, J., Schultz, L., González-Dugo, M., Cammalleri, C., d'Urso, G., Pimstein, A., et al. (2010). Mapping daily evapotranspiration at field to global scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci. Discuss*, 7:5957–5990.

[Anderson et al., 2011] Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., González-Dugo, M., Cammalleri, C., d'Urso, G., Pimstein, A., et al. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*.

[Antognelli, 2020] Antognelli, S. (2020). Ndvi and ndmi vegetation indices: Instructions for use. https://www.agricolus.com/en/indici-vegetazione-ndvi-ndmi-istruzioni-luso/.

[Asfaw et al., 2018] Asfaw, E., Suryabhagavan, K., and Argaw, M. (2018). Soil salinity modeling and mapping using remote sensing and gis: The case of wonji sugar cane irrigation farm, ethiopia. *Journal of the Saudi Society of Agricultural Sciences*, 17(3):250–258.

[Bannari et al., 2018] Bannari, A., El-Battay, A., Bannari, R., and Rhinane, H. (2018). Sentinel-msi vnir and swir bands sensitivity analysis for soil salinity discrimination in an arid landscape. *Remote Sensing*, 10(6):855.

[Bastiaanssen et al., 2012] Bastiaanssen, W., Cheema, M., Immerzeel, W., Miltenburg, I., and Pelgrum, H. (2012). Surface energy balance and actual evapotranspiration of the transboundary indus basin estimated from satellite measurements and the etlook model. *Water Resources Research*, 48(11).

[Bastiaanssen et al., 2005] Bastiaanssen, W., Noordman, E., Pelgrum, H., Davids, G., Thoreson, B., and Allen, R. (2005). Sebal model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of irrigation and drainage engineering*, 131(1):85–93.

[Bastiaanssen et al., 1998] Bastiaanssen, W. G., Menenti, M., Feddes, R., and Holtslag, A. (1998). A remote sensing surface energy balance algorithm for land (sebal). 1. formulation. *Journal of hydrology*, 212:198–212.

[Bégué et al., 2010] Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., and Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using ndvi. *International Journal of Remote Sensing*, 31(20):5391–5407.

[Begue et al., 2008] Begue, A., Todoroff, P., and Pater, J. (2008). Multi-time scale analysis of sugarcane within-field variability: improved crop diagnosis using satellite time series? *Precision Agriculture*, 9(3):161–171.

[Berlin, 2020] Berlin, F. U. (2020). Remote sensing data analysis online course. https://blogs.fu-berlin.de/reseda/sentinel-2/.

[Blatchford et al., 2019] Blatchford, M. L., Mannaerts, C. M., Zeng, Y., Nouri, H., and Karimi, P. (2019). Status of accuracy in remotely sensed and in-situ agricultural water productivity estimates: A review. *Remote Sensing of Environment*, 234:111413.

[Bolton and Friedl, 2013] Bolton, D. K. and Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173:74–84.

[Brownlee, 2019] Brownlee, J. (2019). How to calculate nonparametric statistical hypothesis tests in python. https://machinelearningmastery.com/nonparametric-statistical-significance-tests-in-python/.

[Clements, 1962] Clements, H. F. (1962). The ripening of sugar cane.

[Coenders-Gerrits et al., 2020] Coenders-Gerrits, M., Schilperoort, B., and Jiménez-Rodríguez, C. (2020). Evaporative processes on vegetation: An inside look. In *Precipitation Partitioning by Vegetation*, pages 35–48. Springer.

[den Besten et al., nd] den Besten, N., Kassing, R., Muchanga, B., Earnshaw, C., Muleya, M., de Jeu, R., Karimi, P., and van der Zaag, P. (n.d.). Estimation of daily evaporation at field-scale for sustainable water management in a sugarcane plantation in xinavane, mozambique. Under review.

[den Besten et al., 2019] den Besten, N., Schellekens, J., de Jeu, R., and van der Zaag, P. (2019). The influence of shallow groundwater on the actual transpiration flux of irrigated fields using satellite observations.

[Diker et al., 2004] Diker, K., Heermann, D., Bausch, W., and Wright, D. (2004). Shannon-wiener's diversity index for linking yield monitor and remotely sensed data for corn. *Transactions of the ASAE*, 47(4):1347.

[Dinka and Ndambuki, 2014] Dinka, M. O. and Ndambuki, J. M. (2014). Identifying the potential causes of waterlogging in irrigated agriculture: the case of the wonji-shoa sugar cane plantation (ethiopia). *Irrigation and drainage*, 63(1):80–92.

