MULTI-SENSORY DATA FUSION FOR SOIL MOISTURE CONTENT ESTIMATION

BY

Oualid Yahia

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

Surrey Space Centre

Department of Electrical and Electronic Engineering

Faculty of Engineering and Physical Sciences

University of Surrey

Guildford, Surrey, GU2 7XH, UK

December 2019

Supervised by:

Dr. Raffaella Guida

Declaration of originality

I confirm that the project dissertation I am submitting is entirely my own work and that any material used from other sources has been clearly identified and properly acknowledged and referenced. In submitting this final version of my report to the JISC anti-plagiarism software resource, I confirm that my work does not contravene the university regulations on plagiarism as described in the Student Handbook. In so doing I also acknowledge that I may be held to account for any particular instances of uncited work detected by the JISC anti-plagiarism software, or as may be found by the project examiner or project organiser. I also understand that if an allegation of plagiarism is upheld via an Academic Misconduct Hearing, then I may forfeit any credit for this module, or a more severe penalty may be agreed.

Thesis Title

Multi-sensory data fusion for soil moisture content estimation.

Author Name

Oualid Yahia

Author Signature　　　　　　　　　　　　　　　　　　　　　　　Date: 25/11/2020

Supervisor's name: Dr Raffaella Guida

Word Count

Number of Pages: 129

Number of Words: 31103

ACKNOWLEDGEMENTS

There has been a time when I thought I would never be writing this part of the thesis, yet here I am, writing it. I want to express my sincerest gratitude for anyone who helped the completion of this achievement in any way, shape, or form.

I also want to express my most profound appreciation for my primary supervisor Dr Raffaella Guida for her continuous technical and emotional support throughout the time of my PhD, for her patience, motivation, and immense knowledge, she helped me became the researcher that I am today, and for that, I am eternally grateful. I would also like to thank my secondary supervisor Dr Pasquale IERVOLINO too, for his guidance and encouragement especially in the confirmation exam when it felt all would be lost, his office was always open, and he always had the right advice to give.

I want to express my love and deepest affection for my lovely wife Asma, who stood beside me even in the worse circumstances. She remained patient despite us being worlds apart, and she kept me focused when I lost sight of my objectives, thank you for being the best part of life.

I would also like to thank my family for their endless support, my mother for believing in me, and my father for doing all the leg work for me to come to the UK, their encouragement helped me to believe in myself enough to leave for another country and pursue goals I thought to be unattainable. I especially thank my brother Amjed for being so patient with helping me collect the ground truth I desperately needed.

I could not forget writing this acknowledgement without mentioning the immense support I received from my friends. Thank you Lotfi, Alla, Zaki, Chawki, and Asma, you are the best friends a person can have.

I would like to express my sincerest gratitude for Alexandra Brady, for listening when I needed to talk, for offering unwavering support and for her undying devotion to her profession. You are a godsend, and I am forever in your debt.

Finally, I would like to thank The Algerian Space Agency and Surrey Satellite Technology Limited for sponsoring my PhD.

ABSTRACT

Soil moisture content (SMC) is an important parameter in many fields, especially in agricultural practices. That is the reason that an accurate retrieval of this parameter is of the utmost importance. Point-based measurements of soil moisture while accurate, are expensive in terms of time and effort, not to mention that their inability to depict spatial variability of SMC accurately on a large scale. Soil moisture retrieval methods using remote sensing technologies show great promise but suffer from numerous limitations. To minimize the effects of those limitations, a novel decision level data fusion algorithm for SMC estimation is proposed in this research. Initially, individual estimations are determined from 3 different methodologies; the inversion of Empirically Adapted Integral Equation Model (EA-IEM) which is semi-empirically calibrated using a parameter *Lopt* for Sentinel-1, the Perpendicular Drought Index (PDI), and Temperature Vegetation Dryness Index (TVDI) for LANDSAT-8. Then, three feature level fusions using novel combinations of salient features extracted from each of the method mentioned above are performed using an Artificial Neural Network (ANN). The latter is characterized by the modification of its performance function from absolute error to Root Mean Square Error. Finally, all estimations including the feature level fusions estimation are fused at the decision level using a novel weights-based estimation, which is implemented through a novel Matlab code. The performance of the proposed system is validated and tested using measurements collected from three study areas, an agricultural field in Blackwell farms, Guildford, United Kingdom, and two different agricultural fields in Sidi Rached, Tipasa, Algeria. Those measurements consisted of SMC level, and surface roughness parameters which were extracted using a newly designed laser profilometre. The proposed SMC estimation system produces stronger correlations and lower RMSE values than any individual SMC estimation in the order of at least 0.38%, 1.4%, and 1.09% for Blackwell farms, Sidi Rached 1 and Sidi Rached 2 datasets respectively.

Keywords: soil moisture content; remote sensing; data fusion; empirically adapted integral equation model; Sentinel-1; perpendicular drought index; temperature vegetation dryness index; Landsat-8; feature level fusion; artificial neural network; decision level fusion; Blackwell farms; Sidi Rached.

CONTENTS

CONTENTS VI
LIST OF TABLES IX
LIST OF FIGURES X
LIST OF ABBREVIATIONS XII
LIST OF SYMBOLS XVI
1. INTRODUCTION 1
1.1 BACKGROUND 1
1.2 AIM AND OBJECTIVES 3
1.3 NOVELTY AND CONTRIBUTIONS 4
1.4 PUBLICATION LIST 5
1.5 OUTLINE OF THE THESIS 5
2. LITERATURE REVIEW 7
2.1 INTRODUCTION 7
2.2 SOIL MOISTURE CONTENT 7
2.3 SOIL MOISTURE CONTENT RETRIEVAL USING REMOTE SENSING 11
2.3.1 Soil Moisture Retrieval Using Operational Estimation by Remote Sensing ... 13
2.3.2 Soil Moisture Content Effects on Active Microwave Sensing 14
2.3.2.1 IEM Theory 16
2.3.2.2 Limitations of SMC Estimation Through IEM inversion 20
2.4 SOIL MOISTURE RETRIEVAL USING MULTISPECTRAL AND THERMAL REMOTE SENSORS 21
2.4.1 Perpendicular Drought Index 22
2.4.1.1 Concept 22

2.4.1.2 Limitations of SMC Estimation through PDI 25

2.4.2 Temperature Vegetation Dryness Index 25

2.4.2.1 Vegetation Indices 25

2.4.2.2 TVDI Concept 27

2.4.2.3 Limitations of SMC Estimation Through TVDI 29

2.5 DATA FUSION TECHNIQUES 30

2.6 MULTI-SENSORY DATA FUSION FOR SOIL MOISTURE CONTENT ESTIMATION 31

2.7 CONCLUSION 38

3. SOIL MOISTURE CONTENT ESTIMATION SYSTEM – CONCEPTUAL DESIGN 39

3.1 INTRODUCTION 39

3.2 MOTIVATIONS BEHIND USING DATA FUSION 39

3.3 THE UPDATED VERSION OF THE IEM 42

3.4 THE PROPOSED SOIL MOISTURE CONTENT ESTIMATION SYSTEM 43

3.4.1 Pre-processing 45

3.4.2 PDI and TVDI Determinations 45

3.4.3 Feature Level Fusion 48

3.4.4 Fusion Centre 53

3.5 CONCLUSION 56

4. GENERATING TEST DATASETS 57

4.1 INTRODUCTION 57

4.2 STUDY AREA DETAILS 57

4.3 EARTH OBSERVATION DATA 62

4.3.1 Sentinel-1 62

4.3.2 Landsat-8 63

4.4 GROUND TRUTH MEASUREMENTS 63

4.4.1 ML3 Theta Soil Moisture Probe *65*
4.4.2 Profilometre *67*
4.5 CONCLUSION 69
5. TESTING AND EVALUATION 63
5.1 INTRODUCTION 70
5.2 RESULTS AND ANALYSIS 70
5.2.1 Blackwell Farms *71*
5.2.2 Sidi Rached 1 .... *74*
5.2.3 Sidi Rached 2 .... *78*
5.3 REMARKS AND DISCUSSIONS 82
5.4 COMPARISON OF THE FUSION CENTRE TO METHODOLOGIES IN LITERATURE 83
5.5 CONCLUSION 87
6. CONCLUSIONS AND FUTURE WORK 88
6.1 SUMMARY 88
6.2 ACHIEVEMENTS AND MAIN NOVELTY CONTRIBUTIONS 89
6.3 FUTURE WORK 91
7. REFERENCES 93

LIST OF TABLES

TABLE 1. VEGETATION INDICES. ADAPTED FROM (CANDIAGO, REMONDINO ET AL. 2015). 26
TABLE 2. COMPARISON OF THE DIFFERENT USED METHODS OF SMC RETRIEVAL. 40
TABLE 3. FEATURE LEVEL FUSION JOINT VECTORS. 49
TABLE 4. COMPARISON BETWEEN DIFFERENT FEATURE LEVEL ESTIMATORS APPLIED IN 3 STUDY AREAS. 50
TABLE 5. THE INPUT FEATURE VECTOR FOR ALL OF THE USED ARTIFICIAL NEURAL NETWORKS. 52
TABLE 6. DETAILS OF STUDY AREAS. 58
TABLE 7. SENTINEL-1 ACQUISITION DETAILS 63
TABLE 8. TECHNICAL SPECIFICATIONS OF LANDSAT-8 64
TABLE 9. THE VALUES OF THE MEASURED SURFACE ROUGHNESS PARAMETERS. 68
TABLE 10. THE RESULTS OF EACH OF THE USED ESTIMATION METHODS IN THE BLACKWELL FARMS DATASET. 73
TABLE 11. THE RESULTS OF EACH OF THE USED ESTIMATION METHODS IN THE SIDI RACHED 1 DATASET. 76
TABLE 12. THE RESULTS OF EACH OF THE USED ESTIMATION METHODS IN THE SIDI RACHED 2 DATASET. 80
TABLE 13. COMPARISON BETWEEN THE RESULTS ATTAINED BY RELEVANT STUDIES AND THE PROPOSED SMC ESTIMATION SYSTEM. 84

LIST OF FIGURES

FIGURE 1. THE SATURATED AND UNSATURATED SOIL ZONES. ADAPTED FROM (PENG, LOEW ET AL. 2017). 8
FIGURE 2. USDA SOIL TEXTURE CLASSIFICATION (KELLOGG 1993). 9
FIGURE 3. RELATIONSHIP BETWEEN SOIL MOISTURE LEVELS AND SOIL TEXTURE CLASSES (ZOTARELLI, DUKES ET AL. 2010).... 10
FIGURE 4. THE ELECTROMAGNETIC SPECTRUM (SHIPPERT 2003). 12
FIGURE 5. EVOLUTION OF THE SPECTRAL SIGNATURES BEHAVIOUR DEPENDING ON THE VOLUMETRIC SOIL MOISTURE CONTENT RANGING BETWEEN 0 AND 0.48 $M^3 \cdot M^{-3}$ (OLTRA-CARRI, BAUP ET AL. 2015). 12
FIGURE 6. DIELECTRIC CONSTANT AS A FUNCTION OF SOIL MOISTURE (HALLIKAINEN, ULABY ET AL. 1985).... 15
(2) 15
FIGURE 7. GEOMETRY OF RAYLEIGH CRITERION. ADAPTED FROM (HAJNSEK, PAPATHANASSIOU 2005). 18
FIGURE 8. SCHEMATIC DIAGRAM OF AN MLP. ADAPTED FROM (HECHT-NIELSEN 1988). ... 19
FIGURE 9. NIR/RED TRIANGULAR FEATURE SPACE. ADAPTED FROM (GAO, ZHONGLING, XU ET AL. 2013). 23
FIGURE 10. DEFINITION OF THE *PDI*. ADAPTED FROM (GHULAM, QIN, AND ZHAN 2007). 24
FIGURE 11. PRESCRIPTION MAPS USING VEGETATION INDICES VALUES (CANDIAGO, REMONDINO ET AL. 2015) 27
FIGURE 12. DEFINITION OF THE *TVDI* IN THE *LST/VI* FEATURE SPACE. ADAPTED FROM (SANDHOLT, RASMUSSEN ET AL. 2002, LAMBIN, EHRLICH 1996). 28
FIGURE 13. THE PROPOSED SOIL MOISTURE CONTENT ESTIMATION SYSTEM.... 44
FIGURE 14. SOIL LINES OF THE STUDY AREAS; BLACKWELL FARMS (A), SIDI RACHED 1 (B), SIDI RACHED 2 (C).... 46
FIGURE 15. TRIANGLE *LST/NDVI* FEATURE SPACE OF THE STUDY AREAS; BLACKWELL FARMS (A), SIDI RACHED 1 (B), SIDI RACHED 2 (C). 47
FIGURE 16. THE STRUCTURE OF A FEED-FORWARD NEURAL NETWORK. ADAPTED FROM (ASCE TASK COMMITTEE ON APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY 2000). 51
FIGURE 17. FLOWCHART OF THE FUSION CENTRE. 54

FIGURE 18. IMAGE FROM SENTINEL-1 DATASET OF THE STUDY AREA BLACKWELL FARMS REPRESENTED BY THE GREEN RECTANGLE (10X11 PIXELS). 59
FIGURE 19. IMAGE FROM LANDSAT-8 DATASET OF THE STUDY AREA BLACKWELL FARMS REPRESENTED BY THE GREEN RECTANGLE (10X11 PIXELS). 59
FIGURE 20. IMAGE FROM SENTINEL-1 DATASET OF THE STUDY AREA SIDI RACHED 1 REPRESENTED BY THE GREEN RECTANGLE (6 X 19 PIXELS). 60
FIGURE 21. IMAGE FROM LANDSAT-8 DATASET OF THE STUDY AREA SIDI RACHED 1 REPRESENTED BY THE GREEN RECTANGLE (6 X 19 PIXELS). 60
FIGURE 22. IMAGE FROM SENTINEL-1 DATASET OF THE STUDY AREA SIDI RACHED 2 REPRESENTED BY THE GREEN RECTANGLE (6 X 6 PIXELS). 61
FIGURE 23. IMAGE FROM LANDSAT-8 DATASET OF THE STUDY AREA SIDI RACHED 2 REPRESENTED BY THE GREEN RECTANGLE (6 X 6 PIXELS). 61
FIGURE 24. FIELD PHOTOGRAPHS OF THE LANDSCAPES OF BLACKWELL FARMS (A), SIDI RACHED 1 (B) AND SIDI RACHED 2 (C). 62
FIGURE 25. PROCESS OF MEASUREMENTS OF SOIL MOISTURE CONTENT LEVEL USING ML3 THETA PROBE SOIL MOISTURE SENSOR. 66
FIGURE 26. THE PROCESS OF SURFACE ROUGHNESS PARAMETERS MEASUREMENT USING A NEEDLE PROFILOMETER. 67
FIGURE 27. THE PROCESS OF SURFACE ROUGHNESS PARAMETERS MEASUREMENT USING A LASER PROFILOMETER. 68
FIGURE 28. RESULTS OF EACH OF THE USED SMC ESTIMATION METHODS IN BLACKWELL FARMS DATASETS. 72
FIGURE 29. RESULTS EACH OF THE USED SMC ESTIMATION METHODS IN THE SIDI RACHED 1 DATASETS. 75
FIGURE 30. RESULTS OF EACH OF THE USED SMC ESTIMATION METHODS IN THE SIDI RACHED 2 DATASETS. 79

LIST OF ABBREVIATIONS

Acronym	Definition
ANN	Artificial Neural Networks
ANN_{EA-IEM}	Artificial neural network with features from Integral equation model as input
ANN_{FLF1}	Artificial neural network with Feature Level Fusion 1 as input
ANN_{FLF2}	Artificial neural network with Feature Level Fusion 2 as input
ANN_{FLF3}	Artificial neural network with Feature Level Fusion 3 as input
ANN_{PDI}	Artificial neural network with PDI as input
ANN_{TVDI}	Artificial neural network with TVDI as input
AMSR-E	Advanced Microwave Scanning Radiometer Earth Observation System
ASCAT	the Advanced Scatterometer
BAY	Bayesian Inference Approach
dB	Decibel Units
DN	Digital Number
EA-IEM	Empirically Adapted Integral Equation model
ESA	European Space Agency
ET	Evapotranspiration
FAO	Food and Agriculture Association
FC	Field capacity
FLF1	Feature Level Fusion 1
FLF2	Feature Level Fusion 2
FLF3	Feature Level Fusion 3

GLAI	Green Leaf Area Index
GNDVI	Green Normalized Difference Vegetation Index
GNSS	Global Navigation Satellite Systems
GOES	Geostationary Environmental Operational Satellites
GPS	Global Positioning System
GRNN	Generalised Regression Neural Network
HSMDI	High-Resolution Soil Moisture Drought Index
IEM	Integral inversion Model
IHS	Intensity Hue Saturation transformation
LAI	Leaf Area Index
Lopt	Semi-empirical calibration parameter
LPRM	Land Parameter Retrieval Model
LS	Least Square
LS-SVM	Least Square Support Vector Machine
LST	Land Surface temperature
MIRAS	Microwave Imaging Radiometer with Aperture Synthesis
MLP	Multi-Layer Perceptron
MODIS	Moderate Resolution Imaging Spectrometer
MS	Multi-spectral
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
OLI	Operational Land Imager
PAW	Plant Available Water
PBO	Plate Boundary Observatory
PCA	Principal Component Analysis

PDI	Perpendicular Drought Index
PWP	Permanent Wilting Point
RF	Random Forest
RMSE	Root Mean Square Error
RS	Remote Sensing
SAVI	Soil Adjusted Vegetation Index
SD	Standard Deviation
SMAP	Soil Moisture Active Passive satellite
SMC	Soil moisture content
$SMC_{EA\text{-}IEM}$	Output estimation of $ANN_{EA\text{-}IEM}$
SMC_{FLF1}	Output estimation of ANN_{FLF1}
SMC_{FLF2}	Output estimation of ANN_{FLF2}
SMC_{FLF3}	Output estimation of ANN_{FLF3}
SMC_{Fused}	Final fused SMC estimation
SMC_{PDI}	Output estimation of ANN_{PDI}
SMC_{TVDI}	Output estimation of ANN_{TVDI}
SMOS	Soil moisture and Ocean Salinity satellite
SNAP	Sentinel Application Platform
SNR	Signal to Noise Ratio
SNSs	Sparse Network Stations
STARFM	Spatial and Temporal Adaptive Reflective Fusion Model
TCA	Triple Collocation Analysis
TIRS	Thermal Infrared Sensor
TRMM	Tropical Rainfall Measuring Mission
TSEB	Two-Source Energy Balance

TVDI	Temperature Vegetation Dryness Index
ubRMSE	Unbiased Root Mean Square Error
USDA	United State Department of Agriculture
USGS	United States Geological Survey
VI	Vegetation Indices
WBF	Weight-Based Fusion

List of Symbols

Symbol	Definition
$\mathbf{T_s}$	Land Surface Temperature
$\boldsymbol{m_v}$	Soil moisture content
$\boldsymbol{V}$	soil volume (m^3)
$\boldsymbol{v}$	the volume of water in V (m^3).
$\boldsymbol{\varepsilon}$	Dielectric constant (the relative permittivity)
$\boldsymbol{\varepsilon'}$	the real component of dielectric constant representing relative permittivity
$\boldsymbol{\varepsilon''}$	the imaginary component representing dielectric loss factor
$\boldsymbol{k}$	Wavenumber
$\boldsymbol{s}$	Root Mean Square of surface heights
$\boldsymbol{\theta_i}$	Incidence Angle
$\boldsymbol{\mu_r}$	permeability of the surface
$\boldsymbol{\sigma^0_{pp}}$	Backscattering coefficient (pp polarisation)
$\boldsymbol{R_h}$	The horizontally polarized Fresnel reflection coefficient
$\boldsymbol{R_v}$	The vertically polarized Fresnel reflection coefficient
$\boldsymbol{W^n}$	Fourier transform of the *n*th power
$\boldsymbol{\rho(x)}$	One dimensional correlation function
$\boldsymbol{l}$	the correlation length
$\boldsymbol{\Delta\phi}$	Phase difference
$\boldsymbol{\lambda}$	Wavelength
$\boldsymbol{N}$	Number of samples
$\boldsymbol{\bar{z}}$	Mean of surface heights

$\rho(\xi)$	Autocorrelation function
R_{red}	Reflectance in the Red band
R_{NIR}	Reflectance in the near-infrared band
M	The slope of the soil line
I	The intercept of the soil line
R_G	Reflectance in the Green band
L	Calibration parameter of SAVI
R^2	Coefficient of determination
σ^0_{dry}	Backscattering coefficient of images in a dry season
Lopt	Semi-empirical calibration parameter of The Integral Equation Model
w_{ij}	the connection weight from *i*th node in the input layer to *j*th
v_{jt}	the connection weight from the *j*th node in the hidden layer to *t*th
f(x)	Sigmoid function of the Artificial Neural Network
H_j	the value of the *j*th hidden node
b_j	the bias of the *j*th hidden node
Y_t	the value of *t*th output node
γ_t	the bias of the *t*th output node
P_i	the estimated SMC values
O_i	the measured SMC values

1. INTRODUCTION

1.1 Background

Though agriculture is essential for the nourishment of the overgrowing world's population, it does not come without a price. Agriculture represents approximately 38.5% of Earth's land area (FAO 2011) and 70% of the world's freshwater withdrawal corresponds to irrigation purposes (Nhemachena, Matchaya et al. 2018), which leads to the subject of this research. In order to, effectively reduce agricultural water consumption, it is important to understand an important component of the hydrological cycle, soil moisture (Jackson, Schmugge et al. 1996).

