

**Monitoring vegetation dynamics using
vegetation optical depth retrieved from
L-band single-incidence angle backscatter
observations: A field-based study over corn**

Ge Gao

TITLE

**MONITORING VEGETATION DYNAMICS USING VEGETATION
OPTICAL DEPTH RETRIEVED FROM L-BAND
SINGLE-INCIDENCE ANGLE BACKSCATTER OBSERVATIONS: A
FIELD-BASED STUDY OVER CORN**

Thesis

to obtain the Master degree
at the Delft University of Technology,
on the authority of Prof. dr. ir. S. C. Steele-Dunne, chair of the committee,
to defend in public on Thursday November 14, 2019 at 13:00 p.m.

by

Ge GAO

The committee of this thesis

Prof.dr.ir.S.C.Steele-Dunne,	Delft University of Technology,	supervisor
Dr.ir. F.J. (Freek) van Leijen,	Delft University of Technology	
Dr. M.Hrachowitz,	Delft University of Technology	
S.Khabbazan MSc.,	Delft University of Technology	



Student number: 4682742

Keywords: Vegetation optical depth, vegetation water content, L-band backscatter

Printed by: G. Gao

Front: Corn field in summer agricultural season. Skyline horizon, blue sky background. Photo credit: AGRICOLLEGES international (education to grow)

Copyright © 2019 by G. Gao

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

*Life is like a box of chocolates.
You never know what you're gonna get.*

Forrest Gump

CONTENTS

List of Figures	ix
List of Tables	xv
Abstract	xvii
1 Introduction	1
1.1 The Relation between VOD and Bulk Vegetation Water Content	2
1.2 Aim of the Research	2
1.3 Thesis Outline	4
2 Theoretical Background	5
2.1 Physical Principles of VOD Retrieval	5
2.1.1 Physical Principles of Ground-Based Scatterometers	5
2.2 Vegetation Optical Depth	8
2.2.1 The Application of VOD	9
2.2.2 The Derivation of VOD	9
2.2.3 Estimating VOD from Water Cloud Model(WCM)	11
3 Case Study Area	13
3.1 General Information	13
3.1.1 Meteorological Data	13
3.1.2 The University of Florida L-band Automated Radar System	15
3.1.3 Soil Texture	15
3.1.4 Soil Moisture	16
3.1.5 Growth Stages of the Corn (BBCH stages)	16
3.1.6 Vegetation Water Content of Corn	17
3.1.7 Leaf Area Index (LAI)	17
4 Methodology	21
4.1 Estimating Vegetation Optical Depth	21
4.1.1 Dry and Wet Reference Determining	22
4.1.2 Estimating VOD from Water Cloud Model(WCM)	27
4.2 Other Vegetation Indices	28
4.2.1 Cross-Ratio	28
4.2.2 NDVI	29
4.3 The Relation between VOD and Vegetation Water Content	30
4.4 Identifying the Drivers of Temporal VOD Variations	30
4.4.1 Random Forest Machine Learning	30

5	Results and Discussion	33
5.1	Estimating Vegetation Optical Depth	33
5.1.1	Dry and Wet References Time Series	35
5.1.2	VOD Time Series	36
5.2	Cross-Ratio and NDVI Time Series	38
5.2.1	Cross-Ratio	38
5.2.2	NDVI.	39
5.3	The Relation between VOD and VWC_{bulk}	40
5.4	Sensitivity Analysis of VOD Using Random Forest.	42
5.4.1	The Sensitivity of VOD to Water Content of Different Vegetation Components	42
5.4.2	The Sensitivity of VOD to Stem and Leaf Water Content at Different Heights	44
5.5	Sensitivity Analysis of CR using random forest	50
5.5.1	The Sensitivity of CR to Water Content of Different Vegetation Com- ponents	50
5.5.2	The Sensitivity of CR to Stem and Leaf Water Content at Different Heights	50
6	Conclusion and Recommendation	53
6.1	Conclusion	53
6.2	Recommendation.	56
	References	57
A	Noise in backscatter observations at HH polarization	63
B	Growth stages for corn	65
C	Smoothed slope and intercept time series	69
D	Mean Decrease in Gini	71
E	Linear relation between VOD and VWC in certain growth stages	77
	Acknowledgements	79

LIST OF FIGURES

2.1	Relation between backscatter and incidence angle with relation to soil moisture and vegetation after [1]	11
2.2	Illustration of three components of backscatter, the dry reference, wet reference, and static components from the TU-Wien algorithm.	12
3.1	The location of the case study corn field. Source [2]	14
3.2	The irrigation, precipitation and dew information during the growing season. The blue bars represent the precipitation (mm/d) happened in the date showed in x-axis. The pink and green bars do not represent the amount of dew or irrigation. When pink bar appears it means dew appeared on that day. When green bar appears it just indicates irrigation was conducted on that day.	14
3.3	UF-LARS in corn field mounted on the 25 m Genie platform	15
3.4	Volumetric soil moisture data and LAI data (top) as well as backscatter data at co- and cross-polarization (bottom)	16
3.5	Pictures of corn during different growth stages	18
3.6	Contributions of different vegetation components to VWC_{bulk}	19
4.1	Flow chart of VOD retrieval	23
4.2	In-situ photo of corn field at 2018-05-18	24
4.3	Illustration of sliding window method	25
4.4	Example of how the linear model is applied to a certain sliding window. The window length of this plot is 5-day and the central time point is 2018-05-13 11:29:00 am. The blue dots in the plot are the samples inside this 5-day window with volumetric soil moisture (VSM) as the independent variable and backscatter as the dependent variable. The red line is the linear relation between VSM and backscatter derived by the OLS method. This linear relation with slope=90.87 and intercept=-21.13 is assumed as the linear relation at the central time point of this window	26
5.1	The time series of slope derived by OLS at every radar time step with goodness of fit higher then set threshold. Y-axis is the value of slope(k), x-axis is the time, and the color of each dot indicates the goodness of fit)	34
5.2	The time series of intercept derived by OLS at every radar time step with goodness of fit higher then set threshold. Y-axis is the value of intercept(C), x-axis is the time, and the color of each dot indicates the goodness of fit)	35
5.3	The dry reference, wet reference, and static component at HH (top), VV(middle) and cross-polarization(bottom)	36

5.4	VOD at co- and cross-polarization(top) and transmissivity at co- and cross-polarization(bottom)	37
5.5	Volumetric soil moisture and LAI(top) and backscatter at VV and cross-polarization as well as CR values(bottom)	38
5.6	Blue dots are NDVI values during the growing season obtained from Google Earth Engine	39
5.7	Scatter plots show the relation between VOD and VWC (This figure contains three rows, the y-axis of each row is HH-, VV- and cross-polarization (from top to bottom), respectively. As for the four columns of this figure, the x-axis of each column is bulk, stem, leaf and ear water content (from left to right), respectively. Then in every subplot, the color of each point represents the BBCH growth stage. Because ears emerge in the later part of the growing season. Therefore, the y-axis of the fourth column (with grey background) in this figure does not start from 0)	40
5.8	Feature importance of RF model for VOD and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)	43
5.9	Stem(top) and leaf(bottom) water content distribution with height. The top plot shows the stem water content in kg/m^2 . Each dot represents the total water content in all stems of a 10-cm stem section. The y-axis of the each dot stands for the height at the top of that 10-cm stem section. The bottom plot shows the leaf water content in kg/m^2 . Each dot represents the total water content of leaves with their collar height showed in the y-axis.	45
5.10	Feature importance of RF model for VOD and water content of stem section at different height (Each subplot is the result of one RF model, every bar in the subplots stands feature importance of one input variable(the water content of a stem section between a certain height range) of this RF model. For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)	47
5.11	Feature importance of RF model for VOD and water content of leaf at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)	49

5.12 Feature importance of RF model for CR and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 50

5.13 Feature importance of RF model for CR and water content of stem vegetation content at different height(Each subplot is the result of one RF model, every bar in the subplots stands for one input variable of this RF model (stem water content at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 51

5.14 Feature importance of RF model for CR and water content of leaf vegetation content at different height(Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 52

A.1 The backscatter observations at co- and cross- polarization. The backscatter data inside the red circle in HH polarization is where the noise appears) 64

B.1 Table with the description of different BBCH stages for corn (Superscript 1: A leaf may be described as unfolded when its ligule is visible or the tip of next leaf is visible; Superscript 2: Tillering or stem elongation may occur earlier than stage 19, in this case continue with principal growth stage 3. Superscript 3: In maize, tassel emergence may occur earlier, in this case continue with principal growth stage 5) 66

B.2 Corn growth stage conversions 67

C.1 The slope time series derived by the OLS with R^2 higher than 0.5 (orange dots). And a smoother slope time series using a moving average of 5-day (blue dots) at different polarization mode(HH polarization(top left), VV polarization(top right) and cross-polarization(bottom left)) 69

C.2 The intercept time series derived by the OLS with R^2 higher than 0.5 (blue dots). And a smoother slope time series using a moving average of 5-day (red dots) at different polarization mode(HH polarization(top left), VV polarization(top right) and cross-polarization(bottom left)) 70

- D.1 Feature importance of RF model for VOD and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 72
- D.2 Feature importance of RF model for VOD and water content of stem at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of a stem section between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 73
- D.3 Feature importance of RF model for VOD and water content of leaf at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 74
- D.4 Feature importance of RF model for CR and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 75
- D.5 Feature importance of RF model for CR and water content of stem vegetation content at different height(Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem water content at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 75
- D.6 Feature importance of RF model for CR and water content of leaf vegetation content at different height(Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively) 76

E.1	Linear relation between VOD_{HH} and VWC_{bulk} (top left), VOD_{HH} and VWC_{leaf} (top right), VOD_{HH} and VWC_{stem} (bottom left), VOD_{HH} and VWC_{ear} (bottom right)	77
E.2	Linear relation between VOD_{VV} and VWC_{bulk} (top left), VOD_{VV} and VWC_{leaf} (top right), VOD_{VV} and VWC_{stem} (bottom left), VOD_{VV} and VWC_{ear} (bottom right)	78
E.3	Linear relation between VOD_{cross} and VWC_{bulk} (top left), VOD_{cross} and VWC_{leaf} (top right), VOD_{cross} and VWC_{stem} (bottom left), VOD_{cross} and VWC_{ear} (bottom right)	78

LIST OF TABLES

5.1	The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{bulk} from BBCH=13 to BBCH=51	41
5.2	The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{stem} from BBCH=13 to BBCH=51	41
5.3	The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{leaf} from BBCH=13 to BBCH=51	41
5.4	The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{ear} from BBCH=63 to BBCH=71	41
5.5	The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the Input variables: VWC_{stem} , VOD_{stem} , and VOD_{ear} at vegetative and reproductive stage	42
5.6	The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the input variables: the water content of stem sections between a certain height range at vegetative and reproductive stage	46
5.7	The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the input variables: the water content of leaf lumps between a certain height range at vegetative and reproductive stage	48

ABSTRACT

Vegetation water content (VWC) is an important parameter for sustainable land and water management. In agriculture, VWC can be used to monitor drought and assess crop productivity. Being able to monitor VWC is important to reduce agricultural vulnerability and scientifically manage agricultural water use. Due to the dense in-situ networks are expensive and have difficulties in capturing the large spatial variability of VWC. Therefore, remote sensing has great potential in VWC monitoring. With the application of remote sensing, worldwide data with coarse or fine spatial and temporal resolution can be extracted with relatively low cost. Vegetation Optical Depth (VOD), extracted from the radar remote sensing observations, is a dimensionless parameter that highly related to VWC. Thus, VOD can be used as an indicator for VWC. The present research aims to thoroughly analyze the relation between VOD and VWC. High temporal backscatter data and detailed field experiment data of soil moisture and vegetation water content during a full growing season of corn (between 18 April 2018 to 13 June 2018) were used. Correlations between VOD and VWC were analyzed. The result shows that VOD is highly related to VWC_{bulk} . However, the linear relation between VWC_{bulk} and VOD is only valid before the heading stage at both co- and cross-polarization. Then, random forest machine learning was conducted to determine the sensitivity of VOD to the water content of different parts of the plant. This sensitivity analysis contains two parts: a) the sensitivity of VOD to the water content of different vegetation components and b) The sensitivity of VOD to stem and leaf water content at different heights. The results of a) show that VOD is more sensitive to stem and leaf water content in the vegetative stage whereas more sensitive to ear water content during the reproductive stage. Besides, stem, leaf and ear water content can better capture the VOD variation during the vegetative stage. The result from b) suggests that VOD can provide information about the vertical distribution of moisture inside the canopy. Finally, a cross-comparison was conducted between VOD and other commonly used vegetation indicators, which includes NDVI and Cross-Ratio. VOD is available regardless of cloud conditions and is, therefore, more reliable than NDVI. Compared with cross-ratio, VOD is better related to vegetation moisture dynamics.

1

INTRODUCTION

Vegetation Water Content (VWC) is an important parameter in sustainable land and water management. It can be used in assessing agricultural productivity as well as to monitor drought and to avoid bushfire [3]. Besides, vegetation water content is critical for soil moisture retrieval based on microwave remote sensing. The ability to estimate VWC will lead to improved soil moisture estimations, which is important for quantifying the global energy, water, and carbon cycles [4]. There are many components of bulk vegetation water content (VWC_{bulk}), like leaf water content and stem water content. The bulk vegetation water content is the total mass of water in the above-ground components of the plant per unit area [5]. Take this definition as a format, we can easily define the other components of vegetation water content.

To describe vegetation water content with remote sensing, the Normalized Vegetation Difference Index (NDVI) is often used in optical remote sensing. NDVI can also be related to canopy attributes like the Leaf Area Index (LAI). However, NDVI is measured in the optical (visible) part of the electromagnetic spectrum, so there are several factors that may limit NDVI for the vegetation water content study. Signal received by sensors are affected by atmosphere, cloud, geometric misregistration, anisotropic reflectance effects as well as electronic errors [6]. Thus, products with spectral reflectance and index usually contain noise from the background (e.g., soil and exposed water bodies) [7]. Besides, NDVI only senses the top of the canopy [8][9][10] and measures mostly the canopy greenness [11].

In microwave remote sensing, backscatter observations from a vegetated surface is influenced by vegetation water content [12]. Compared with optical remote sensing, microwaves with frequencies between 0.3 GHz and 100 GHz can get rid of the influence of clouds and measure overnight. Thus, these observations have the potential to be used for all-weather monitoring of regional vegetation canopy. In the microwave domain, Vegetation Optical Depth (VOD) is often used to describe the transparency of vegetation cover, which is highly related to the total water content [13]. In passive microwave remote sensing, the Land Parameter Retrieval Model (LPRM) is often used to retrieve the VOD [14]. However, the resolutions of the passive remote sensing are

low (up to 50*30km), which is better suited for large scale vegetation monitoring. For field-scale applications, active microwave remote sensing with finer resolution is preferred [15][16][17]. In active remote sensing, the TU-Wien VOD estimation method can be used for VOD retrieval [1]. This method consists of two main models, the TU-Wien soil retrieval algorithm and the Water Cloud Model (WCM). The vegetation correction is estimated by the TU-Wien soil moisture retrieval algorithm. This algorithm retrieves soil moisture from multi-incidence angle backscatter data. The slope and curvature obtained from the relationship of backscatter and incidence angle can be used to estimate the vegetation correction. Then, convert the vegetation correction into VOD values with the Water Cloud Model (WCM).

1.1. THE RELATION BETWEEN VOD AND BULK VEGETATION WATER CONTENT

VWC_{bulk} is assumed proportional to microwave VOD of the above-ground biomass canopy [18].

$$VOD = b * VWC_{bulk} \quad (1.1)$$

The parameter b in Equation 1.1 depends on the vegetation structure and microwave frequency as well as polarization [19]. A b -factor is often assumed as constant and determined by crop type. Thus, in different growth stages of the plant, the static b -factor may induce errors. So, can this linear model describe the relationship between VWC_{bulk} and VOD in the entire growing season? However, our understanding of the relation between VOD and VWC_{bulk} is not sufficient. Therefore, this is a problem that needs to be solved.

1.2. AIM OF THE RESEARCH

The aim of this research is to improve the understanding of the relationship between VOD and bulk vegetation water content. In order to better understand this relation, estimated VOD from tower-based high temporal backscatter observations and detailed in-situ field measurements are used in this study. The radar backscatter data used in this research is from the University of Florida L-band Automated Radar System (UF-LARS). UF-LARS returns normalized radar backscatter (σ^0) data at four polarization combinations (HH, VV, HV, and VH) and an azimuth range of 0° to 180° with an azimuthal resolution of 9° every 15 minutes. Besides the dense radar data, from 18 April 2018 to 13 June 2018 (an entire growing season of corn), a detailed field experiment in vegetation water content, soil moisture measurement of the corn field was conducted. This high temporal resolution of radar backscatter observations as well as the detailed soil moisture and vegetation water content data give us an opportunity to study the relation between VOD and VWC_{bulk} . Further, sensitivity analysis of VOD to a) water content of different vegetation components (stem, leaf, ears) as well as b) water content of leaf and stem located at different height of the canopy are conducted. The research questions are therefore defined as follow:

1. *Can the TU-Wien VOD estimation method be adapted to estimate VOD from backscatter at single-incidence angle?*

In the TU-Wien VOD estimation method, the vegetation correction is determined by multi-incidence angle backscatter observations. However, the radar data from UF-LARS only contains observations from incident angle at 40°. But, in this research, we have the soil moisture measurement that the TU-Wien VOD estimation does not have. Because of the difference between initial data sets, the TU-Wien VOD estimation method needs to be adapted in this research. In order to let the adaptation reasonable, we explain and compare the theoretical principles and assumptions of the TU-Wien VOD estimation method before and after adaptation. Then, we discuss whether the adapted version is credible and feasible.

2. *What is the relation between VOD and VWC_{bulk} in the entire growing season? Does the linear relation hold true in the entire growing season?*

By making a scatter plot in different growth stages, a thorough study of the relation between VOD and VWC_{bulk} is conducted. Statistical analysis is performed when the linear relation holds true, the Ordinary Least Squares (OLS) method is used to derive the slope and intercept.

3. *Bulk vegetation water content comprises many components. Here we only focus on stem, leaf, and ears. Thus, which vegetation components are the main drivers for VOD variation? Will this result change in different growth stages? Will this result also change in different polarization?*

This sensitivity analysis is done by using random forest machine learning techniques. Random Forest (RF) models are built in different growth stages. In each period, the input variables of each RF model are the same, which are stem, leaf, and ear water content while the response variable of each RF model is different, which is VOD at HH, VV, and cross-pol, respectively.

4. *The moisture inside the canopy is not uniformly distributed. The backscatter will be influenced by the vertical distribution of moisture. Is the VOD more sensitive to the particular part of the canopy layer that contains more moisture?*

Detailed destructive measurements of the water content profile will be used as input variables to a RF model to identify which attributes at which height have the most influence on VOD variation.

5. *Apart from VOD, Cross-Ratio is also a commonly used indicator for vegetation water content in the microwave frequency region. Besides, NDVI in the spectrum region is also an often used vegetation water content indicator. What is the difference between VOD, CR, and NDVI? Can they all sufficiently describe the vegetation water content within the canopy?*

Here, we compare the sensitivity of VOD, CR and NDVI to vegetation water content.

1.3. THESIS OUTLINE

In this thesis, a theoretical background is given in Chapter 2. This chapter provides relevant background knowledge and theories for vegetation optical depth retrieval. Then, models in TU-Wien VOD estimation method are explained. In Chapter 3, a general description of the project site is given. Besides, the in-situ measurements which are needed for the VOD retrieval are mentioned. In Chapter 4, the methodology used during this research is discussed. After this, the results of this research are presented in Chapter 5. Finally, the research questions will be answered and a conclusion on this thesis is given in Chapter 6.

2

THEORETICAL BACKGROUND

The main focus of the first part of Chapter 2 provides relevant background knowledge and theories for vegetation optical depth retrieval. Then, in the second part of this chapter, models used to estimate VOD in this study are explained, which helps to understand the result and discussion in the following chapter.

2.1. PHYSICAL PRINCIPLES OF VOD RETRIEVAL

2.1.1. PHYSICAL PRINCIPLES OF GROUND-BASED SCATTEROMETERS

The interaction of electromagnetic radiation with matter is fundamental to remote sensing. The measure of received signal strength is the most important in remote sensing devices. Here, in this section, we just attempt to provide an overview of some most relevant knowledge that will be sufficient for understanding this research.

THE RADAR EQUATION

The fundamental relation between the characteristic of the radar, the target, and the received signal is called the *radar equation*:

$$P_r = \frac{P_t G_t A_r \sigma}{(4\pi)^2 R^4} \quad (2.1)$$

The power P_t is transmitted by an antenna with gain G_t , the spreading loss $\frac{1}{(4\pi R_t^2)}$ is the reduction of power density associated with spreading of the power over a sphere of radius R surrounding the antenna. Radar has two spreading loss factors, one loss happens when the power transmitted toward the scatterer $\frac{1}{(4\pi R_t^2)}$ and the other loss happens when the power reradiated by the scatterer $\frac{1}{(4\pi R_r^2)}$. If we assume the $R_t = R_r$, then the total distance is $2R$ and the total loss factor is $\frac{1}{((4\pi)^2 R^4)}$. A_r is the effective aperture of the receiving antenna. σ is the scattering cross-section which is associated with the individual scatterer. Scatter cross-section is a function of the directions of the incident wave and the wave towards the receiver, as well as the scatterer shape and dielectric properties.

Further, the differential scattering coefficient (or backscatter cross-section) is defined as the average value of individual scatterer cross-section per unit area. The value of backscatter cross-section in the linear domain is given by:

$$\sigma^0 = \left\langle \frac{\sigma_i}{\Delta A_i} \right\rangle \quad (2.2)$$

where the angle brackets represent the operation of calculating the integral mean of the variables within it.

Because the backscatter cross-section can vary by several orders of magnitude, the backscatter cross-section is converted to the power domain and explained in dB, the converting function is given by:

$$\sigma^0(dB) = 10 * \log\left(\sigma^0 \left(\frac{m^2}{m^2}\right)\right) \quad (2.3)$$

DIELECTRIC PROPERTIES

The **dielectric constant** (denoted as ϵ) is one of the properties of the medium used to describe the propagation of electromagnetic radiation in the material. For the case of a uniform homogeneous medium that does not absorb energy from an electromagnetic wave propagating through it, a dielectric constant is a real number. However, for the materials neither perfect conductive nor perfect transparent to the electromagnetic radiation, energy loss happens when the wave propagating through the medium, the dielectric constant is represented as a complex number [20] [21]:

$$\epsilon = \epsilon' + \epsilon'' \quad (2.4)$$

where ϵ' and ϵ'' are referred to as the real and imaginary part of the dielectric constant. The real part of the dielectric constant is the permittivity of the material, which determine the velocity and wavelength of the refracted wave. The imaginary part is the dielectric loss factor, which describes the absorbed energy by the medium when the electromagnetic wave propagating through.

The **dielectric constant of soil** depends on moisture content, bulk density, soil type, temperature, salinity, and microwave frequency [22]. Dry soil has a dielectric constant of $\epsilon \approx 4$, whereas liquid water has a dielectric constant of $\epsilon \approx 80$ for frequencies lower than 10 GHz[23]. With the large contrast between them, the increase of the moisture content will lead to increase in soil dielectric constant. The relationship between volumetric soil moisture content and soil dielectric constant is almost linear, except at low moisture content. The non-linearity at low moisture content is caused by the strong bond between the surfaces of soil particles and the thin film of water (bond water). These bonds will limit the free rotation of the water molecules [24]. When soil moisture content increases, both the real and imaginary part of the dielectric constant increases. However, the increasing rates of these two components are frequency dependent. The increasing rate of the real component is higher at a lower frequency, while the increasing rate of the imaginary component is higher at a higher frequency [25]. The **dielectric constant of vegetation** is influenced by water content, microwave frequency, salinity, and temperature [26] [27]. ϵ' of water is more than one order larger than ϵ' of bulk vegetation

material, and ϵ'' of water is several orders of magnitude larger than ϵ'' of bulk vegetation. Therefore, the water content of vegetation will influence the dielectric constant of vegetation [27]. A dual-dispersion model describe the dielectric constant of vegetation ϵ_v as a additive of three components: (a) a non-dispersive residual component ϵ_r , (b) a free water component $v_{fw}\epsilon_f$, where v_{fw} is the volume fraction of free water, ϵ_f is the dielectric constant of free water and (c) a bulk vegetation-bond water component $v_b\epsilon_b$, where v_b is the volumetric fraction of the bulk vegetation-bond water mixture and ϵ_b is its dielectric constant [28].