[Dong and Crow, 2017] Dong, J. and Crow, W. T. (2017). An improved triple collocation analysis algorithm for decomposing autocorrelated and white soil moisture retrieval errors. *Journal of Geophysical Research: Atmospheres*, 122(24):13–081.

[eLEAF, 2020] eLEAF (2020). Final draft text for some of the wapor v2.1. data components. Copied by Karin Viergever, eLEAF, for Kathelijne Beenen on 20/1/2020; personal correspondence.

[ESA, 2015] ESA (2015). Sentinel-2 user handbook.

[ESA, 2020a] ESA (2020a). Level-2a product formatting. https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/product-formatting.

[ESA, 2020b] ESA (2020b). Proba-v - esa eo missions. https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/proba-v.

[FAO, 2016] FAO (2016). *AQUASTAT Country Profile – Mozambique.* Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

[FAO, 2018] FAO (2018). Wapor database methodology: Level 2. remote sensing for water productivity technical report: Methodology series. Licence: CC BY-NC-SA 3.0 IGO.

[FAO and IHE DELFT, 2019] FAO and IHE DELFT (2019). Wapor quality assessment: Technical report on the data quality of the wapor fao database version 1.0. http://www.fao.org/3/ca4895en/CA4895EN.pdf.

[Fisher et al., 2017] Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., et al. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, 53(4):2618–2626.

[Francois et al., 2014] Francois, M., Santandrea, S., Mellab, K., Vrancken, D., and Versluys, J. (2014). The proba-v mission: The space segment. *International Journal of Remote Sensing*, 35(7):2548–2564.

[G4AW, 2014] G4AW (2014). Mozambique quickscan. https://g4aw.spaceoffice.nl/en/projects/country-visits/mozambique/.

[Gao, 1996] Gao, B.-C. (1996). Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266.

[Gerrits, 2010] Gerrits, A. (2010). The role of interception in the hydrological cycle. http://resolver.tudelft.nl/uuid:7dd2523b-2169-4e7e-992c-365d2294d02e.

[Grieve et al., 2012] Grieve, C. M., Grattan, S. R., and Maas, E. V. (2012). Plant salt tolerance. *Agricultural salinity assessment and management*, 2:405–459.

[Gruber et al., 2016] Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, 45:200–211.

[Gu et al., 2008] Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., and Verdin, J. P. (2008). Evaluation of modis ndvi and ndwi for vegetation drought monitoring using oklahoma mesonet soil moisture data. *Geophysical Research Letters*, 35(22).

[Hobbins and Huntington, 2016] Hobbins, M. and Huntington, J. (2016). Evapotranspiration and evaporative demand.

[Huang et al., 2009] Huang, J., Chen, D., and Cosh, M. (2009). Sub-pixel reflectance unmixing in estimating vegetation water content and dry biomass of corn and soybeans cropland using normalized difference water index (ndwi) from satellites. *International Journal of Remote Sensing*, 30(8):2075–2104.

[Hurst et al., 2004] Hurst, C. A., Thorburn, P. J., Lockington, D., and Bristow, K. L. (2004). Sugarcane water use from shallow water tables: implications for improving irrigation water use efficiency. *Agricultural Water Management*, 65(1):1–19.

[IFAD, 2016] IFAD (2016). Investing in rural people in mozambique.

[Inamura et al., 2004] Inamura, T., Goto, K., Iida, M., Nonami, K., Inoue, H., and Umeda, M. (2004). Geostatistical analysis of yield, soil properties and crop management practices in paddy rice fields. *Plant Production Science*, 7(2):230–239.

[IPCC, 2014] IPCC (2014). *Climate change 2014:Impacts, adaptation and vulnerability: Regional aspects. Part A: global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

[IWACA-TECH, 2020] IWACA-TECH (2020). The iwaca-tech project. http://www.iwacatech.com/.

[Khalili, 2007] Khalili, B. (2007). *Monitoring of Incomati River Basin with Remote Sensing*. Lund University.

[Kitchen et al., 2003] Kitchen, N., Drummond, S., Lund, E., Sudduth, K., and Buchleiter, G. (2003). Soil electrical conductivity and topography related to yield for three contrasting soil–crop systems. *Agronomy journal*, 95(3):483–495.

[Koenders et al., 2019] Koenders, D., van den Munckhof, G., van Egmond, J., Brummelkamp, J., van Dongen, K., and Schurer, S. (2019). Improving water efficiency and crop yield on a sugarcane planta- tion in xinavane, mozambique.