Accurate Surface Soil Moisture Content (SMC) levels estimations are instrumental not only for agricultural applications but for a deeper understanding of a variety of hydrological processes as well. On the global level, SMC can help determine a variety of land-atmosphere interactions, not to mention its crucial role in recent climate change studies (Hauser, Orth et al. 2016, Trenberth, Dai et al. 2014). SMC is also a significant factor for medium to small level applications such as natural resources management, drought assessments (Amani, Salehi et al. 2017), and most importantly, agricultural practices like irrigation scheduling (Rodríguez-Fernández, de Souza et al. 2017).

In light of the highlighted importance of Soil Moisture Content (SMC), numerous attempts have been made in the literature to devise methods for its retrieval using different sensing platforms (Rahimzadeh-Bajgiran, Berg et al. 2013, Byun, Liaqat et al. 2014). Direct in-situ measurements of SMC offer the best estimation possible in terms of accuracy, but it comes at the expenses of time and effort, especially due to the fact that those discrete measurements are point-based, which makes them specific to particular locations only and does not depict the spatial distribution and variability of soil moisture realistically (Byun, Liaqat et al. 2014). These limitations can be overcome by the use of indirect measurements or, in other words, remote sensing (Jackson 1993).

Remote sensing is capable of offering a continuous spatial and temporal coverage of SMC at all levels, and operational SMC satellites from various space agencies are an excellent example (Petropoulos, George P., Griffiths et al. 2013). Mission purposed satellites like Soil Moisture and Ocean Salinity (SMOS) (Kerr, Waldteufel et al. 2001a) or Soil Moisture Active Passive (SMAP) (Entekhabi, Yueh et al. 2014), do provide accurate SMC estimations (4% error) at a depth of 0-5 cm every 3 days (Al-Yaari, Wigneron et al. 2017).

Introduction

However, their respective spatial resolution (30-50 km for SMOS and 10-40 km for SMAP) limits their usefulness at the regional level (Entekhabi, Njoku et al. 2010). Conversely, for small scale agriculture or family farms (< 2 ha), which happens to represent 75% of the agricultural land of the world (Lowder, Skoet et al. 2016), a different set of sensors with the significantly better spatial resolution is required, namely Synthetic Aperture Radars (SAR) (Moran, Peters-Lidard et al. 2004), thermal infrared and multispectral imagers (Hassan-Esfahani, Torres-Rua et al. 2014).

High-resolution SAR imagers are independent of weather conditions, have night and day imaging capability, and offer surface penetration at various depths depending on their frequencies (from a few cms in X-band to tens of cm in the L-band in dry soil conditions) (Zribi, Muddu et al. 2019). SMC estimation methods pertaining to high-resolution SAR utilise the fact that the backscattered radar signal is directly influenced by the dielectric constant of the upper few centimetres of the surface, the latter is also sensitive to soil roughness, soil texture, soil moisture (Kornelsen, Coulibaly 2013a). The relationship of the dielectric constant and SMC is best described as a polynomial (Hallikainen, Ulaby et al. 1985). This relationship has been successfully and consistently exploited by myriad models, whether be them semi-empirical methods like Oh (Oh, Sarabandi et al. 1992) and Dubois (Dubois, Van Zyl et al. 1995), or theoretical models such as the inversion of the Integral Equation Model (IEM) (Koyama, Liu et al. 2017, Baghdadi, Holah et al. 2006). The IEM is used extensively to determine soil moisture content and surface roughness parameters (Fung, Li et al. 1992), but its application in the presence of medium to intense vegetation covers is difficult since the sensitivity of the radar response to SMC is significantly reduced in these areas, especially at very short radar wavelength (Khabazan, Motagh et al. 2013).

Estimations using multispectral and thermal infrared synergies, on the other hand, are not affected by the presence of partial or even intense vegetation covers (Lambin, Ehrlich 1996). The concept of these synergies takes advantage of the fact that surface radiant temperatures are correlated with the distribution and variability of SMC levels and vegetation (Du, Song et al. 2017). In remote sensing terms, surface radiant temperatures can be represented using Land Surface Temperature (LST) derived from atmospherically corrected thermal infrared images (with the wavelength in the range from 8 to 13 microns), and vegetation cover intensity (Yang, Y., Guan et al. 2015). Vegetation cover intensity can be represented by various Vegetation Indices (VI), which are essentially derived from algebraic combinations of the visible red (380-760 nm) and near-infrared (760 nm-1 microns) (Petropoulos, Ireland et al. 2015). (Sandholt, Rasmussen et al. 2002) realized that the relationship between LST and VI can indicate SMC levels. VI/LST data points, when represented as two-dimensional scatter plot, form a triangle/trapezoid feature space, which could be later used to determine extreme boundaries (dry/wet edges) to calculate an index called Temperature Vegetation Dryness Index (*TVDI*). TVDI, in turn, has a linear relationship with SMC (Petropoulos, G., Carlson et al. 2009). This method, however, suffers from uncertainty and subjectivity, especially when atmospheric conditions are not uniform. There are also limitations imposed by the current satellite technology involved in this synergy, namely coarse temporal resolution and susceptibility to cloudy conditions (Yang, Y., Guan et al. 2015).

Since all the aforementioned sensors and methods produce variable results under different conditions, data fusion techniques are widely considered a suitable solution since they entail the compensation of the limitation of each sensor by the advantages of the other. To achieve the most accurate estimation possible, information is extracted from multiple sensors and combined instead of inferring estimations from a single sensor (Dong, Zhuang et al. 2009). A comprehensive description of each method, as well as their respective limitations, will be provided in breadth in the Literature Review section.

1.2 Aim and Objectives

This research aims to address the shortcomings of the previously highlighted remote sensing technologies by designing a system capable of using the advantages of each sensor to compensate for the limitations of another in the SMC retrieval sense. The goal of this research is to design a novel system that incorporates data fusion techniques to achieve soil moisture content determination (validated by point-based ground measurements) with better accuracy (compared to using one single technology on its own) and said system would need to ensure:

- The retrieval of SMC levels using Synthetic Aperture Radar (SAR) by inverting an updated version of the Integral Equation Model.
- The retrieval of SMC levels using a multispectral index called the Perpendicular Drought Index.
- The retrieval of SMC levels using a synergy of thermal and multispectral images by exploiting the relationship between land surface temperature and the intensity of vegetation covers. This relationship culminates in an index called Temperature Vegetation Dryness Index which, in turn, estimates soil moisture content.
- The retrieval of SMC levels using a feature level fusion, by feeding features extracted from each of the retrieval methods to an estimator. Different combinations of features have been used in this method.

- The retrieval of SMC levels using a decision level fusion by fusing all achieved estimations in a weight-based system.
- Better accuracy of SMC estimation than that of each method is achieved with a fusion scheme.

1.3 Novelty and contributions

The novelty of this research lies in the data fusion aspect of the proposed SMC estimation system, especially the feature and decision level fusions, more specifically the weight-based nature of this system.

All of the previously described methods have already been investigated extensively in numerous studies (Dawson, Fung et al. 1997, Barrett, Dwyer et al. 2009, Sahebi, Angles 2010, Kornelsen, Coulibaly 2013a, MirMazloumi, Sahebi 2016, Mirsoleimani, Sahebi et al. 2019, Huang, S., Ding et al. 2019, Ghulam, Qin, and Zhan 2007, Zhang, J., Zhou et al. 2014, GE, ZHANG et al. 2018, Chen, Sun, Wang et al. 2019, Koyama, Liu et al. 2017, Baghdadi, Holah et al. 2006, Aisyah, Kusratmoko et al. 2019, Gherboudj, Magagi et al. 2011). Surveying those investigations produced a comprehensive knowledge of the advantages and limitations of each of those methods. The survey confirmed to the author the existence of a strong case for the use of data fusion techniques to ameliorate the accuracy of SMC estimation. However, employing said techniques for SMC estimation is hardly a novel contribution, as various iterations of those techniques have already been researched by authors in (Kurucu, Sanli et al. 2009, Bai, L., Long et al. 2019, Notarnicola, Posa 2001, Posa, Notarnicola et al. 2004, Yuan, Xu et al. 2020, Zaman, McKee et al. 2012). This lead the author of this research to explore a novel approach to data fusion techniques as an attempt to achieve a more SMC estimation. The elements of the novelty of this approach can be summarized by the following points:

- The use of an updated version of IEM suggested by (Song, Zhou et al. 2009), and incorporating a more accurate surface roughness parameter suggested by (Baghdadi, Holah et al. 2006) into that updated version.
- The design and implementation of a new Laser profilometre to measure surface roughness parameters.
- The design, planning, and execution of field campaigns to collect SMC and surface roughness ground measurement for tests and validation of the proposed system in 3 distinct study areas.
- The replacement of the performance function of a Back Propagation Neural Network from absolute error to Root Mean Square Error.

- The design and implementation of a novel feature level fusions using different combinations of the extracted salient features from each method. Specifically, to the time of writing of this thesis, the use of features from IEM inversion (especially this updated version of the IEM), Perpendicular Drought Index, and Temperature Vegetation Dryness Index has never been used in the same SMC estimation system.
- The design of a novel weight-based system, where the weights are inferred from ground truth measurements, and those weights can be updated if future measurements are introduced.
- The implementation of all components of the proposed SMC estimation system through the conception of a novel Matlab code.
- Achieving an acceptable accuracy of estimation compared to the more similar studies in the literature.

1.4 Publication List

In the time elapsed doing this research, two conference papers were published:

- YAHIA, O., GUIDA, R., and IERVOLINO, P., 2018. Sentinel-1 and Landsat-8 feature level fusion for soil moisture content estimation. EUSAR 2018, which was presented in the form of a poster.

- YAHIA, O., GUIDA, R., and IERVOLINO, P., 2018. Weights Based Decision Level Data Fusion of Landsat-8 and Sentinel-1 for Soil Moisture Content Estimation, IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium 2018, IEEE, pp. 8078-8081, which was presented in an oral presentation.

1.5 Outline of The Thesis

This thesis is composed of six chapters: Introduction, Literature Review, Soil Moisture Content Estimation System- Conceptual design, Generating Test Datasets, Testing and Evaluation, and Conclusions and Future Work.

Chapter 2, named Literature Review, provides a definition of Soil Moisture Content. Then it describes methods pertaining to Soil Moisture Content Retrieval Using Remote Sensing. Those methods include SMC Retrieval Using Operational Estimation by Remote Sensing, SMC Effects on Active Microwave Sensing, of which the chapter goes into finer details describing the Integral Equation Model theory as well as the limitations of SMC estimation through IEM inversion. Those methods also include SMC Retrieval Using Multispectral and Thermal Remote Sensors, most specifically Perpendicular Drought Index and Temperature

Vegetation Dryness Index, and their respective concepts, and their limitations when it comes to SMC estimations. Chapter 2 also defines Data fusion techniques and their massive potential to solve the limitation of each of the methods mentioned above supported by the evidence of past literature concerning data fusion specifically for SMC Estimation. Chapter 2 delineates the motivations and the rationale behind the novel SMC estimation system. Chapter 2 offers a detailed description of each component of the proposed system as well.

Chapter 3 is named Soil Moisture Content Estimation System – Conceptual Design. Chapter 3 will justify the motivations behind using data fusion techniques to maximize the accuracy of SMC estimation. Then, the proposed SMC estimation system is introduced with all its various components. Components consist of the updated version of the Integral Equation Model and the reason behind its use, Pre-processing, the Perpendicular Drought Index and Temperature Vegetation Dryness Index respective determinations, Feature level fusion and the rationale behind it, and finally, more importantly, the most salient item of novelty in this research, the Fusion Centre, the reasoning behind it, as well as the proposed index to measure the accuracy of its estimation.

Chapter 4 is called Generating Test Datasets. Chapter 4 is dedicated to delineating all details concerning the used study areas, as well as the corresponding earth observation data of these areas of interest (whether be it Sentinel-1 or Landsat-8 data). Chapter 4 incorporates descriptions of the process of ground truth measurement collection. Those measurements consist of two different information, SMC using the ML3 Theta Soil Moisture Probe and surface roughness parameters through two different types of Profilometres.

Chapter 5 is named Testing and Evaluation. It contains all of the achieved results and analysis by the proposed SMC estimation system in study areas: Blackwell Farms, Sidi Rached 1, and Sidi Rached 2. The results are analyzed and evaluated to produce key remarks and discussion points about the proposed system.

Chapter 6 is named Conclusions and Future Work. It provides a summary of the thesis, and a list of the attained achievement in the course of this research, as well as the future research, plans the author has.

2. LITERATURE REVIEW

2.1 Introduction

The design of a soil moisture retrieval system requires an intimate knowledge of soil moisture content, and that includes its basic concepts and the technologies used for its measurement. It is quite important to be familiarized with remote sensing to gather relevant information for the aforementioned system. This chapter will begin by providing key definitions of SMC and remote sensing and how soil moisture was retrieved in remote sensing data according to the state of the art. Different soil moisture content retrieval methods based on different sensors are explained in terms of theory and limitations of performance, methods such as the operational estimation by remote sensing, retrieval using active microwaves sensors (SAR), or more specifically, the use of Integral equation model inversion, and retrieval using multispectral and thermal remote sensors, where a multispectral index in the Perpendicular Drought Index was explained as well as another synergetic index (multispectral and thermal infrared synergy) in the Temperature Vegetation Dryness Index.

This chapter concludes with the definition of data fusion techniques along with its different processing levels, and the proposal of data fusion as a solution to the discussed limitations.

2.2 Soil Moisture Content

Soil moisture content (SMC) can be defined as the amount of water present within unsaturated soil particles (Hillel 1998). Based on the depth from the surface, soil moisture is split into two zones: the first zone represents the soil moisture in the upper 10 cm soil layer, and it is named the surface soil moisture, the second zone is immediately beneath the first is named the root zone soil moisture, where the water is available down to 200 cm below the soil surface, and it contains the groundwater available to plants as illustrated by Figure 1:

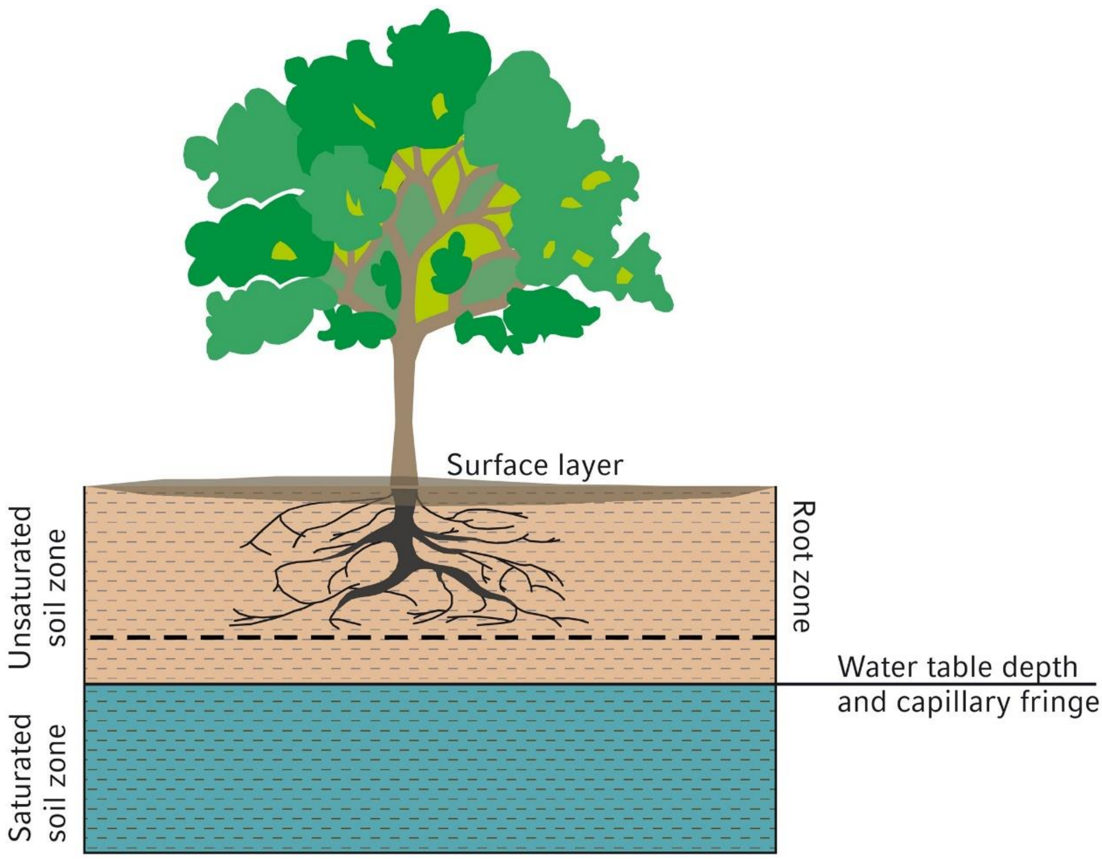

Figure 1. The saturated and unsaturated soil zones. Adapted from (Peng, Loew et al. 2017).