SURFACE AND VOLUME SCATTERING FROM A VEGETATED SURFACE

When electromagnetic wave impinges from above on the boundary between two different mediums, part of the incident wave is scattered back and the others are transmitted into the lower medium. On one hand, if the lower medium is homogeneous or can be considered as such, the scattering will only happen at the surface boundary. This is called surface scattering. On the other hand, if the lower medium is inhomogeneous or is a mixture of different dielectric properties, then part of the transmitted wave will be scattered back by the inhomogeneities. The scattered wave may cross the boundary and enter into the upper-medium. In volume scattering, the scatter takes place within the volume of the lower medium. The mechanism of volume scattering will cause a redistribution of the energy in the transmitted wave into other directions and result in a loss from the transmitted wave [29].

There are several patterns of **surface scattering** when the boundary between two mediums is a smooth-surface boundary. The reflection is called specular reflection and can be described with Fresnel reflection laws. When the boundary surface is a little bit rougher, part of the incident wave will be reflected in the specular direction and the rest will be scattered in all directions. When the surface continues getting rougher, the scattered part is getting more and more important. Finally when the surface approaches the Lambertian surface, which scatters all the radiation incident on it. The Rayleigh criterion shown below can be used to identify a rough surface:

$$h > \frac{\lambda}{8 * \cos\theta} \quad (2.5)$$

where h is the mean height of surface variations, λ is wavelength, and θ is the incident angle. Roughness depends on the wavelength and incident angle. A surface may be considered as rough by an electromagnetic wave with shorter wavelength like an optical wave, while it may seem very smooth by an electromagnetic wave with longer wavelength like a microwave.

Volume scattering can be observed if microwave radiation penetrates into a medium. The penetration depth is defined as the distance when the incident power attenuates to $\frac{1}{e}$ (exponential coefficient). Effective penetration depth can be determined by dielectric properties of the medium. The wavelength (or, we can say the frequency) determines the value of penetration depth (δ_p). The formula is given by:

$$\delta_p \cong \frac{\lambda\sqrt{\epsilon'}}{2\pi\epsilon''} \text{ (for materials with } \frac{\epsilon''}{\epsilon'} < 0.1) \quad (2.6)$$

This approximate formula is applicable to most natural materials except water. For electromagnetic wave with a longer wavelength, the penetration depth tends to be longer. Besides wavelength, moisture content (represented by dielectric constant) also influences the penetration depth. The penetration depth will decrease with increasing moisture content. Direct estimation of penetration depth is hard because the size of vegetation components such as stems and leaves are mostly of the order of magnitude of the wavelength or even larger. Besides, the effective dielectric constant of the vegetation canopies inside the equation is difficult to estimate. However, indirect estimation of the penetration depth of the vegetation canopy can be obtained by comparing the backscattering coefficient at incidence angle close to nadir for bare soil and vegetated soil surfaces. These observations show that the penetration depth of mature crops during the green stage is several meters at frequencies around 1 GHz and decreases to one meter or even less at frequencies above 10 GHz [29]. The significant factors that influence the penetration depth of the vegetation canopy at a given frequency are the geometry of the canopy, the moisture content of the canopy and the volume of vegetation material per unit volume of the canopy [29]. It is obvious that the penetration depth is a dynamic value during the growing season of the plants. Take corn as an example. When the corn is during its fruit-filling stage with lower water content, the penetration depth can be higher compared with the period when the corn is green, fresh and lush.

The discussion above introduces two main scattering mechanism, the total backscatter from a vegetated soil surface is generally a combination of volume scattering in the canopy layer and the surface scattering by the underlying soil surface when the air-vegetation boundary is not important [30]. In addition to penetration depth just discussed, the contribution of these two components to total backscattering depends on some other key factors, which include the height of the canopy, frequency, and incidence angle of the radar signal. For surface scattering, backscattering decreases rapidly with increasing incidence angles close to nadirs (under this condition, the decreasing on a very rough surface is less rapid compared with the smooth surface). Whereas, volume scattering decreases very slowly with increasing incidence angle. Therefore, at incidence angle close to nadirs, the relative importance of surface scattering and volume scattering are expected to be comparable. When the incidence angle increasing, unless the frequency is in the lower part of the microwave spectrum and/or the water content of the vegetation canopy is small, volume scattering is dominant [29]. When the incidence angle is between 35° to 40°, the path length of the microwave propagate through the vegetation canopy becomes longer. Thus the backscatter from vegetation contributes more to the total backscatter [31]. Whereas having the incidence angle below 30° reduces the path length and maximize the backscatter from the soil surface, which is preferable for soil moisture retrieval [32].

2.2. VEGETATION OPTICAL DEPTH

The main focus of this research is on Vegetation Optical Depth (VOD). This section describes the TU-Wien VOD estimation method applied in this study. The VOD is a measure of the opacity of a vegetation canopy for radiation passing through. Besides, it can also be seen as an ecological indicator that is closely related to the above ground biomass water content [24] and provides information complementary to visible-near

infrared vegetation indices [33], which are often limited by the effects of atmosphere, background soil conditions, and saturation at high levels of vegetation [34]. The microwave measurements are sensitive not only to the leafy part of vegetation properties but also to the properties of the overall vegetation canopy when the microwave sensor can penetrate it [34].

2.2.1. THE APPLICATION OF VOD

Microwave remote sensing is not only sensitive to vegetation water content, but also to soil moisture content. However, the way to separate these two components is not straightforward. VOD can help to parameterize the influence of the canopy layer. When the effect of VOD is taken into account properly, the prediction accuracy of a soil moisture retrieval model will increase. For example, the transmissivity of the vegetation which is defined in terms of the vegetation optical depth is used to describe the vegetation effect in [24]. Besides, VOD is directly related to bulk vegetation water content, and this relation is approximate as a linear [18], in which the slope is considered constant and depends on microwave frequency, canopy type, and structure [19]. VOD is also a useful indicator of vegetation state and variability since the biomass of vegetation provides available storage of water, VOD can be used as a proxy of biomass [35]. Also, VOD is helpful for crop monitoring. It can help determine vegetation phenology [36], and the crop water stress [12].

2.2.2. THE DERIVATION OF VOD

The TU-Wien VOD estimation method is used in this study. We convert the vegetation correction of TU-Wien soil moisture retrieval algorithm into VOD estimations using a Water Cloud Model.

WATER CLOUD MODEL

Previous researches [25] [30] [37] suggest that the total backscatter from the vegetated soil is a combination of the soil backscatter attenuated by the vegetation cover, direct backscatter from the canopy layer and the interaction between the canopy and the underlying soil surface. The backscatter from the vegetation canopy is affected by vegetation 3-D structure and water content (related to biomass) [38]. It is impossible for a scattering model to describe each vegetation component's geometry, location and dielectric constant inside the canopy volume in details. Also, it is difficult for a model to describe the scattering properties of each element inside the canopy. Therefore, approximations are needed to simplify the problem as well as to establish a model that can be applied to a class of canopies with similar geometries [29].

A rather simple model called Water Cloud Model (WCM) that models the vegetation backscatter coefficient of a canopy layer was established by Attema and Ulaby in 1978. In this model, the vegetation volume is assumed to be filled with uniformly distributed identical scatterers. In WCM, the scattering properties of each vegetation element are treated as a whole and are related to the volumetric water content of the canopy. However, the multiple scattering which describes the interaction between soil and vegetation cover is generally neglected in WCM [30]. Therefore, the WCM is a zeroth-order radiative transfer solution which the total backscatter (σ^0) from a canopy consists of backscatter

from vegetation cover (σ_{veg}^0) and backscatter from the underlying soil ($\gamma^2\sigma_{soil}^0$) [17].

$$\sigma^0 = \sigma_{veg}^0 + \gamma^2\sigma_{soil}^0 \quad (2.7)$$

$$\sigma_{veg}^0 = B * \cos\theta(1 - \gamma^2) \quad (2.8)$$

$$\gamma^2 = \exp\left(\frac{-2\tau}{\cos\theta}\right) \quad (2.9)$$

Where γ^2 in Equation 2.7 is an attenuation factor that represents the two-way vegetation transmissivity. This γ^2 can be calculated by Equation 2.9 with vegetation optical depth (τ) as one of the variables. B in Equation 2.8 is a parameter related to the single scattering albedo ω , which for small isotropic scatterers take the value of $\frac{3\omega}{4}$. The σ_{soil}^0 in Equation 2.7 is the backscatter from bare soil. One way to estimate σ_{soil}^0 is by the linear model between soil moisture and the backscatter coefficient. The University of Kansas used truck-mounted scatterometer systems to observe bare soil field with different surface roughnesses and soil textures. When the roughness and soil texture were set, a linear relation was found between radar backscatter coefficient and volumetric soil moisture (M_v) in the upper 2 to 5 cm of soil. The linear correlation coefficients ρ was typically in the order of 0.9.

$$\sigma_{soil}^0 = K * M_v + C \quad (2.10)$$

Where K and C are the slope and intercept, M_v is volumetric soil moisture (because the dielectric constant of wet soils is proportional to the number of water dipoles per unit volume, the preferred measure for soil moisture is volumetric) [29].

TU-WIEN SOIL MOISTURE RETRIEVAL ALGORITHM

TU-Wien soil moisture retrieval Algorithm is an empirical model based on the change detection method (refers to the method of analyzing sufficient long-term data of the interested area in order to identify differences in the state of land features). This algorithm retrieves soil moisture from multi-incidence angle backscatter data. One basic assumption for the change detection model is that the backscatter coefficient expressed in dB is linearly related to surface soil moisture content [39]. Besides, the slope and curvature of $\sigma^0(\theta)$ are assumed only influenced by vegetation density [1]. The radiation energy received by the scatterometer sensor increases with incident angle approaching nadir. However, for the vegetation-cover soil, the change of vegetation density, vegetation type, as well as the change of vegetation geometry will influence the shape of incidence angle dependency of backscatter (see in Figure 2.1). At lower incidence angles, the backscatter is lower for vegetated-soil compared with bare soil condition. This is because the vegetation cover above tends to attenuate the underlying soil backscatter. At higher incidence angles, the volume scattering resulting in a larger total backscatter than for bare soil. Therefore, at some incident angles, the backscatter from the dormant vegetation layer and fully grown canopy layer will cross over (see in Figure 2.1). These incidence angles are the so-called cross-over angles. In the TU-Wien soil moisture retrieval algorithm, it assumes the effect of vegetation is minimum at these cross-over angles. Then, the dry and wet references (denoted as σ_{dry}^0 and σ_{wet}^0) which represent the historically driest and wettest conditions with minimum vegetation effects are derived using the slope,

curvature and cross-over angles obtained from $\sigma^0(\theta)$ relationship. Since soil moisture and vegetation both lead to an increase in backscatter, σ_{wet}^0 is more or less independent to vegetation status. Besides, it represents the saturated soil condition. Therefore, the value of σ_{wet}^0 can be set as a constant value through the entire growing season. As for σ_{dry}^0 , it increases and decreases simultaneously with vegetation growth and senescence. Hence, the difference between dry and wet references shows the sensitivity to soil moisture dynamics. And the change of this sensitivity can be used as vegetation correction.

In simple radiative transfer models, for example, the Water Cloud Model introduced in the section 2.2.2, the influence of vegetation cover are governed by the two-way vegetation transmissivity which weights the relative contribution of surface scattering and volume scattering. Although the TU-Wien model uses a different way of parameterization, the aim and function of the vegetation correction are the same.

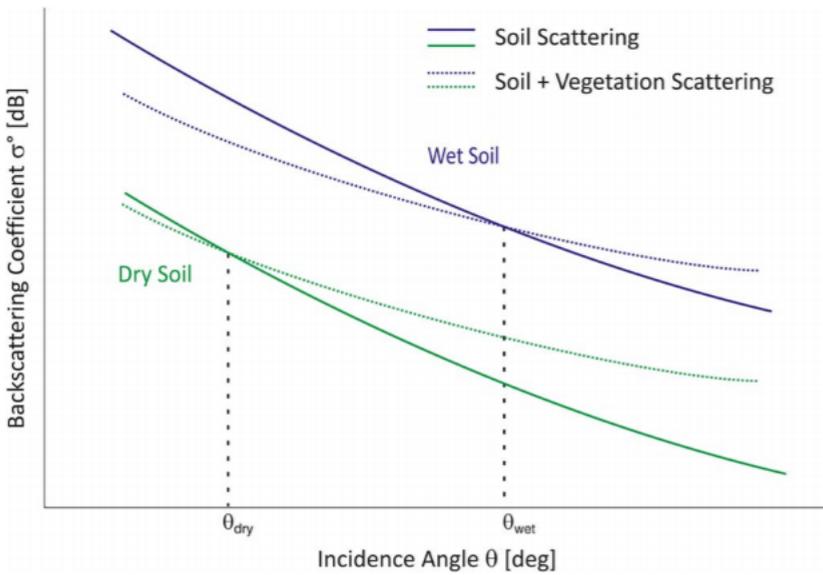


Figure 2.1: Relation between backscatter and incidence angle with relation to soil moisture and vegetation after [1]

2.2.3. ESTIMATING VOD FROM WATER CLOUD MODEL(WCM)

The difference between dry and wet reference (calculated with Equation 2.11) is the sensitivity to changes in surface soil moisture content.

$$\Delta\sigma^0(t) = \sigma_{wet}^0(t) - \sigma_{dry}^0(t) \quad (2.11)$$

Where $\sigma_{wet}^0(t)$ and $\sigma_{dry}^0(t)$ represent the wet and dry reference values at a certain time point t , respectively.

Since the backscatter in decibels has been found to be linearly related to soil moisture content and the backscatter noise is independent of the absolute backscatter, all the

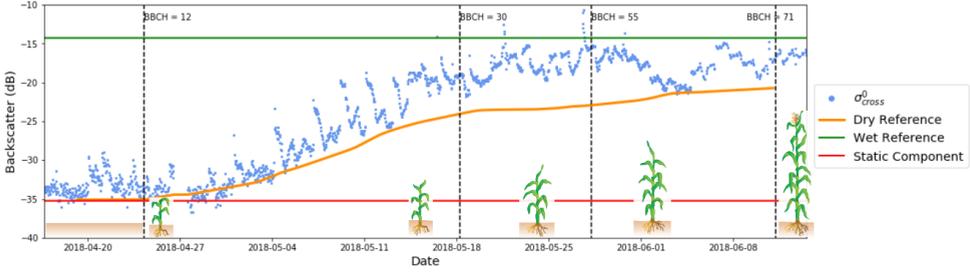


Figure 2.2: Illustration of three components of backscatter, the dry reference, wet reference, and static components from the TU-Wien algorithm.

calculation based on the idea of the TU-Wien soil moisture algorithm is calculated in the decibels domain.

Vegetation optical depth from the active microwave observations can be computed by bringing the dry and wet references into the Water Cloud Model [1]. Here, we call this VOD retrieval method as the TU-Wien VOD estimation method. Since dry and wet references intrinsically include the changes in vegetation density, the difference between dry and wet references ($\Delta\sigma^0$) shows the sensitivity to soil moisture dynamics. And the change of $\Delta\sigma^0$ can be used as vegetation correction. While in WCM, the vegetation correction is represented by γ^2 which is directly related to VOD (τ). Thus, if we replace γ^2 with $\Delta\sigma^0$, we can relate VOD to the dry and wet references. Notice that, when bring σ_{wet}^0 and σ_{dry}^0 into the Water Cloud Model, both of them need to be firstly converted into the linear domain.

$$\Delta\sigma^0(t) = \Delta\sigma_s^0 * \exp\left(\frac{-2\tau}{\cos\theta}\right) \quad (2.12)$$

$$\Delta\sigma_s^0 = \text{Max}(\sigma_{s,wet}^0(t)) - \text{Min}(\sigma_{s,dry}^0(t)) \quad (2.13)$$

Which can be solved for τ as

$$\tau(t) = \frac{\cos\theta}{2} * \ln\left(\frac{\Delta\sigma_s^0}{\Delta\sigma^0(t)}\right) \quad (2.14)$$

Where, in Equation 2.12, $\Delta\sigma^0(t)$ (m^2/m^2) is the difference between $\sigma_{wet}^0(t)$ (m^2/m^2) and $\sigma_{dry}^0(t)$ (m^2/m^2) at certain time point t . $\Delta\sigma_s^0$ (also in the linear domain) in Equation 2.13 represents the maximum range in bare soil backscatter values which is only related to soil moisture dynamics. This parameter is assumed constant through time. $\text{Min}(\sigma_{s,dry}^0(t))$ is also known as the static component.

As shown in Figure 2.2, at the beginning of the growing season, there is no vegetation cover on the soil surface, the dry reference equals to the static component. The difference between dry and wet references ($\Delta\sigma^0$) is the largest which means the sensitivity to soil moisture is the highest. Then, the dry reference increases simultaneously with the corn canopy growth which leads to the decrease of $\Delta\sigma^0$. On one hand, the change of $\Delta\sigma^0$ reflects the vegetation growth reduces the sensitivity to soil moisture. On the other hand, this shows the vegetation state and its dynamic.

3

CASE STUDY AREA

In this chapter, the general information of the exact study area is shown. Besides, the data sets collected during the field campaign which are relevant to our research are explained. All measurements displayed in this section are necessary for the following VOD retrieval.

3.1. GENERAL INFORMATION

The study was performed at the UF/IFAS Extension Plant Science Research and Education Unit (PRSEU), Citra, Florida, USA (29.4°N, 82.17°W). Florida is a hot-humid region. The rainy season runs from around May 1 to as late as November [40]. The dimension of the study cornfield is 100 m * 65 m and the location is shown in Figure 3.1. The sweet corn (type BSS0977 ATTRIBUTE 100M) was planted on 13 April 2018 and harvested on 18 June 2018. The growing season lasted for 72 days. The plant density of the cornfield is 6.6 plants/ m^2 , the plant spacing is 7 plants/m, and the row spacing is 0.925 m.

3.1.1. METEOROLOGICAL DATA

The meteorological data were obtained from a weather station from Florida Automated Weather Network (FAWN) at the Plant Science Research & Education Unit. From this weather station, meteorological data is obtained every 15 minutes. The weather station collects data such as soil temperature (°C) at 10 cm depth, air temperature (°C) at a height of 60 cm, 2 m and 10 m, relative humidity (%), rainfall (inches). In this research, we mainly focused on precipitation data, soil moisture data and dew point. The dew point temperature at 2 m (°C) has been calculated with the air temperature (°C) at 2 m and the relative humidity (%) [2]. Figure 3.2 shows some information about the irrigation, precipitation and dew. The exact time of irrigation, precipitation and dew of a day can be found in [41].



Figure 3.1: The location of the case study corn field. Source [2]

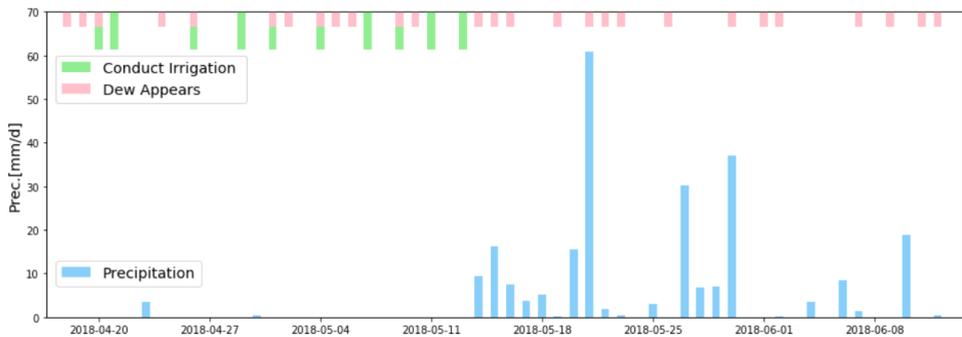


Figure 3.2: The irrigation, precipitation and dew information during the growing season. The blue bars represent the precipitation (mm/d) happened in the date showed in x-axis. The pink and green bars do not represent the amount of dew or irrigation. When pink bar appears it means dew appeared on that day. When green bar appears it just indicates irrigation was conducted on that day.

3.1.2. THE UNIVERSITY OF FLORIDA L-BAND AUTOMATED RADAR SYSTEM

The University of Florida L-band Automated Radar System (UF-LARS) returns radar backscatter (σ^0) data at four polarization combinations (HH, VV, HV, and VH) and an azimuth range of 0° to 180° with an azimuthal resolution of 9° every 15 minutes resulting in 21 collected spatial samples. The antenna incidence angle was set to 40° , while the height was at 16.2m giving a footprint of an arc which has a size of $1033.77m^2$, based upon the 6 dB beamwidth (see in Figure 3.3). Since theoretically, HV equals VH, here, the cross-polarization signal stands for the averaged value of HV and VH observations (See in Figure 3.4).



Figure 3.3: UF-LARS in corn field mounted on the 25 m Genie platform

Notice that, the in-situ backscatter data at HH polarization contains some noise at the beginning of the growing season (see magnified detail in Appendix A). The noise is caused by some currently unsolved issues during the backscatter data processing. This will influence the flowing analysis in HH polarization mode. Therefore, compared with VV and cross-polarization, the results from HH polarization is less credible.

3.1.3. SOIL TEXTURE

At the Plant Science Research & Education Unit of the University of Florida, the composition of the top layer of the soil consists of approximately 94% sand, 2% silt, 3% clay, and 2% organic matter, where the percentages are given on a weight basis [42].

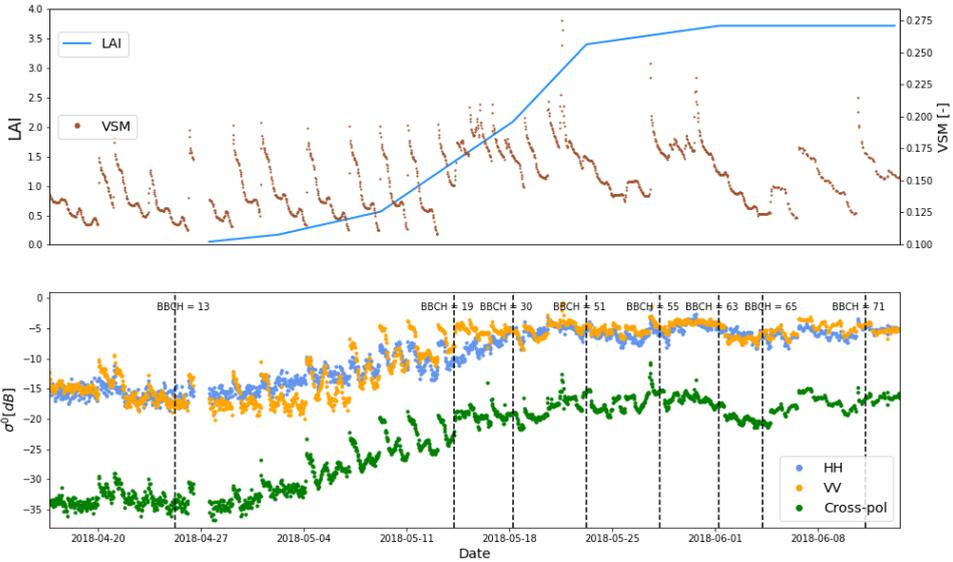


Figure 3.4: Volumetric soil moisture data and LAI data (top) as well as backscatter data at co- and cross-polarization (bottom)

3.1.4. SOIL MOISTURE

The root zone soil moisture is measured with 10 calibrated EC-5 sensors. The sensors were installed in each pit at 5, 10, 20 40 and 80 cm depth. The root zone moisture was measured throughout the season with an interval of 15 minutes. Previous investigations [43] [44] assumed that the radar response is governed by the moisture in the surface 0 to 5 cm layer of the soil. While the effective soil depth responsible for the radar return is shallower than 5 cm for very wet soil and deeper than 5 cm for very dry soil. Therefore, characterizing the moisture content by a fixed layer, 5 cm in depth, has been found to be an adequate approximation [43] [44]. Thus, in this study, averaged 5 cm depth soil moisture of two pits are used.

3.1.5. GROWTH STAGES OF THE CORN (BBCH STAGES)

Corn is a broadleaf plant and has different vegetation structures in the entire growing season. The campaign was conducted from 13 April to 18 June. The BBCH-scale is used to identify the phenological development stages of plants ("Growth stages of mono- and dicotyledonous plants") [45]. Appendix B shows the detail description of each stage [46]. The growing season of a corn consists of many stages, which can be divided into two major parts: (1) the vegetative stage (BBCH 10 – BBCH 59), and (2) the reproductive stage (BBCH 61 – BBCH 99). During the vegetative stage, the plant elongates and the leaves develop (tasseling is the last phase of the vegetative stage). During the reproductive stage, the fruit develops and ripens. In the case study area, from 13 April to 18 June, the main growth stages of corn include leaf development, stem elongation, inflorescence emergence and development of fruit, which is BBCH 13 to BBCH 73. The growth stage of

the corn was determined with the use of the BBCH staging Manual [46]. The dates of different BBCH stages are shown in Figure 3.6.