[Kottek et al., 2017] Kottek, M., Grieser, J., Rudolf, B., and Rubel, F. (2017). World map of the köppen-geiger climate classification updated. http://koeppen-geiger.vu-wien.ac.at/present.htm.

[Kravchenko and Bullock, 2000] Kravchenko, A. N. and Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, 92(1):75–83.

[Laerd Statistics, 2018a] Laerd Statistics (2018a). Pearson's product-moment correlation. https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php.

[Laerd Statistics, 2018b] Laerd Statistics (2018b). Spearman's rank-order correlation. https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php.

[Leal et al., 2017] Leal, D. P., Coelho, R. D., Barbosa, F. d. S., Júnior, F., Eusimio, F., Mauri, R., and Santos, L. d. C. (2017). Water productivity for sugar and biomass of sugarcane varieties. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 21(9):618–622.

[Lieuw, 2001] Lieuw, S. (2001). Optical remote sensing. https://crisp.nus.edu.sg/ research/tutorial/optical.htm.

[Lobell, 2013] Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143:56–64.

[Lobell et al., 2010] Lobell, D. B., Ortiz-Monasterio, J. I., and Lee, A. S. (2010). Satellite evidence for yield growth opportunities in northwest india. *Field crops research*, 118(1):13–20.

[Lofton et al., 2012] Lofton, J., Tubana, B. S., Kanke, Y., Teboh, J., Viator, H., and Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors (Basel, Switzerland)*, 12(6):7529–47.

[Margulis, 2017] Margulis, S. A. (2017). Introduction to hydrology.

[Martens et al., 2018] Martens, B., de Jeu, R., Verhoest, N., Schuurmans, H., Kleijer, J., and Miralles, D. (2018). Towards estimating land evaporation at field scales using gleam. *Remote Sensing*, 10(11):1720.

[McCabe et al., 2019] McCabe, M. F., Miralles, D. G., Holmes, T. R., and Fisher, J. B. (2019). Advances in the remote sensing of terrestrial evaporation.

[McGlinchey and Inman-Bamber, 1996] McGlinchey, M. and Inman-Bamber, N. (1996). Predicting sugarcane water use with the penman-monteith equation.

[Miralles, 2011] Miralles, D. (2011). Evaporation in the Global Water Cycle. Analysing Land Evaporation Using Satellite Observations.

[Miralles et al., 2010] Miralles, D. G., Crow, W. T., and Cosh, M. H. (2010). Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology*, 11(6):1423–1429.

[Mitášová and Hofierka, 1993] Mitášová, H. and Hofierka, J. (1993). Interpolation by regularized spline with tension: Ii. application to terrain modeling and surface geometry analysis. *Mathematical Geology*, 25(6):657–669.

[Mladenova et al., 2017] Mladenova, I. E., Bolten, J. D., Crow, W. T., Anderson, M. C., Hain, C. R., Johnson, D. M., and Mueller, R. (2017). Intercomparison of soil moisture, evaporative stress, and vegetation indices for estimating corn and soybean yields over the us. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4):1328–1343.

[Molden et al., 2010] Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M. A., and Kijne, J. (2010). Improving agricultural water productivity: Between optimism and caution. *Agricultural Water Management*, 97(4):528–535.

[Molijn et al., 2016] Molijn, R. A., Iannini, L., Hanssen, R. F., and Rocha, J. V. (2016). Sugarcane growth monitoring through spatial cluster and temporal trend analysis of radar and optical remote sensing images.

[Molijn et al., 2018] Molijn, R. A., Iannini, L., Rocha, J. V., and Hanssen, R. F. (2018). Ground reference data for sugarcane biomass estimation in são paulo state, brazil. *Scientific data*, 5:180150.

[Molijn et al., 2019] Molijn, R. A., Iannini, L., Vieira Rocha, J., and Hanssen, R. F. (2019). Sugarcane productivity mapping through c-band and l-band sar and optical satellite imagery. *Remote Sensing*, 11(9):1109.

[Morel et al., 2014] Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J.-F., and Petit, M. (2014). Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on reunion island. *Remote Sensing*, 6(7):6620–6635.

[Mulianga et al., 2015] Mulianga, B., Bégué, A., Clouvel, P., and Todoroff, P. (2015). Mapping cropping practices of a sugarcane-based cropping system in kenya using remote sensing. *Remote Sensing*, 7(11):14428–14444.