Figure 1 depicts the surface later and root zone, as well as the saturated and unsaturated soil zones. Where soil surface water (whether from irrigation or rain) drains downwards into deeper soil layers to eventually the permanently saturated layer, the top of the permanently saturated layer is called the water table depth (or groundwater depth). Where the capillary fringe is the layer of variable thickness that lies directly beneath the water table. The water in this layer moves upwards by capillary action (Petropoulos, Ireland et al. 2015).

In this research, the focus would be on the estimation of surface soil moisture for agricultural practices.

Soil moisture content can vary from one type of soil to another. Different soils hold different amounts of water depending on their structure and texture. Speaking in the microscopic sense, the soil is composed of particles, and its classification depends entirely on the dimensions and void spaces of those particles. The described particles are clay, silt, and sand (Kellogg 1993). The United State Department of Agriculture (USDA) classifies soil types according to a soil triangle which includes all possible combination of clay, silt, and sand as illustrated by Figure 2:

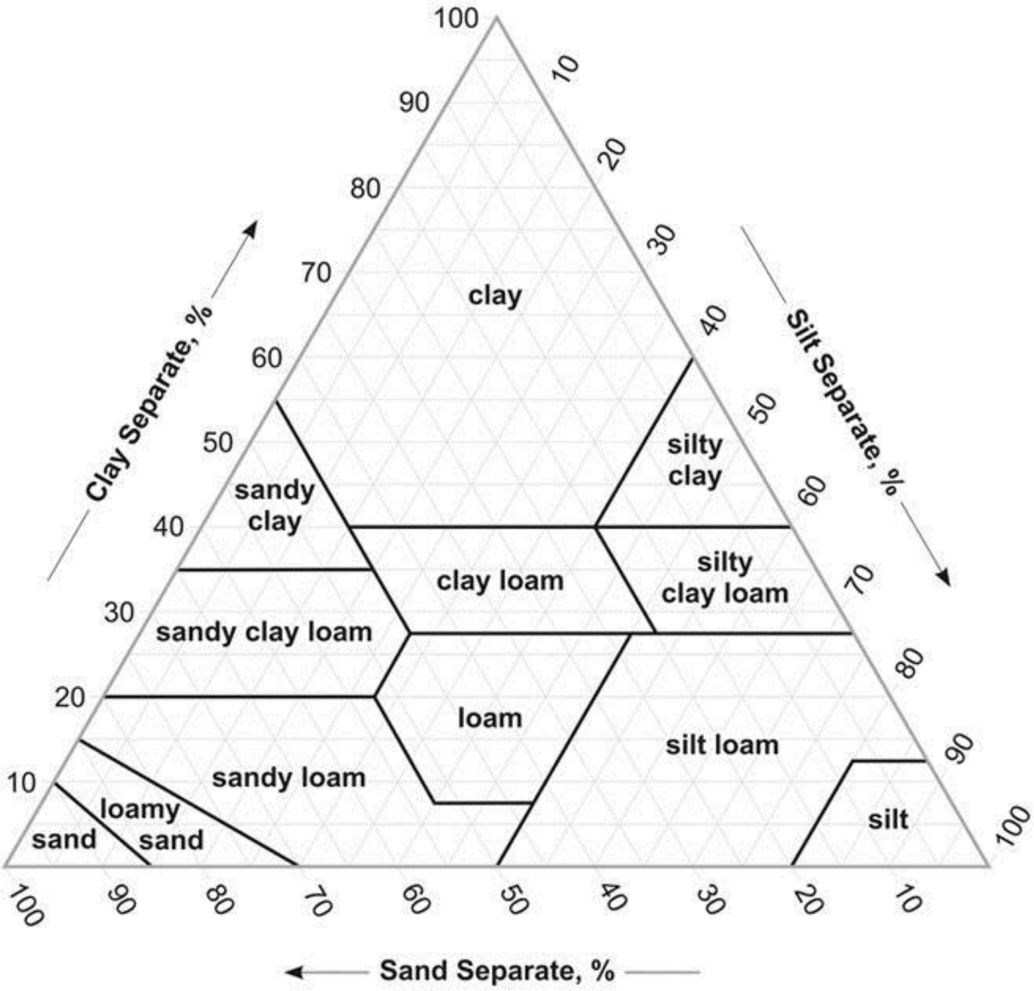

Figure 2. USDA soil texture classification (Kellogg 1993).

Soil samples are attributed to one of twelves classes depending on the percentages of sand, silt, and clay, which represent the bottom, right and left axes, respectively, of the triangle in Figure 2. The distinction between the different classes is ascertained through methods such as the pipette method, the hydrometer method, and field estimates. Field estimates are the simplest method to determine a soil texture class. Those estimates rely on the feel of the soil texture (gritty, smooth, or sticky) and how it responds to rubbing it between the fingers to form a ribbon. Sand particles have a gritty feel, silt particles feel smooth and silky (Fenton, Vero et al. 2015). While a sticky feel corresponds to the presence of clay. Field estimates are not entirely too accurate and laboratory determinations are necessary for validation. Please note that the criteria of this texture classification is relevant only to the mineralogical composition of the soil, and not the percentage of the organic matter present within it, which is a whole different criterion of classification that entails different methods of determination (Ditzler, Scheffe et al. 2017).

Soil saturation occurs when soil pores are completely filled with water (no space for air within the soil particles). After the source of water whether from irrigation or rainfall dries up, water present in the larger pores moves downwards (drainage) which allows for air to replaces some of that water in which case the soil is called at Field Capacity (FC). At FC, the proportions of water and air are considered optimal for plant growth. As the water in the soil eventually dries out, the soil is at its lower limit of soil moisture and it is said at the Permanent Wilting Point (PWP), which signifies that soil moisture level is so low plants cannot absorb it. The difference between FC and PWP is the total amount of water available for vegetation and is called Plant Available Water (PAW). Figure 3 elucidates the correlation between plant available water, field capacity, and permanent wilting point (Zotarelli, Dukes et al. 2010):

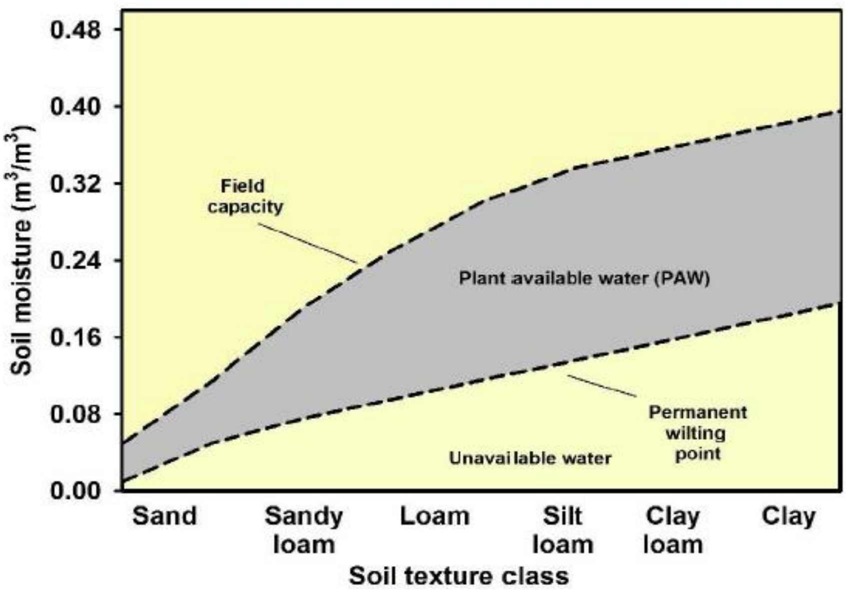

Figure 3. Relationship between soil moisture levels and soil texture classes (Zotarelli, Dukes et al. 2010).

Figure 3 demonstrates that different texture classes of soil have different FC and PWP, which in turn means different PAW. Coarse textured soil (sand) has the least PAW compared to the finer-textured soil (loam), with Clay being the texture possessing the most PAW. It has to do with the size of pores which corresponds to faster drainage in the sand (in a matter of hours), and relatively slower drainage for clayey soils (2-3 days) (C. Brouwer, A. Goffeau et al. 1985).

Soil moisture content importance is quite apparent, especially if the task at hand is to detect water deficit. The latter takes place when the evaporative demand of a plant is greater than the water supply in the soil. It has been established that short-term water deficit may affect the plant growth processes (Shao, Chu et al. 2008). Water deficit can cause wilting, closure of stomata, and decrease in cell enlargement and growth. Severe water stress may result in an arrest of photosynthesis, a disturbance of metabolism, and finally death (Berry, Kalra et al. 1988). Therefore, in the agricultural sense, precise soil moisture content retrieval is crucial for plant health. The soil moisture content in a soil volume V is expressed by equation 1:

$$\boldsymbol{m_v = \left(\frac{v}{V}\right) \times 100}\% \tag{1}$$

Where m_v = soil moisture content (%), V= soil volume (m^3) and v = is the volume of water in V (m^3). The equation can be applicable in many scales, depending on the measurement method or the research at hand, the scales can range from cubic centimetres to cubic kilometres (Seneviratne, Corti et al. 2010).

2.3 Soil Moisture Content Retrieval Using Remote Sensing

Remote sensing is the process of inferring information from indirect measurements collected by sensors on-board various platforms (aircraft and satellites), without being in direct contact with the observed object, phenomenon, and or environment (Schowengerdt 2007). As opposed to measurements collected from hand-held sensors, input collected from the latter is also called proximal sensing (Mulla 2013). This information is usually the measurement of reflected or emitted electromagnetic radiation from soil or plant material, in which case we speak of "passive" sensing. Conversely, when sensors transmit their energy, they are called "active" sensors (such as Synthetic Aperture Radar), and they will measure the backscattered energy (the echo) (Sikdar, Glavic et al. 2004).

Spectral remote sensing consists of the determination of the wavelength of each photon of light by its energy level. Light and other electromagnetic radiations are designated in terms of their wavelengths. For instance, visible light possesses wavelengths between 0.4 and 0.7 microns, whereas radio waves have wavelengths greater than 30 cm (Shippert 2003) as displayed in Figure 4:

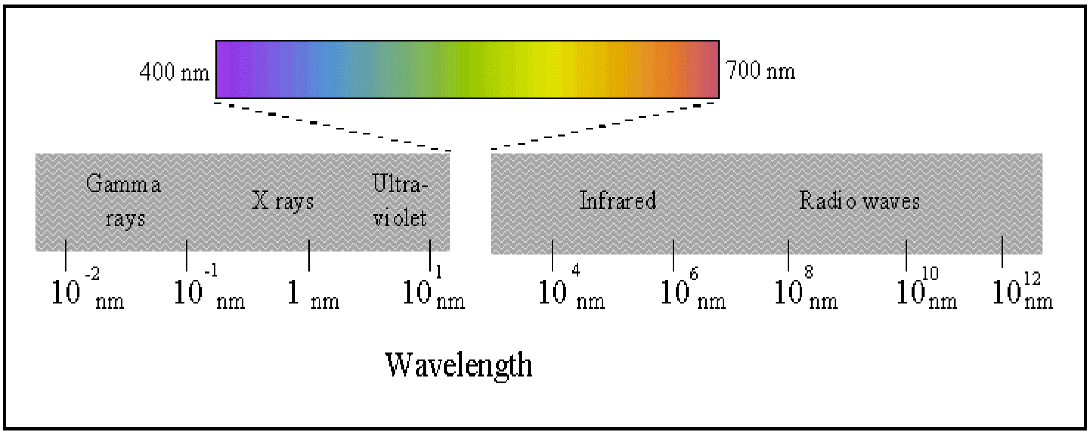

Figure 4. The electromagnetic spectrum (Shippert 2003).

It is well established in the literature that SMC has a tremendous effect on soil reflectance. Soil reflectance decreases with the increased presence of SMC volume and vice versa (Mulla 2013, Randall B. Smith 2013, Lobell, Asner 2002, Oltra-Carri, Baup et al. 2015, Somers, Gysels et al. 2010). Figure 5 further illustrates the effect of SMC levels on the spectral reflectance of the soil:

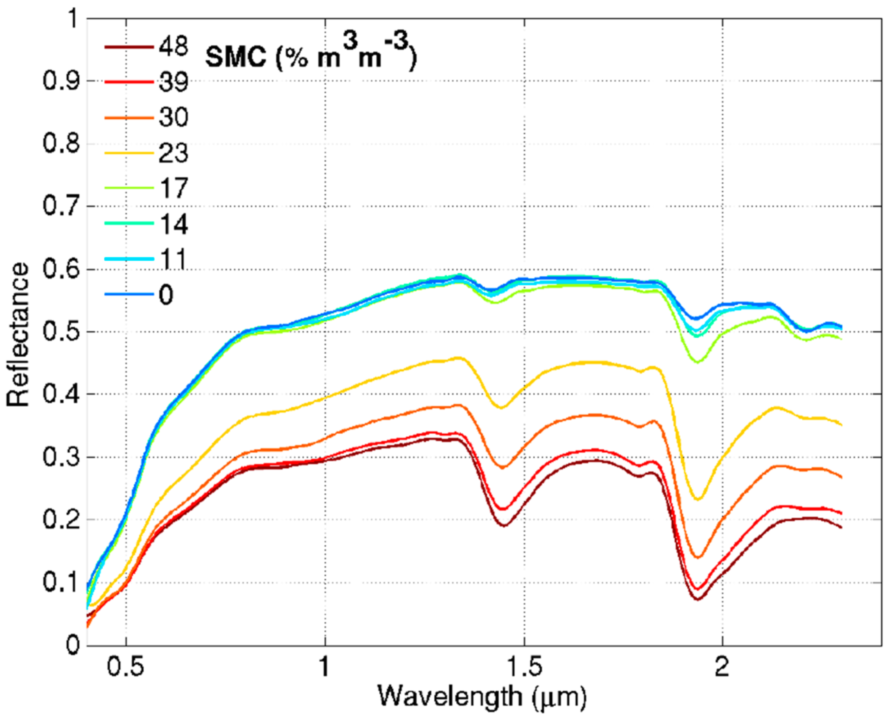

Figure 5. Evolution of the spectral signatures behaviour depending on the volumetric soil moisture content ranging between 0 and 0.48 $m^3 \cdot m^{-3}$ (Oltra-Carri, Baup et al. 2015).

As opposed to passive sensors, Active microwave instruments produce their own artificial radiant energy source for illumination which grants them penetration capabilities through clouds, dust, water vapour, and rain cells (Karthikeyan, Pan et al. 2017). Active microwave sensors, more specifically radars, are equipped with transceiver antennas capable of emitting modulated electromagnetic pulses and measuring the strength and the time between the transmitted and reflected pulses (Mansourpour, Rajabi et al. 2006). Those measurements allow the determination of both the type of ground target and its distance from the transmitter. Radar Images are composed of pixels containing intensity values expressed as uncalibrated digital numbers which are often converted to a physical quantity named the Backscattering Coefficient measured in decibel units (dB) (Mo, Schmugge et al. 1984, Döring, Schwerdt 2013). Radars can be categorized into two groups: non-imaging and imaging radars. The output of non-imaging radars is not an image, but rather one-dimensional data such as the data produced by altimeters and scatterometer. The output of imaging radars is comprised of two measurements: slant range and azimuth (Iervolino 2015). Synthetic Aperture Radars (SAR) are included within the latter group. SARs are operational in myriad frequency bands of the microwave spectrum, the most recurrent being L, C, and X bands (Ulaby, F. T., Long et al. 2014).

A more detailed review of techniques for soil moisture retrieval from SAR images will be provided in section 2.3.2.

2.3.1 Soil Moisture Retrieval Using Operational Estimation by Remote Sensing

Many satellite missions and remote sensing radiometers have been used to gain a comprehensive understanding of the global hydrological processes, and this is maybe apparent by the emergence of new SMC operational products from different space agencies (Petropoulos, Ireland et al. 2015). Soil Moisture and Ocean Salinity (SMOS) (Kerr, Waldteufel et al. 2001a) and Soil Moisture Active Passive (SMAP) (Entekhabi, Yueh et al. 2014), are great examples of such efforts.

SMOS is a European Space Agency (ESA) satellite mission launched on the 2nd of November 2009 with the objective of providing soil moisture and ocean salinity maps. Mapping these two important components in the water cycle clears our understanding of the exchange processes between the surface of the Earth and the atmosphere which improves weather and climate models (Kerr, Waldteufel et al. 2001b). SMOS uses a novel interferometric radiometer called Microwave Imaging Radiometer with Aperture Synthesis

(MIRAS). Microwave radiometry at the L-band (1400 -1427 MHz) is used to estimate surface soil moisture and ocean salinity (Kerr, Waldteufel et al. 2010). The produced images in the span of 3 days form a global soil moisture map at the depth of 0-5 cm with 4% accuracy, however, the spatial resolution of 30 to 50 km while it suits global measurements, is not suitable for mid-scale to small scale applications (Kerr, Font et al. 2012).

The SMAP mission is another example of mission solely purposed for SMC estimation. SMAP was one of the four missions recommended by the U.S. National Research Council Committee on Earth Science and Applications from Space. Similar to SMOS, the goal is to measure soil moisture and with the same accuracy (4%). The difference is that it uses an L-band radiometer (1.41 GHz) and an L-band radar (1.26 GHz) measurements (at a spatial resolution of 40 km and 3 km respectively) to extract combined information about near-surface soil moisture at 0 to 5 cm depth, the data products in the revisit time of 3 days at hydrometeorology and hydro-climatology scales are 10 km and 40 km respectively (Entekhabi, Njoku et al. 2008). Minimizing the effect of vegetation on soil parameters is the reason behind the simultaneous use of both active and passive sensors. While the radiometer can provide better soil moisture measurements under vegetation conditions, radar has a far better spatial resolution. Therefore the combination of information from both provides an enhanced estimation of soil moisture measurements in terms of spatial capabilities and accuracy (Entekhabi, Njoku et al. 2010, Brown, Escobar et al. 2013, Reichle, Ardizzone et al. 2018). However, even with those enhancements, the spatial resolution is too poor to be considered useful for many uses, especially for agriculture.

2.3.2 Soil Moisture Content Effects on Active Microwave Sensing

SAR is a popular active microwave technique due to its large potential for SMC retrieval at the regional scales (Oldak, Jackson et al. 2003, Mattia, Balenzano et al. 2018), the backscattered radar signal is influenced by SMC levels as well as several other surface characteristics such as surface roughness profile, mineralogical composition of the soil, and dielectric features of the soil, and radar characteristics like the incidence angle, the working frequency of the SAR and polarization (Khabazan, Motagh et al. 2013).