Around April 25th and 27th, BBCH=13 is reached, when three leaves are visible and the plant height is around 12.9 cm. The leaf development stage is ended around May 16th when 9 or more leaves are unfolded (BBCH=19), the height of the plant is around 71.6 cm. The stem elongation stage begins on May 18th (BBCH=30), the plant height is around 84.1 cm. The inflorescence stage begins on May 23th (BBCH=51) with plant height around 125.9 cm, and the end of this stage is reached on May 30th (BBCH=59) with the plant height around 170.9. Then, the flowering stage begins on June 1st (BBCH=63) and ends around June 8th (BBCH=67), the height of the plant is around 189.7 cm and 198.9 cm respectively. Finally, the development of fruit stage is reached on June 11th (BBCH=71) with the plant height of 205.9 cm and ends on June 18th (BBCH=79) with the plant height of 208.8 cm (see in Figure 3.5).

3.1.6. VEGETATION WATER CONTENT OF CORN

The VWC was measured 4 times per week during the experiment. The measurement schedule was 6 A.M. on Monday, Wednesday and Friday, 6 P.M. on Wednesday from 27 April to 13 June 2018. The samples were weighted, both fresh and after drying. All the samples were put in the oven at 60°C for one week. Bulk vegetation water content values were determined from the difference between fresh and dry biomass by Equation 3.1. Where η is the plant density, m_{dry} and m_{wet} stand for the measured weight of dry and wet biomass respectively. Besides, destructive sampling was performed by dividing the samples into stem, leaf, ear, tiller and tassel. Then the water content of these destructive parts of the plant were measured using the same procedure like bulk vegetation water content. Finally, the vegetation water content was interpolated in order to get a complete temporal variation in the growing season (see in Figure 3.6).

$$VWC = \eta * (m_{fresh} - m_{dry}) \quad (3.1)$$

Where the unit of η is the number of plants /m². The unit of m_{fresh} and m_{dry} is kg.

3.1.7. LEAF AREA INDEX (LAI)

The leaf area index (LAI) is a dimensionless quantity that describes the characteristics of plant canopies quantitatively. It is generally defined as the one-sided green leaf area of plants per unit ground surface area (m²/m²) in broadleaf canopies. Once a week, canopy geometry measurements were conducted on four to eight plants, including the maximum leaf length (l_{max}) and width (w_{max}) for all leaves. The assumption was made that a leaf had an ellipse shape, the leaf area can be calculated by Equation 3.2. The leaves and the rest of the plant were dried in the oven at 60°C for one week to measure the dry weight of the leaf and the dry biomass of the sample. The fraction of leaf (F_{leaf}) was determined by calculating the ratio of leaf dry weight to sample dry biomass. The specific leaf area (SLA) was determined by calculating the ratio of leaf area to the dry mass of the leaves. Then, the total dry biomass ($DM = \eta * m_{dry}$) measured in section 3.1.6 was used to calculate LAI with Equation 3.3. The values of LAI is shown in Figure 3.4.

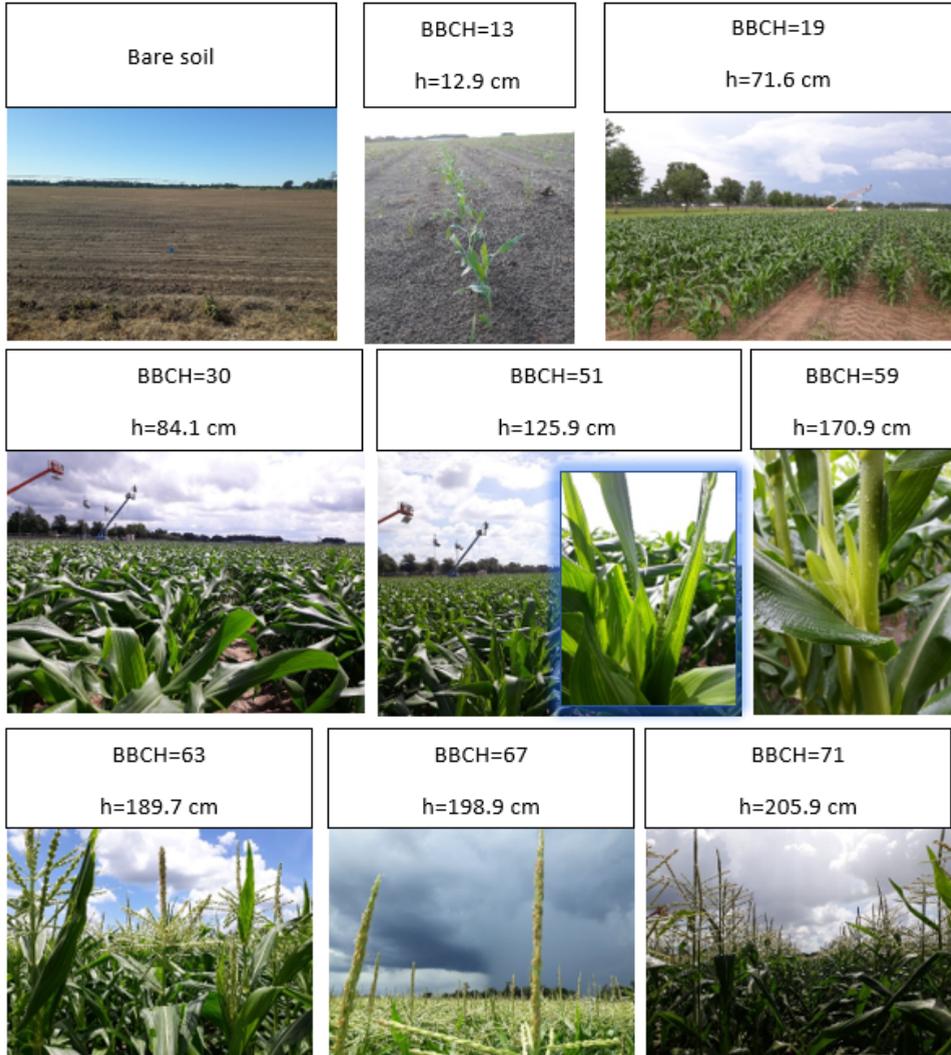


Figure 3.5: Pictures of corn during different growth stages

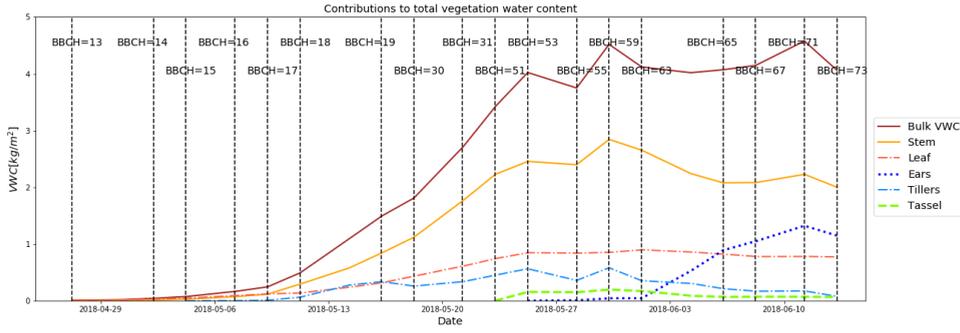


Figure 3.6: Contributions of different vegetation components to VWC_{bulk}

$$A_{leaf} = l_{max} * w_{max} * \pi \quad (3.2)$$

$$LAI = DM * F_{leaf} * SLA \quad (3.3)$$

Where the unit of l_{max} and w_{max} is m. The unit of F_{leaf} is kg/kg . The unit of DM is kg/m^2 and the unit of SLA is m^2/kg .

4

METHODOLOGY

In this chapter, methods that are used to achieve the research aim are addressed. This chapter will begin with how the TU-Wien VOD estimation method is adapted to estimate VOD from backscatter observations at a single incidence angle. Then, the retrieval method of other vegetation indicators selected by this research is explained. After that, the method used to investigate the relation between VOD and VWC is explained, which helps to improve the understanding of the relationship between VOD and water content of the canopy layer in the next chapter.

4.1. ESTIMATING VEGETATION OPTICAL DEPTH

In ground campaigns, tower- and truck-based scatterometers are applied to study the sensitivity of backscatter to soil moisture, vegetation structure, and vegetation water content dynamics under different frequency, polarization, and incidence angle. The data collected by ground-based scatterometers are typically in plot scale. Moreover, because of the high temporal resolution (diurnally, daily, or over the entire growth cycle) of the collected data, the ground-based scatterometers are ideal for the collection of multi-temporal datasets with high temporal resolution [17]. The ground-based UF-LARS was used to obtain backscatter observations every 15 minutes. The high temporal backscatter observations collected over a growing season of corn are used to estimate VOD here.

The aim of this section is to obtain VOD value at every radar retrieval time step at different polarization modes. The general idea of VOD retrieval in this study is firstly using the TU-Wien soil moisture retrieval method to obtain the vegetation correction which is represented by the dry and wet references. Secondly, vegetation optical depth from active microwave observations is calculated by ingesting the derived vegetation correction into the Water Cloud Model. However, in the TU-Wien soil moisture retrieval algorithm, multi-looking direction ability of scatterometer is used to describe the incidence angle behavior of the backscatter signal as a seasonal function. The vegetation correction is determined based on the slope and curvature of $\sigma^0(\theta)$, and cross-over angles. Thus, this cannot directly applied to the single-incidence backscatter data we have. Thus, the TU-Wien VOD estimation method needs adaptations. However, unlike the TU-Wien VOD

estimation method, we have the auxiliary in-situ soil moisture data which will help us to estimate the dry and wet references without the multi-incidence angle data. The flow chart (see in fig 4.1) provides you a general understanding of how the VOD time series is derived. More details are provided in the following sections.

4.1.1. DRY AND WET REFERENCE DETERMINING

Some adaptations are needed in order to apply the idea of using dry and wet references to represent the vegetation correction with single-incidence angle backscatter observations. In this study, with the auxiliary soil moisture data the simple linear model introduced in section 2.2.2 can be applied, which relates the backscatter coefficient with volumetric soil moisture (M_v). In this model, we assume the corn field as bare soil condition and the variation of backscatter is only caused by soil moisture change. Once the linear model at a certain time step is built, the dry and wet references can be determined by substituting the smallest and largest M_v into the linear model. The purpose of estimating dry and wet references using the slope and curvature of $\sigma^0(\theta)$ as well as cross-over angles is also to reduce the influence of vegetation cover. Therefore, when we have single-incidence angle backscatter data as well as the in-situ soil moisture data, we can build the linear model between the backscatter coefficient and volumetric soil moisture to estimate dry and wet references.

EXCLUDING THE LESS DYNAMIC SOIL MOISTURE DATA

When the soil is extremely dry, the value of volumetric soil moisture will concentrate on a certain range. Thus, the scatter plot of soil moisture and backscatter will look like a 'cloud' of points. From the mathematical point of view, if we try to fit a line through this 'point cloud', the result can be arbitrary. Thus, in order to avoid this situation, we need to exclude the soil moisture and the corresponding backscatter data when the soil moisture is less dynamic.

EXCLUDING BACKSCATTER INFLUENCED BY VEGETATION SURFACE WATER

Surface soil moisture changes rapidly within hours to days whereas the vegetation water content varies within several days to weeks [21]. In this study, we assume that the short term variation of backscatter is mainly caused by soil moisture variation. However, the presence and disappearance of interception on the canopy layer will also cause the backscatter fluctuates which is not related to soil moisture dynamics. Especially when the vegetation canopy is getting lush and dense, the interception on the canopy layer cannot be neglected. To avoid mistaking the short term variations of backscatter caused by vegetation surface water as the variations caused by soil moisture dynamics, data in the period that precipitation, irrigation, and dew happen when the vegetation cover is dense enough is also excluded before all the calculations begin. In this study, the canopy is considered as dense when it covers the soil surface. We call it canopy closure. The date of canopy closure is determined based on LAI, BBCH stages, and in-situ photo at the corn field. Here, the closure date is set as 18 May, with LAI = 2.09, BBCH stage = 30 and the in-situ photo of the canopy is shown in Figure 4.2).

With single-incidence angle backscatter observations, before the determination of dry and wet references, backscatter and soil moisture data need to be selected based on

Soil Moisture Data Selection: Exclude soil moisture as well as backscatter data at dry soil condition when the soil moisture is less dynamic (because in the following steps we will find the linear relation between the Volumetric Soil Moisture (independent variable) and the backscatter (dependent variable). If the soil is dry, the value of VSM will concentrate on a small range as well as the backscatter. Thus, the scatter plot of backscatter against VSM will look like a cloud of points. And if we fit a line through this 'cloud', the result will be arbitrary).

Backscatter Data Selection: Assuming that the short term variation of backscatter is mainly caused by soil moisture variation. However, the presence and disappearance of vegetation surface water will also cause the backscatter fluctuates which is not related to soil moisture dynamics. Especially when the vegetation canopy is getting lush and dense. Thus, identify a closure date of the corn canopy layer. After this date, the canopy is thick enough to cover the soil surface. Then, exclude the backscatter as well as the soil moisture data when dew, precipitation, and irrigation happens.

Exclude the $K(t) < 0$ and corresponding $C(t)$ values, which is against the physical theory. Also, according to the goodness of fit, exclude the poor fitting result (the goodness of fit is evaluate by R-squared, here we set the threshold=0.5, when the R-squared > 0.5 , we assume the linear relation is credible).

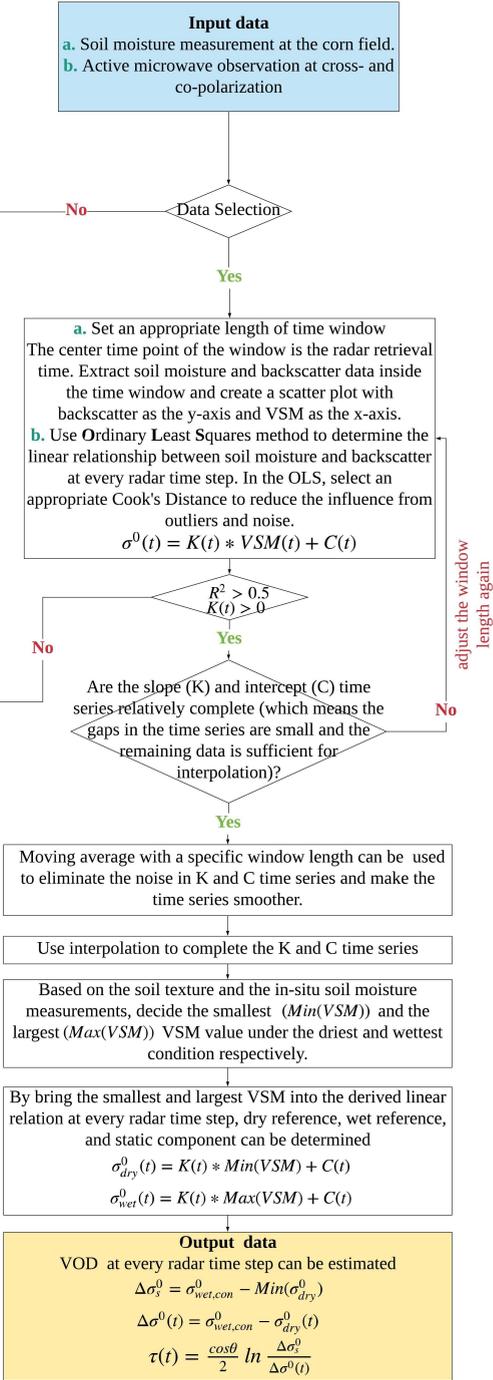


Figure 4.1: Flow chart of VOD retrieval

the criteria above. Only when the data exclusion is appropriate, the determined dry and wet references are reasonable.



Figure 4.2: In-situ photo of corn field at 2018-05-18

ORDINARY LEAST SQUARES (OLS) METHOD

After all the assumptions are met, the OLS method is used to derive the linear model (see in Equation 4.1) between M_v and σ_t^0 for every radar retrieval time. When the linear model at a certain time is built, bring the smallest and largest soil moisture of the entire growing season into the model, with the derived slope and intercept, the dry and wet references at this time point are calculated with Equation 4.2 and Equation 4.3.

$$\sigma^0(t) = K(t) * M_V + C(t) \quad (4.1)$$

$$\sigma_{dry}^0(t) = K(t) * M_v(min) + C(t) \quad (4.2)$$

$$\sigma_{wet}^0(t) = K(t) * M_v(max) + C(t) \quad (4.3)$$

Where $\sigma_{wet}^0(t)$ and $\sigma_{dry}^0(t)$ are the wet and dry references at each radar retrieval time point t , $K(t)$ and $C(t)$ are the slope and intercept of the linear relation at time point t . $M_v(min)$ and $M_v(max)$ are the minimum and maximum volumetric soil moisture content in the entire growing season. In this study, the value of $M_v(min)$ and $M_v(max)$ is 0.11 and 0.26 respectively.

In order to have one dry and one wet reference value every 15 minutes (radar system returns signal every 15 minutes), the linear model is applied to a sliding window with specific size.

SLIDING WINDOW METHOD

The sliding window method of finite window size, making the algorithm a finite impulse response filter and is suitable for analyzing a statistic over a finite duration of data. In the sliding window method, a window of specified length (denoted as Len), moves over the radar backscatter data and the corresponding volumetric soil moisture data with a step length (denoted as Δt) of 15 minutes (same with the radar retrieval interval). Then the linear relation between M_v and the σ^0 is derived over the data within the window. The output relation is derived by all the input data within the window, which are samples $\frac{Len-\Delta t}{2}$ before the central time t as well as samples $\frac{Len+\Delta t}{2}$ after the central time t . The output linear relation for each window is considered to represent the linear relation between M_v and σ^0 at the central time t of the current window. However, at some central time t (located at the start and the end of the time series), they do not have enough previous or subsequent data to fill the Len , then the algorithm neglects these central time points and no linear relations are derived. Figure 4.3 can help to understand the sliding window method. And Figure 4.4 shows how the linear model is applied to a certain window.

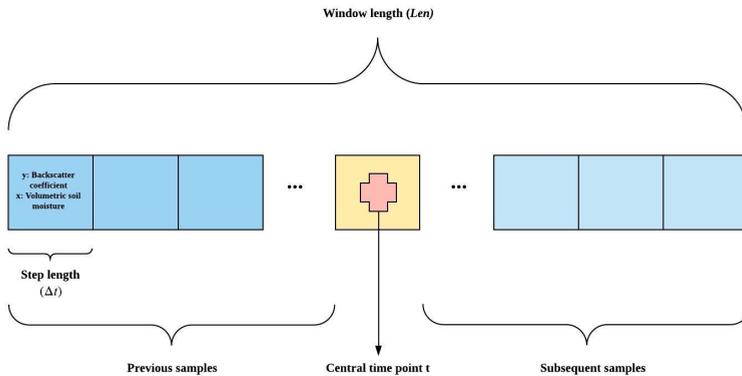


Figure 4.3: Illustration of sliding window method

DETERMINATION OF THE WINDOW LENGTH (Len)

The window length (Len) defines the length of M_v and σ^0 data over which the ordinary least squares algorithm derives the linear relation. If Len is large, the result calculated is closer to the stationary static of the data. Thus, for data without rapid changes, the longer window size can help to obtain a smoother static. Here, the purpose is using the derived linear relation to estimate dry and wet references at each central time point. Recall that the dry and wet references in the TU-Wien soil moisture algorithm are the backscatter that minimizes the influence of vegetation and assume the change of backscatter is mainly a result of soil moisture variations. Therefore, the temporal patterns of computed $\sigma_{dry}^0(t)$ and $\sigma_{wet}^0(t)$ are expected to better reflect the changes in backscatter due to soil moisture variations. According to the assumption that the soil moisture dynamics in

At VV polarization, linear function at radar time: 2018-05-13 11:29:00 .
 $R^2 = 0.8435502525745393$ slope= 90.87152353545466 intercept= -21.131070977166342

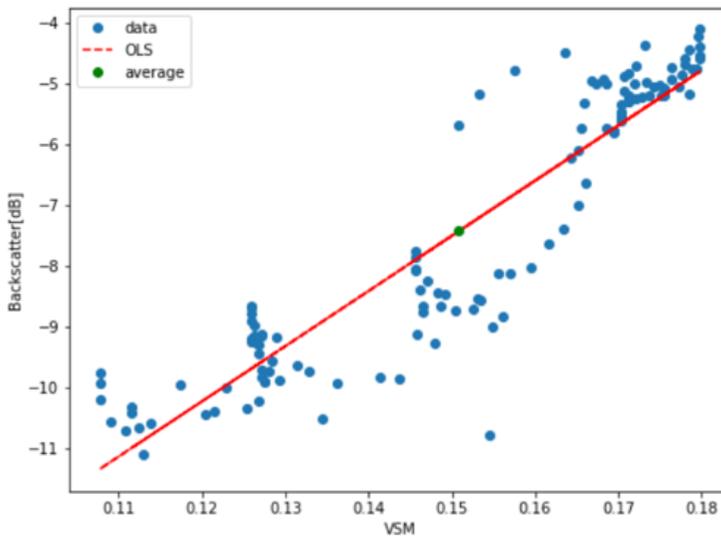


Figure 4.4: Example of how the linear model is applied to a certain sliding window. The window length of this plot is 5-day and the central time point is 2018-05-13 11:29:00 am. The blue dots in the plot are the samples inside this 5-day window with volumetric soil moisture (VSM) as the independent variable and backscatter as the dependent variable. The red line is the linear relation between VSM and backscatter derived by the OLS method. This linear relation with slope=90.87 and intercept=-21.13 is assumed as the linear relation at the central time point of this window

short time period, the expected size of window length needs to be as short as possible. However, from a mathematical perspective, if Len is too short, M_p values do not have enough variations, the regression will be fitting a line through a ‘cloud’ of points, the linear relation is not credible. It is clear that choosing an appropriate Len will improve the accuracy of the linear regression and further produce a more reasonable estimation of the temporal pattern of dry and wet references. In this research, after trial and error, a 5-day Len is set.

GOODNESS OF FIT

When the linear relation of every radar retrieval time point is obtained, the goodness-of-fit test is used to examine how well the linear model fits the data. In OLS, a test based on the coefficient of determination (denoted as R^2) is often used [47]. In this study, in order to make the estimation of dry and wet reference more credible, a threshold of R^2 is set as 0.5 (while for HH polarization, the threshold is set as 0.2, because the noise appears in the beginning period of the growing season). The linear relation at a certain time step with the R^2 less than 0.5 is excluded from the subsequent calculation. However, for the purpose of keeping the time series of dry and wet reference as complete as possible, Cook’s distance D is introduced in the OLS to detect and remove the outliers inside the window. Each element in the Cook’s distance D is the normalized change in the fitted response values due to the deletion of an observation [48]. The Cook’s distance of observation i is calculated by:

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_{j(i)})^2}{p * MSE} \quad (4.4)$$

where \hat{y}_j is the j th fitted response value, $\hat{y}_{j(i)}$ is the j th fitted response value, where the fit does not include observation i . MSE is the mean squared error. p is the number of coefficients in the regression model.

If the Cook’s distance D value of one sample point is larger than $\frac{4}{p}$, this point is usually considered as an outlier. Here, the outliers in backscatter data are more likely representing the fluctuation and noise caused by the vegetation surface water.

After removing the outliers from the sampled window, the R^2 of that derived linear relation at a certain time point will be improved. Thus, more time point can meet the threshold and be used for further estimation.

TIME SERIES OF SLOPE AND INTERCEPT OF THE LINEAR RELATION

After filtering the linear relations obtained in the growing season with a set threshold, those linear relations with R^2 higher than threshold are retained. Then, we can get the time series of both slope and intercept from the retained linear relations. To smooth the slope and intercept time series, a 5-day moving average is applied. Then, with the smallest and largest M_p of the entire growing season as well as the obtained slope and intercept time series, dry and wet references can be determined.

4.1.2. ESTIMATING VOD FROM WATER CLOUD MODEL(WCM)

After determining the dry and wet references, VOD can be estimated with Equation 2.11 to Equation 2.14. Recall the previous discussion of setting wet reference as a constant

value through the growing season, in this study, we exclude the largest 5% and the smallest 5% of the $\sigma_{wet}^0(t)$, then the average value of the rest $\sigma_{wet}^0(t)$ is set as the representative value of the wet reference of the growing season (denoted as $\sigma_{wet,con}^0$). Then, Equation 2.11 can be written as Equation 4.5.

$$\Delta\sigma^0(t) = \sigma_{wet,con}^0 - \sigma_{dry}^0(t) \quad (4.5)$$

Besides, in this study, at the beginning of the growing season, the corn field is in bare soil condition. Therefore, the $\Delta\sigma_s^0$ can be computed with Equation 4.6.

$$\Delta\sigma_s^0 = \sigma_{wet,con}^0 - Min(\sigma_{dry}^0(t)) \quad (4.6)$$

Unlike multi-incidence angle backscatter observations using incidence angle behavior of backscatter signal as a seasonal function, single-incidence backscatter observations use the variation of backscatter sensitivity to soil moisture to indicate the vegetation growth and senescence. At the beginning of the growing season with less vegetation, the backscatter is more sensitive to soil moisture. When the vegetation cover grows thicker, the sensitivity reduces. When the vegetation cover fades, the sensitivity of backscatter to soil moisture recovers. The dry and wet references are determined based on the slope and intercept time series and can be used as vegetation correction. Then, VOD is estimated by bringing this vegetation correction into the Water Cloud Model with Equation 2.14.