[Mulianga et al., 2013] Mulianga, B., Bégué, A., Simoes, M., and Todoroff, P. (2013). Forecasting regional sugarcane yield based on time integral and spatial aggregation of modis ndvi. *Remote Sensing*, 5(5):2184–2199.

[Nearing et al., 2017] Nearing, G. S., Yatheendradas, S., Crow, W. T., Bosch, D. D., Cosh, M. H., Goodrich, D. C., Seyfried, M. S., and Starks, P. J. (2017). Nonparametric triple collocation. *Water Resources Research*, 53(7):5516–5530.

[Nobre et al., 2011] Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S. (2011). Height above the nearest drainage–a hydrologically relevant new terrain model. *Journal of Hydrology*, 404(1-2):13–29.

[Pagán et al., 2019] Pagán, B. R., Maes, W. H., Gentine, P., Martens, B., and Miralles, D. G. (2019). Exploring the potential of satellite solar-induced fluorescence to constrain global transpiration estimates. *Remote Sensing*, 11(4):413.

[Panda et al., 2010] Panda, S. S., Ames, D. P., and Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(3):673–696.

[Pereira et al., 2002] Pereira, L. S., Oweis, T., and Zairi, A. (2002). Irrigation management under water scarcity. *Agricultural water management*, 57(3):175–206.

[Petersen, 2018] Petersen, L. K. (2018). Real-time prediction of crop yields from modis relative vegetation health: A continent-wide analysis of africa. *Remote Sensing*, 10(11):1726.

[Pinheiro Lisboa et al., 2018] Pinheiro Lisboa, I., Melo Damian, J., Roberto Cherubin, M., Silva Barros, P. P., Ricardo Fiorio, P., Cerri, C. C., and Eduardo Pellegrino Cerri, C. (2018). Prediction of sugarcane yield based on ndvi and concentration of leaf-tissue nutrients in fields managed with straw removal. *Agronomy*, 8(9):196.

[Price, 1982] Price, J. C. (1982). Estimation of regional scale evapotranspiration through analysis of satellite thermal-infrared data.

[Priestley and Taylor, 1972] Priestley, C. H. B. and Taylor, R. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, 100(2):81–92.

[Putman and Suarez, 2011] Putman, W. M. and Suarez, M. (2011). Cloud-system resolving simulations with the nasa goddard earth observing system global atmospheric model (geos-5). *Geophysical Research Letters*, 38(16).

[Qian et al., 2019] Qian, T., Tsunekawa, A., Peng, F., Masunaga, T., Wang, T., and Li, R. (2019). Derivation of salt content in salinized soil from hyperspectral reflectance data: A case study at minqin oasis, northwest china. *Journal of Arid Land*, 11(1):111–122.

[Rabus et al., 2003] Rabus, B., Eineder, M., Roth, A., and Bamler, R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS journal of photogrammetry and remote sensing*, 57(4):241–262.

[Rao et al., 2002] Rao, P. K., Rao, V. V., and Venkataratnam, L. (2002). Remote sensing: A technology for assessment of sugarcane crop acreage and yield. *Sugar Tech*, 4(3-4):97–101.

[Rhoades et al., 1992] Rhoades, J., Kandiah, A., and Mashali, A. (1992). The use of saline waters for crop production. *Irrigation and Drainage Paper*, 48.

[Robertson and Muchow, 1994] Robertson, M. and Muchow, R. (1994). Future research challenges for efficient crop water use in sugarcane production.

[Sanches et al., 2020] Sanches, G. M., Magalhães, P. S. G., dos Santos Luciano, A. C., Camargo, L. A., and Franco, H. C. (2020). Comprehensive assessment of spatial soil variability related to topographic parameters in sugarcane fields. *Geoderma*, 362:114012.

[Scipal et al., 2010] Scipal, K., Dorigo, W., and de Jeu, R. (2010). Triple collocation—a new tool to determine the error structure of global soil moisture products.

[Shilo, 2018] Shilo, R. (2018). What's the matter with ndvi? you are! https://www.planetwatchers.com/whats-the-matter-with-ndvi-you-are/.

[Simões et al., 2005] Simões, M. d. S., Rocha, J. V., and Lamparelli, R. A. C. (2005). Spectral variables, growth analysis and yield of sugarcane. *Scientia Agricola*, 62(3):199–207.

[Simões et al., 2016] Simões, W. L., Calgaro, M., Coelho, D. S., Santos, D. B. d., and Souza, M. A. d. (2016). Growth of sugar cane varieties under salinity. *Revista Ceres*, 63(2):265–271.