The dielectric features of the surface soil are often referred to as dielectric constant (ε) (Verhoest, Lievens 2013). It is important to note that the relationship between the dielectric constant and SMC is of polynomial nature. The polynomial relationship of SMC and dielectric constant summarizes the dependence of the latter on the mineralogical composition of the soil as well as the SAR frequency and SMC levels, as demonstrated by (Hallikainen, Ulaby et al. 1985, Verhoest, Lievens 2013), empirically calculated by regression analysis using frequencies of 1.4 GHz and 18 GHz. The dependence is illustrated in Figure 6:

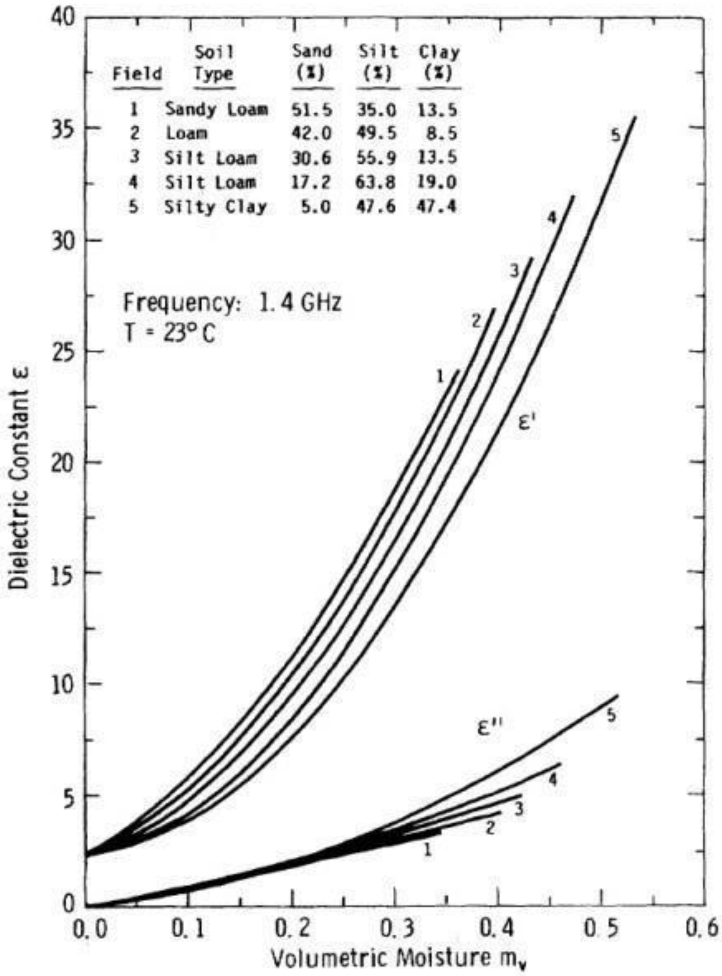

Figure 6. Dielectric constant as a function of soil moisture (Hallikainen, Ulaby et al. 1985)

Where ε' is the real component of dielectric constant representing relative permittivity, and ε'' is the imaginary component representing dielectric loss factor (Hallikainen, Ulaby et al. 1985).

(Hallikainen, Ulaby et al. 1985) expressed that relationship by Equation 2:

$$\varepsilon = (a_0 + a_1 S + a_2 C) + (b_0 + b_1 S + b_2 C) m_v + (c_0 + c_1 S + c_2 C) m_v^2 \quad (2)$$

Where S and C are sand and clay textural components of the soil in percentage, and a_0 to c_2 are the corresponding coefficients of the polynomial expression depending on the frequency.

2.3.2.1 IEM Theory

The soil moisture estimation model used for this study is the single scattering IEM (Fung, Li et al. 1992).

IEM inversion is a theoretical model capable of representing backscattered radar signal with proven ability to estimate SMC and surface roughness parameters as it has been investigated in numerous studies (Bai, X., He et al. 2016, Baghdadi, Nicolas, Chaaya et al. 2011, Chen, K. S., Yen et al. 1995, Mao, Tang et al. 2008, Paloscia, Pampaloni et al. 2008). The IEM backscattering model is valid within a wide range of different roughness values often encountered in agricultural surfaces (as long as $k{\cdot}s \leq 3$, where k is the wavenumber and s is the Root Mean Square (RMS) of surface heights) (Baghdadi, Nicolas, Gherboudj et al. 2004).

The following expressions represent the backscatter coefficient of the surface contribution (Fung, Li et al. 1992) :

$$\boldsymbol{\sigma^0_{pp}} = \frac{\boldsymbol{k^2}}{\boldsymbol{2}} \boldsymbol{e}^{(-2k_z^2 s^2)} \sum_{\boldsymbol{n=1}}^{\infty} \boldsymbol{s^{2n}} \left|\boldsymbol{I^n_{pp}}\right|^2 \frac{\boldsymbol{W^n(-2k_x, 0)}}{\boldsymbol{n}!} \tag{3}$$

$$I^n_{pp} = (2k_z)^n f_{pp}\, exp[-2k_z^2 s^2] + \frac{1}{2}\left\{k_z^n \left[F_{pp}(-k_x, 0) + F_{pp}(k_x, 0)\right]\right\} \tag{4}$$

$$f_{vv} = \frac{2R_v}{\cos\theta_i} \tag{5}$$

$$f_{hh} = \frac{-2R_h}{\cos\theta_i} \tag{6}$$

$$F_{vv}(-k_x, 0) + F_{vv}(k_x, 0) = \frac{2\sin^2\theta_i(1+R_v)^2}{\cos\theta_i}\left[\left(1-\frac{1}{\varepsilon_s}\right) + \frac{\mu_r\varepsilon_s - \sin^2\theta_i - \varepsilon_s\cos^2\theta_i}{\varepsilon_s^2\cos^2\theta_i}\right] \tag{7}$$

$$F_{hh}(-k_x, 0) + F_{hh}(k_x, 0) = \frac{2\sin^2\theta_i(1+R_h)^2}{\cos\theta_i}\left[\left(1-\frac{1}{\varepsilon_s}\right) + \frac{\mu_r\varepsilon_s - \sin^2\theta_i - \mu_r\cos^2\theta_i}{\mu_r^2\cos^2\theta_i}\right] \tag{8}$$

$$W^{(n)}(a, b) = \frac{1}{2\pi}\iint \rho^n(x, y) e^{-i(ax+by)} dxdy \tag{9}$$

Where σ_{pp}^{0} is the backscattering coefficient with pp signifying the polarization state; θ_i is the incident angle; $k_z = kcos\theta_i$, $k_x = ksin\theta_i$; R_h and R_v are the horizontally and vertically polarized Fresnel reflection coefficients respectively; ε_s and μ_r are the relative permittivity and permeability of the surface; W^n is the Fourier transform of the *n*th power of the surface correlation function $\rho(x, y)$. The latter presents an exponential distribution (equation 10**)** for low surface roughness values and a Gaussian (equation 11) for high surface roughness values (Baghdadi, N., Gaultier et al. 2002), For one-dimensional surface roughness profiles, the correlation functions are expressed in equations 10 and 11:

$$\boldsymbol{\rho(x) = e^{-\frac{|x|}{l}}} \tag{10}$$

$$\rho(x) = e^{-\frac{x^2}{l^2}} \tag{11}$$

Where *l* is the correlation length.

The degree of the roughness of a given surface has a massive impact on the backscattered signal, this highlights the significance of the correct identification of a surface roughness profile (Frei, Henkel 2002). The Rayleigh criterion is widely used to establish the degree of smoothness of a given surface. A soil surface is considered rough if the phase difference (Δϕ) between two rays scattered from a separate point on the surface exceeds π/2 (Baghdadi, Nicolas, Zribi et al. 2008).

Figure 7 illustrates the geometry of the phase difference of two paralleled waves scattered from different points on a rough surface. A surface is considered rough if it satisfies Inequation 12:

$$\boldsymbol{s > \frac{\lambda}{8cos\theta}} \tag{12}$$

Where $\boldsymbol{\lambda}$ is the wavelength.

The phase difference Δϕ is calculated using equation 13:

$$\boldsymbol{\Delta\phi \; = \; 2k \cdot s \cdot cos\theta} \tag{13}$$

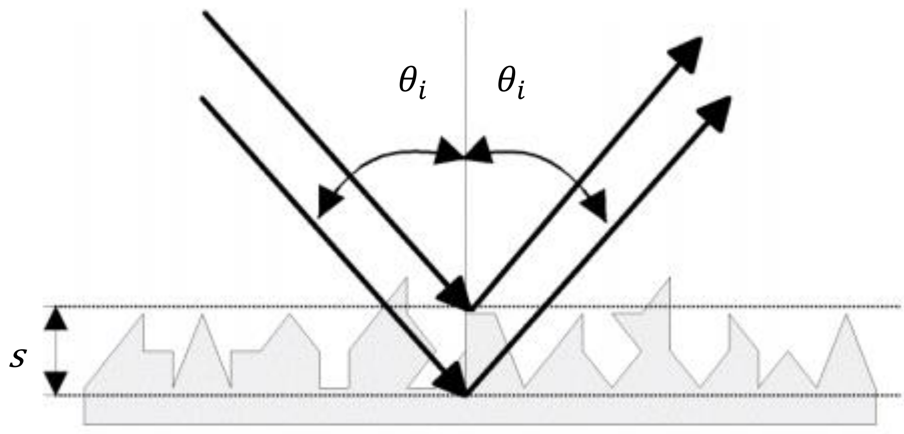

Figure 7. The geometry of Rayleigh criterion. Adapted from (Hajnsek, Papathanassiou 2005).

Where s, and θ_i are RMS of surface heights and incident angle respectively.

The Fraunhofer criterion is a stricter criterion proposed by (Ulaby, Fawwaz Tayssir 1982). Instead, a soil surface is considered rough if the phase difference between two rays scattered from a separate point on the surface exceeds $\frac{\pi}{8}$, which leads to inequation 14:

$$\boldsymbol{s > \frac{\lambda}{32cos\theta}} \tag{14}$$

The RMS height can be calculated using **equations 15** and **16**:

$$\boldsymbol{s = \sqrt{\frac{1}{N}\left[\left(\sum_{i=1}^{N} Z_i^2\right) - N\,\overline{Z}^2\right]}} \tag{15}$$

where

$$\boldsymbol{\overline{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i} \tag{16}$$

where N is the number of points, and $\bar{Z}$ is the mean of heights (Bryant, Moran et al. 2007).

As for the approximation of the correlation length *l*, it is achievable once the normalised autocorrelation function $\boldsymbol{\rho(\xi)}$ is defined. The latter is calculated using **equation 17**:

$$\boldsymbol{\rho(\xi)} = \frac{\sum_{i=1}^{N-j} Z_i Z_{i+j}}{\sum_{i=1}^{N} Z_i^2} \tag{17}$$

The surface correlation length *l* is the horizontal distance over which the surface profile is auto-correlated with a value larger than *1/e* (Verhoest, Lievens et al. 2008).

Due to the mathematical complexity of IEM, an alternative method is used to invert it to calculate SMC; Artificial Neural Networks (ANN) (Hecht-Nielsen 1988). ANNs have been used extensively and successfully to invert IEM to infer SMC and roughness parameters in numerous studies (Baghdadi, N., Gaultier et al. 2002, Baghdadi, N., Cresson et al. 2012, Baghdadi, Nicolas, El Hajj et al. 2018, Yahia, Guida et al. 2018a) [78]. ANN is a parallel distributed information processing structure that consists of processing elements interconnected together with unidirectional signal channels referred to as connections or weights (Gardner, Dorling 1998). The ANN used for the IEM inversion is a multi-layer perceptron (MLP) (Hecht-Nielsen 1988), It is a feed-forward network, characterized by a unique unidirectional data flow, without loop; in particular, the MLP is made up of several layers: an input layer, at least one or several hidden layers, and an output layer, and the training algorithm uses ground measured data to minimise error (Gardner, Dorling 1998, Daliakopoulos, Coulibaly et al. 2005). An example of an MLP is Figure 8:

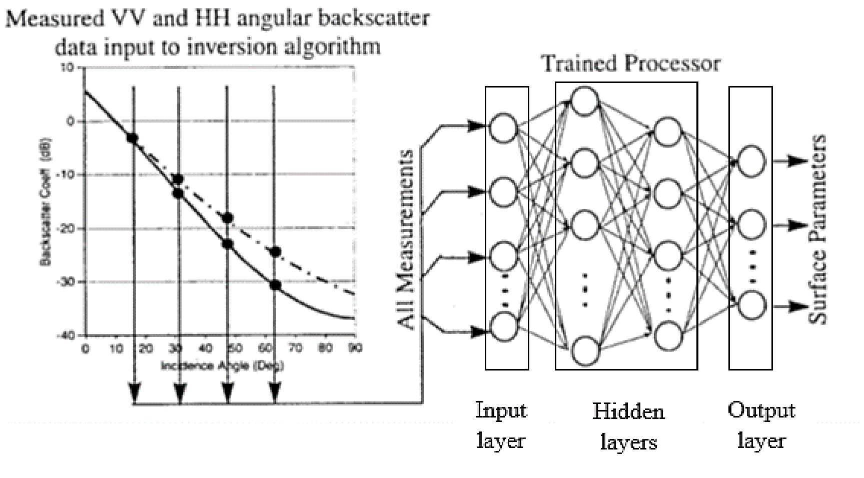

Figure 8. Schematic diagram of an MLP. Adapted from (Hecht-Nielsen 1988).

In this instance (Figure 8), the MLP is used to infer surface roughness parameters using measurements of the angular backscattering coefficient in polarisations VV and HH and incidence angles as input data.

More details about the inner workings of the IEM inversion using ANN will be provided in the methodology section.

2.3.2.2 Limitations of SMC Estimation Through IEM inversion

SMC retrieval using SAR, most specifically while using the IEM model, suffers from limitations in terms of performance due to several factors as summarized by the following bullet points:

- Speckle noise is an interference that plagues active microwave sensors characterised as multiplicative noise (Moreira, Prats-Iraola et al. 2013). It represents variations in backscatter from inhomogeneous cells which can be a consequence of multiple scattering events caused by the nature of the surface of the target (volume scattering). Consequently, that leads to a granular appearance of SAR images (Thoma, Moran et al. 2008), such interference, requires the use of filters to help attain a better soil moisture content retrieval which comes at the expense of soil moisture heterogeneity in the filtered pixels (Thoma, Moran et al. 2008).
- The effect of SMC on radar signals is less discernible when SMC levels exceed 35%, especially at the HH polarisation (Baghdadi, Nicolas, Zribi 2006).
- Dielectric behaviour of the soil (which is a key indicator of SMC) is heavily influenced by the distribution of grain size, which in turn determines the amount of free space for available water in the soil, which highlights the importance of the accurate identification of the mineralogical composition of the soil in question (Kornelsen, Coulibaly 2013a).
- The accuracy of SMC retrieval using IEM is largely dependent on the characterization of surface roughness parameters, as well as the accuracy of the measurements of those parameters (Verhoest, Lievens et al. 2008). However, (Zribi, Dechambre 2003, Baghdadi, Nicolas, Chaaya et al. 2011) introduced semi-empirical calibrations of RMS height and correlation length to improve the characterisation of surface roughness parameters which has yielded promising results.
- The poor temporal resolution of high-resolution SARs, which makes tracking SMC temporal variations difficult (Cenci, Pulvirenti et al. 2018).

- Susceptibility to intense vegetation covers, which can cause volume scattering, the latter has a direct negative impact of the accuracy of SMC retrieval (Bindlish, Barros 2000).

Due to the limitations described above, alternative methods to estimate SMC using a different group of sensors were explored, namely, multispectral, and thermal remote sensors.

2.4 Soil Moisture Retrieval Using Multispectral and Thermal Remote Sensors

Despite the poor signal penetration abilities of multispectral and thermal sensors (compared to active microwave sensors) (Petropoulos, Ireland et al. 2015), methods of SMC retrieval using those sensors offer a broad range of satellites with decent spatial resolution and a multitude of bands to choose from (Yang, Guan et al. 2015)

Numerous multispectral indices of SMC estimations have been researched by a multitude of studies (Zhang, J., Zhou et al. 2014, Chen, Wen et al. 2015, Zhang, D., Zhou 2016, GE, ZHANG et al. 2018, Sha, Hu et al. 2018, Casamitjana, Torres-Madroñero et al. 2020). Therefore, the author of this research had to consider several constraints before choosing suitable methods. Indices that requires historical information were not considered due to the scale and nature of the study areas as well as their corresponding earth observation data (Landsat-8 data). That excluded monitoring methods based on vegetation indices such as Vegetation Condition Index (Kogan 1995), and Anomaly Vegetation Index (Weiying, Qianguang et al. 1994), as well as monitoring methods based on land surface temperature such as Thermal Inertia (Lei, Bian et al. 2014), and Vegetation Temperature Condition Index (Wang, Peng-xin, Li et al. 2001). On the other hand, vegetation water indices like Normalised Difference Water Index (Gao, Bo-Cai 1996), and the improved Normalised Multi-band Drought Index (Wang, Lingli, Qu 2007), were not considered as well due to their reliance on simple combinations of bands reflectance which offers a limited representation of the effect of SMC levels on reflectance (Zhang, D., Zhou 2016). Instead, the author of this research considered SMC estimation methods through surface reflectivity feature space as they have shown a lot of promise, especially for drought monitoring, not to mention that these methods offer a more nuanced representation of the relationship of SMC and surface reflectance. These methods are also a valid indicator of SMC, and their validation has been the subject of numerous studies (Ghulam, Qin, and Zhan 2007, Ghulam, Qin, Teyip et al. 2007, Chen, Sun, Wang et al. 2019, Chen, Sun et al. 2019). The same can be said about SMC estimation methods through vegetation indices/land surface temperature feature space. These methods exploit the relationships of SMC with vegetation cover intensity and land surface temperature to produce indices characterised by a proven linear relationship with SMC (Wang, Changyao, Qi et al. 2004, Gao, Zhiqiang, Gao et al. 2011, Chen, J., Wang et al. 2011, Rahimzadeh-Bajgiran, Omasa et al. 2012, Du, Song et al. 2017, Aisyah, Kusratmoko et al. 2019). This inherent relationship has made this category of methods extremely useful especially in the presence of an intense vegetation covers (Zhang, X., Zhao et al. 2019).

It was due to these constraints, that the investigated SMC retrieval methods of this research were methods using surface reflectivity feature space, and methods using vegetation indices/land surface temperature feature space. Perpendicular Drought Index (*PDI*) and Temperature Vegetation Dryness Index (*TVDI*) are the most commonly used indices for each respective category.