4.2. OTHER VEGETATION INDICES

4.2.1. CROSS-RATIO

The UF-LARS radar system collects radar backscatter at four polarization combinations (HH, VV, HV, and VH). Polarization may contain information about the form and the orientation of small scattering elements which compose the surface or target. Ulaby and Moore used the University of Kansas microwave active and passive spectrometer (MAPS) from 4 to 8 GHz in the experiment, the result shows that lower frequencies horizontally polarized backscatter is more sensitive to soil moisture [17]. *Srivastava et.al* [49] found in their experiment that L-band, backscatter in HV-polarization mode was most sensitive to vegetation water content and can be used as a good estimator of VWC. HH polarization is more sensitive to surface scattering, while cross-polarization is more sensitive to volume scattering and VV polarization is the combination of HH and cross-polarization [50]. Therefore, backscatter at HH-polarization mode is better for ground parameter retrieval, and cross-polarization is better for vegetation parameters retrieval.

However, Brown's study at C-band [51] shows that soil backscatter returns at VH polarization are probably caused by double scattering (stem-ground), the double-bounce scattering may exceed the direct backscatter from the soil [52]. Under this situation, the Cross-Ratio is introduced. The determination of CR is based on Equation 4.7 in the linear domain. Then, all CR values are converted to the logarithmic domain.

$$CR (dB) = 10 * \log\left(\frac{VH (m^2/m^2)}{VV (m^2/m^2)}\right) \quad (4.7)$$

Compared with co-polarization, backscatter at cross-pol. increases more strongly due to the volume scattering, which means CR will also increase due to the increase of volume scattering. Besides, CR can reduce the double-bounce effect. Apart from that, CR probably reduces errors associated with the radar system (e.g. error caused by the radar system stability) or environmental influence (e.g. error associated with soil moisture variability). This will make CR a more stable parameter compared with either VV- or cross-polarization backscatter coefficient [50]. Thus, CR can be used as an estimator for crop monitoring. However, a challenge of using the Cross-Ratio is the soil roughness and vegetation structure. Rough soil surface can lead to depolarization which will have the same CR value as a vegetated surface. Vegetation structure influence on backscatter signals and the structure of crops changes throughout the growing season [53]. Thus, there are more drivers than vegetation water content that cause the variation of CR.

4.2.2. NDVI

The Normalized Difference Vegetation Index (NDVI) is a simple graphical indicator that can be used to analyze remote sensing measurements. Because of its simplicity, efficiency and high spatial resolution, the NDVI is the most commonly used vegetation index currently [54]. NDVI is defined as:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (4.8)$$

Where NIR and RED represent the amounts of near-infrared and red light, respectively, reflected by the above-ground biomass and captured by the sensors installed on the satellite. For healthy vegetation, chlorophyll inside the green vegetation reflects more near-infrared and green lights and absorbs more red and blue light. Whereas, the mesophyll leaf structure absorbs the near-infrared light. Thus, the value of NDVI indicates the amount of photosynthetically active vegetation. NDVI values range between -1 and 1, the higher the NDVI, the healthier the vegetation will be, while the negative value indicates the absence of vegetation [55].

Although the NDVI is not an intrinsic physical quantity, it is indeed correlated with certain physical properties of the vegetation canopy: leaf area index (LAI), fractional vegetation cover, vegetation condition and biomass [8]. There are some drawbacks of NDVI. Firstly, NDVI is limited by atmospheric effects, NDVI is sensitive to the attenuation by cloud, cloud shadows or aerosols. Secondly, the sensitivity between NDVI to LAI becomes weaker with LAI increasing and beyond a threshold which typically between 2 and 3. Then, the effects of soil brightness and color may produce a large NDVI variability. Finally, the NDVI only sense the top layer of the canopy [8] [9] [10]. Although the VOD and NDVI show some correlation across spatial and temporal scales, they are fundamental measures of different vegetation properties. The VOD measured by microwave observations is an integral measure of vegetation water content and structural effect [11] while NDVI measures the canopy greenness.

In this study the NDVI values are obtained by combining the NIR and Red band of the Sentinel 2 data from the Google Earth Engine. The high cloud occurrence in the case study will influence the value of NDVI. To avoid the effect of the cloud, function to mask clouds uses the cloud mask information in the Sentinel-2 QA60 band. The cloud mask

enables cloudy and cloud-free pixels to be identified. The mask includes both dense clouds and cirrus clouds with an indicator specifying the cloud type. When the cloud mask is set as 0, it represents the cloud-free pixel [56][57].

4.3. THE RELATION BETWEEN VOD AND VEGETATION WATER CONTENT

As mentioned before, VOD is closely related to water content in the above ground plant components. Previous studies suggest that VOD and bulk vegetation water content are linearly related to each other. However, in this study we try to find out whether this linear relation holds true through the entire growing season of the corn. VOD at co- and cross-polarization are analyzed here.

First, the vegetation water content of different vegetation components (denoted as VWC_i) are interpolated into a 15 minutes interval dataset as shown in Figure 3.6. Then, we have one VOD value and one VWC_i at every radar retrieval time. The relation between VWC_i and VOD at different polarization modes will be clear after making a scatter plot that VWC_i as x-coordinate and VOD as y-coordinate. If the linear relation holds true in certain growth stages, the OLS method is used to derive the linear relation. R^2 of each linear relation is reported to show how well the linear model fits the data. R^2 equals the square of the Pearson correlation coefficient between the observed and modeled values of the dependent variable.

4.4. IDENTIFYING THE DRIVERS OF TEMPORAL VOD VARIATIONS

4.4.1. RANDOM FOREST MACHINE LEARNING

In sensitivity analysis (a) The sensitivity of VOD to water content of different vegetation components, (b) The sensitivity of VOD to the stems water content at different height and (c) The sensitivity of VOD to the leaf water content at different height are further quantified using a supervised random forest (RF) machine learning technique. In order to compare VOD with CR, the sensitivity analysis of CR using the RF model is conducted in the same way as VOD.

A random forest classifier is an ensemble classifier that consists of many decision trees. RF technique generates a random sample of the original data with replacement (bootstrapping) and uses a user-defined number of the randomly selected subset of variables to determine node splitting. In the supervised mode, RF will run as a classifier and will measure the importance of all variables as well as the ability of each variable to classify the data appropriately. Supervised classifiers are more robust compared to the model-based approaches [58]. The supervised classifiers can learn and find out the characteristics of the response variable from the provided training samples and then apply the learned characteristics in the unclassified data. The RF has the advantage that it does not make any assumptions about the relation between the input and the output variables and can be used for nonlinear multiple regression [1]. Besides, RF can limit over-fitting without substantially increasing error due to bias.

Here, the additional RF functions called feature importance are used to identify the

most important input variables for the response variable. There are two popular approaches for feature ranking. Mean Decrease in Gini (MDG) and Mean Decrease Accuracy (MDA). The **MDG** measures how much a variable reduces the Gini impurity metric in a particular class [59][60]. However, when using MDG, several things need to keep in mind. Firstly, when the data set has two or more correlated features, any of them can be used in the RF model as a predictor, with no preference of one over another. But once they are all selected as predictors, the use of one of them will significantly reduce the importance of others since the impurity they can effectively remove is already removed by the first feature. As a result, the reported feature importance of them will be lower. Secondly, MDG is biased toward preferring variables with more categories. In this study, the input variables are not independent of each other. Thus, we used MDA as the feature selection method. **MDA** measures the impact of each feature on the accuracy of the model. The general idea in MDA is first to randomly permute (simply shuffles the array along the timeline. The order of the sub-arrays has changed, but its contents remain unchanged) the values of each feature and then take account of the decrease of the model accuracy after permutation. Obviously, permutation of the unimportant feature will result in small impact on model accuracy while the permutation of the important feature will lead to clearly decrease in model accuracy. In this study we use function `r2_score` in python to estimate the accuracy between the modeled and actual value. Best possible score is 1.0 and it can be negative because the model can be arbitrarily worse. Thus, when negative value appears, the input can not explain the target at all. A constant model that always predicts the expected value of response variable, disregarding the input features, would get a R^2 score of 0.0. The MDA between the original data and permuted data is calculated by:

$$MDA = \frac{accuracy - accuracy_{permute}}{accuracy} \quad (4.9)$$

In the sensitivity analysis, we want to use the MDA as ranking method to find out which component of the vegetation water content is more important to VOD. Three sensitivity analysis are conducted in both vegetative and reproductive stage.

- (a) Three RF models are built. The input variables of all three models are the stem water content (VWC_{stem}), leaf water content (VWC_{leaf}), and ear water content (VWC_{ear}). The response variable of the three models is VOD at HH polarization, VV polarization and cross-polarization respectively.
- (b) Three RF models are built. The input variables of all three models are stem height 0-20 cm, stem height 20-40 cm, stem height 40-60 cm, stem height 60-80 cm, stem height 80-110 cm, stem height 110-140 cm, and stem height 140-170 cm. The response variable of the three models is VOD at HH polarization, VV polarization and cross-polarization respectively.
- (c) Three RF models are built. The input variables of all three models are leaf 1-4, leaf 5-7, leaf 8-10, leaf 11-13, leaf 14-15. The response variable of the three models is VOD at HH polarization, VV polarization and cross-polarization respectively.

In order to compare the differences between VOD and CR, sensitivity analysis to identify the drivers of CR variations are conducted as well in both vegetative and reproductive stage.

- (a) The input variables are the stem water content (VWC_{stem}), leaf water content (VWC_{leaf}), and ear water content (VWC_{ear}). The response variable is CR.
- (b) The input variables are stem height 0-20 cm, stem height 20-40 cm, stem height 40-60 cm, stem height 60-80 cm, stem height 80-110 cm, stem height 110-140 cm, and stem height 140-170 cm. The response variable is CR.
- (c) The input variables are leaf 1-4, leaf 5-7, leaf 8-10, leaf 11-13, leaf 14-15. The response variable is CR.

4

Within all these random forest models, 300 decision trees with a maximum depth of 30 are trained. After building the RF, the analysis of the feature importance can be conducted, which quantifies the relative importance of the different variables for estimating VOD at different polarization modes in different growth stages. The variable importance quantifies the reduce of model accuracy when this variable is permuted. It can give us a general idea that which input variables have more effect on VOD. Notice that, the water content of different vegetation components is correlated in some ways, therefore, the MDG is not suitable to rank the features. In this study, the feature importance is calculated by MDA. The feature importance calculated by MDG is shown in appendix D, the ranking is different from the result using MDA.

The Out-Of-Bag (OOB) score is calculated to assess the performance of the models. In the RF, every tree uses bootstrap sampling, this means that every tree in the forest is built on about 63% of the available data, leaving the remaining approximately 37% which is the OOB data. The OOB score is the average R^2 calculated from the trees that do not contain a certain value for response variable in the respective bootstrap sample [53]. In this study, RF is not recommended for VOD prediction because the sample size is not large enough. However, the feature importance function of RF can be used to identify the drivers of the VOD variation.

5

RESULTS AND DISCUSSION

This chapter describes and discusses the results. First, the time series of dry, wet references as well as the time series of VOD of the corn field are shown. Second, the temporal variation of VOD during the growing season is discussed. Next, the time series of other vegetation indicators is compared with VOD. Then, the relation between VOD and VWC is explored. Lastly, we will move to the sensitivity analysis which tries to discover which component of bulk vegetation water content has more influence on VOD variations at different BBCH stages.

5.1. ESTIMATING VEGETATION OPTICAL DEPTH

SLOPE AND INTERCEPT DERIVATION

In this study, the R-squared returned by the OLS method is used to evaluate the goodness of fit. We set a threshold of R-squared and assume that a linear relation is credible when its R-squared is higher than this threshold. The threshold at VV and cross-polarization is set as 0.5 while the threshold at HH polarization is set as 0.2. This lower threshold is because we considered the noise contains in HH polarization (see in Appendix A) will affect the fitting result. Thus, compared with VV and cross-polarization we have to lower the standard at HH polarization in order to keep the temporal variation of slope and intercept relatively complete.

In Figure 5.1 and Figure 5.2, with threshold = 0.5, slope and intercept time series at VV polarization is more complete compared with other polarization modes. This means the linear relation can apply better in VV polarization. When the vegetation cover is less dense (before May 11th, LAI \leq 2.1, BBCH before 18), the slope value of VV polarization is higher compared with other polarization modes, which means during this period VV polarization is more sensitive to M_v . However, when vegetation cover is thicker (After May 30th, LAI \geq 3.4, BBCH after 59), the slope value of cross-polarization is higher compared with other polarization modes, which means during this period cross-polarization is more sensitive to M_v . There is an obvious gap that starts around May 17th and ends around the end of May. In this period, precipitation happens in most time of a day (the

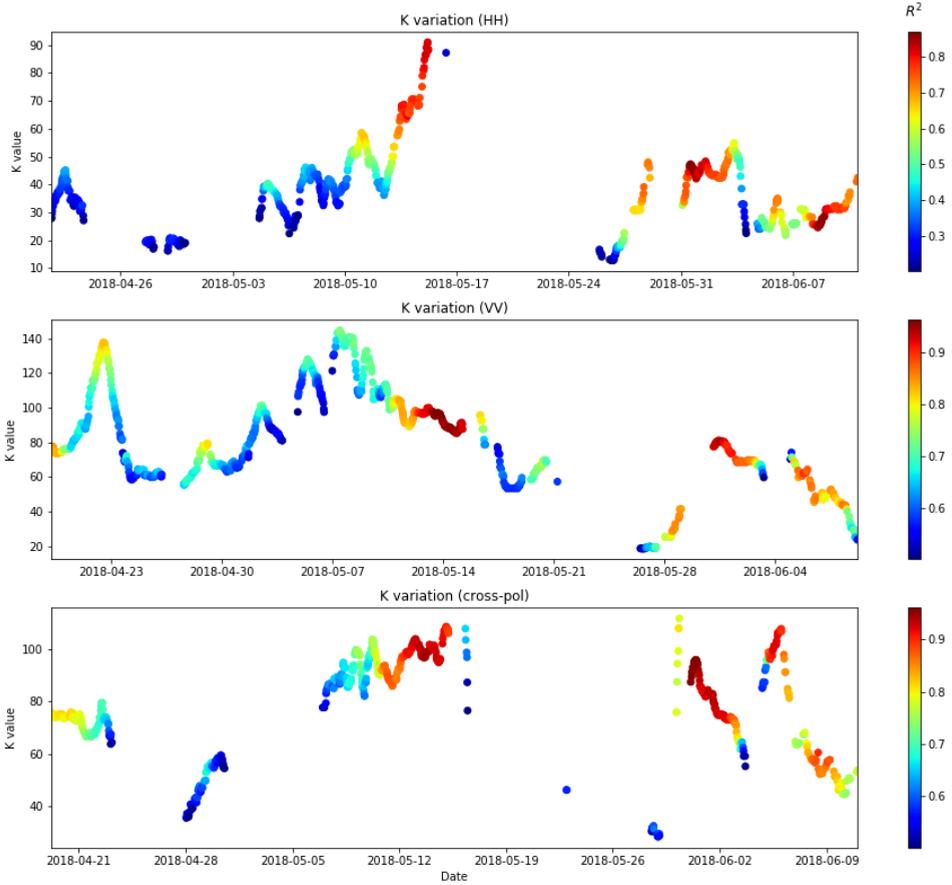


Figure 5.1: The time series of slope derived by OLS at every radar time step with goodness of fit higher than set threshold. Y-axis is the value of slope(k), x-axis is the time, and the color of each dot indicates the goodness of fit)

exact time of precipitation can be found in [41]), this influence the linear relation between backscatter and M_v . However, VV polarization is less vulnerable in this situation. In Figure 5.2, the intercept (C) time series shows a rising trend from the start to the end of the growing season. The C value is an indicator of the total backscatter. Clearly, the growth of vegetation cover will result in an increase in backscatter. Therefore, as the vegetation becomes more luxuriant, the time series of C also shows an upward trend.

In Figure 5.1 and Figure 5.2, there are fluctuations as well as unreasonable spikes with extreme large or small values in the slope and intercept time series. The slope and intercept time series are affected by the linear model prediction accuracy, continuous precipitation, and the uncertainty of radar or soil moisture sensor. Therefore, in order to get a smoother time series, a 5-day moving average is applied to reduce the noise in slope and intercept time series. Smoothed time series is shown in Appendix C).

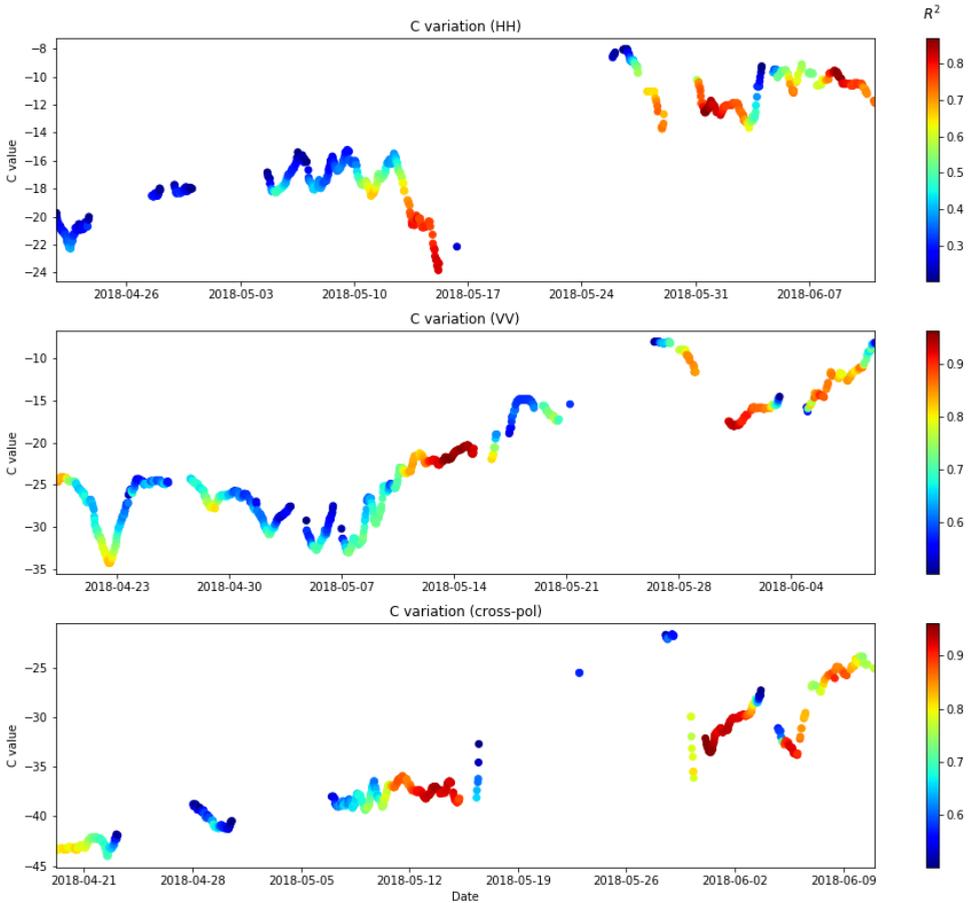


Figure 5.2: The time series of intercept derived by OLS at every radar time step with goodness of fit higher than set threshold. Y-axis is the value of intercept(C), x-axis is the time, and the color of each dot indicates the goodness of fit

5.1.1. DRY AND WET REFERENCES TIME SERIES

After obtaining the slope and intercept of the linear relation at every radar retrieval time, we can determine the dry and wet references by substituting the smallest and largest volumetric soil moisture into these linear relations.

The dry and wet references are shown in Figure 5.3, some backscatter values are much higher than the wet reference line and some backscatter values are slightly lower than the dry reference. As discussed in section 4.1.1, in order to meet the assumption of this study, backscatter data under the extremely dry condition and backscatter during precipitation, irrigation and dew after the closure date are all excluded before the calculation starts.

The variation of the dry reference is related to the growth of the vegetation layer. At HH polarization, the dry reference keeps growing before BBCH=63 and becomes sta-

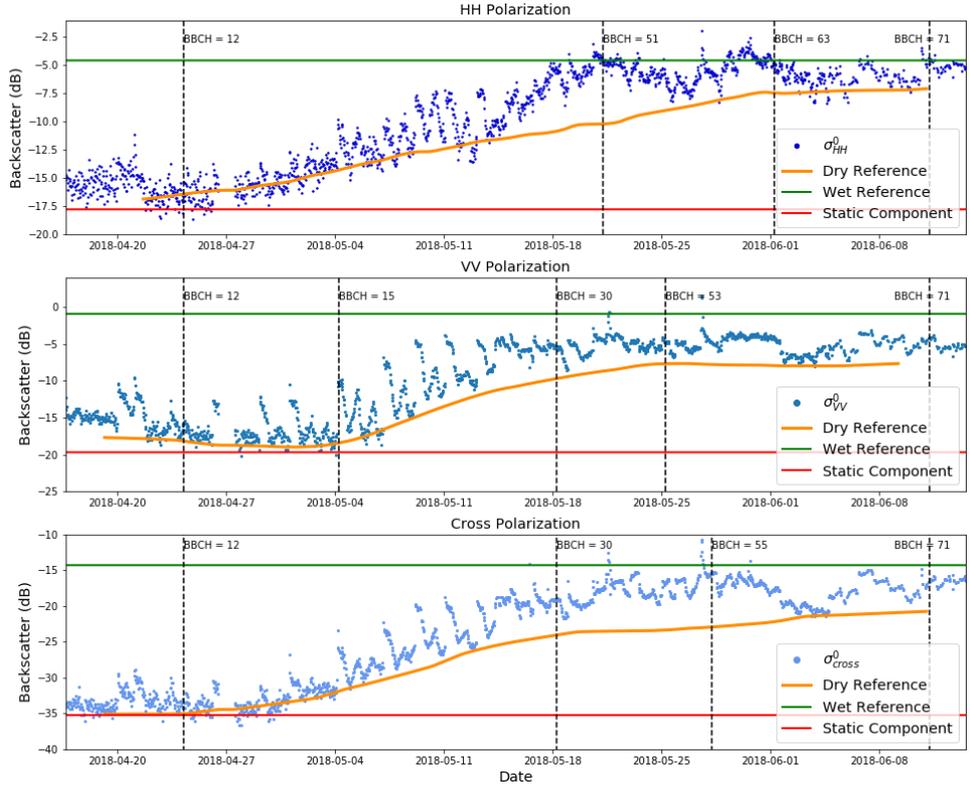


Figure 5.3: The dry reference, wet reference, and static component at HH (top), VV(middle) and cross-polarization(bottom)

ble after that. At VV polarization, the dry reference rises after BBCH=15 and keeps the growing trend until BBCH=53. Then it stays relatively stable after BBCH=53. At cross-polarization, the dry reference keeps rising before BBCH=55 and stays relatively stable after that. The rapid growth of dry reference at co- and cross-polarization all happens within the vegetative stage, while the variation of dry reference is small during the reproductive stage. Before 11 May, the difference between the dry and wet references are larger in VV polarization. Besides, the rising of the dry reference at VV polarization appears later compared with other polarization modes. Thus, backscatter at VV polarization is more sensitive to soil moisture. From Figure 5.3, the difference between the dry and wet references at co- and cross-polarization can indicate the variation of vegetation cover reasonably.