[Steduto et al., 2012] Steduto, P., Hsiao, T., Fereres, E., and Dirk Raes, D. (2012). Crop yield response to water. http://www.fao.org/3/i2800e/i2800e00.htm.

[Steele-Dunne et al., 2017] Steele-Dunne, S. C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P.-W., and Papathanassiou, K. (2017). Radar remote sensing of agricultural canopies: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2249–2273.

[Stoffelen, 1998] Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of geophysical research: oceans*, 103(C4):7755–7766.

[Talebnejad and Sepaskhah, 2015] Talebnejad, R. and Sepaskhah, A. (2015). Effect of different saline groundwater depths and irrigation water salinities on yield and water use of quinoa in lysimeter. *Agricultural water management*, 148:177–188.

[Tukimat et al., 2012] Tukimat, N. N. A., Harun, S., and Shahid, S. (2012). Comparison of different methods in estimating potential evapotranspiration at muda irrigation scheme of malaysia. *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, 113(1):77–85.

[UN, 2019] UN (2019). Un data mozambique. https://data.un.org/en/iso/mz.html.

[USGS, 2020a] USGS (2020a). Mcd43a3 v006. https://lpdaac.usgs.gov/products/mcd43a3v006/.

[USGS, 2020b] USGS (2020b). Mod11a1 v006. https://lpdaac.usgs.gov/products/mod11a1v006/.

[Van Beek et al., 2015] Van Beek, J., Tits, L., Somers, B., Deckers, T., Verjans, W., Bylemans, D., Janssens, P., and Coppin, P. (2015). Temporal dependency of yield and quality estimation through spectral vegetation indices in pear orchards. *Remote Sensing*, 7(8):9886–9903.

[Van der Zaag and Carmo Vaz, 2003] Van der Zaag, P. and Carmo Vaz, Á. (2003). Sharing the incomati waters: cooperation and competition in the balance. *Water Policy*, 5(4):349–368.

[Vanella et al., 2019] Vanella, D., Ramírez-Cuesta, J. M., Intrigliolo, D. S., and Consoli, S. (2019). Combining electrical resistivity tomography and satellite images for improving evapotranspiration estimates of citrus orchards. *Remote Sensing*, 11(4):373.

[Vilanculos and Mafalacusser, 2012] Vilanculos, M. F. and Mafalacusser, J. M. (2012). Soil survey report of açucareira de xinavane, sa.

[Vogelzang and Stoffelen, 2012] Vogelzang, J. and Stoffelen, A. (2012). Triple collocation. *EUMETSAT Report.[Accessed at https://nwpsaf. eu/deliverables/scatterometer/Triple Collocation_NWPSAF_TR_KN_021_v1_0. pdf*.

[Von Hoyningrn-Huene, 1983] Von Hoyningrn-Huene, J. (1983). Die interzeption des niederschlages in landwirtschaftlichen pflanzenbeständen.

[WAAA, 2020] WAAA (2020). Estimating soil texture by hand. https://www.agric.wa.gov.au/soil-constraints/soil-texture-estimating-hand.

[Wahid and Ghazanfar, 2006] Wahid, A. and Ghazanfar, A. (2006). Possible involvement of some secondary metabolites in salt tolerance of sugarcane. *Journal of plant physiology*, 163(7):723–730.

[Wolters et al., 2018] Wolters, E., Dierckx, W., Iordache, M.-D., and Else, S. (2018). Proba-v products user manual v3.01.

[Xu et al., 2020] Xu, C., Qu, J. J., Hao, X., Cosh, M. H., Zhu, Z., and Gutenberg, L. (2020). Monitoring crop water content for corn and soybean fields through data fusion of modis and landsat measurements in iowa. *Agricultural Water Management*, 227:105844.

[Yang et al., 2018] Yang, Y., Anderson, M. C., Gao, F., Wardlow, B., Hain, C. R., Otkin, J. A., Alfieri, J., Yang, Y., Sun, L., and Dulaney, W. (2018). Field-scale mapping of evaporative stress indicators of crop yield: An application over mead, ne, usa. *Remote Sensing of Environment*, 210:387–402.

[Yao, 1974] Yao, A. Y. (1974). Agricultural potential estimated from the ratio of actual to potential evapotranspiration. *Agricultural Meteorology*, 13(3):405–417.

[Zhangzhong et al., 2016] Zhangzhong, L., Yang, P., Li, Y., and Ren, S. (2016). Effects of flow path geometrical parameters on flow characteristics and hydraulic performance of drip irrigation emitters. *Irrigation and Drainage*, 65(4):426–438.