2.4.1 Perpendicular Drought Index

2.4.1.1 Concept

(Vermote, Tanré et al. 1997, Ghulam, Qin, and Zhan 2007) have found that incident radiances in the violet, blue, and red wavelengths are potently absorbed by vegetation lamina tissues, whereas, the latter, actually reflects the near-infrared (NIR) wavelengths. High vegetation cover intensity signifies small reflectance in the Red band and high reflectance in the NIR bands. Due to the fact that absorption of the red range is saturated rapidly, the increase of vegetation covers intensity can only be reflected by the increase of reflectance in the NIR region. The reflectance of bare soil is typically high in red to NIR spectral region; however, the presence of water content in bare soil results into a decrease in said reflectance, especially in the NIR domain. (Vermote, Tanré et al. 1997, Yang, Guan et al. 2015, Gao, Zhongling, Xu et al. 2013) have also found that plotting atmospherically corrected red bands pixels against their NIR counterparts results in a triangular spectral feature space that would represent vegetation cover and SMC conditions as depicted in Figure 9:

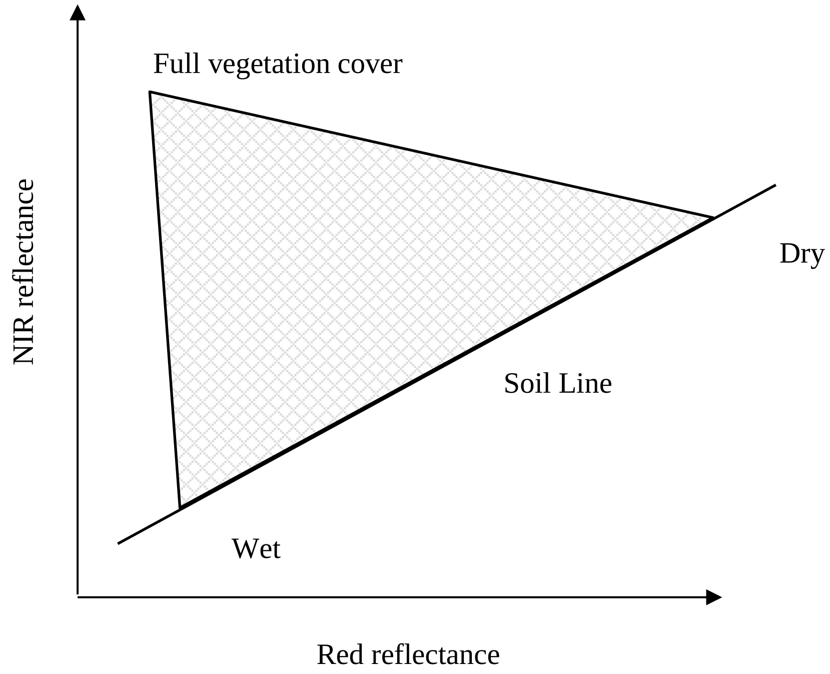

Figure 9. NIR/Red triangular feature space. Adapted from (Gao, Zhongling, Xu et al. 2013).

The soil line (bare soil) can be expressed using equation 16 (Ghulam, Qin, and Zhan 2007):

$$\boldsymbol{R_{NIR} = MR_{red} + I} \tag{16}$$

Where R_{red}, R_{NIR} are atmospherically corrected surface reflectance derived from red and NIR bands respectively, and M and I are the slope and the intercept of the soil line respectively in the NIR-red feature space (Shahabfar, Eitzinger 2011).

(Ghulam, Qin, and Zhan 2007) maintained that any mathematical operation that strengthens the contrasts between NIR and red could be used to express the vegetation surface drought status and distinguish bare soil pixels information from that of vegetated pixels. (Ghulam, Qin, and Zhan 2007) proposed designing an orthogonal axes system, above the aforementioned triangular feature space, expressed by an index named Perpendicular Drought Index (*PDI*). In order to fully comprehend the concept of *PDI*, the reader needs to observe Figure 10, where:

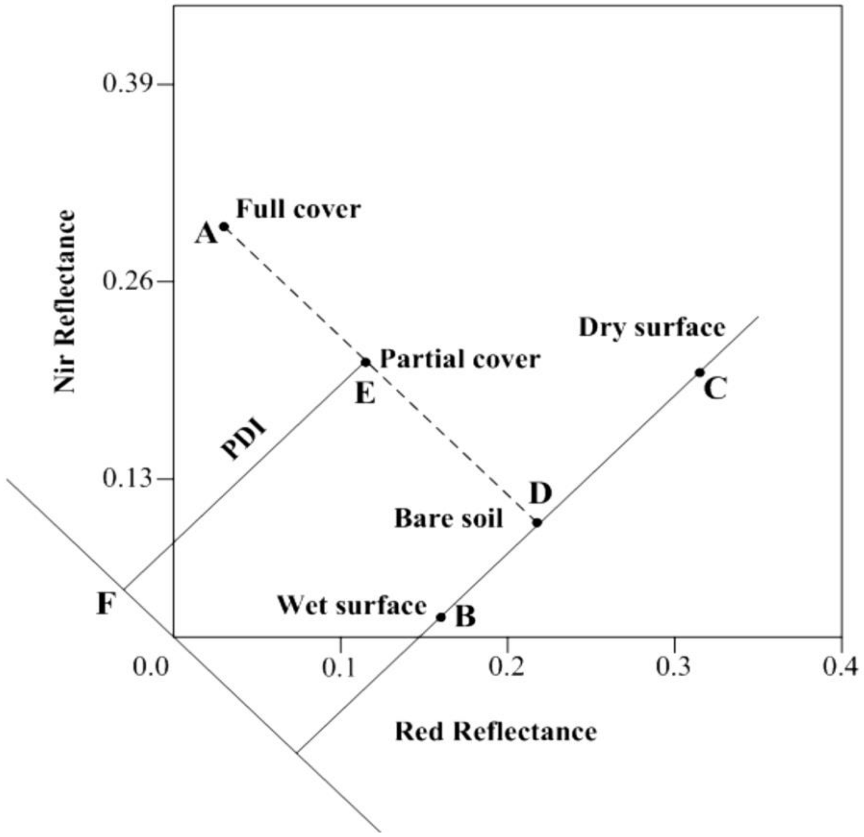

Figure 10. Definition of the *PDI.* Adapted from (Ghulam, Qin, and Zhan 2007).

A D line is a representation of the change in terms of vegetation cover intensity from full (A), partial (E) to bare soil in (D). BC is a line depicting SMC levels from a wet surface (B) and semi-arid (D) to completely dry surface in (C). BC is also referred to as the soil line as it demonstrates the direction of drought severity. *F* is the line perpendicular to the soil line while dissecting the coordinate origin and parallel to the AD line. *PDI* is the vertical distance from any random pixel point from to line F and the mathematical formula for *PDI* can be written using equation 17 (Ghulam, Qin, and Zhan 2007):

$$\boldsymbol{PDI} = \frac{\mathbf{1}}{\sqrt{\boldsymbol{M}^2 + \mathbf{1}}}(\boldsymbol{R_{red}} + \boldsymbol{MR_{NIR}}) \tag{17}$$

PDI can be a great descriptor of the levels and distribution of SMC in the NIR/Red triangular feature space with points far from the normal line $\boldsymbol{F}$ represent dry surfaces, and points near said line are correspondent to wet surfaces (Shahabfar, Eitzinger 2011). *PDI* is normalized, and it varies between 0 and 1 with 0 being akin to low water stress and one being extreme water stress (Ghulam, Qin, and Zhan 2007).

2.4.1.2 Limitations of SMC Estimation through PDI

Since *PDI* is heavily dependent on the NIR-Red reflectance, any variability caused by biophysical features of soil, i.e. soil surface colour, and vegetation types and conditions, would have a significant effect on the index, which means each study area would require its own local calibration to obtain its correspondent coefficient ***M*** (slope of the soil line) (Ghulam, Qin, and Zhan 2007, Ghulam, Qin, Teyip et al. 2007)

PDI performs at its best at low vegetation presence/ bare soil applications whereas, in areas with surface covers types varying from bare soil to densely vegetated surfaces, its performance seems to suffer, not to mention its susceptibility to cloud presence and to surfaces with non-flat topography (Shahabfar, Eitzinger 2011).

PDI, even with its previously mentioned limitations, is still a very valid indicator of SMC levels due to its linear relationship to SMC. However, its limitations have prompted the author of this research to seek another index to counterbalance those limitations, along with the limitations of the IEM inversion. The chosen index for that is the Temperature Vegetation Dryness Index.

2.4.2 Temperature Vegetation Dryness Index

To fully understand the synergetic use of remote sensing observations made by thermal and multispectral imageries to indirectly measure surface SMC from surface temperature (LST or T_s) and vegetation indices, it is imperative to define vegetation indices as well as a theoretical basis and biophysical properties of the *LST*/*VI* relationship.

2.4.2.1 Vegetation Indices

Vegetation Indices (VI) are arithmetical combinations of certain spectral bands with the main goal being the distinction between various vegetation properties (canopy biomass, absorbed radiation, chlorophyll content). It is notable that the vegetation reflectance while low at the Blue and Red regions of the visible spectrum, it is at its peak in the Green region, and even greater in the invisible Near Infrared (NIR) (Purevdorj, Tateishi et al. 1998).

The most popular and derivable combinations of tri-band multispectral sensors are Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI) and Soil Adjusted Vegetation Index (SAVI) (Candiago, Remondino et al. 2015). Table 1 gives us a brief preview of these vegetation indices:

Table 1. **Vegetation indices. Adapted from (Candiago, Remondino et al. 2015).**

Index	Calculation	Use
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{R_{NIR-}\ R_R}{R_{NIR+}\ R_R}$	Distinguishes between vegetated and non-vegetated features.
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{R_{NIR-}\ R_G}{R_{NIR+}\ R_G}$	Detects Chlorophyll concentration, Leaf Index Area and biomass.
Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{R_{NIR-}\ R_R}{R_{NIR}\ +\ R_R + L} * (1 + L)$	Eliminates the effect of soil while observing vegetation cover (where soil surface is exposed).

R_{NIR}, R_R and R_G represent the reflectance in NIR, Red, Green bands, respectively, L is a constant empirical value related to the vegetation density on the ground (Candiago, Remondino et al. 2015).

NDVI values range from -1 to 1, negative values represent non-vegetation features like water, barren areas, ice, snow or clouds, however, the common range for green vegetation is 0.2 to 0.9, values from 0.2 to 0.3 are bushes and grasslands while values 0.4 to 0.9 are forests and crops (Pettorelli, Vik et al. 2005).

GNDVI values range from 0 to 1; it displays a higher level of sensitivity to chlorophyll concentration than NDVI (Hunt, Hively et al. 2008).

SAVI is used to minimise soil noise while observing the vegetation, it is especially useful in areas where soil surface is exposed, and vegetative cover is minimal, its values range from -1 to 1 with lower values signifying poor cover of green vegetation. L is a calibration parameter with values also range from 0 to 1, with L =0 indicating very high vegetation cover (SAVI = NDVI), while L= 1 points are representative of areas with no green vegetation at all. In general, L=0.5 is optimal in most situations (Huete 1988).

Figure 10 displays prescription maps using NDVI, GNDVI and SAVI (L = 0.5) values of a tomato field near the village of San Bartolo, in the municipality of Ravenna (Italy). Areas A and B show zones with high and low VI values, respectively:

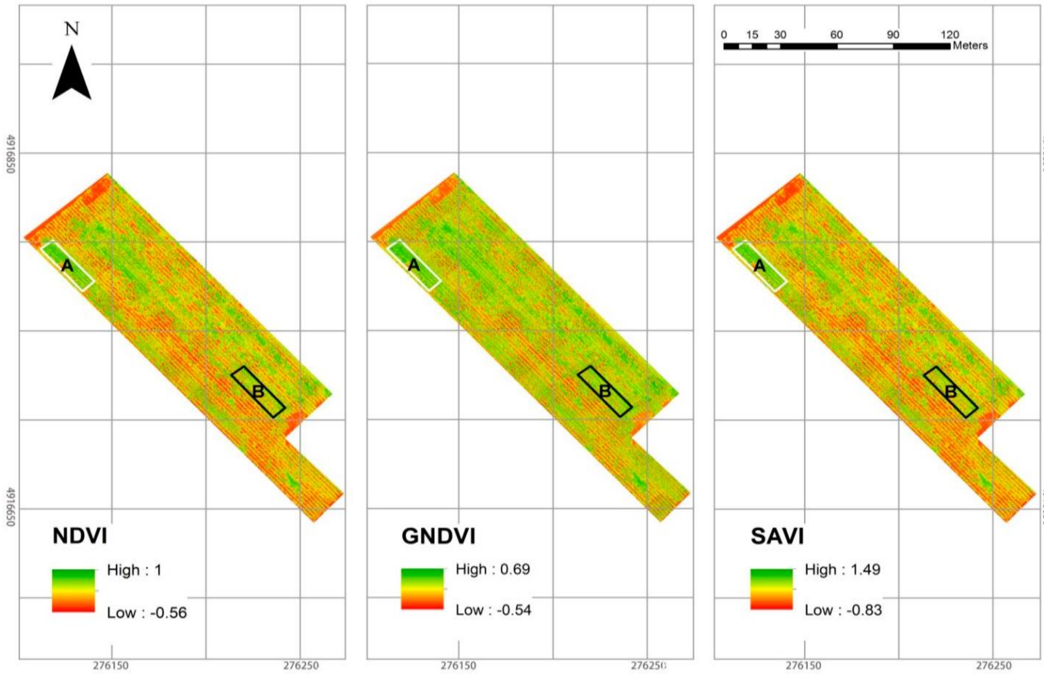

Figure 11. Prescription maps using vegetation indices values (Candiago, Remondino et al. 2015)

Figure 10 further highlights the effectiveness of vegetation indices in determining the intensity vegetation cover, which is instrumental in SMC retrieval in the *LST*/*VI* space.

2.4.2.2 TVDI Concept

The use of this index requires the assumption that the relationship between SMC levels, the intensity of fractional vegetation covers and *LST* is simplified and minimized into a two-dimensional scatter plot in which data points take the form of triangle/trapezoid (Sandholt, Rasmussen et al. 2002, Yang, X., Wu et al. 2008). Changes in SMC levels are plotted as a function of surface temperature and fractional vegetation cover (which can be expressed by Vegetation Indices) (Lambin, Ehrlich 1996). The difference in radiative temperatures between soil and vegetation canopy affects *LST*. Evapotranspiration is another factor influencing surface temperature through the energy balance at the surface (Wang, Chengbin, Chen et al. 2019). The available energy for sensible heating of the surface increases whenever there is a decrease in evapotranspiration due to stomatal resistance to transpiration which is controlled by soil moisture availability (Gao, Zhiqiang, Gao et al. 2011). Consequently, the combination of fractional vegetation cover and surface temperature allows the estimation of SMC from bare soil to full vegetated covers (Petropoulos, G., Carlson et al. 2009). This research uses the

Temperature Vegetation Dryness Index (*TVDI*) to obtain information on surface SMC via the *LST*/VI triangular space as depicted in Figure 12:

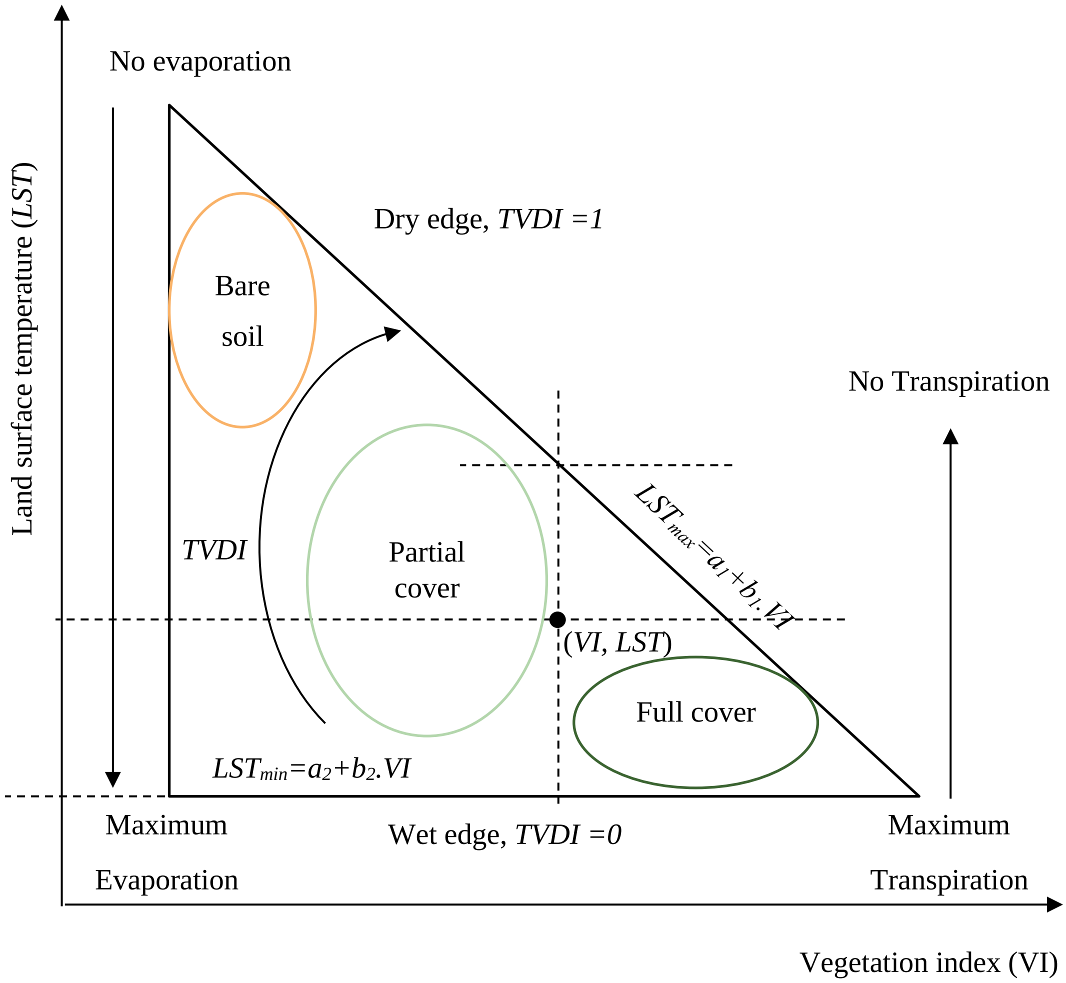

Figure 12. Definition of the *TVDI* in the *LST/VI* feature space. Adapted from (Sandholt, Rasmussen et al. 2002, Lambin, Ehrlich 1996).

LST_{max} is the maximum surface temperature observation for a given *VI* value. 'a_1' and 'b_1' are respectively the intercept and the slope of the linear dry edge, and 'a_2' and 'b_2' are respectively the intercept and the slope of the linear wet edge, *TVDI* is expressed by equation 18:

$$\boldsymbol{TVDI = \frac{LST - LST_{min}}{LST_{max} - LST_{min}}} \tag{18}$$

Where *LST* is the observed surface temperature (in Kelvin), at a random pixel, and LST_{min} represents the wet edge denoted $LST_{min} = a_2 + b_2VI$, and LST_{max} represents the dry edge denoted $LST_{max} = a_1 + b_1VI$.

TVDI values range from 0 to 1, where 1 indicates low levels of SMC, and 0 indicates maximum evapotranspiration and water access which signify high SMC levels. (Sandholt, Rasmussen et al. 2002), compared *TVDI* values to simulated soil moisture levels from the distributed hydrological model based on the MIKE SHE distributed hydrological model (Abbott, Bathurst et al. 1986), finding that SMC and *TVDI* have a relationship that could be represented by a linear function (m_v= x*TVDI*+y) easily calculated using linear regression (R^2=0.7). Since then, different versions of *TVDI* were validated using in situ measurements in various studies which produced promising results (RW.ERROR - Unable to find reference:doc:5a520907e4b08e15c00cb525, Zhu, Jia et al. 2017, Chen, Wen et al. 2015). Therefore, it is safe to assume that *TVDI* is considered as a valid indicator of SMC levels, which can be used as an addition to the previously discussed methods (*IEM* inversion and *PDI*).