5.1.2. VOD TIME SERIES

Typically, values of VOD for agriculture crops have generally be given as less than one [61]. While, in this study, the magnitude of estimated VOD at both co- and cross-polarization seems less than expected. However, VOD values are very sensitive to dry and wet refer-

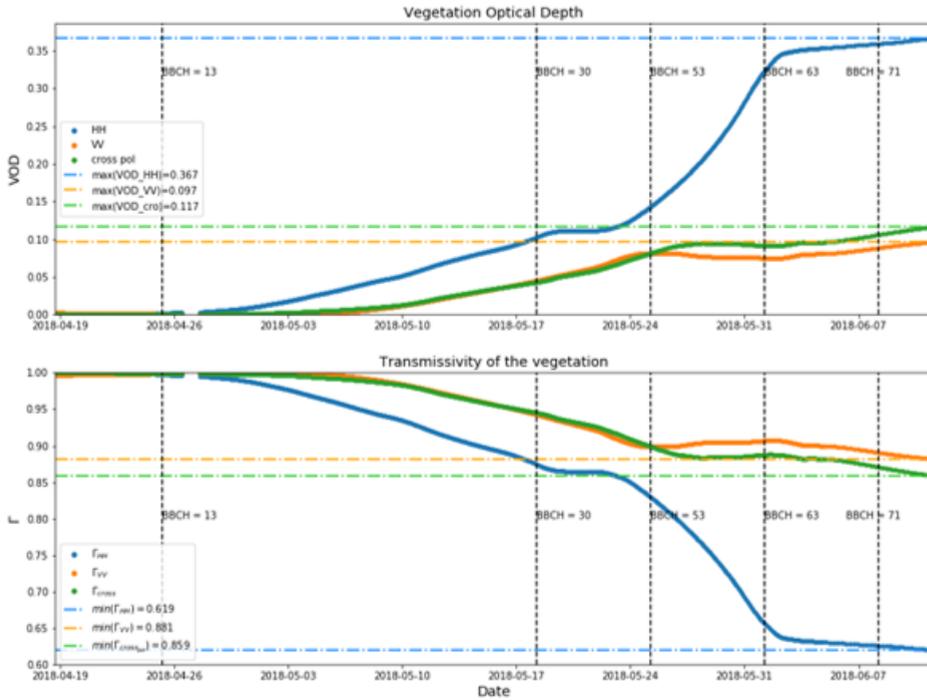


Figure 5.4: VOD at co- and cross-polarization(top) and transmissivity at co- and cross-polarization(bottom)

ence values. Based on the discussion in section 5.1.1, dry and wet references are reasonable and can indicate the vegetation variability. Besides, we use the variation of VOD to reflect the water content variation in the vegetation canopy. Hence the temporal variation of VOD is more important than its actual values.

From Figure 5.4, at HH polarization, VOD begins to increase at BBCH=13 until BBCH=30 while another increasing period is from BBCH=53 until BBCH=63 at a larger rate. The former is in the leaf development stage which coincides with the increase of the bulk vegetation water content(see in Figure 3.6). The latter is in the heading stage where no rapid increasing of bulk vegetation water content happens. Thus, this rapid change of VOD may be related to the moisture redistribution inside the canopy in the heading stage. Then, VOD is relatively stable from BBCH=30 to BBCH=53, while the bulk vegetation water content is still increasing. From BBCH=63 to the end, VOD only increases slightly, and the bulk vegetation water content also increases a bit in this period. At VV and cross-polarization, the VOD has similar trend, which can be divided into two parts. From BBCH=13 to BBCH around 53, VOD grows more rapidly. This period is in the leaf development stage and stem elongation stage and coincides with the rapid increase of bulk vegetation water content. From BBCH around 53 to the end, the rate becomes lower. This rate change happens at the beginning of the heading stage of the corn. However, the time that the rate becomes lower appears later at the cross-pol. than in the VV pol., which indicates the VOD at VV polarization is more likely to be affected by heading.

For VOD at VV and cross-polarization, their variation correlated well with bulk vegetation water content before BBCH=53. While VOD at HH polarization, its variation correlates well only before BBCH=30. This indicates that the structure change of corn has more influence on VOD at HH polarization.

5.2. CROSS-RATIO AND NDVI TIME SERIES

In this study, except for VOD, we also select CR and NDVI as vegetation indicators. The temporal variation of CR and NDVI can be seen in Figure 5.5 and Figure 5.6.

5.2.1. CROSS-RATIO

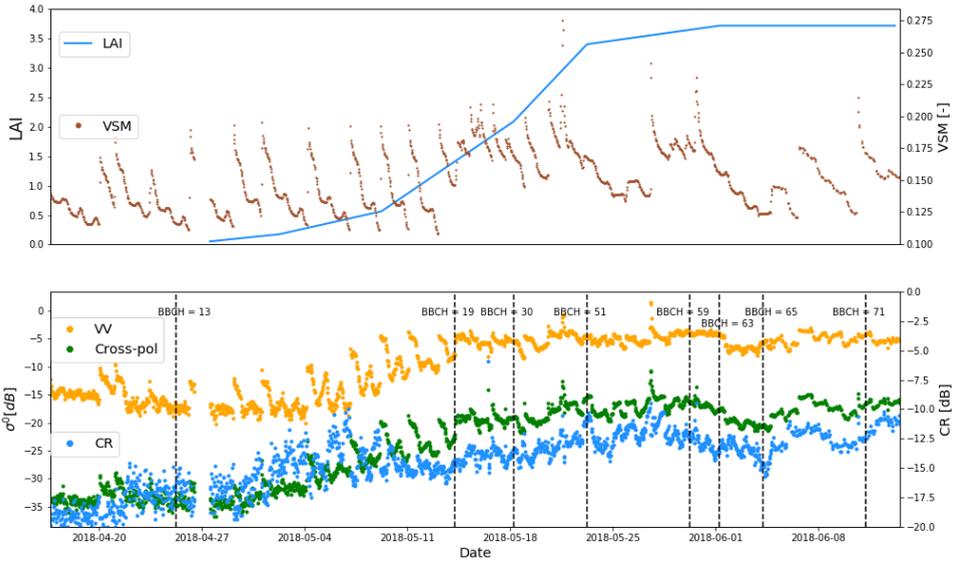


Figure 5.5: Volumetric soil moisture and LAI(top) and backscatter at VV and cross-polarization as well as CR values(bottom)

In Figure 5.5, we can observe that the CR increases with vegetation. CR starts to increase from BBCH=13 and this growth ends at BBCH=51. The value of CR increases from -18 dB to -12 dB in this period. This increasing period covers the leaf development stage and the stem elongation stage. The average VWC_{bulk} at BBCH=51 is 3.5 kg/m^2 and LAI of this day is 3.4. Then from BBCH=51 to BBCH=55, CR grows a little. After BBCH=55, the increasing trend stops and starts to decrease rapidly. And from BBCH=65 to BBCH around 71, the increasing trend of CR re-emerges. However, the time series of CR also shows fluctuations due to soil moisture dynamics clearly. Thus, at L-band, the effect of soil moisture is not reduced much as expected.

In BBCH=13 to BBCH=51, both LAI and VWC_{bulk} grows rapidly. The increasing CR correlated to LAI and VWC_{bulk} in this period. After BBCH=51, the LAI exceeds 3, the sensitivity of CR to LAI is lost. A similar threshold of LAI was also found by Jiao *et.al* [62]. From BBCH=53 to BBCH=71, unlike VOD time series in Figure 5.4, CR time series

has both increase and decrease periods. Bulk vegetation water content in Figure 3.6 also shows the similar increase and decrease almost in the same period. Thus, CR correlates better to VWC_{bulk} in this late growing period, which firstly decreases from BBCH=59 to BBCH=65 and then increases from BBCH=65 to BBCH=71 (see in Figure 3.6). This shows that after the tassel fully emerged and separate, CR can respond to small changes in bulk VWC.

The time series of CR indicates: firstly, the soil moisture effect on CR is not reduced much, this may due to the L-band has larger penetration depth and the backscatter is more sensitive to soil moisture content. Secondly, from BBCH= 13 to BBCH=55, the sensitivity of CR to vegetation growth (from -18 dB to -12 dB) is less compared with backscatter at cross-polarization (from -35 dB to -10 dB). Thirdly, compared with VOD, after the heading stage, CR responses to small changes in VWC_{bulk} . Fourthly, when LAI exceeds 3, the sensitivity of CR to LAI is lost. Lastly, the noise of CR at the beginning of the growing period may be caused by the structural change due to leaf development.

5.2.2. NDVI

Because of the cloud influence in the case study area, only six valid values of NDVI were obtained during the growing season (see in Figure 5.6). In the NDVI time series, there are two strange points. One is on April 19th, one day after the corn is planted. The NDVI value of this day is even larger than the NDVI value after the corn emergence. The other point is on May 29th, the NDVI value at that day is similar to the NDVI value at the start of the growing season. Theoretically, the variation of vegetation canopy with time is in a small range, the time series of NDVI is expected to be smooth. However, the signal received by sensors is affected by the atmosphere, cloud, geometric misregistration, anisotropic reflectance effects, electronic errors [6]. Hence, the NDVI time series may contain some noise. If the time series is dense enough we could use a smoothing method to remove the noise. However, within this short growing season, the amount of NDVI samples is far from sufficient. Thus, it is hard for us to draw valuable conclusions from this NDVI time series. In this case, the NDVI is less reliable compared with radar data.

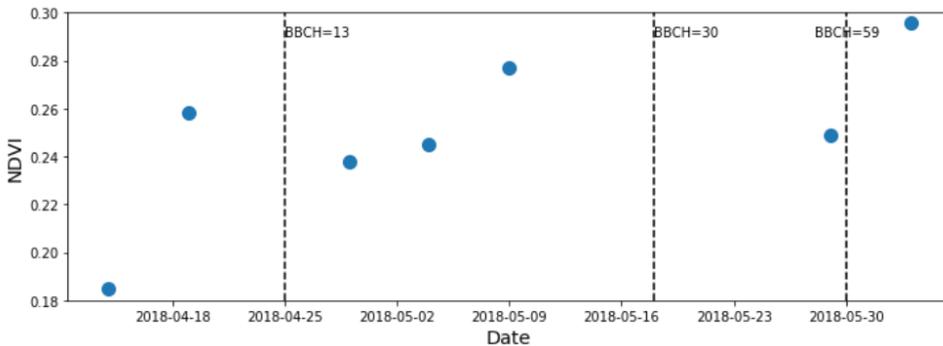


Figure 5.6: Blue dots are NDVI values during the growing season obtained from Google Earth Engine

5.3. THE RELATION BETWEEN VOD AND VWC_{bulk}

From Figure 5.7, it is clear that the linear relation between VOD and VWC_{bulk} does not always hold true during the entire growing season of the corn at both co- and cross-polarization. The relation between VWC_{leaf} and VOD, VWC_{stem} and VOD hold true at co- and cross-polarization before BBCH=51. For VOD_{ear} and VOD, the linear relation holds true from BBCH=63 to BBCH=71. The non-linearity of VOD and VWC_{bulk} after BBCH=51 is mainly due to the ear emergence. In this period, the influence of moisture redistribution inside the canopy on VOD variation is obvious. Notice, at HH polarization, before BBCH=51, the relation between VOD and VWC_{bulk} (also stem, leaf water content) is less linear visually. This is caused by the large range of the y-axis. If we only focus on the data before BBCH=51, then the linear relation is obvious (see the scatter plot in appendix E). Table 5.1 to Table 5.4 show the linear relation derived by ordinary least square method.

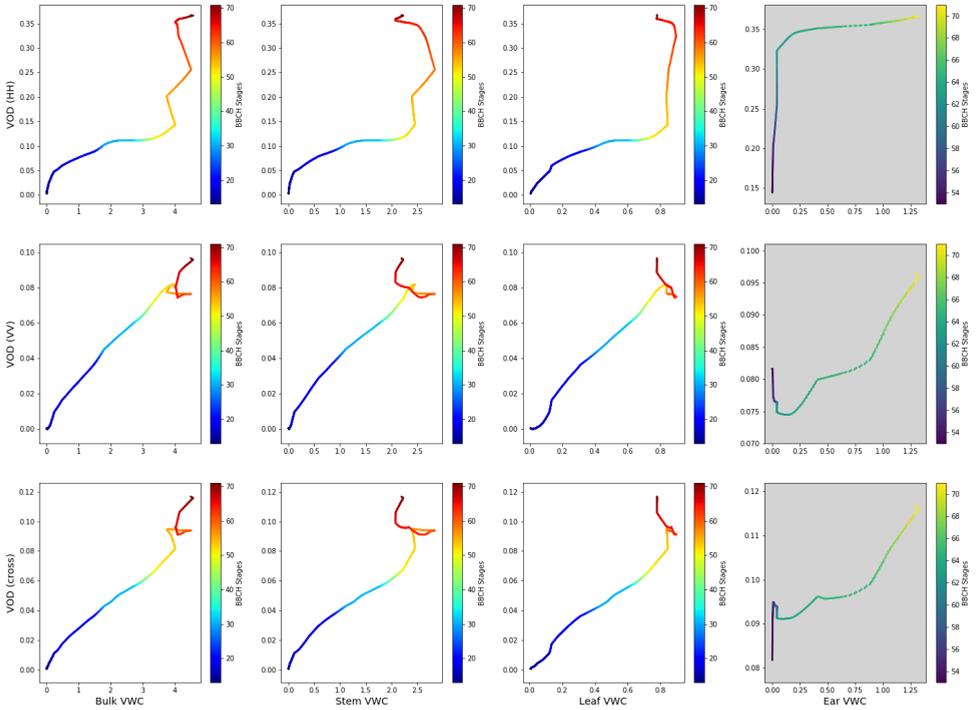


Figure 5.7: Scatter plots show the relation between VOD and VWC (This figure contains three rows, the y-axis of each row is HH-, VV- and cross-polarization (from top to bottom), respectively. As for the four columns of this figure, the x-axis of each column is bulk, stem, leaf and ear water content (from left to right), respectively. Then in every subplot, the color of each point represents the BBCH growth stage. Because ears emerge in the later part of the growing season. Therefore, the y-axis of the fourth column (with grey background) in this figure does not start from 0)

From Figure 5.7, for bulk vegetation water content, when it hits 3.64 kg/m^2 , the linear relation is not valid. For stem and leaf water content, the ending point of the

Polarization	Slope	Intercept	R^2
HH	0.033	0.027	0.858
VV	0.022	0.002	0.989
Cross	0.020	0.004	0.983

Table 5.1: The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{bulk} from BBCH=13 to BBCH=51

Polarization	Slope	Intercept	R^2
HH	0.050	0.030	0.827
VV	0.034	0.003	0.977
Cross	0.030	0.005	0.968

Table 5.2: The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{stem} from BBCH=13 to BBCH=51

Polarization	Slope	Intercept	R^2
HH	0.159	0.023	0.875
VV	0.104	-0.001	0.985
Cross	0.093	0.002	0.981

Table 5.3: The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{leaf} from BBCH=13 to BBCH=51

Polarization	Slope	Intercept	R^2
HH	0.016	0.343	0.964
VV	0.017	0.071	0.942
Cross	0.020	0.086	0.899

Table 5.4: The slope, intercept and R^2 calculated by OLS to describe the linear relation between VOD and VWC_{ear} from BBCH=63 to BBCH=71

linear relation is 2.31 and 0.78 kg/m^2 respectively. As for the ear water content, the linear relation is not valid when ear water content lower than 0.17 kg/m^2 . Therefore, at both co- and cross-polarization, the linear relation between VWC_{bulk} and VOD is valid only before the heading stage.

The slopes of these linear relationships indicate the sensitivity of VOD to water content of different vegetation components. From Table 5.1, VOD at HH polarization is more sensitive to bulk vegetation water content. However, the R^2 of the linear model at HH polarization is smaller compared with linear model either at VV or cross-polarization. This may be caused by the noise contains in the backscatter observations at HH polarization.

In summary, the VOD is highly related to VWC_{bulk} . The linear relation between VWC_{bulk} and VOD is not completely wrong. Before BBCH=51, their relationship can be described by a linear model with a static slope. However, after BBCH=51 especially when the ear emergence, this simple relation cannot be used anymore. During the period when the linear relation is true, VOD at HH-polarization is more sensitive to VWC_{bulk} .

5

5.4. SENSITIVITY ANALYSIS OF VOD USING RANDOM FOREST

5.4.1. THE SENSITIVITY OF VOD TO WATER CONTENT OF DIFFERENT VEGETATION COMPONENTS

The Random Forest model is used to elucidate the sensitivity of VOD to water content in stems, leaves, and ears. According to the result from section 5.3, the linear relation between VOD and VWC_{bulk} does not always hold true in the growing season. Therefore, we conduct the sensitivity analysis in the vegetative stage and the reproductive stage respectively in order to see the drivers of VOD in different growth stages. In each period, three RF models are built. The input variables are VWC_{stem} , VWC_{leaf} , and VWC_{ear} and the response variable is VOD at HH, VV and cross-polarization respectively. The result of the Out-of-Bag scores (OOB score) and values of feature importance are displayed in Table 5.5 and Figure 5.8.

BBCH=13~BBCH=59			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.999476	0.999667	0.999810
BBCH=59~BBCH=71			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.949115	0.989968	0.985738

Table 5.5: The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the Input variables: VWC_{stem} , VOD_{stem} , and VOD_{ear} at vegetative and reproductive stage

In the vegetative stage, stem water content and leaf water content are the most important features at both co- and cross-polarization. This is obvious because there is almost no moisture in ear in this period. Although they are equally important, the feature

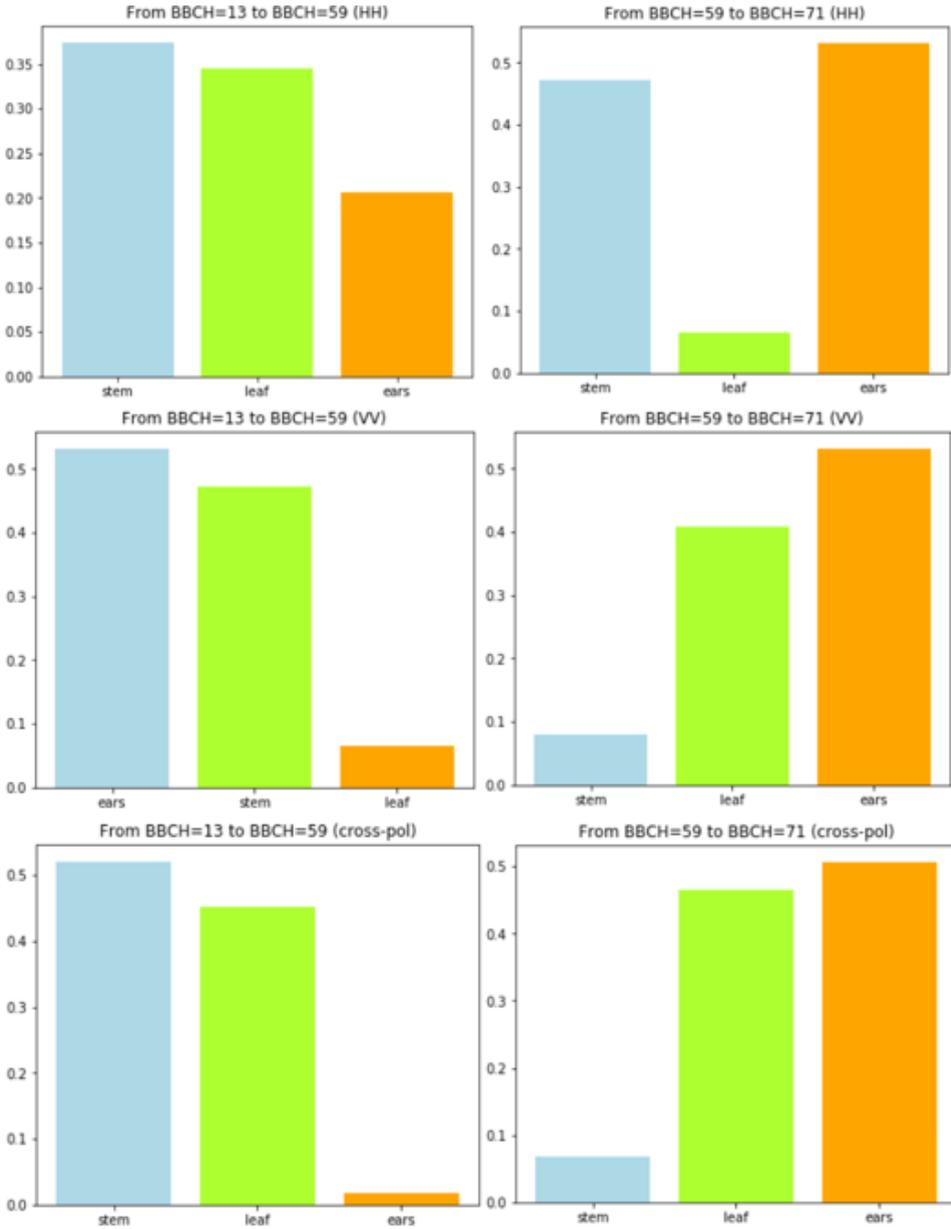


Figure 5.8: Feature importance of RF model for VOD and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

importance of stem water content is slightly larger. However, at HH polarization, the feature importance of ear is unexpectedly large. From Figure 5.4, we can see that unlike VV and cross-polarization, a rapid increase of VOD at HH polarization happens between BBCH=53 to BBCH=63. In this period, ear emerged and the water content gradually increases though the amount is negligible compared with stem and leaf (see in Figure 3.6). However, in this period the water content of leaf is almost stable and the stem water content decreases from BBCH=53 to BBCH=59. From BBCH=53 to BBCH=59, the increasing trend of ear water content coincides with the rapid growth of VOD at HH polarization. This may cause the high value of the feature importance of ear water content at HH polarization.

In the following reproductive stage, the ear water content becomes the most important feature at both co- and cross-polarization. At HH polarization mode, the importance of leaf water content reduces massively, while at VV- and cross-polarization, the importance of stem water content decrease sharply. Here, the RF results also prove that, ear plays an important role in VOD variations.

The OOB scores of all RF models are higher than 0.9 (see in Table 5.5) which demonstrates that these vegetation water content have the ability to explain the variation of VOD. However, the OOB score at the reproductive stage is relatively small compared with the vegetative stage. This may indicate that other variable also shows influence on VOD besides the vegetation water content. It is the variation of the structure of the corn, which reduces the sensitivity of VOD to vegetation water dynamics.

In summary, although the VOD is highly related to VWC_{bulk} , the main driver of VOD can be any of the components of the VWC_{bulk} . In the vegetative stage, VWC_{stem} and VWC_{ear} are undoubtedly the most important driver for VOD variations, because the other components are not emerging or mature yet. Then, in the reproductive stage, VWC_{ear} becomes the most important driver although it is not the main contributor to the VWC_{bulk} at the very beginning. The above situation happens at both co- and cross-polarization modes. However, in the reproductive stage, the second important driver for VOD variation is different at co- and cross-polarization. VWC_{stem} occupy this place at HH polarization whereas at VV and cross-polarization VWC_{leaf} occupy this place.

5.4.2. THE SENSITIVITY OF VOD TO STEM AND LEAF WATER CONTENT AT DIFFERENT HEIGHTS

The WCM assumes that the moisture is distributed evenly in the canopy layer. Although this assumption makes the parameterization of the vegetation layer easier, it is a huge simplification of reality. In Figure 5.9, it clearly shows that the moisture distribution in vegetation canopy is not uniform. The stem water content decreases with height. In the vegetative stage, the stem water content is higher than 0.10 kg/m^2 below 110 cm. In the reproductive stage, the stem water content between 70 to 110 cm becomes relatively dry and most water is concentrated in the lower part of the stem. In general, when the ear emerges and separates from the stem, the water content of stem at and above that ear forming point becomes comparatively dry. As for the leaf water content, before BBCH=30, the leaf water content of every leaf is relatively small. After BBCH=30, leaf 1

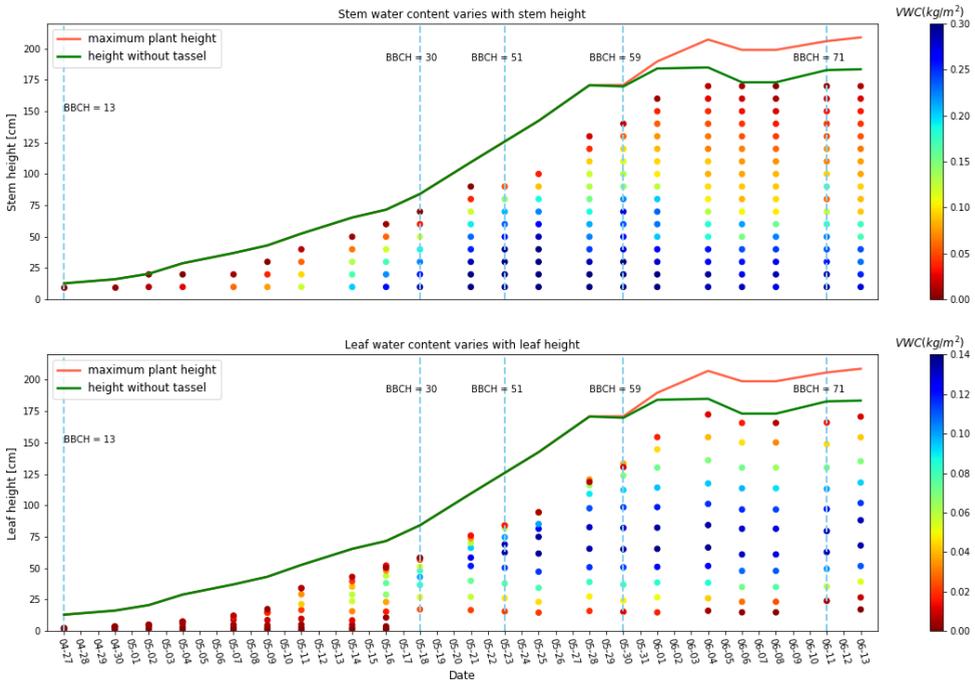


Figure 5.9: Stem(top) and leaf(bottom) water content distribution with height. The top plot shows the stem water content in kg/m^2 . Each dot represents the total water content in all stems of a 10-cm stem section. The y-axis of the each dot stands for the height at the top of that 10-cm stem section. The bottom plot shows the leaf water content in kg/m^2 . Each dot represents the total water content of leaves with their collar height showed in the y-axis.

to leaf 4 die. After BBCH=51, leaf 8 to leaf 11 contain the most part of the moisture. In general, since larger leaves often occur at the midsection of the corn, most of the leaf moisture concentrates at this part as well.