2.4.2.3 Limitations of SMC Estimation Through TVDI

TVDI suffers from few sources of error that can reduce the accuracy of its SMC retrieval abilities, and those sources can lead to a few performance issues such as:

- The loss of the spatial and temporal variability of SMC due to the poor spatial and temporal resolution of the satellites suitable for this method (Chen, Wen et al. 2015).
- The potential incorrect determinations of the “triangle” from satellite data without huge data grids of large scale areas; the observed area of interest may not always include the full range of spatial variability in terms of land surface conditions such as dry bare soil, wet bare soil, vegetation exhibiting water stress and well-watered vegetation (Sandholt, Rasmussen et al. 2002). That can lead to difficulties calculating the ideal dry and wet edge due to local specific factors like vegetation species, topography, net radiation and cloud presence, which makes the aforementioned edges subjective to their datasets (area of interest) (Yang, X., Wu et al. 2008, Cho, Lee et al. 2014).
- The susceptibility to errors of estimation in terms of *LST* due to atmospheric effects and illumination effects (shadows) not to mention that *TVDI* only accounts for SMC in the top surface layer (no surface penetration) (Petropoulos, G., Carlson et al. 2009).

Despite the preceding limitations, the *TVDI* methodology can still be considered as another valid descriptor of SMC level for this research. Its robustness for applications over large areas as well as its insensitivity to surface cover type (Sandholt, Rasmussen et al. 2002) can be valuable in SMC estimation pipeline proposed by this research, as it will reduce any inaccuracies caused by the limitations of IEM inversion and *PDI* in intensely vegetated areas of interest. To achieve that, the author of this research has opted to investigate if the use of data fusion techniques would successfully ameliorate SMC retrieval accuracy.

2.5 **Data Fusion Techniques**

Multi-sensor data fusion allows the combination of data gathered from different sensors and related information, to achieve improved accuracy and better specific estimations unattainable by the use of a single sensor (Hall, Llinas 1997). The resulted perception can be instrumental for optimal control of information for informed decision making. Data fusion potential for pattern recognition, visual enhancement, object detection and area surveillance (Dong, Zhuang et al. 2009). Multi-sensor data fusion in the remote sensing sense can be performed on 4 different processing levels, depending on the phase that fusion takes place: signal level, pixel level, feature level and decision level (Dai, Khorram 1999).

In the signal level, fusion Signals from different sensors are merged to form a new signal with better signal to noise ratio than the initials signals (Stathaki 2011).

In the pixel level fusion, each pixel of the fused image is determined from a set of pixels from different images with the goal of improving spectral or spatial resolution. Popularly used techniques are Intensity Hue Saturation (IHS) transformation (Nasr, Ramadan 2008), Gaussian Pyramid (Olkkonen, Pesola 1996), Wavelet-based image fusion (Amolins, Zhang et al. 2007), Principal Component Analysis (PCA) and Brovey Transform (Nikolakopoulos 2008).

In the feature level fusion, salient features are extracted from each sensor in question to create what is called a feature vector, and the latter can be used later for classification or decision making. Some popular techniques employed at this level are Artificial Neural Networks (ANN) (Jiang, Yang et al. 2004a), Cluster Analysis, Bayesian Inference (Zeng, Zhang et al. 2006).

In the decision level fusion, each sensor image is processed independently, features are extracted, and decisions are made separately. These decisions can be used to negate or validate each other. In the end, they are fused into a final decision using a few techniques like Fuzzy Logic, Expert Systems, Dempster-Shafer theory, Voting Strategies (Zeng, Zhang et al. 2006).

2.6 Multi-Sensory Data Fusion for Soil Moisture Content Estimation

Data fusion in the soil moisture content sense has been the subject of a plentiful amount of research. (Kurucu, Sanli et al. 2009) performed an image fusion, or more specifically the IHS-transform method, of images from multiple multispectral bands (SPOT-2) and radar images from Radarsat-1. This fusion was validated using 135 soil samples, 80 samples of bare soil and 55 samples of soil containing wheat and barley. SMC values were extracted using the oven method, the relationship of reflectance to SMC presence as well as soil texture was analysed using the Hydrometer method. The authors found that the reflectance corresponding to soil samples with high clay composition was correlated the most with SMC (0.72), pixels corresponding to silt dominant soil samples exhibited a weak correlation, and finally, pixels of soil samples containing the high presence of sand has a strong negative correlation (-0.7). The authors concluded that the contribution of SAR has decreased the inherent sensitivity of reflectance of the pixel corresponding to the samples containing plants.

(Bai, L., Long et al. 2019) have used a different approach to image fusion. Thermal and multispectral data from MODIS and LANDSAT-8 satellites had been downscaled and merged to be used in the trapezoidal method, the latter was then used to infer SMC. The downscaling methods in question were High-resolution Urban Thermal Sharpener and Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model. That study concluded that the resulted downscaled LST was highly consistent with in-situ measurements, producing less RMSE ranging from 0.73 to 2.75 K, which in turn, has yielded an SMC estimation with a decreased RMSE (from 4.8% to 3.8%).

(Moran, Hymer et al. 2000) proposed a feature level data fusion as well. The fused data in question were features derived from ERS-2 C-band SAR data and optical data from Landsat TM. The proposed methodology was tested and validated on 3 semiarid regions in 8 different acquisition dates. The authors in (Moran, Hymer et al. 2000) elaborated on a model proposed by (Sano, Qi et al. 1998) which consisted of exploring the sensitivity of radar backscattered signal to surface roughness, vegetation cover presence which is expressed by Green Leaf Area Index (GLAI), and SMC through a linear regression estimation. The elaboration was comprised of the idea of investigating the empirical relationship between the difference between any given SAR image and the backscatter from a dry season image $\left(\sigma^0 - \sigma^0_{dry}\right)$, and

GLAI. The $(\sigma^0 - \sigma^0_{dry})$ feature was found to be in better agreement with SMC (R^2=0.93) than its corresponding σ^0 counterpart (R^2=0.27). Furthermore, that empirical relationship was demonstrated through plotting $(\sigma^0 - \sigma^0_{dry})$/GLAI regression line. The vertical distance of any given point from said line was found independent of surface roughness, and it had a linear relationship with surface SMC of each study area. This was especially true for pixels representing areas with sparse vegetation cover (GLAI<0.35). Conversely, the relationship was not sensible for pixels with values as a change of 25% in SMC resulted only in a change of 3 dB in σ^0. The authors in (Moran, Hymer et al. 2000) reported that this methodology yielded an absolute error of an average of 2.51% across all of datasets. The authors also raised concerns about overall insensitivity of SAR to low SMC values, and about the overall accuracy of GLAI estimation using standard optical remote sensing algorithms.

(Notarnicola, Posa 2001) explored another avenue, the authors proposed a Bayesian fusion, which can be safely considered as a feature level fusion of passive and active microwave data to estimate soil moisture in bare soil, features were derived from RASAM truck-mounted radiometer-scatterometer operating at a frequency of 4.6 GHz. The features in question were the backscattering coefficient (IEM) and emissivity through the Wang model for emissivity (WANG, JAMES R., CHOUDHURY 1995), the features were introduced to Bayesian parameter estimation, and the estimated parameters were the dielectric constant and surface roughness parameters, with an emphasis on the dielectric constant (given its polynomial relationship to SMC). This specific fusion has yielded a 10% estimation error.

(Posa, Notarnicola et al. 2004) elaborated on their approach by comparing the efficiency of their Bayesian approach (BAY) to Artificial Neural Networks (ANN). Two distinct sets of data were used in this study. The first set is composed of backscattering coefficient and emissivity were extracted from several configurations of the data measured by a truck-mounted radiometer-scatterometer. The second set is composed of the backscattering coefficient which was extracted from C-band scatterometer data. The accuracy of the results was expressed by the coefficient of correlation (R), Root Mean Squared Error (RMSE), and the standard deviation (SD). The most accurate estimation for the first set was achieved using a configuration of 2 frequencies (4.6 GHz and 2.5 GHz for the scatterometer and radiometer respectively) and one co-polarisation (HH). ANN outperformed the Bayesian approach in terms of RMSE (0.48%) but seems to slightly underperform in terms of correlation (0.83 for ANN and 0.84 for BAY). As for SD, since the authors did not provide the relevant metrics of the ground truth, it is difficult to analyse the resulted SD, however, the SD of the Bayesian method seems to be more plausible 2.73 % than that of the ANN 0.45%. it was also noted that that there is a major discrepancy when it comes to Bay method with values of the dielectric constant exceeding 18-20 (there was a major focus on the real part of the dielectric constant). The results of the second set of data indicate a better performance from ANN in terms of both R and RMSE (0.84 and 3.28% respectively) than those of the Bayesian approach (0.68 and 5.08%). As for the SD, the same as in the first group, it is quite difficult to discern whether which method performs better. It was also found that Bay overestimate the dielectric constant for the whole dataset while ANN only overestimates data associated with rough fields with an average bias of 5%.

(Van der Schalie, De Jeu et al. 2018) investigated the effect of three different data fusion approaches for SMC retrievals. These approaches consisted of the fusion of features from extracted 10 years of passive microwave data (2003-2011), which were generated from Advanced Microwave Scanning Radiometer Earth Observation System (AMSR-E) at multiple polarisations and frequencies and validated by data from SMOS satellite (2010-2013). The approaches in question were ANN, regression, and Land Parameter Retrieval Model (current baseline algorithm for passive microwave component in the ESA-Climate Change Initiative-Soil Moisture). For comparison and evaluation, an active microwave satellite was also used for SMC estimation, the satellite in question was the Advanced Scatterometer (ASCAT) and algorithm used for SMC retrieval was the Change Detection Algorithm. To address the limitations of each of the proposed methods in this study, two methods were used, a large-scale precipitation-based validation technique which represents the anomaly correlation-based skills in satellite SMC estimation through data assimilation with precipitation data derived from Tropical Rainfall Measuring Mission (TRMM), and the Triple Collocation Analysis (TCA). NDVI data derived from MODIS were also used for analysis and evaluation. In terms of correlation, the ANN approach seems to outperform land parameters retrieval algorithm and regression, especially at intense vegetation cover (R=0.7 at NDVI=0.7). On the other hand, ANN also outperforms all other approaches when it comes to Unbiased Root Mean Square Difference across the whole range of vegetation intensity with its best performance recorded at NDVI=0.3 to 0.7, ubRMSE=1.9%. While LPRM seems to produce its lowest RMSE at low vegetated areas (NDVI ≤0.45, ubRMSE=2%). The regression approach, on the other hand, performs its best at bare to sparsely vegetated areas (NDVI≤0.1, ubRMSE=3%). SMC estimation using the ASCAT approach is consistent with SMOS throughout a wide range of

NDVI values, it does, however, records its best performance at NDVI values around 0.48 to 0.6 with an ubRMSE=2%.

(Semmens, Anderson et al. 2016) went in a different direction, the authors were interested in using a multi-sensor feature level fusion approach to monitor daily evapotranspiration (ET) over two California vineyards using Landsat-8 during 2013 growing season, leading into the drought in early 2014. The authors proposed using Spatial and Temporal Adaptive Reflective Fusion Model (STARFM) to merge data from multiple satellites with a multi-scale ET retrieval algorithm based on Two-Source Energy Balance (TSEB) and land surface representation for daily ET computation at 30 m resolution. The proposed system by (Semmens, Anderson et al. 2016) consist of running TSEB using thermal imagery band in the Geostationary Environmental Operational Satellites (GOES, 4 km spatial resolution, hourly temporal sampling), data from Moderate Resolution Imaging Spectrometer (MODIS, 1 km spatial resolution, daily acquisition), and data from Landsat-8 satellite (resampled to 30 m resolution, 16 days revisit time). The features suggested by this study were LST, LAI, and Albedo from Landsat-8, LST, geolocation, LAI, albedo, and NDVI from MODIS, LST, LAI from GOES, and meteorological features like vapour pressure, wind speed, air temperature and Insolation. All features were disaggregated to 30 m spatial resolution than fed to the STARFM. This fusion approach produced RMSE in the order of 0.92 and 0.96 mm/day compared to ground measurements from flux tower sites in irrigated fields with 8 and 5-year-old pinot noir vines, respectively. RMSE was then reduced to 4.93 and 5.76 mm/week at the weekly timestep which was relevant for the irrigation process of both fields. The authors in (Semmens, Anderson et al. 2016) reported a model overestimation of ET especially in the early season in the younger vineyard which was believed to be due to issues with the model parameterisation of canopy architecture which is commonly caused by inter-row grass cover crop.

Authors in (Park, Im et al. 2017) suggested another version of a feature level fusion. The goal was to design a High-Resolution Soil Moisture Drought Index (HSMDI) for drought monitoring in the Korean Peninsula. The proposed approach consisted of downscaling Tropical Rainfall Measuring Mission satellite data (precipitation 25 km), AMSR-E data during the period of 2003-2011 (soil moisture 25 km), and combining features from MODIS (LST, NDVI, Enhanced Vegetation Index, Albedo, LAI, and ET 1 km). The downscaling was achieved using machine learning, more specifically a random forest algorithm, and the approach was validated by in-situ soil measurements. The resulted SMC estimation was normalised for each pixel to produce the proposed index. The downscaled 1 km soil moisture estimation (up to 1 km) was correlated to both AMSR-E and in-situ measurements with a mean coefficient of determinations R^2=0.29 and 0.59, respectively. HSMDI produced encouraging results, it exhibited a high correlation with crop yield data, especially in non-irrigated regions containing the highland radish and Napa cabbage cultivated with a mean R^2 of 0.77.

Authors in (Portal, Vall-Llosscra et al. 2018) another feature level fusion of features derived from microwave, optical, and thermal data to map SMC at relatively high resolution. The microwave data were generated from SMOS (at 40 km spatial resolution), and optical and thermal data were generated from MODIS (1 km). The features in question were normalised brightness temperatures at horizontal and vertical polarizations along with their corresponding incident angles from SMOS, and NDVI and LST from MODIS. The SMC ground truth data used for validation were from two in-situ stations: REMEDHUS in Spain and OzNet in Australia. The proposed approach consisted of downscaling the low-resolution SMOS data through the multiple linear regression of the aforesaid features to achieve SMC data at 1 km. The values of LST pixels of MODIS containing clouds are replaced by values from the fifth generation of the ECMWF atmospheric reanalysis (ERA5). The approach produced an SMC estimation with an average R of 0.8 and RMSE of 7%.

(Xu, Yuan et al. 2019) have also proposed a feature level fusion. The authors of this study developed a system to estimate regional SMC in the continental U.S by employing a Generalised Regression Neural Network (GRNN). The latter was trained using a sparse ground-based measurement from Sparse Network Stations (SNS). The use of GRNN allowed establishing a nonlinear relationship between passive microwave observations from SMAP satellite and the measurements mentioned above in the period from April 2015 to March 2018. The scale mismatch occurring from the small spatial support of ground-based measurements was rectified by the exploitation of the extended triple collocation method, which ensured the reliability of generated data from individual sparse network stations at SMAP coarse footprint scale could be verified before fed into GRNN. The data associated with the collocation included Ground-based SMC (SNSs), Model-based SMC (ERA-Interim SMC simulations at the top 7 cm), and Satellite based SMC (SMAP). To guarantee the most accurate validation measurement possible, only ground station with a correlation coefficient of 0.7 and up were considered reliable. Then, the extracted features were introduced to the GRNN, with the features being SMAP brightness temperatures in both polarizations, surface soil temperature from GEOS-5 model, vegetation water content from MODIS, Month, Latitude, and longitude.

The choice of GRNN was justified by the authors due to its improved ability when it comes to overfitting compared to a conventional feed-forward neural network. the authors employed a 10-fold cross-validation method. The GRNN model had a promising performance in terms of R=0.88 and ubRMSE=5%, and R=0.74 and ubRMSE =7.1% when samples were cross-validated. Authors in (Yuan, Xu et al. 2020) elaborated on this exact approach by developing a point-surface collaborative method inversion to estimate regional SMC instead. For SNSs, ground stations with a correlation coefficient of 0.7 were the only ones considered in the calculation, which resulted in only 40% of the overall number of those stations fitted that threshold (372). The approach was compared to a traditional Back Propagation Neural Network and GRNN model had a better performance in terms of accuracy when cross-validated (R=0.88 and RMSE=5%) than its traditional counterpart (R = 0.8, RMSE= 6.3%).

(Huang, Liang et al. 2019) proposed the fusion of multiple Global Navigation Satellite Systems (GNSS) through a multiple Least Square regressions. The authors investigated the relationship of SMC and Signal to Noise Ratio (SNR) which is the quality of the signal received by the antenna. The latter is influenced by numerous factors which include antenna gain parameters, multipath effect, and random noise of the receiver. The idea is to benefit from the aforesaid relationship of the SNR of multipath reflections with SMC by considering the relative phase delay of multipath reflections as measures to estimate the fluctuations in SMC levels. GNSS monitoring data were generated from the P041 from the Plate Boundary Observatory (PBO) network. The station provides a high sampling rate and plentiful meteorological data from the GPS carrier (L1 and L2). In 2011, the Global Positioning System (GPS) carrier L2 observation data included high-quality L2 band SNR observation data. The authors fused features from data derived from a combined total of 5 GPS satellites in the duration of 220 days (from day 70 to 290 in 2011). Then, those features were used as the input of a multivariant linear regression model. The fusion has produced improvements in terms of correlation coefficient R, from 0.73 in a single satellite to 0.89 and using the fusion all 5 respectively. However, in terms of RMSE, one of the GPS satellite has produced an estimation with less RMSE (7%) than that of the fusion of all 5 GPS satellites (13.1%).

Authors in (Ren, Liang et al. 2019) have elaborated on this approach by proposing using a different fusion model, Least Square Support Vector Machine (LS-SVM). The authors, this time, fused features from data generated from 32 GPS satellites in 222 days (from day 73 to 294 in 2015). Then, those features were fed into a sliding method LS-SVM estimator. This particular estimator has produced better results as the inclusion of data from multiple satellites has improved the overall performance of the SMC estimation by minimising the multi-path effect. The fusion has produced improvements in terms of correlation coefficient R and RMSE, from 0.74 and 7.2% in a single GPS satellite to 0.94 and 4.4% using all 32 respectively.

It is quite observable that there is no shortage of SMC estimation related research, especially through a feature level fusion. The fusion of multiple parameters relevant to SMC almost always produces better accuracy than estimations produced by a single parameter. However, the author of this research had identified that the most salient gap in terms literature concerning multisensory data fusion for soil moisture content estimation, lies fundamentally in the decision level fusion aspect of it, or in this case, estimation level. This research also concentrates at the regional application of such estimations, as the eventual purpose of this research is to be used for agricultural practices, which renders the use of data from mission purposed satellites such SMOS, SMAP, and GNSS data irrelevant. That will also rule out the use of in-situ stations for validations for the same reason. When these constraints are considered, there has not been a piece of research dedicated to the design of an estimation level fusion system for soil moisture content estimation especially on the regional scale. Also, to a lesser extent, the combination set of the selected features for this research, i.e. *IEM* inversion parameters (this specifically modified version of IEM), *PDI*, and *TVDI* have never been fused in the same SMC estimation system. Not to mention, that all of the studies discussed above have used ground-based stations for the validation of their respective estimations, which makes the fashion of the validation using this specific set of ground-level measurements (spatial resolution at 30 m) also another novelty, especially at these study areas in the UK, and especially in Algeria, which has no studies dedicated to SMC estimation using earth observation data.