RF modeling is used to unravel the sensitivity of VOD to water content in stem sections, leaves lumps at different heights respectively. We conduct the sensitivity analysis both in the vegetative and the reproductive stage.

STEM

In the vegetative stage, according to section 5.4.1, the stem water content is an very important driver for VOD variation at both co- and cross-polarization. At HH and cross-polarization, stem water content at height between 80 and 110 cm has a feature importance higher than 0.2. At VV polarization, stem water content at height 60-80 cm has feature importance higher than 0.2 (see in Figure 5.10). Thus, HH and cross-polarization are more sensitive to moisture content in the upper canopy compared with VV polarization.

In the reproductive stage, according to section 5.4.1, for both VV and cross-polarization, stem water content is not an important driver for VOD variation anymore. Stem water content has a large influence on VOD variation only at HH polarization. Then from Figure 5.10, we can see that the feature importance ranking at HH polarization is more reasonable. The water content in the stem section between 0 to 80 cm is the most important driver to VOD variation. This section of stem also concentrates most moisture of the stem(see in Figure 5.9). Same as the vegetative stage, HH-polarization is still more sensitive to the higher canopy. Notice that the feature importance of SH60-80 shows a value greater than 1 at cross-polarization, which means the accuracy of the model becomes negative after randomly permuting the values of that feature. This indicates that the change of this feature cannot explain VOD variation. Thus, the water content in 60-80 cm stem section is not important for VOD variation. The OOB scores in the vegetative stage is slightly larger compared with the reproductive stage (see in Table 5.6). This indicates that RF model has better performance at the vegetative stage.

BBCH=13~BBCH=59			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.999076	0.999744	0.999798
BBCH=59~BBCH=71			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.980094	0.954216	0.980536

Table 5.6: The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the input variables: the water content of stem sections between a certain height range at vegetative and reproductive stage

LEAF

In the vegetative stage, according to section 5.3, leaf water content is an important driver for VOD variation at both co- and cross-polarization. At HH polarization, VOD is more

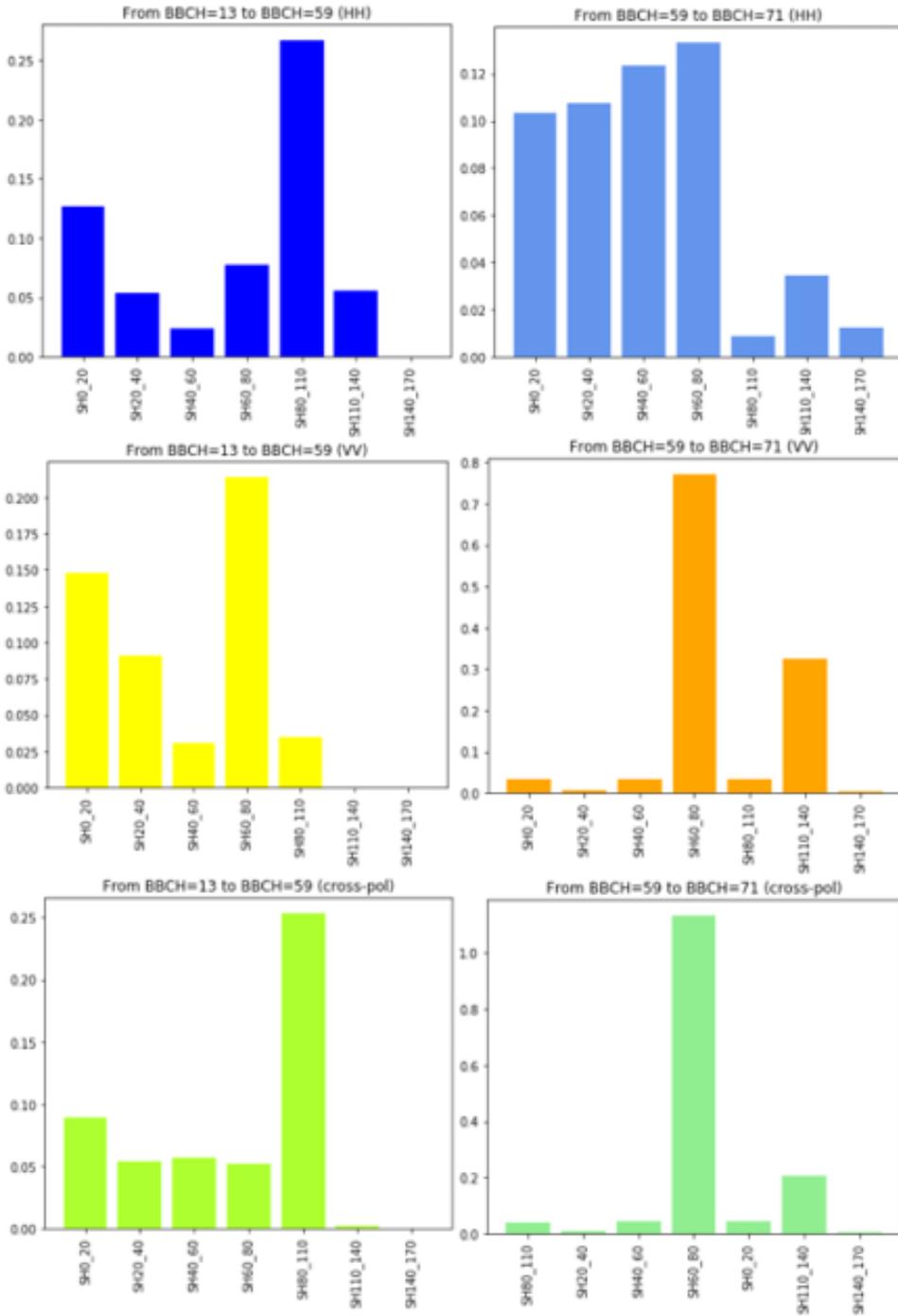


Figure 5.10: Feature importance of RF model for VOD and water content of stem section at different height (Each subplot is the result of one RF model, every bar in the subplots stands feature importance of one input variable(the water content of a stem section between a certain height range) of this RF model. For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively

sensitive to the moisture content in leaf 14 and leaf 15 which locates at the upper canopy. Besides, the water content in leaf 14 and leaf 15 is relatively dry (see in Figure 5.9). The feature ranking of moisture content in leaf 14 and leaf 15 is done by firstly permuting its moisture content then calculating the change in model accuracy. Leaf 14 and leaf 15 locate around 110 cm in the canopy layer, therefore, we can also say VOD is more sensitive to moisture change in this part of the canopy. Besides, in this period, stem water content is also a important driver for VOD variation(see in Figure 5.8). Thus, if we focus on canopy layer around 110 cm, in Figure 5.10, VOD is more sensitive to moisture in stem section from 80 to 110 cm. Thus, the higher feature importance value of moisture content in leaf 14-15 may caused by the VOD sensitivity to stem water content at this height. Next, at VV and cross-polarization, VOD is more sensitive to the moisture content in leaf 8 to leaf 13. This is also where most part of the moisture in the canopy concentrates.

In the following reproductive stage, according to Figure 5.8, leaf water content is not an important driver for VOD variation at HH polarization anymore. At VV and cross-polarization, the moisture content in leaf 5, leaf 6, and leaf 7 is the highest, even higher than 0.8. Then, the moisture content of leaf 8, leaf 9 and leaf 10 is also an important driver but with much smaller feature importance. However, the moisture content in leaf 5-7 is much smaller compared with the moisture content in leaf 8-10. Leaf 5-7 locate between 15 cm and 40 cm in the canopy layer, therefore, we can also say VOD is more sensitive to moisture change in this part of the canopy. From Figure 5.8, the stem is already not the main driver of VOD variation at VV and cross-polarization in this period. The remaining vegetation component that has large influence on VOD variation at VV and cross-polarization is the ear. Therefore, the canopy layer between 15 cm to 40 cm may contain ear. From Table 5.7, OOB scores in the vegetative stage are larger compared with reproductive stage. Thus, RF models have better performance at the vegetative stage.

BBCH=13~BBCH=59			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.999007	0.999770	0.999834
BBCH=59~BBCH=71			
Polarization	<i>HH</i>	<i>VV</i>	<i>cross-pol</i>
OOB score	0.984048	0.954216	0.983554

Table 5.7: The OOB score result of RF models with the response variable: VOD at co- and cross-polarization respectively and the input variables: the water content of leaf lumps between a certain height range at vegetative and reproductive stage

In summary, VOD is indeed more sensitive to the canopy part that contains more moisture. Thus, it can tell us some information about the vertical moisture distribution of the canopy. According to the result, at the vegetative stage, VOD is more sensitive to the stem moisture below 110 cm. At the reproductive stage, VOD is more sensitive to stem moisture below 80 cm. Moisture redistribution between 80 to 110 cm stem section may cause by the ear emergence in this part of the canopy. As for leaf, in both vegetative and

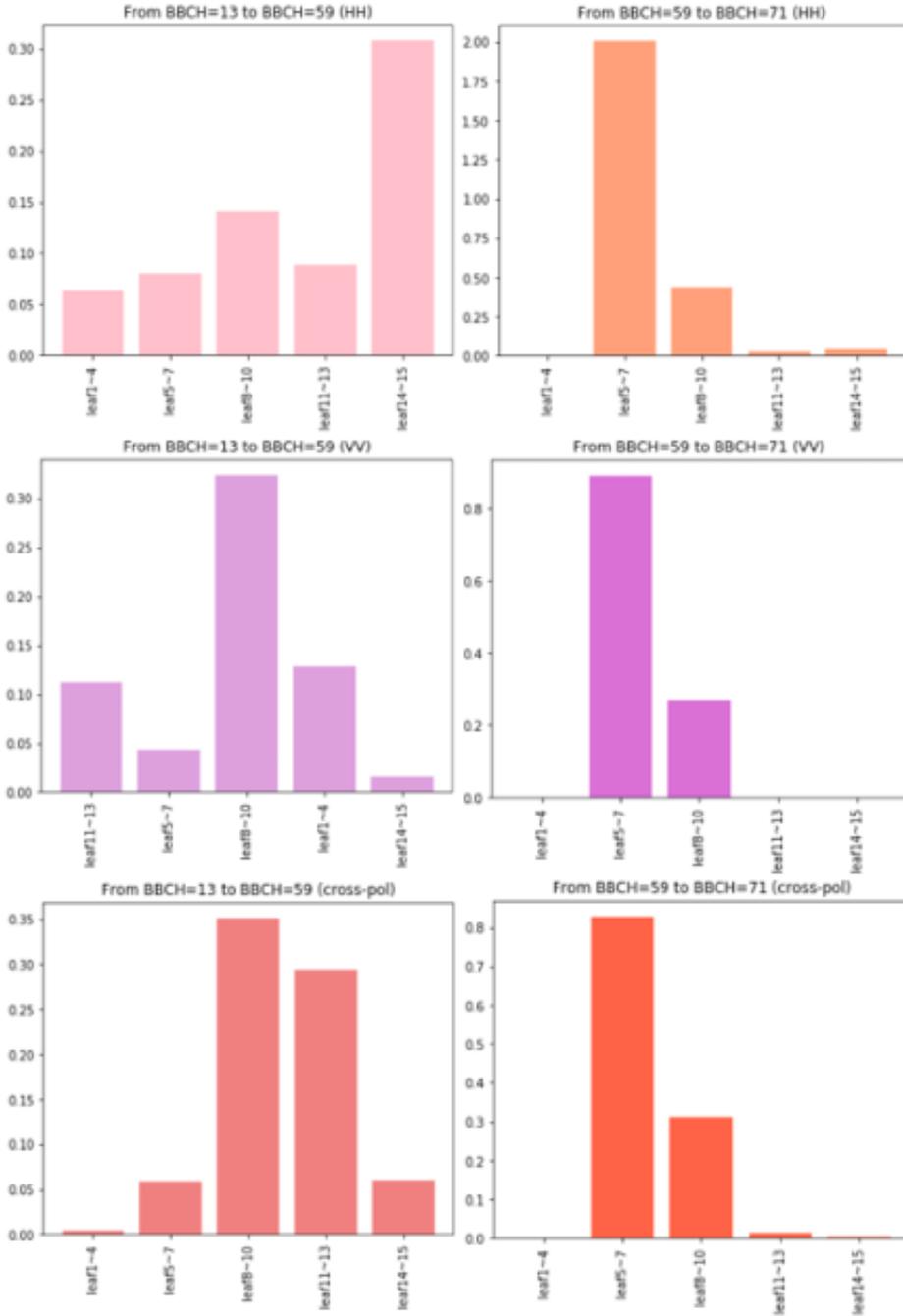


Figure 5.11: Feature importance of RF model for VOD and water content of leaf at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH}, VOD_{VV}, and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively

reproductive stages, VOD is more sensitive to leaf in the middle part of the canopy (no.8, 9, 10). Although VOD shows sensitivity to the canopy layer that contains more moisture, VOD is more sensitive to the higher part of this layer, especially at HH polarization.

5.5. SENSITIVITY ANALYSIS OF CR USING RANDOM FOREST

5.5.1. THE SENSITIVITY OF CR TO WATER CONTENT OF DIFFERENT VEGETATION COMPONENTS

In the vegetative stage, stem water content and leaf water content are the most important features though the leaf water content is slightly larger. Then, in the following reproductive stage, the stem and ear water content are the most important driver for CR variations. The OOB score of the vegetative stage RF model is 0.796312 and the OOB score of the reproductive stage is 0.765552. Although the OOB score is high, it still cannot compare with RF models take VOD as the response variable. From Figure 5.12, in the vegetative stage, similar to VOD as the response variable, stem and leaf water content are the variables best explaining the variability in CR. In the reproductive stage, unlike VOD as the response variable, the stem water content has the largest feature importance. Thus, compared with VOD, CR is less sensitive to ear water content.

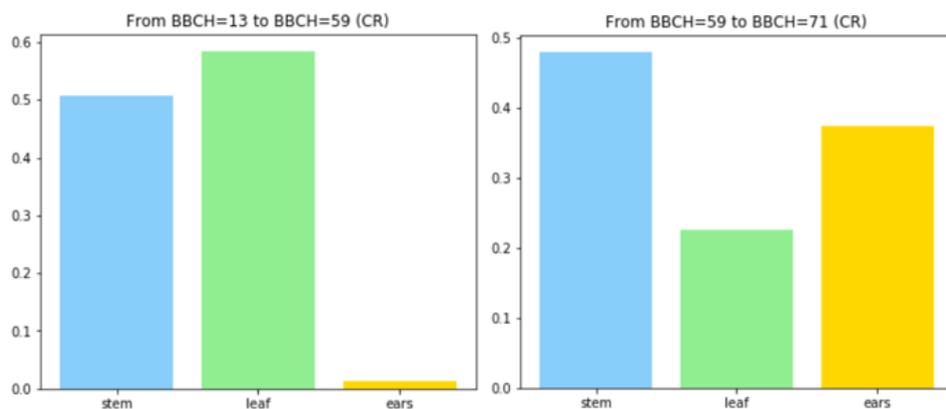


Figure 5.12: Feature importance of RF model for CR and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

5.5.2. THE SENSITIVITY OF CR TO STEM AND LEAF WATER CONTENT AT DIFFERENT HEIGHTS

STEM

In the vegetative stage, according to Figure 5.12, stem water content is important to CR variations. Then, in Figure 5.13, the feature importance of moisture content in the stem section between 0 and 20 cm is much higher than other input variables. Thus, the mois-

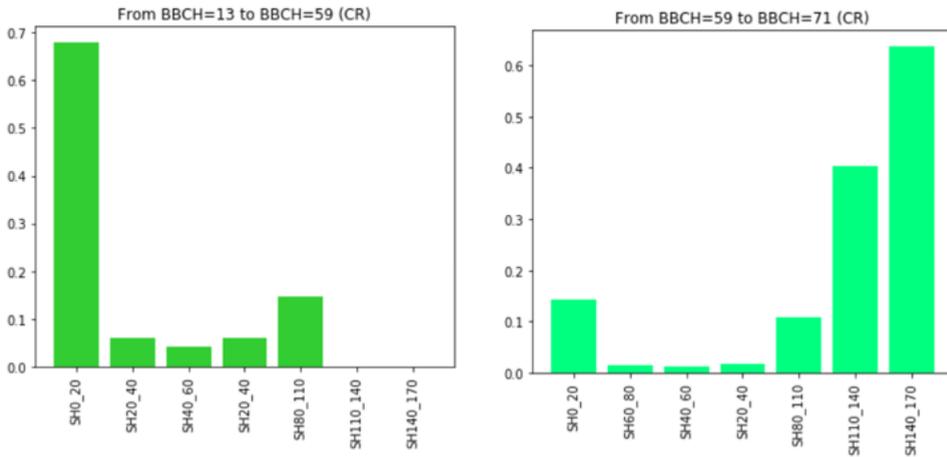


Figure 5.13: Feature importance of RF model for CR and water content of stem vegetation content at different height (Each subplot is the result of one RF model, every bar in the subplots stands for one input variable of this RF model (stem water content at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

ture content at this part of the stem section is the main driver of CR variation. From Figure 5.9, part of moisture concentrates in this stem section.

In the reproductive stage, according to Figure 5.12, stem water content is still important to CR variations. However, unlike the vegetative stage, the larger values of feature importance appear in the upper part (110 cm to 170 cm) of the stem section. From Figure 5.9, the moisture content of this part is relatively small compared with the lower part of the stem. Besides, the canopy layer at this height contains less moisture as well. Thus, this may indicate CR values are more sensitive to moisture in the upper part of the canopy. The OOB score at the vegetative and reproductive stage is 0.795884 and 0.726938 respectively. RF model has better performance at the vegetative stage.

LEAF

In the vegetative stage, according to Figure 5.12, leaf water content is the main driver of CR variations. Then, in Figure 5.14, the moisture content in leaf 5-13 has more effect on the CR changes. The value of feature importance of leaf 5-7 is the highest. Although the moisture content in leaf 5-7 is relatively small, the moisture content of leaf at the beginning of the growing period is small as well. Thus, moisture in leaf 5-7 still plays an important role in the beginning. Then in the reproductive stage, according to Figure 5.12, leaf water content is not an important driver for CR anymore. The OOB score of the vegetative stage and reproductive stage is 0.788487 and 0.695279 respectively.

In summary, the OOB scores of RF models show that CR is sensitive to VWC_{bulk} but VOD can do better. Besides, VOD can do better in providing information about moisture vertically distribution as well. In the vegetative stage, the main driver for CR variation

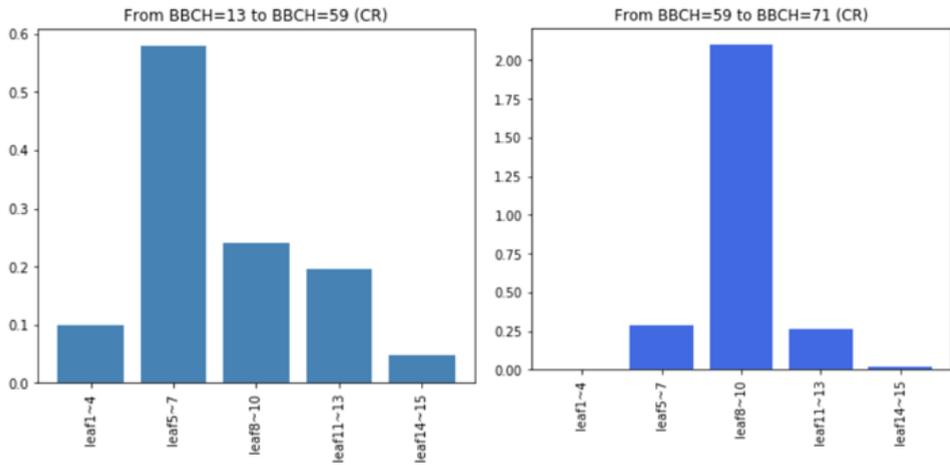


Figure 5.14: Feature importance of RF model for CR and water content of leaf vegetation content at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

5

is also the VWC_{stem} and VWC_{leaf} but VWC_{leaf} is slightly more important. While in the reproductive stage, VWC_{ear} has less effect on CR.

6

CONCLUSION AND RECOMMENDATION

6.1. CONCLUSION

1. ***Can the TU-Wien VOD estimation method be adapted to estimate VOD from backscatter at single-incidence angle?***

In the TU-Wien VOD estimation method, multi-incidence angle backscatter data is required to derive the relation between incidence angle and backscatter ($\sigma^0(\theta)$). Then, the slope and curvature of $\sigma^0(\theta)$ as well as the cross-over angles are used to determine the dry and wet references with the assumption that the effects of vegetation cover are minimized and the variation of backscatter coefficient is mainly caused by the soil moisture variation. In our study, we do not have multi-incidence angle backscatter data. However, we have in-situ soil moisture data. Thus, we applied the linear model between volumetric soil moisture and backscatter with the assumption that the soil moisture is the only variable that causes the backscatter variation. Notice that, when the soil moisture is low and does not vary much in days, the scatter plot of backscatter against soil moisture will look like a cloud of points. From the mathematical point of view, the result can be arbitrary when fitting a line through a cloud of points. Thus, soil moisture data and backscatter data during the long dry-down period need to be excluded. Especially when the dry-down period is almost or even longer than the sliding window length we set.

Apart from that, vegetation water content varies within several days to weeks. Therefore, in the TU-Wien VOD estimation method, it assumes that the short-term variation of backscatter is mainly caused by soil moisture variation. However, vegetation surface water will cause the short-term backscatter variation which is not related to soil moisture dynamics. The effect of surface water on the canopy layer cannot be neglected especially when the vegetation canopy is getting lush and dense. Therefore, we use LAI as well as in-situ photograph of the field to determine when the canopy is thick enough to cover the soil surface and we call this

day as the closure of the canopy. After the closure of the canopy, in order to avoid the influence from vegetation surface water, we exclude the soil moisture as well as the backscatter data when irrigation, precipitation and dew happens.

In summary, if you have soil moisture data, after the adaptation made above, the TU-Wien VOD estimation method can be applied to single-incidence angle backscatter observations. In single-incidence backscatter observations, the variation of sensitivity of backscatter to soil moisture are used to indicate the vegetation growth and senescence. At the beginning of the growing season, vegetation cover is less dense, the backscatter is more sensitive to soil moisture. When the vegetation cover grows thicker, the sensitivity reduces. When the vegetation cover fades, the sensitivity of backscatter to soil moisture recovers. However, this adaptation can only be used when the auxiliary soil moisture data is available.

2. *What is the relation between VOD and VWC_{bulk} in the entire growing season? Does the linear relation hold true in the entire growing season?*

The linear relation between VOD and VWC_{bulk} does not always hold true during the entire growing season of the corn at both co- and cross-polarization. This linear relation can be applied during the vegetative stage when stem and water content are the main driver of the VOD variation. And the slope factor of this linear relation can be set as a constant value. When the heading stage (BBCH=51) begins, although the value of VWC_{ear} is small, the importance of VWC_{ear} cannot be neglected. When VWC_{ear} becomes the dominant driver of VOD variability in the reproductive stage, the previous derived linear relation between VWC_{bulk} and VOD is not valid anymore.

3. *Bulk vegetation water content comprises many components. Here we only focus on stem, leaf, and ears. Thus, which vegetation components are the main drivers for VOD variation? Will this result change in different growth stages? Will this result also change in different polarization?*

VWC_{bulk} has many components. VOD is highly related to VWC_{bulk} but the main driver of VOD variability can be any of these components. In the vegetative stage, the canopy layer is not very dense. Thus, the estimated VOD values are relatively small and the total backscatter is dominated by the surface scattering from the soil surface at both co- and cross-polarization. During this period, the main driver of the VOD variation is the VWC_{stem} and VWC_{leaf} . In the following reproductive stage with a denser canopy, the estimated VOD increases and contribution of volume scattering from the vegetation cover becomes dominated in the total backscatter at both co- and cross-polarization. This time, the most important driver of VOD variation is the VWC_{ear} .

In summary, the VWC_{stem} and VWC_{leaf} have more influence on VOD variation in the vegetative stage while the VWC_{ear} affects VOD more during the reproductive stage. This happens at both co- and cross-polarization. However, during the reproductive stage, the second main driver of VOD variation can be different at different polarization. At HH-polarization, the VWC_{stem} is the second driver while at VV and cross-polarization VWC_{leaf} is the second driver.

Besides, from the OOB scores of the RF models in different growing stages, VOD is better related to VWC_{stem} , VWC_{leaf} and VWC_{ear} in the vegetative stage. In the reproductive stage, except for the vegetation water content, other factors such as the structure of the plant begin to affect the VOD variation and reduce the sensitivity of VOD to vegetation water content. This happens at both co- and cross-polarization.

4. ***The moisture inside the canopy is not uniformly distributed. The backscatter will be influenced by the vertical distribution of moisture. Is the VOD more sensitive to the particular part of the canopy layer that contains more moisture?***

In the vegetative stage, stem and leaf water content are important for VOD variations. Therefore, we can continue to find which section is the main driver for VOD variation. In this period, VOD shows more sensitivity to the leaf lump (located in the middle of the canopy) and the stem section (below 110 cm) with more moisture content. In the reproductive, stem water content is more important for HH pol. while leaf water content is more important for VV and cross-polarization. Thus, VOD at HH polarization can better describe the vertical distribution of stem water content while VOD at VV and cross-polarization can better describe the vertical distribution of leaf water content. VOD is more sensitive to the stem section below 80 cm and more sensitive to leaf lump in the middle part of the canopy.

In summary, VOD is indeed more sensitive to the canopy layer which contains more moisture. However, VOD is less sensitive to the lower part of this layer than to the upper part of this layer, especially at HH polarization.

5. ***Apart from VOD, Cross-Ratio is also a commonly used indicator for vegetation water content in the microwave frequency region. Besides, NDVI in the spectrum region is also an often used vegetation water content indicator. What is the difference between VOD, CR, and NDVI? Can they all sufficiently describe the vegetation water content within the canopy?***

Because of the cloud influence in the case study area, only a limited number of NDVI values can be obtained during the growing season. Besides, the NDVI time series smoothing methods need to be used for reducing noise in NDVI time series because the signal received by sensors is affected by the atmosphere, cloud, geometric misregistration, anisotropic reflectance effects, electronic errors. However, during the growing season the NDVI data is far from sufficient. It is hard to draw valuable conclusions from this time series. Therefore, NDVI is more suitable for analysis in a longer period like yearly data. In this case, compared with radar data (CR and VOD), NDVI is less reliable.

Although CR data is sufficient, the soil effect is not effectively removed from the CR time series. Compared with VOD, firstly, CR is less sensitive to ear water content at the reproductive stage. Secondly, after tassel fully emerged and separate, CR can response to small changes in bulk vegetation water content. Besides, although the OOB scores of all RF models take CR as the response variable is higher than 0.6, it still cannot compare with OOB scores of RF models take VOD as the response variable (all higher than 0.9). This indicates that compared with CR, VOD is better related to vegetation water dynamic.

6.2. RECOMMENDATION

We made adaptations to the TU-Wien VOD estimation method because we have auxiliary soil moisture data. Although this adaptation is feasible, this cannot be used except for experimental settings. Because, in most cases, we cannot get the soil moisture data from the area we monitored.

Another recommendation is in the application of the random forest machine learning technique. The additional function called feature importance is used to rank the driver of the VOD variation. In our study, the input variables are not independent of each other. Therefore, it is better to use Mean Decrease Accuracy as the ranking method instead of the Mean Decrease in Gini. The results based on MDA and MDG can be totally different. Sometimes the feature ranking results from MDA show that a group of leaves or a section of the stem may have relatively small moisture content whereas its importance to VOD variation is extremely large. When a situation like this happens, it does not always mean errors, we need to check whether there are other vegetation components in the canopy layer that is analyzed. Although we only focus on estimating feature importance of leaf lumps and stem sections in the RF models, the way we use is actually estimate the influence of the moisture of one part of the canopy.

Besides, the random forest model in this study is not recommended for VOD prediction. We only use the feature importance function of RF to identify the drivers of the VOD variation. However, if in the future the sample size is large enough (for example, samples in different years), then the RF model can be build and used for prediction.

REFERENCES

- [1] M. Vreugdenhil, W. A. Dorigo, W. Wagner, R. A. De Jeu, S. Hahn, and M. J. Van Marle, "Analyzing the vegetation parameterization in the TU-Wien ASCAT soil moisture retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3513–3531, 2016.
- [2] K. Bremer, "Capturing the plant-water dynamics of corn: A study on the stomatal conductance and the leaf water potential of corn during the growing season," 2019.
- [3] Y. Huang, J. P. Walker, Y. Gao, X. Wu, and A. Moneris, "Estimation of vegetation water content from the radar vegetation index at L-band," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 981–989, 2015.
- [4] E. R. Hunt Jr, L. Li, M. T. Yilmaz, and T. J. Jackson, "Comparison of vegetation water contents derived from shortwave-infrared and passive-microwave sensors over central Iowa," *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2376–2383, 2011.
- [5] A. G. Konings, K. Rao, and S. C. Steele-Dunne, "Macro to micro: microwave remote sensing of plant water content for physiology and ecology," *New Phytologist*, 2019.
- [6] Z. Cai, P. Jönsson, H. Jin, and L. Eklundh, "Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data," *Remote Sensing*, vol. 9, no. 12, p. 1271, 2017.
- [7] R. Liu, J. Wen, X. Wang, Z. Wang, Z. Li, Y. Xie, L. Zhu, and D. Li, "Derivation of vegetation optical depth and water content in the source region of the yellow river using the FY-3B microwave data," *Remote Sensing*, vol. 11, no. 13, p. 1536, 2019.
- [8] T. N. Carlson and D. A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote sensing of Environment*, vol. 62, no. 3, pp. 241–252, 1997.
- [9] Y. Y. Liu, A. I. van Dijk, M. F. McCabe, J. P. Evans, and R. A. de Jeu, "Global vegetation biomass change (1988–2008) and attribution to environmental and human drivers," *Global Ecology and Biogeography*, vol. 22, no. 6, pp. 692–705, 2013.
- [10] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of Sensors*, vol. 2017, 2017.
- [11] A. G. Konings, M. Piles, N. Das, and D. Entekhabi, "L-band vegetation optical depth and effective scattering albedo estimation from SMAP," *Remote Sensing of Environment*, vol. 198, pp. 460–470, 2017.
- [12] T. van Emmerik, S. C. Steele-Dunne, J. Judge, and N. van de Giesen, "Impact of diurnal variation in vegetation water content on radar backscatter from maize during water stress," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3855–3869, 2015.

- [13] Y. Li, J. Shi, and T. Zhao, "Effective vegetation optical depth retrieval using microwave vegetation indices from WindSat data for short vegetation," *Journal of Applied Remote Sensing*, vol. 9, no. 1, p. 096003, 2015.
- [14] M. Owe, R. de Jeu, and T. Holmes, "Multisensor historical climatology of satellite-derived global land surface moisture," *Journal of Geophysical Research: Earth Surface*, vol. 113, no. F1, 2008.
- [15] R. Parinussa, M. Yilmaz, M. Anderson, C. Hain, and R. De Jeu, "An intercomparison of remotely sensed soil moisture products at various spatial scales over the Iberian Peninsula," *Hydrological processes*, vol. 28, no. 18, pp. 4865–4876, 2014.
- [16] E. Santi, "An application of the SFIM technique to enhance the spatial resolution of spaceborne microwave radiometers," *International Journal of Remote Sensing*, vol. 31, no. 9, pp. 2419–2428, 2010.
- [17] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P.-W. Liu, and K. Papathanassiou, "Radar remote sensing of agricultural canopies: A review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2249–2273, 2017.
- [18] T. Jackson, "Laboratory evaluation of a field-portable dielectric/soil-moisture probe," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 2, pp. 241–245, 1990.
- [19] A. A. Van de Griend and J.-P. Wigneron, "The b-factor as a function of frequency and canopy type at H-polarization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 786–794, 2004.
- [20] W. G. Rees, *Physical principles of remote sensing*. Cambridge University Press, 2013.
- [21] V. Naeimi, *Model improvements and error characterization for global ERS and METOP scatterometer soil moisture data*. na, 2009.
- [22] J. Cihlar and F. T. Ulaby, "Dielectric properties of soils as a function of moisture content," 1974.
- [23] T. J. Schmugge, "Remote sensing of soil moisture: Recent advances," *IEEE Transactions on Geoscience and Remote Sensing*, no. 3, pp. 336–344, 1983.
- [24] R. A. de Jeu, "Retrieval of land surface parameters using passive microwave remote sensing," 2003.
- [25] F. Ulaby, R. Moore, and A. Fung, "Microwave remote sensing: Active and passive: Volume III. from theory to applications. Norwood," 1986.
- [26] M. A. El-Rayes and F. T. Ulaby, "Microwave dielectric spectrum of vegetation-Part I: Experimental observations," *IEEE Transactions on Geoscience and Remote Sensing*, no. 5, pp. 541–549, 1987.

- [27] F. T. Ulaby and R. Jedlicka, "Microwave dielectric properties of plant materials," *IEEE Transactions on Geoscience and Remote Sensing*, no. 4, pp. 406–415, 1984.
- [28] F. T. Ulaby and M. A. El-Rayes, "Microwave dielectric spectrum of vegetation-Part II: Dual-dispersion model," *IEEE Transactions on Geoscience and Remote Sensing*, no. 5, pp. 550–557, 1987.
- [29] F. Ulaby, R. Moore, and A. Fung, *Microwave remote sensing: Active and passive. Volume 2-Radar remote sensing and surface scattering and emission theory*, 1982.
- [30] E. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio science*, vol. 13, no. 2, pp. 357–364, 1978.
- [31] X. Blaes, P. Defourny, U. Wegmuller, A. Della Vecchia, L. Guerriero, and P. Ferrazzoli, "C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model," *IEEE transactions on geoscience and remote sensing*, vol. 44, no. 4, pp. 791–800, 2006.
- [32] F. Mattia, G. Satalino, L. Dente, and G. Pasquariello, "Using a priori information to improve soil moisture retrieval from ENVISAT ASAR AP data in semiarid regions," *IEEE transactions on geoscience and remote sensing*, vol. 44, no. 4, pp. 900–912, 2006.
- [33] N. J. Rodríguez-Fernández, A. Mialon, S. Mermoz, A. Bouvet, P. Richaume, A. Al Bitar, A. Al-Yaari, M. Brandt, T. Kaminski, T. Le Toan *et al.*, "An evaluation of SMOS L-band vegetation optical depth (L-VOD) data sets: High sensitivity of L-VOD to above-ground biomass in Africa," *Biogeosciences*, vol. 15, no. 14, pp. 4627–4645, 2018.
- [34] J. Shi, T. Jackson, J. Tao, J. Du, R. Bindlish, L. Lu, and K. Chen, "Microwave vegetation indices for short vegetation covers from satellite passive microwave sensor AMSR-E," *Remote sensing of environment*, vol. 112, no. 12, pp. 4285–4300, 2008.
- [35] Y. Y. Liu, A. I. Van Dijk, R. A. De Jeu, J. G. Canadell, M. F. McCabe, J. P. Evans, and G. Wang, "Recent reversal in loss of global terrestrial biomass," *Nature Climate Change*, vol. 5, no. 5, p. 470, 2015.
- [36] M. O. Jones, L. A. Jones, J. S. Kimball, and K. C. McDonald, "Satellite passive microwave remote sensing for monitoring global land surface phenology," *Remote Sensing of Environment*, vol. 115, no. 4, pp. 1102–1114, 2011.
- [37] B. A. Bouman and D. H. Hoekman, "Multi-temporal, multi-frequency radar measurements of agricultural crops during the Agriscatt-88 campaign in The Netherlands," *Title REMOTE SENSING*, vol. 14, no. 8, pp. 1595–1614, 1993.
- [38] M. A. Karam, A. K. Fung, R. H. Lang, and N. S. Chauhan, "A microwave scattering model for layered vegetation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 767–784, 1992.

- [39] W. Wagner, G. Lemoine, and H. Rott, "A method for estimating soil moisture from ERS scatterometer and soil data," *Remote sensing of environment*, vol. 70, no. 2, pp. 191–207, 1999.
- [40] R. J. Black, *Florida climate data*. Citeseer, 1993.
- [41] P. C. Vermunt, S. Khabbazan, S. C. Steele-Dunne, and J. Judge, "Understanding diurnal and seasonal variations of L-band backscatter over a corn canopy using intensive field measurements. Manuscript in preparation," 2019.
- [42] T. Bongiovanni, P.-W. Liu, K. Nagarajan, D. Preston, P. Rush, T. H. van Emmerik, R. Terwilleger, A. Monsivais-Huertero, J. Judge, S. Steele-Dunne *et al.*, "Field Observations during the Eleventh Microwave Water and Energy Balance Experiment (MicroWEX-11): from April 25, 2012, through December 6," 2015.
- [43] F. T. Ulaby, G. A. Bradley, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part II-vegetation-covered soil," *IEEE Transactions on Geoscience Electronics*, vol. 17, no. 2, pp. 33–40, 1979.
- [44] M. C. Dobson and F. Ulaby, "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part III-soil tension," *IEEE Transactions on Geoscience and Remote Sensing*, no. 1, pp. 51–61, 1981.
- [45] U. Meier, *Growth stages of mono-and dicotyledonous plants*. Blackwell Wissenschafts-Verlag, 1997.
- [46] *Crop Identification and BBCH Staging Manual: SMAP-12 Field Campaign*, Earth Observation Research Branch Team Agriculture and Agri-Food Canada, Oklahoma State University, 2011.
- [47] R. B. D'Agostino, *Goodness-of-fit-techniques*. CRC press, 1986, vol. 68.
- [48] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied linear statistical models*. Irwin Chicago, 1996, vol. 4.
- [49] P. K. Srivastava, P. O'Neill, M. Cosh, R. Lang, and A. Joseph, "Evaluation of radar vegetation indices for vegetation water content estimation using data from a ground-based SMAP simulator," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1296–1299.
- [50] A. Veloso, S. Mermoz, A. Bouvet, T. Le Toan, M. Planells, J.-F. Dejoux, and E. Ceschia, "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sensing of Environment*, vol. 199, pp. 415–426, 2017.
- [51] S. C. Brown, S. Quegan, K. Morrison, J. C. Bennett, and G. Cookmartin, "High-resolution measurements of scattering in wheat canopies-implications for crop parameter retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 7, pp. 1602–1610, 2003.

- [52] S. Brown, G. Cookmartin, K. Morrison, A. McDonald, S. Quegan, C. Anderson, R. Cordey, and P. Dampney, "Wheat scattering mechanisms observed in near-field radar imagery compared with results from a radiative transfer model," in *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*, vol. 7. IEEE, 2000, pp. 2933–2935.
- [53] M. Vreugdenhil, W. Wagner, B. Bauer-Marschallinger, I. Pfeil, I. Teubner, C. Rüdiger, and P. Strauss, "Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study," *Remote Sensing*, vol. 10, no. 9, p. 1396, 2018.
- [54] S. Sruthi and M. M. Aslam, "Agricultural drought analysis using the NDVI and land surface temperature data; a case study of Raichur district," *Aquatic Procedia*, vol. 4, pp. 1258–1264, 2015.
- [55] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, "Using the satellite-derived NDVI to assess ecological responses to environmental change," *Trends in ecology & evolution*, vol. 20, no. 9, pp. 503–510, 2005.
- [56] Sentinel-2 MSI: MultiSpectral Instrument, Level-1C. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2
- [57] Cloud masks. [Online]. Available: <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks>
- [58] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of LIDAR data and building object detection in urban areas," *ISPRS journal of photogrammetry and remote sensing*, vol. 87, pp. 152–165, 2014.
- [59] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.
- [60] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [61] T. Mo, B. Choudhury, T. Schmugge, J. Wang, and T. Jackson, "A model for microwave emission from vegetation-covered fields," *Journal of Geophysical Research: Oceans*, vol. 87, no. C13, pp. 11 229–11 237, 1982.
- [62] X. Jiao, H. McNairn, J. Shang, E. Pattey, J. Liu, and C. Champagne, "The sensitivity of RADARSAT-2 polarimetric sar data to corn and soybean leaf area index," *Canadian Journal of Remote Sensing*, vol. 37, no. 1, pp. 69–81, 2011.

A

NOISE IN BACKSCATTER OBSERVATIONS AT HH POLARIZATION

The vegetation cover is less dense at the beginning of the growing season, however, the influence of soil moisture change does not show in the backscatter at HH polarization (the data in red circle in Figure A.1). The backscatter observations at VV and cross-polarization shows the influence from soil moisture (the data in green circle in Figure A.1).

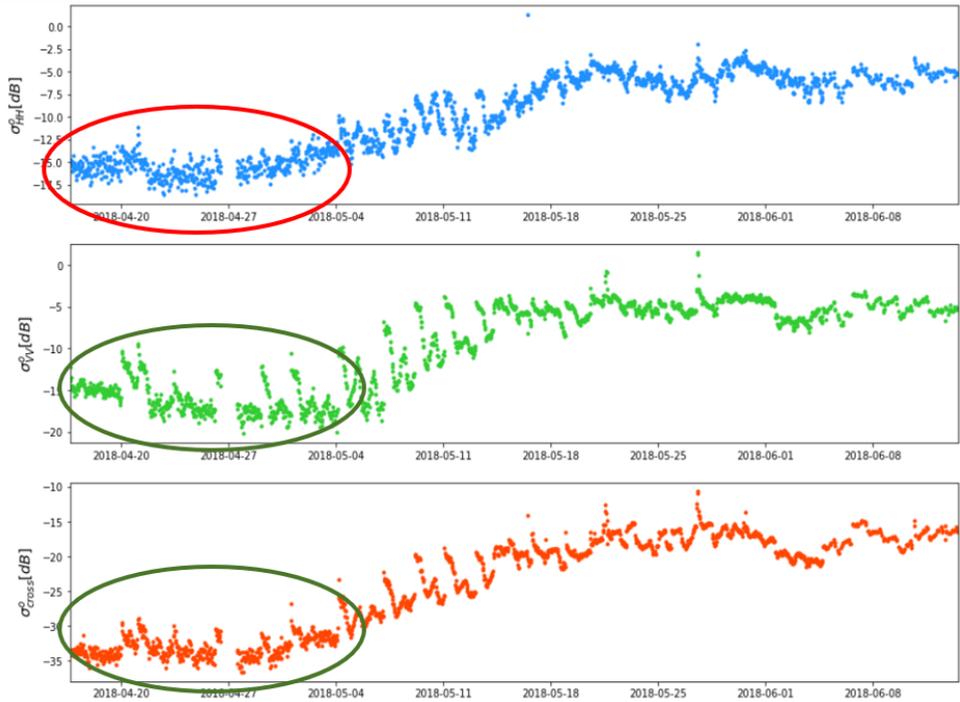


Figure A.1: The backscatter observations at co- and cross- polarization. The backscatter data inside the red circle in HH polarization is where the noise appears)

B

GROWTH STAGES FOR CORN

The table in Figure B.1 contains the detail information about different BBCH growth stages of corn. This table is from the Crop Identification and BBCH Staging Manual: SMAP-12 Field Campaign. Then, the table in Figure B.2 tells information of corn growth stage conversions. Source from: <https://naicc.org/wp/wp-content/uploads/2017/01/Industry-Recommendations-for-Implementing-EFSA-Guidance-Document-for-Conducting-GM.pdf>

Growth stage	Code	Description
0: Germination	00	Dry seed (caryopsis)
	01	Beginning of seed imbibition
	03	Seed imbibition complete
	05	Radicle emerged from caryopsis
	06	Radicle elongated, root hairs and/or side roots visible
	07	Coleptile emerged from caryopsis
	09	Emergence: coleoptile penetrates soil surface (cracking stage)
1: Leaf development ^{1, 2}	10	First leaf through coleoptile
	11	First leaf unfolded
	12	2 leaves unfolded
	13	3 leaves unfolded
	1 .	Stages continuous till . . .
	19	9 or more leaves unfolded
3: Stem elongation	30	Beginning of stem elongation
	31	First node detectable
	32	2 nodes detectable
	33	3 nodes detectable
	3 .	Stages continuous till . . .
	39	9 or more nodes detectable ³
5: Inflorescence emergence, heading	51	Beginning of tassel emergence: tassel detectable at top of stem
	53	Tip of tassel visible
	55	Middle of tassel emergence: middle of tassel begins to separate
	59	End of tassel emergence: tassel fully emerged and separated
6: Flowering, anthesis	61	Male: stamens in middle of tassel visible
		Female: tip of ear emerging from leaf sheath
	63	Male: beginning of pollen shedding
		Female: tips of stigma visible
	65	Male: upper and lower parts of tassel in flower
		Female: stigma fully emerged
67	Male: flowering completed Female: stigma drying	
69	End of flowering: stigma completely dry	
7: Development of fruit	71	Beginning of grain development: kernels at blister stage, about 16% dry matter
	73	Early milk
	75	Kernels in middle of cob yellowish-white (variety-dependent), content milky, about 40% dry matter
	79	Nearly all kernels have reached final size
8: Ripening	83	Early dough: kernel content soft, about 45% dry matter
	85	Dough stage: kernels yellowish to yellow (variety dependent), about 55% dry matter
	87	Physiological maturity: black dot/layer visible at base of kernels, about 60% dry matter
	89	Fully ripe: kernels hard and shiny, about 65% dry matter
9: Senescence	97	Plant dead and collapsing
	99	Harvested product

Figure B.1: Table with the description of different BBCH stages for corn (Superscript 1: A leaf may be described as unfolded when its ligule is visible or the tip of next leaf is visible; Superscript 2: Tillering or stem elongation may occur earlier than stage 19, in this case continue with principal growth stage 3. Superscript 3: In maize, tassel emergence may occur earlier, in this case continue with principal growth stage 5)

Conversions from V/R to BBCH developmental stages of corn

Developmental Stage	BBCH	BBCH DESCRIPTION
Corn VE	09	Emergence
Corn V1	11	First leaf unfolded
Corn V2	12	2 leaves unfolded
Corn V3	13	3 leaves unfolded
Corn V4	14	4 leaves unfolded
Corn V5	15	5 leaves unfolded
Corn V6	16	6 leaves unfolded
Corn V7	17	7 leaves unfolded
Corn V8	18	8 leaves unfolded
Corn V9-Vn	19	9 or more leaves unfolded
Corn VT	55	Middle of tassel emergence: middle of tassel begins to separate
Corn R1	63	Male: beginning of pollen shed Female: tips of stigmata visible
Corn R1	65	Male: upper and lower parts of tassel in flower; Female: stigmata fully emerged
Corn R2	71	Kernels at blister stage
Corn R3	75	Kernels in middle of cob yellowish-white, content milky
Corn R4	85	Dough stage: kernels yellowish to yellow
Corn R5	No equivalent BBCH code in reference	
Corn R6	87	Physiological maturity: black dot/layer visible at base of kernels
	89	Fully ripe: kernels hard and shiny
Corn - several days after R6	99	Harvested seed

Figure B.2: Corn growth stage conversions

C

SMOOTHED SLOPE AND INTERCEPT TIME SERIES

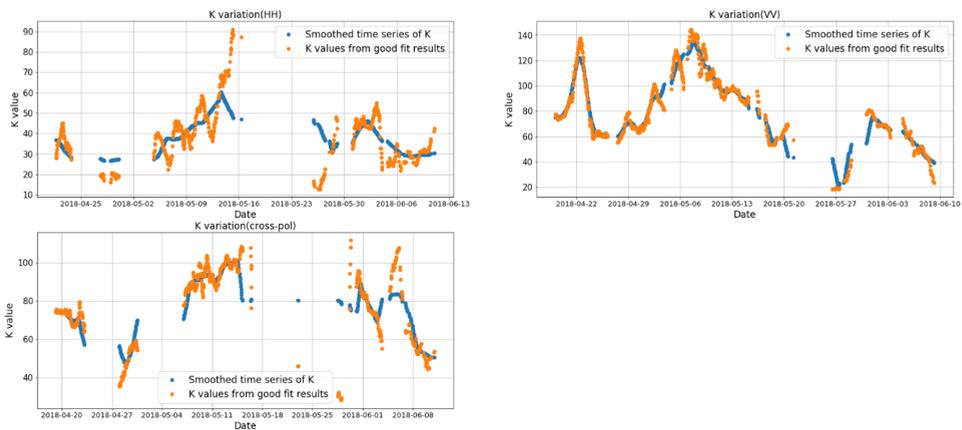


Figure C.1: The slope time series derived by the OLS with R^2 higher than 0.5 (orange dots). And a smoother slope time series using a moving average of 5-day (blue dots) at different polarization mode(HH polarization(top left), VV polarization(top right) and cross-polarization(bottom left))

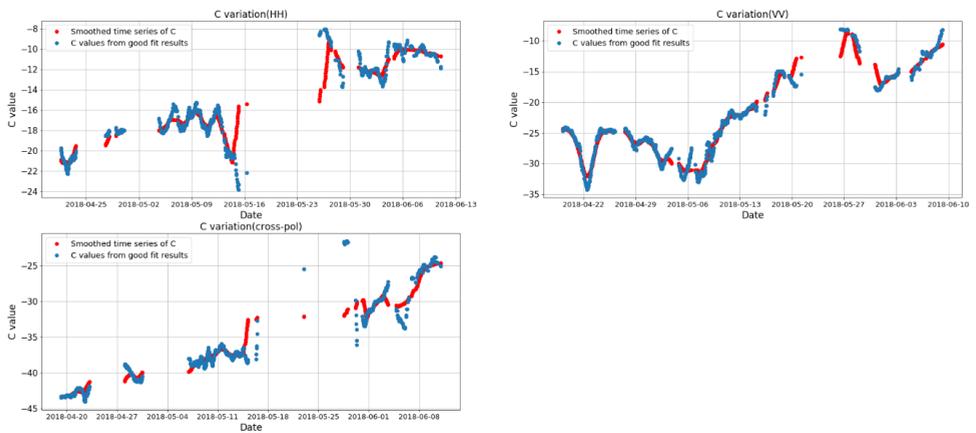


Figure C.2: The intercept time series derived by the OLS with R^2 higher than 0.5 (blue dots). And a smoother slope time series using a moving average of 5-day (red dots) at different polarization mode(HH polarization(top left), VV polarization(top right) and cross-polarization(bottom left))

D

MEAN DECREASE IN GINI

These plots show the feature importance calculated with the Mean Decrease in Gini method. Because the input variables of RF models in this study are not independent from each other. Thus, MDG is not suitable for feature ranking. The result calculated by MDG is different compared with the results calculated with MDA.