To reiterate, this research proposes a novel SMC estimation pipeline which is comprised, simultaneously, of multiple processing levels of data fusion:

- Novel feature level fusion, where parameters from IEM inversion, *PDI*, *TVDI* are extracted and fed to an ANN to estimate SMC.
- Decision level fusion, where SMC estimations from each methodology are fused using a weight-based system.

The full comprehensive description of the proposed SMC estimation system will be provided in chapter 3, Soil Moisture Content Estimation System – Conceptual Design.

2.7 Conclusion

Chapter 2 featured a literature review of the SMC retrieval methodologies relevant to the design of the SMC estimation system proposed in this research.

First, definitions of SMC and different aspects of its effects on remote sensors in terms of dielectric constant, reflectance and land surface temperature were provided.

Then, operational estimations using mission purposed systems like SMOS and SMAP demonstrated limitations in terms of spatial resolution and limited their applicability to global to large scale applications rather than the use intended in this research (Agriculture).

Afterwards, various SMC retrieval methods were explained, each of those methods, whether be it the *IEM* inversion, *PDI* or *TVDI*, have exhibited few factors impacting their performance in terms of accuracy.

Finally, this chapter concluded by the depiction of data fusion techniques as a possible solution to minimise the inaccuracies caused by those performance-related issues. The relevant studies investigating these approaches, as well as the contrast between the approach proposed by the author of this research to the approaches proposed by these studies.

3. SOIL MOISTURE CONTENT ESTIMATION SYSTEM – CONCEPTUAL DESIGN

3.1 Introduction

Chapter 3 outlines the motivations behind the proposed SMC estimation system, not to mention the different limitations encountered by each used method in the aforesaid system, and it also provides a detailed description of an alternative version of IEM inversion as well as the reasons behind the inclusion of a semi-empirical parameter *Lopt*. Then, the soil moisture content estimation is introduced along with a comprehensive explanation of each component of the system, from pre-processing and *PDI* and *TVDI* determinations to the feature level fusion and decision level fusions.

3.2 Motivations Behind Using Data Fusion

In order to understand the reasons behind the selection of the used data fusion techniques, it is important to recall the corresponding limitation of each methodology by observing Table 2:

Table 2. **Comparison of the different used methods of SMC retrieval.**

Group of sensors	Method	Advantages	Disadvantages
Microwave active	Integral Equation Model inversion	- The high spatial resolution of SAR sensors and their independence to clouds presence and night time (Sikdar, Glavic et al. 2004). - Offer SMC information at the deeper surface layers (depending on frequency) (Verhoest, Lievens 2013).	- Speckle (Moreira, Prats-Iraola et al. 2013). - Insensitivity to SMC when it exceeds 35% (Baghdadi, Nicolas, Zribi 2006). - Sensitivity to surface roughness (Verhoest, Lievens et al. 2008). -Sensitivity to the intensity of vegetation cover (Bindlish, Barros 2000). - Coarse temporal resolution (Bindlish, Barros 2000).
Multispectral	Perpendicular Drought Index	- Good spatial resolution. - Simple implementation. - Good performance in bare soil and low vegetated areas (Ghulam, Qin, and Zhan 2007).	- Course temporal resolution and susceptibility to cloudy conditions and surfaces with dense vegetation cover (Shahabfar, Eitzinger 2011). - The subjectivity of the soil line to its corresponding dataset (Ghulam, Qin, Teyip et al. 2007).

Synergetic method	Temperature Vegetation Dryness Index	-Good spatial resolution. -Simple implementation. - Robustness to the effects of partial to full vegetation cover (Yang, X., Wu et al. 2008).	-Susceptibility to cloudy conditions, illumination effects and atmospheric effects (Wang, Changyao, Qi et al. 2004, Yang, Guan et al. 2015). - The coarse temporal resolution (Chen, Wen et al. 2015). - The subjectivity of wet edges to their corresponding datasets (Sandholt, Rasmussen et al. 2002).

Table 2 can only further reinforce the argument for the necessity of data fusion techniques application. In addition to the IEM inversion, a multispectral index (*PDI*) will offer an additional source of SMC estimation for the proposed fusion scheme to minimise the effects of high SMC levels (>35%) and surface roughness values, and the addition of *TVDI* will offer an index resistant to vegetation covers. According to the definition of data fusion techniques provided in section 2.5, data fusion at this instance is only possible at the feature level and decision level (estimation level).

Prior to the presentation of the proposed SMC estimation system, it is noteworthy that a different version of the IEM from the one presented in section 2.3.2.1 had to be used. It was due to the fact that the values of surface roughness parameters (k.s>3) collected from one of the study areas are not within the range of valid surface roughness values for IEM as described in section 4.2. Therefore, certain measures had to be taken to ensure that the range of validity is increased in case of encountering study areas with similar surface roughness values in the future. Those measures were:

- The addition of a semi-empirical calibration parameter *Lopt* (Baghdadi, Holah et al. 2006).

- Implementation of an updated version of the IEM called the Empirically Adapted Integral Equation (EA-IEM) in (Song, Zhou et al. 2009).

3.3 The Updated Version of the IEM

(Baghdadi, Holah et al. 2006) proposed a semi-empirical calibration to IEM by replacing the estimation of correlation length *l* by an optimal calibration parameter called *Lopt*. The goal was to improve the agreement between the backscattering coefficients generated by SAR sensors and those estimated by the IEM. *Lopt* is dependent on RMS of the surface heights, incidence angle and polarization, and the parameters relevant to the configuration of this research are expressed by Equation 19 which was proposed by (Baghdadi, Nicolas, El Hajj et al. 2018) for Sentinel-1:

$$\boldsymbol{Lopt(s, \theta_i, VV) = 1.281 + 0.134(sin0.19\theta_i)^{-1.59}s} \qquad (19)$$

Where s is RMS of surface heights, θ_i is the incidence angle at pixel i and *VV* is the polarisation.

Lopt proved to achieve a better agreement between radar signal inferred from the IEM model and SAR data in the C-band (at 5.6 cm wavelength) for HH and VV polarization as well as for incidence angles (20° to 48°), with an improved validity range of (s < 4 cm) (Baghdadi, Nicolas, Chaaya et al. 2011), which in turn, would lead to a more accurate IEM inversion with a reduced bias and root mean square error in terms of the SMC estimation.

It is worth noting that the availability of ground truth for SMC is necessary to apply the model designed in (Hallikainen, Ulaby et al. 1985) (equation 2). However, the whole purpose of this research is to estimate SMC independent from the presence of ground truth. Therefore, the author was motivated to investigate the Empirically Adapted Integral Equation (EA-IEM) (Song, Zhou et al. 2009), especially the equations related to VV-polarization since it is the only co-polarized configuration available in the Sentinel-1 datasets. The idea is to infer dielectric constant directly from the active microwave backscattering coefficient using knowing that (Song, Zhou et al. 2009):

$$F_v = \frac{\sigma_{vv}^0}{\frac{k^2}{2}\, exp(-2k_z^2 s^2) \sum_{n=1}^{\infty} \frac{(2sk_z)^{2n} W^n(-2k_x, 0)}{n!}} \qquad (20)$$

$$\varepsilon_r = \frac{1}{\left(0.5 - \left(\frac{F_v s^{0.05} sin^{3.35}(\theta + 1.1)(Lopt - 0.049)^{[0.042+0.06 sin\,(\theta-1)]}}{106 exp\,(-1.996 s^2 k_z^2) tan^{0.46}(\theta + 0.32)}\right)^{\frac{5}{27}}\right)^{\frac{1}{\cos\,(1.02\theta - 0.2)}}} - 3 \tag{21}$$

Where F_v in equation 20, is calculated using the calibrated σ_{vv}^0 extracted from Sentinel-l, and the dielectric constant ε_r in equation 21, is calculated for the Gaussian surface correlation function, due to the rough nature of the surface height measurements, where the correlation length was replaced by *Lopt*.

3.4 The Proposed Soil Moisture Content Estimation System

In this research, a novel soil moisture content estimation system is proposed, and the novelty of this system lies primarily in the fusion aspect of different estimations provided by different methods in the feature and the decision level. The reason for the selection of the decision level fusion is to have different independent estimations in case of the absence of a data source.

Figure 13 elucidates a flowchart of the different components of such a system:

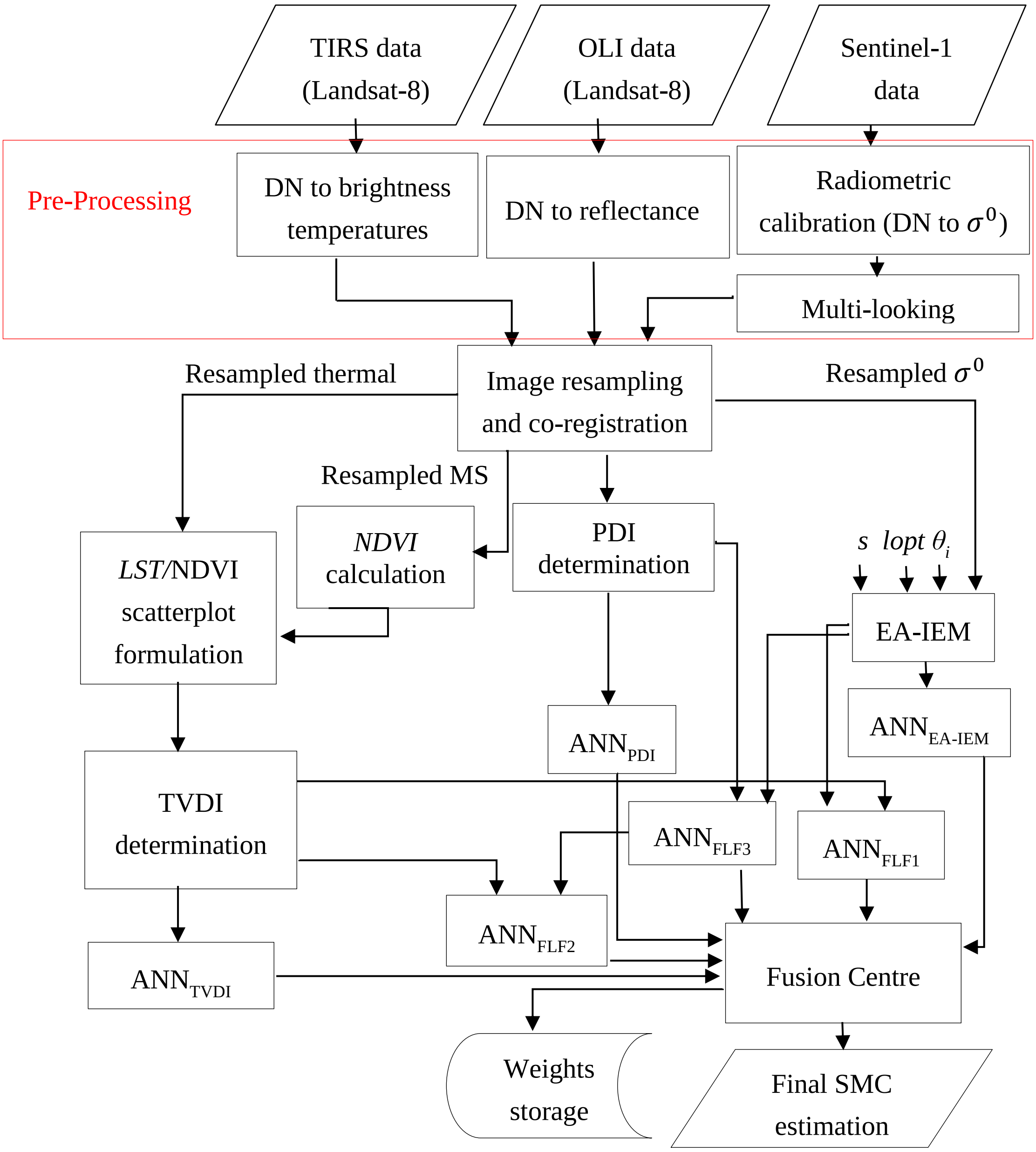

Figure 13. The proposed soil moisture content estimation system.

The components of this system are regrouped by functionality to facilitate their description; the groups are pre-processing, *PDI* and *TVDI* determinations, Feature Level Fusion and Fusion centre.

3.4.1 Pre-processing

All of the extracted EO data from Landsat-8 and Sentinel-1, respectively, are pre-processed using Sentinel Application Platform (SNAP). SNAP is a common architecture for all Sentinel toolboxes being jointly developed by Brockmann Consult, SkyWatch, and C-S. This architecture is ideally suitable for the analysis and processing of earth observation due to its inherent technological innovations which include, Extensibility, Portability, Modular Rich Client Platform, Generic EO Data Abstraction, Tiled Memory Management, as well as a Graph Processing Framework (STEP ESA 2015). Its various functionalities facilitated the module of pre-processing significantly.

Initially, SAR data undergoes radiometric calibration as well as multi-looking before it can be co-registered with multispectral and thermal data, which themselves, are transformed from digital numbers (DN) to reflectance and temperatures, respectively. Then, all resulting images are resampled to 30 metres, which is both the spatial resolution of OLI and the sampling distance of the ground truth SMC measurements points. The 30 metres is a good compromise between the spatial resolution of TIRS (100 metres) and the spatial resolution of Sentinel-1, which is (20.4 m x 24.5 m). Then, the resampled MS image is used to calculate *NDVI* and *PDI*, and to formulate the *LST/NDVI* feature space scatterplot when combined with the resampled thermal image. On the other hand, the resampled image of the backscattering coefficient σ^0, along with the RMS height (s), the calibrated parameter (*lopt*) and incidence angle θ_i, are used to calculate simulated backscattered coefficient, which is the product of EA-IEM.

3.4.2 PDI and TVDI Determinations

For *PDI*, the slope of the soil line is determined using the least-squares linear regression; the corresponding soil lines for each study area are elucidated in Figure 14:

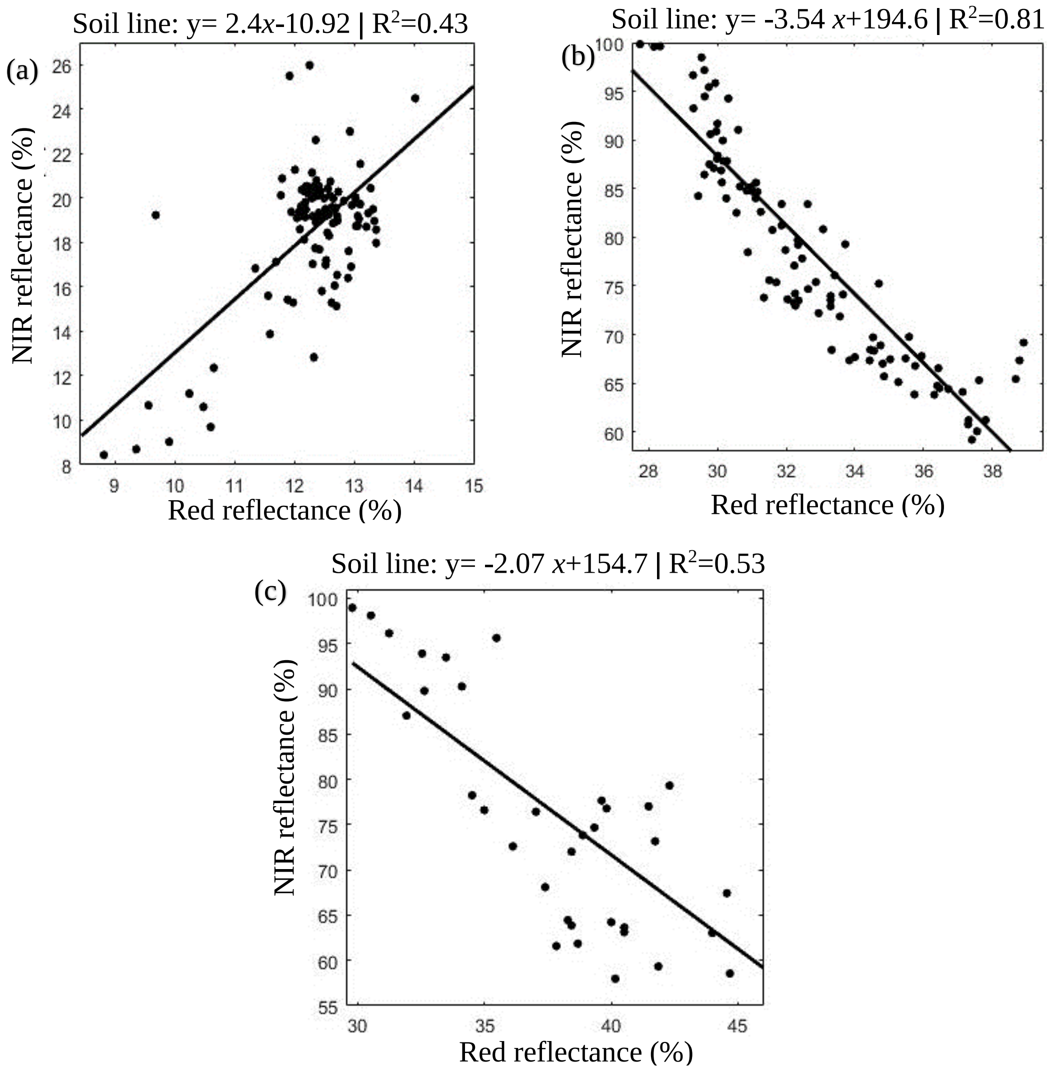

Figure 14. Soil Lines of the study areas; Blackwell farms (a), Sidi Rached 1 (b), Sidi Rached 2 (c).

Initial analysis of Figure 14 confirms that the soil line in the study area (a) represents a weak positive correlation (R^2=0.43) despite that the low-intensity vegetation cover in that area. Whereas, the soil lines in study areas (b) and (c) do not accurately depict the soil line as described by (Ghulam, Qin, and Zhan 2007). Both soil lines depict a negative correlation between reflectance in the red and NIR bands, respectively, with various degrees of strength, as in fair negative correlation in study area (b) (R^2=0.81), weak negative correlation in study area (c) (R^2=0.53). That could be due to the presence of intense vegetation cover in those study areas (b) and (c), as well as their corresponding low number of pixels representing bare soil.

As for the determination of *TVDI*, the intercept and slope of the dry edge of the *LST/NDVI* feature space are inferred by selecting the maximum *LST* for each *NDVI* value, then applying least-squares linear regression to those temperatures. Conversely, to infer the intercept and slope of the wet edge, the minimum *LST* for each of the *NDVI* values is used for linear regression instead.