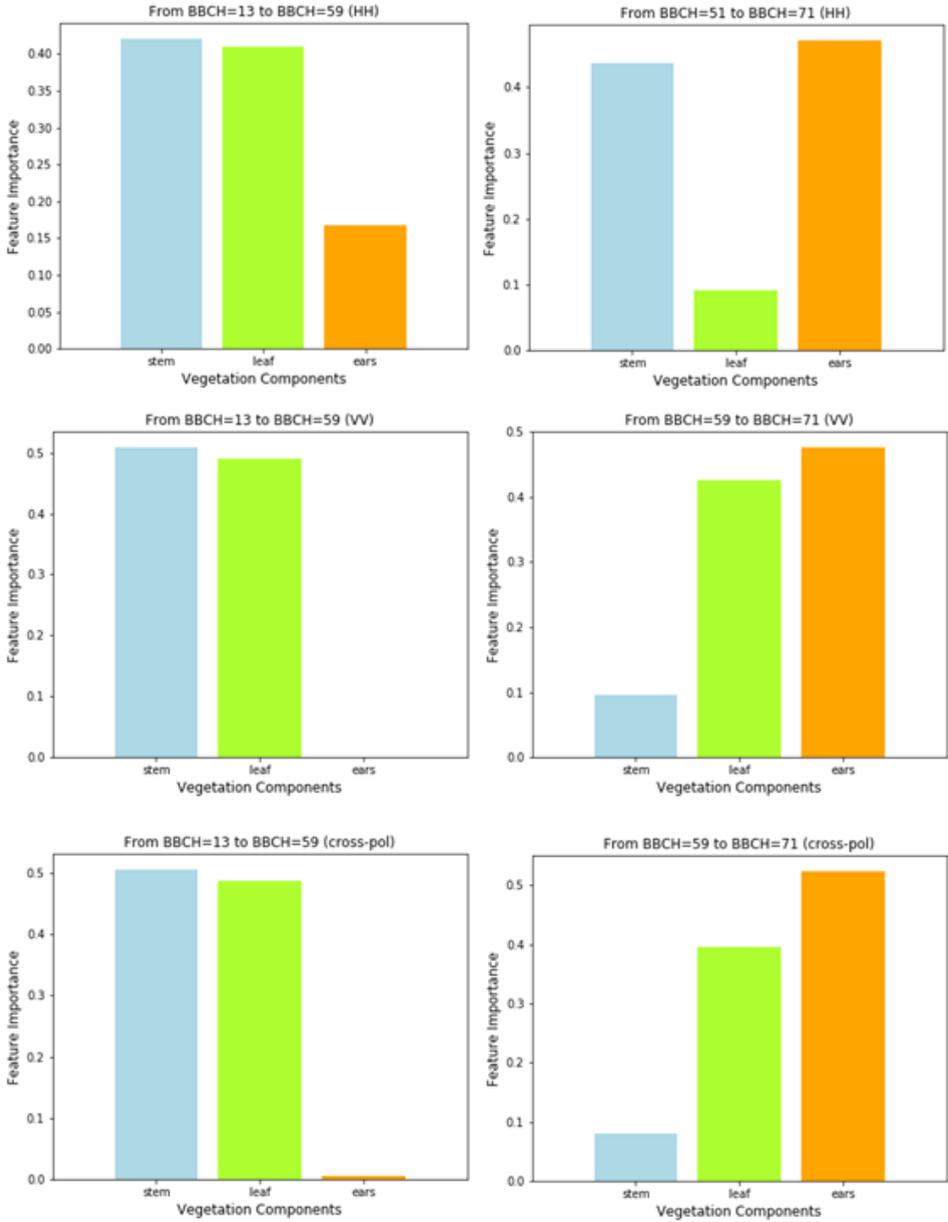


Figure D.1: Feature importance of RF model for VOD and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

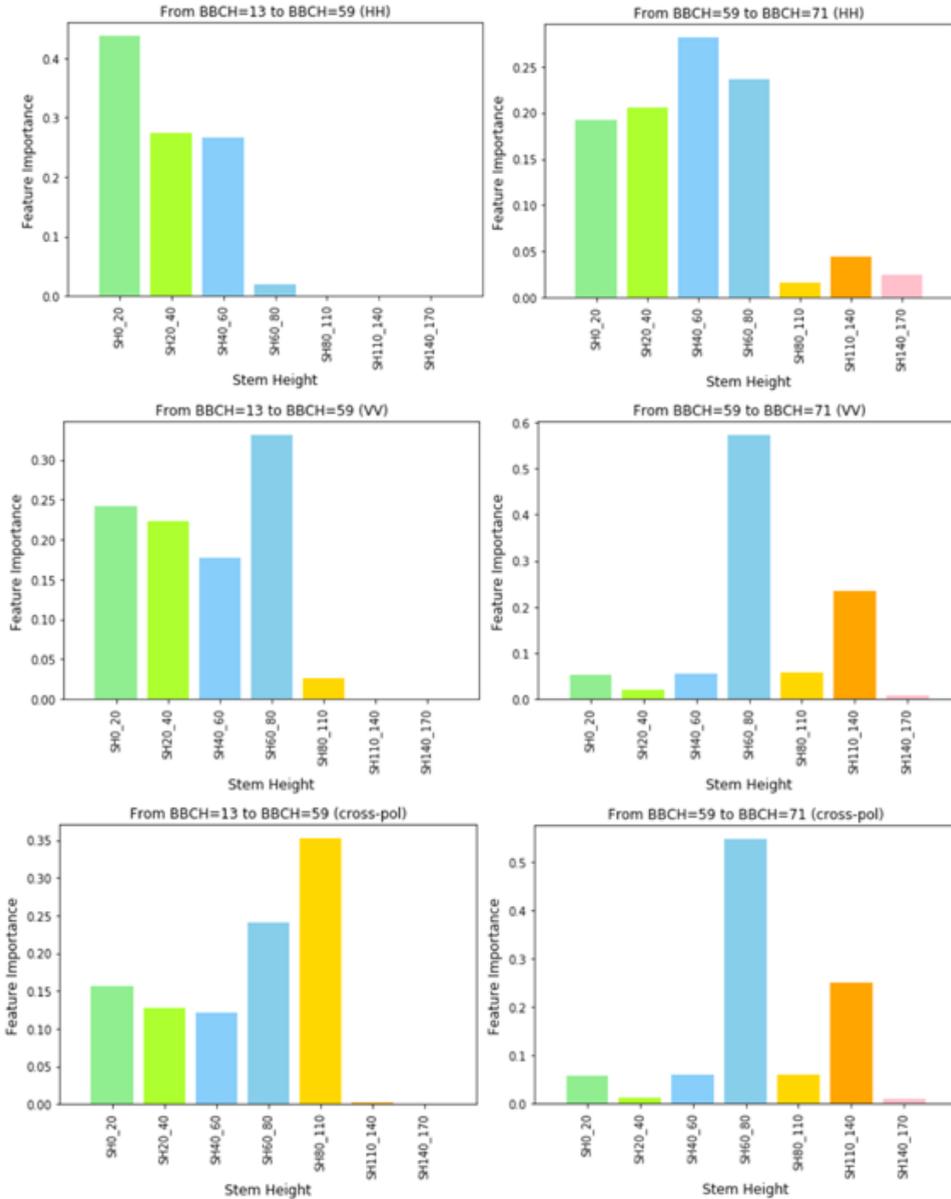


Figure D.2: Feature importance of RF model for VOD and water content of stem at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of a stem section between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively

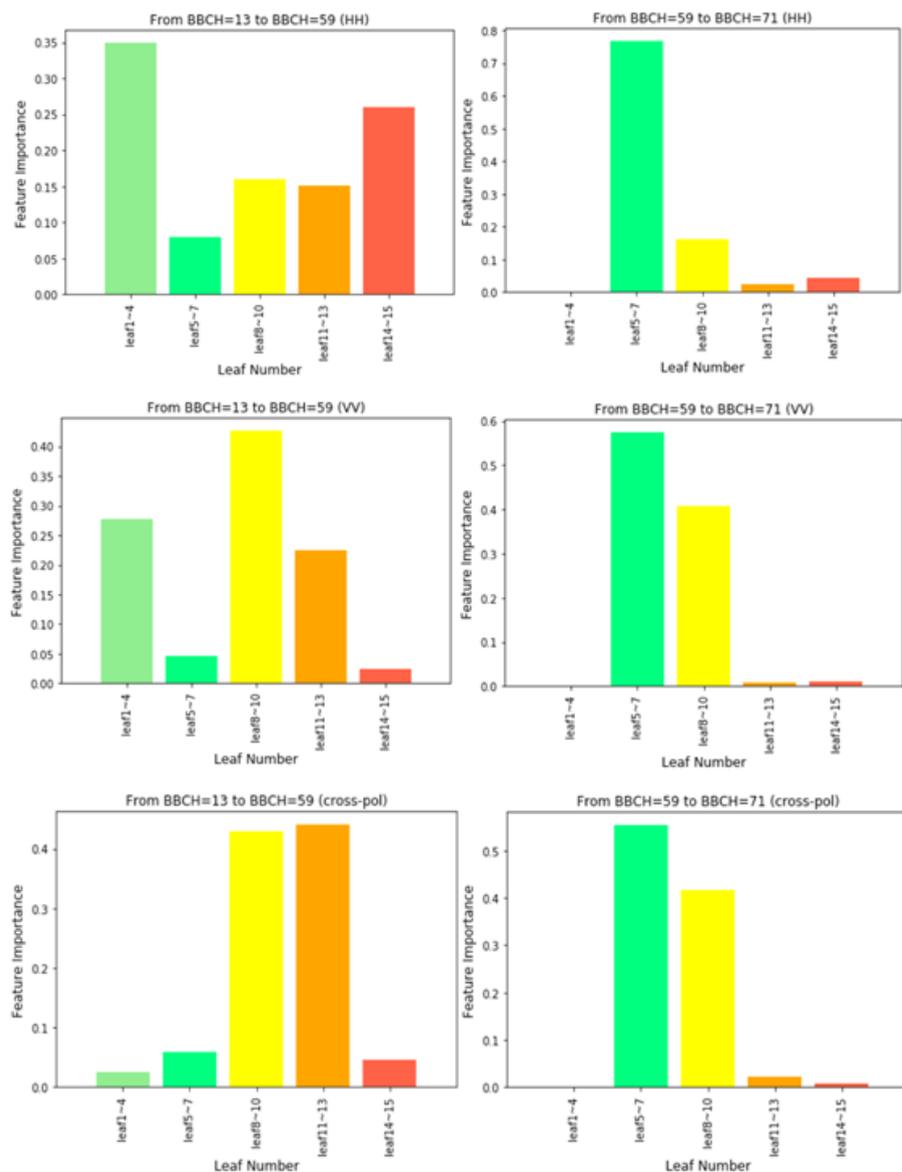
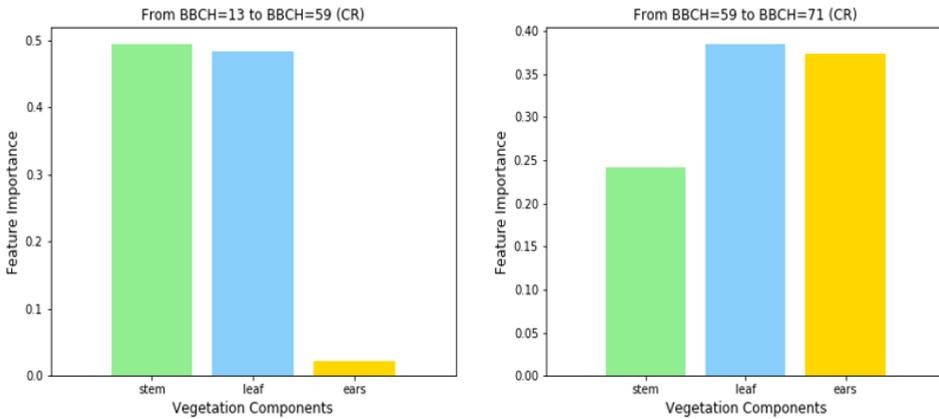


Figure D.3: Feature importance of RF model for VOD and water content of leaf at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves between a certain height range). For the three rows of this figure, the response variable for each row is VOD_{HH} , VOD_{VV} , and VOD_{Cross} (from top to bottom), respectively. As for the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively



D

Figure D.4: Feature importance of RF model for CR and water content of different vegetation components (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem, leaf and ear water content). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

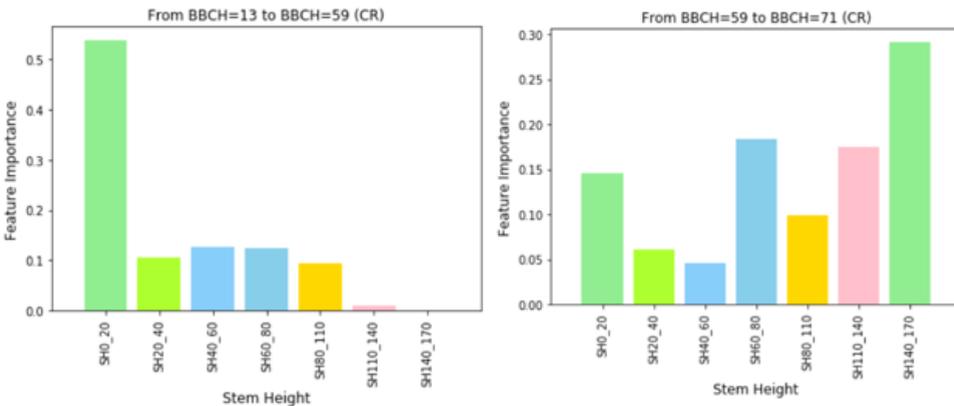


Figure D.5: Feature importance of RF model for CR and water content of stem vegetation content at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (stem water content at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

D

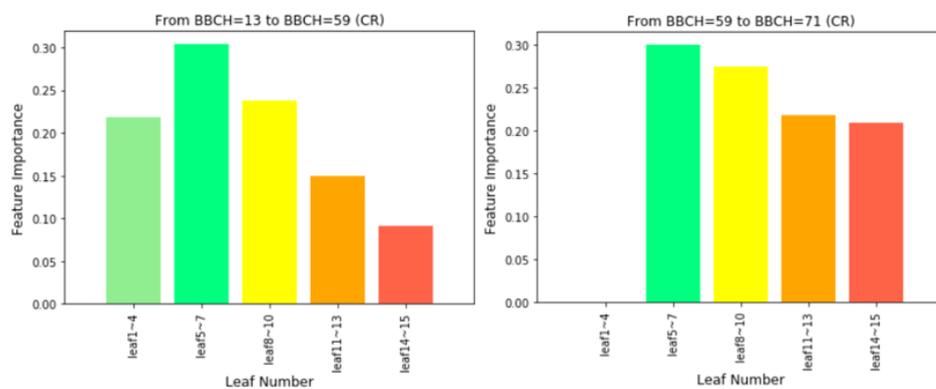


Figure D.6: Feature importance of RF model for CR and water content of leaf vegetation content at different height (Each subplot is the result of one RF model, every bar in the subplots stands for the feature importance of one input variable of this RF model (water content of several leaves at a certain height range). For the two columns of this figure, the growing stage for each column is BBCH=13 to BBCH=59, BBCH=59 to BBCH=71 (from left to right), respectively)

E

LINEAR RELATION BETWEEN VOD AND VWC IN CERTAIN GROWTH STAGES

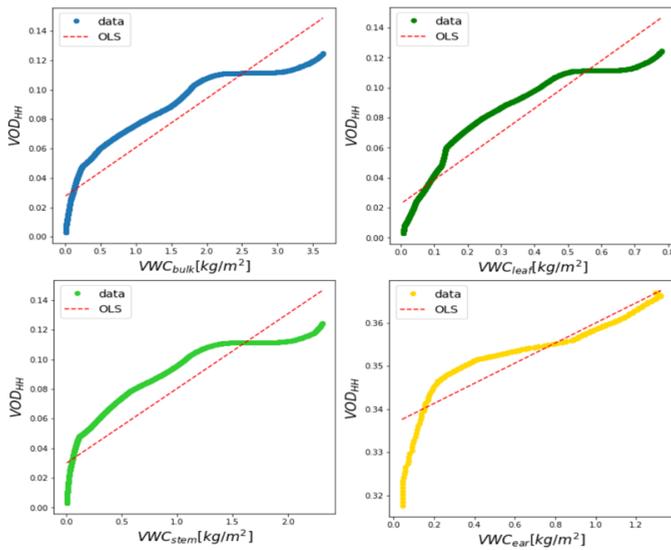


Figure E.1: Linear relation between VOD_{HH} and VWC_{bulk} (top left), VOD_{HH} and VWC_{leaf} (top right), VOD_{HH} and VWC_{stem} (bottom left), VOD_{HH} and VWC_{ear} (bottom right)

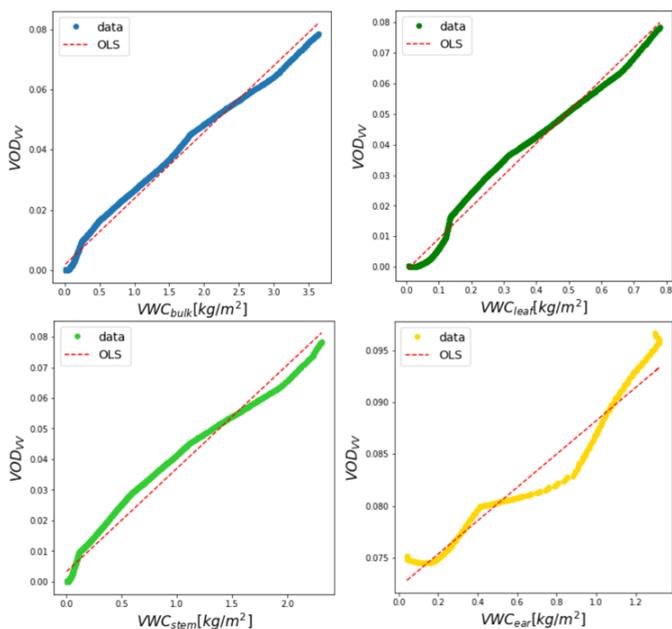


Figure E.2: Linear relation between VOD_{VV} and WVC_{bulk} (top left), VOD_{VV} and WVC_{leaf} (top right), VOD_{VV} and WVC_{stem} (bottom left), VOD_{VV} and WVC_{ear} (bottom right)

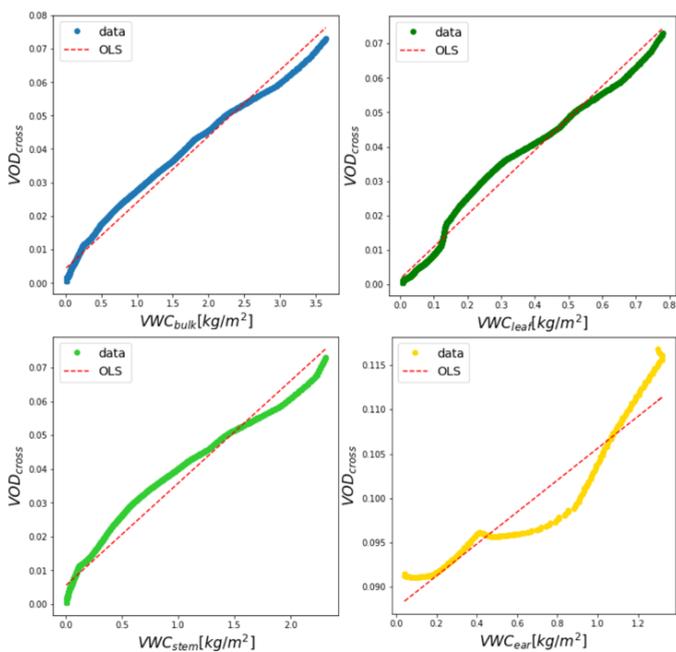


Figure E.3: Linear relation between VOD_{cross} and WVC_{bulk} (top left), VOD_{cross} and WVC_{leaf} (top right), VOD_{cross} and WVC_{stem} (bottom left), VOD_{cross} and WVC_{ear} (bottom right)

ACKNOWLEDGEMENTS

First of all, I would like to thank Professor Susan Steele-Dunne, my daily supervisor and the chair of my committee. You always find time for listening to my questions and helping me out when roadblocks appear in the course of performing the research. I have had the moment that just completely did not know how to express myself clearly because of the nervousness, you always full of patience and encourage me. I have learned many valuable lessons from you, not only the knowledge about radar and the way to structure a thesis but also pay attention to detail and strive for excellence. I could not have imagined having a better supervisor for my Master study. Many thanks for your support, patience, and encouragement throughout my studies. Without your precious guidance and persistent help it would not be possible to finish this thesis.

My appreciation also goes to Jasmeet Judge at the University of Florida, who provides me all the radar data and soil moisture data and answers my questions about the data patiently. And Mariette Vreugdenhil at TU-Wien, who gives me advice on vegetation optical depth and random forest. Besides, I would like to thank all the members in my committee, for giving me the valuable feedbacks and helping me improve the draft.

Saeed, you have offered me a lot of help since I started the additional thesis. You spent hours helping me debug my code, you gave me references that helped me learn basic knowledge about microwave remote sensing, you also took me to the field experiment site to actually see what happens in-situ. Your hard-working and unassuming attitude gives me a positive influence. Paul, thank you for helping me with the vegetation data and patiently answering all my questions about it. It is really nice to work with you two. Also, I want to thank Haoyang Lv for introducing me to the Google Earth Engine, I save a lot of time because of your help. My sincere thank goes to all of you, it is because of your wholehearted support that the research can go smoothly.

Besides, I want to thank all my friends here. Two years ago, I arrived in the Netherlands alone, excited but nervous. Now, I'm lucky to be surrounded by friends. Jessie, Fei Liang, Xixi Tong, and Wouter, you always listen to my complaints and cheer me up when I was anxious and down. Mengya Wei, Chang Gao, Qiaodan Liu, thank you for providing me with unfailing support throughout my years of study. Dongyu Gao, I moved THREE times within two months, thanks for helping me drag all the heavy luggage from one place to another. Besides, thanks for providing me with continuous encouragement throughout the process of writing this thesis. I want all of you to know that I'm appreciated all the times we were able to spend together. Also, my dear old friends, Lluvia Jing and Xiaochen Zhang, although we are in three different countries and far away from each other right now, the conversation between us is always a delight and warms me. Thank you for being with me since high school.

Last but not least, my dearest mom and dad, who can think that your little girl is now in a place 8000 kilometers away from home, expressing her gratitude to you in her Master's thesis with a language that not even taught by you? My loving mom and dad,

thanks for your moral encouragement, financial assistance as well as spirituality support in every path I take.