Figure 15 illustrates the triangle *LST/NDVI* feature space for each of the used study areas:

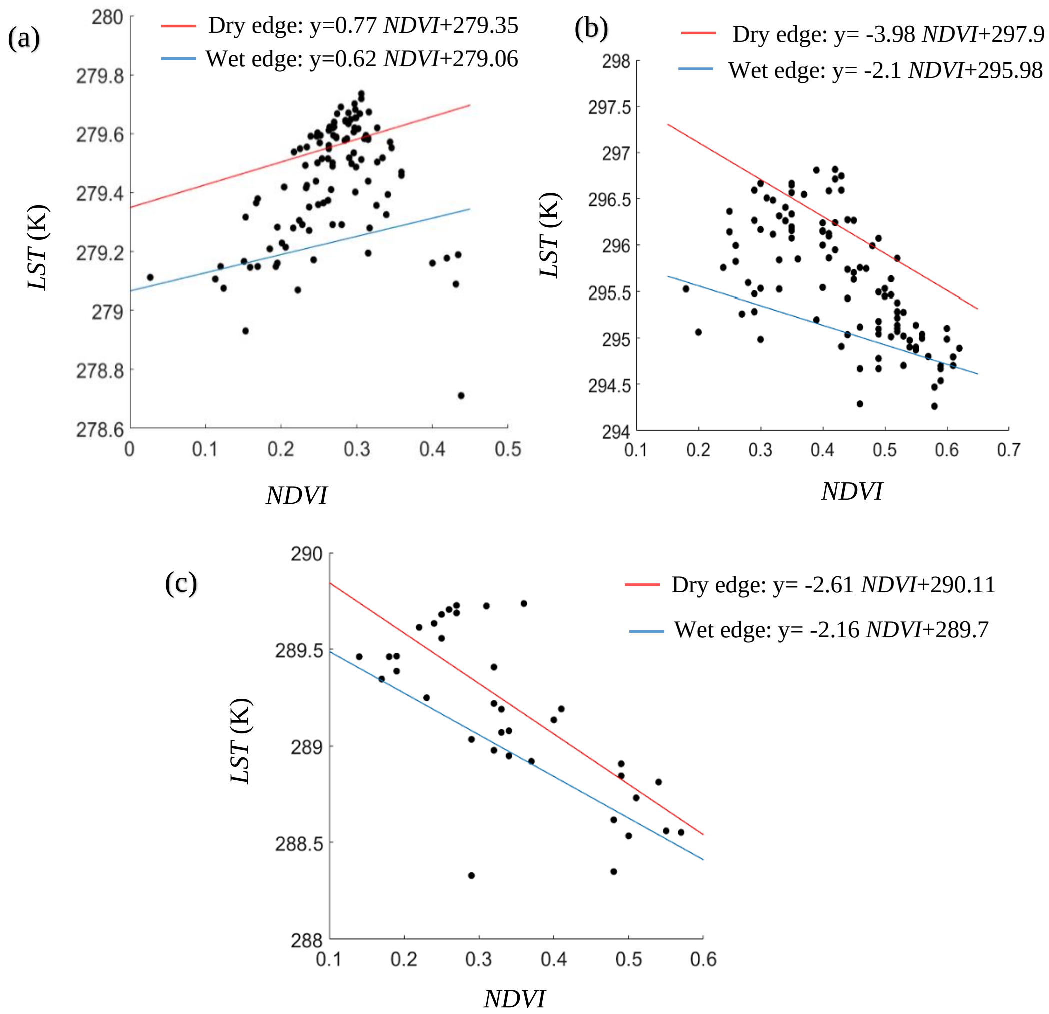

Figure 15. Triangle *LST/NDVI* feature space of the study areas; Blackwell farms (a), Sidi Rached 1 (b), Sidi Rached 2 (c).

Preliminary analysis of Figure 15 reveals that *LST/NDVI* feature space in the study area (a), has a significant number of the pixels with low *NDVI* values (0.1 to 0.3) which is denoted as bare soil and low vegetation cover intensity. It is because of that reason, that the triangular shape of *LST/NDVI* feature space was not achieved, with the triangle being better formed in the study area (b) due to its apparent heterogeneity in terms of vegetation cover intensity with most pixels having *NDVI* values from (0.3 to 0.5). The triangle is also achieved in *LST/NDVI* feature space in the study area (c) despite the low number of its data samples with pixels with NDVI values ranging from 0.15 (bare soil) to 0.5 (fully vegetated areas).

3.4.3 Feature Level Fusion

Feature level fusion is achieved in this research by simple concatenation of the feature vectors obtained from the previously discussed methods of SMC estimation.

Let $X = \{x_1, x_2 \ldots x_n\}$ and $Y = \{y_1, y_2, \ldots y_m\}$ denote feature vectors ($X \in R^n$ and $Y \in R^m$), and the goal is to merge vectors X and Y to generate a new joint feature vector Z with an improved accuracy SMC estimation, where Z is obtained using equation 22 (Ross, Govindarajan 2005):

$$Z = X \cup Y = \{x_1, x_2 \ldots x_n, y_1, y_2, \ldots y_m\},\ Z \in R^{n+m} \tag{22}$$

Studies conducted by (Alexakis, Mexis et al. 2017), used this concept to concatenate features extracted from radar parameters $\{\sigma^0_{VV}, \theta_i\}$, vegetation index $\{NDVI\}$ and thermal image $\{LST\}$ where the addition of the latter parameter has increased the SMC estimation accuracy (with a lowered RMSE values in the order of 2.7% across all study areas).

Similarly, implemented a feature level fusion for SMC estimation by merging a feature vector containing radar and surface features $\{s, l, \theta_i, \sigma^0_{VV}\}$ with the synergetic index $\{TVDI\}$. The implementation yielded less bias and smaller RMSE values by order of 0.474%.

In this research, several joint feature vectors with different combinations of features were used to investigate whether increasing the dimensionality of features will have any effect on the overall accuracy of SMC estimations. The feature vectors were grouped into three joint feature vectors FLF1, FLF2 and FLF3 and Table 3 describe the corresponding features for each vector:

Table 3. **Feature level fusion joint vectors.**

Joint feature vector	Fused features
FLF1	$\{s,l,\theta_i,\sigma_{VV}^0\} \cup \{TVDI\}$
FLF2	$\{TVDI\} \cup \{PDI\}$
FLF3	$\{s,l,\theta_i,\sigma_{VV}^0\} \cup \{TVDI\} \cup \{PDI\}$

FLF1 is the same feature vector used in (Yahia, Guida et al. 2018a). The goal behind that specific selection of features was to introduce a new parameter resistant to vegetation covers presence (*TVDI*) to the EA-IEM inversion.

FLF2 is comprised only of *PDI* and *TVDI* in cases the surface roughness parameters are too high to be in the valid range for EA-IEM inversion. Furthermore, the multispectral aspect of this vector (*PDI*) allows for future exploitation of several potential multispectral imagers (such as Sentinel-2) if the temporal gap between Sentinel-1 and Landsat-8 is too long.

FLF3 is the joint feature vector containing all of the available features to explore whether increasing the dimensionality of feature space even further would result in enhanced SMC estimation in terms of accuracy.

The choice of the feature level fusion estimator was a subject of discussion, numerous different methodologies have been suggested in the literature as seen in section 2.6. The most notable proposed methods were LS-regression, Support Vector Machine (SVM), Random Forests (FR), and ANN. So naturally, to choose the most suitable estimator for this research, these estimation methods were experimented with. To get a preliminary assessment, the author of this research used the MATLAB regression learner tool to compare the performance of each one of the methods mentioned above against the approach proposed by the author of this research in a past study (Yahia, Guida et al. 2018a) in terms of accuracy, which was expressed using RMSE and R^2. It consists of feeding FLF1 feature vector to an estimator and comparing the resulting estimations against ground truth measurements. Table 4, elucidates the performance of each methodology of each of the study areas:

Table 4. **Comparison between different feature level estimators applied in 3 study areas.**

Study areas	**Method**	**RMSE**	**R^2**
Blackwell farms	LS- Regression	2.01	0.07
	SVM	2.1	0.07
	RF	1.96	0.11
	ANN	**1.54**	**0.43**
Sidi Rached 1	LS- Regression	4.87	0.04
	SVM	4.8	0.07
	RF	4.86	0.04
	ANN	**3.19**	**0.58**
Sidi Rached 2	LS- Regression	3.36	0.01
	SVM	3.26	0.06
	RF	3.32	0.01
	ANN	**1.97**	**0.62**

It is clear from Table 4, that ANN is the most consistent estimator in terms of performance. Its estimations seem to offer the strongest correlations and the lowest RMSE values out of all other methodologies. Therefore, the author of this research decided to opt to perform all feature level fusion estimations through Artificial Neural Networks.

As described in section 2.3.2.1, ANN is a system that consists of artificial neurons interconnected by weights, and its basic structure is composed of an input layer, one (or more) hidden layer, and an output layer as depicted in Figure 15:

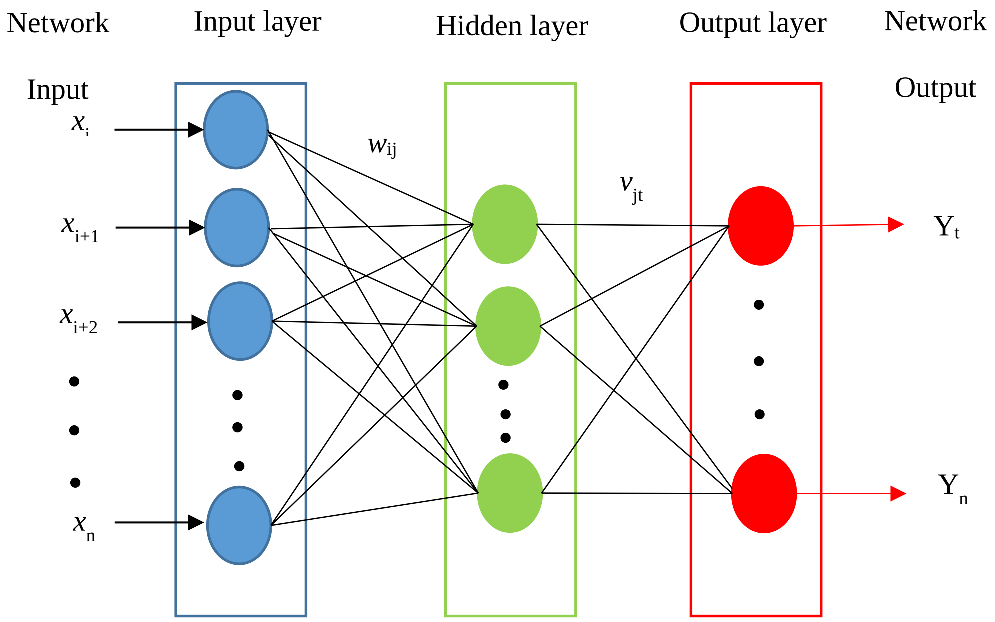

Figure 16. The structure of a feed-forward neural network. Adapted from (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000).

X (x_i, x_{i+1}…x_n) is the input vector, w_{ij} denotes the connection weight from *i*th node in the input layer to *j*th node in the hidden layer, v_{jt} denotes the connection weight from the *j*th node in the hidden layer to *t*th node in the output layers.

Moreover, each neuron has a threshold value, also called bias, and its value has to be exceeded before its activation through a function *f(x)*. The latter determines the response of any given node to the total input signal it receives. In this research, the used activation function (or transfer function) of each neuron is a sigmoid function given by equation 23 (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000):

$$\boldsymbol{f(x) = \frac{1}{1 + exp\ (-x)}} \tag{23}$$

Using equation 23 value of from the *j*th node in a hidden layer is calculated using equation 24 (Jiang, Yang et al. 2004b):

$$\boldsymbol{H_j = \frac{1}{1 + exp\ [-(\sum_{i=1}^{n} w_{ij}x_i - b_j)]}} \tag{24}$$

Where H_j is the value of the jth hidden node, w_{ij} is the weight between the ith input node and jth hidden note, x_i is the input value, n is the number of input nodes, and b_j is the bias of the jth hidden node. The output results or values of note in the output layer are calculated using equation 25 (Jiang, Yang et al. 2004b):

$$\boldsymbol{Y_t} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{exp}\left[-\left(\sum_{i=1}^{m} \boldsymbol{v_{jt} H_j} - \boldsymbol{\gamma_t}\right)\right]} \tag{25}$$

where Y_t is the value of tth output node, v_{jt} is the weight connecting the jth hidden node, m is the number of hidden layers and the tth output node, and γ_t is the bias of the tth output node.

For this research, six distinct ANNs were used with each ANN having a different input vector. However, all of these ANNs input vectors are linked to the same ground truth SMC measurements in the training phase, and Table 5 clarifies the input vectors for each ANN:

Table 5. **The input feature vector for all of the used artificial neural networks.**

ANN	**Input feature vector**
ANN_{TVDI}	*TVDI*
ANN_{PDI}	*PDI*
$ANN_{EA\text{-}IEM}$	$(s,l,\theta_i,\sigma_{VV}^0)$
ANN_{FLF1}	$(s,l,\theta_i,\sigma_{VV}^0, TVDI)$
ANN_{FLF2}	$(TVDI, PDI)$
ANN_{FLF3}	$(s,l,\theta_i,\sigma_{VV}^0,TVDI,PDI)$

The backpropagation training algorithm for all ANNs in question is the Levenberg-Marquardt (Marquardt 1963). Out of the three study areas, a total of 260 available samples were collected, 70% of the samples were used for training, 20% for validation and 10% for testing. The process of validation is extremely important for the generalisation of the ANN due to its role of minimising the effect of overfitting (when an ANN becomes too specific to a data sample) (Hachani, Ouessar et al. 2019). The size of the hidden layer size (10 nodes) and the specific division of the training sample was ascertained after multitudes of experimentations, and this specific configuration produced the best results in terms of accuracy.

After the necessary training of all ANN, all corresponding estimations are sent to the fusion centre, where a weight-based fusion is performed to improve the accuracy of SMC estimation.

3.4.4 Fusion Centre

All estimations produced by the ANNs work as input arguments for the fusion centre function. The latter is an extension of a novel weight-based system designed by the author of this research in (Yahia, Guida et al. 2018b).

Instead of solely respectively assigning weights w_1, w_2 and w_3 to the estimations achieved by ANN_{TVDI}, ANN_{PDI}, $ANN_{EA\text{-}IEM}$, the proposed weight-based system also includes weights w_4, w_5 and w_6 for estimations produced by ANN_{FLF1}, ANN_{FLF2} and ANN_{FLF3} respectively. The idea behind the inclusion is to use the improved accuracy of the feature level fusion estimations to influence the accuracy of the weight-based fusion estimation using equation 26:

$$\begin{aligned}\boldsymbol{SMC_{fused}} &= \boldsymbol{w_1 SMC_{TVDI} + w_2 SMC_{PDI} + w_3 SMC_{EA-IEM} + w_4 SMC_{FLF1}} \\ &\quad \boldsymbol{+\ w_5 SMC_{FLF2} +\ w_6 SMC_{FLF3}}\end{aligned} \tag{26}$$

where SMC_{fused} is the weight-based decision level fusion estimation and SMC_{TVDI}, SMC_{PDI}, $SMC_{EA\text{-}IEM}$, SMC_{FLF1}, SMC_{FLF2} and SMC_{FLF3}, are the output estimations of ANN_{TVDI}, ANN_{PDI}, $ANN_{EA\text{-}IEM}$, ANN_{FLF1}, ANN_{FLF2} and ANN_{FLF3} respectively, and $w_1+w_2+w_3+w_4+w_5+w_6=1$.

Figure 17 illustrates the inner workings of this fusion centre:

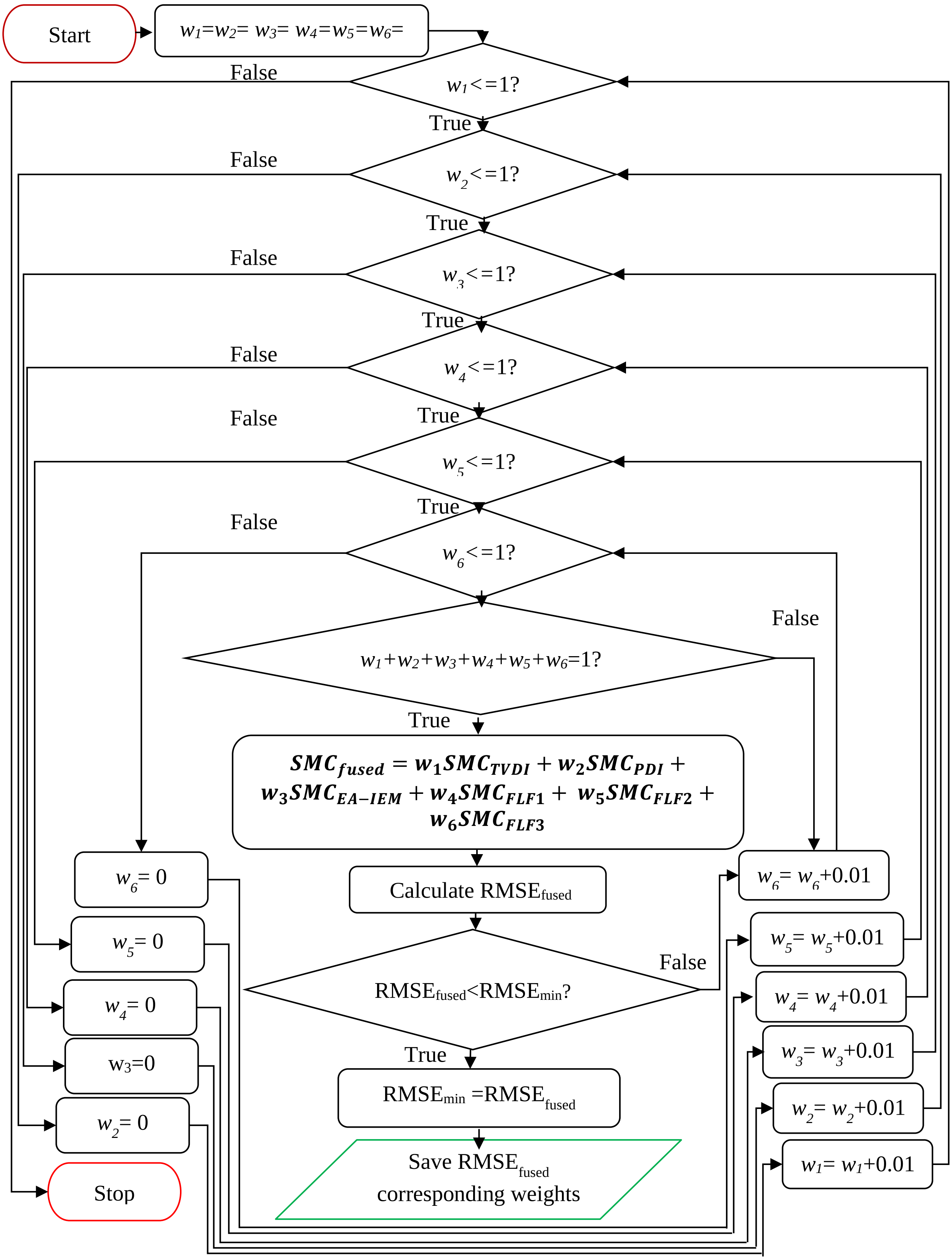

Figure 17. Flowchart of the Fusion Centre.

Initially, all weights are put in loops where they are incremented from 0 to 1 with 0.01 increments, and SMC_{fused} is only calculated if the sum of all of the weights is equal to 1. The Root Mean Square Error (RMSE) of that particular estimation is calculated using equation 27 (Willmott 1982):

$$\boldsymbol{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{1}{N} (P_i - O_i)^2} \tag{27}$$

Where N is the number of data samples, P*i* is the estimated SMC values, or in this case, $SMC_{fused,}$ and O*i* is the measured SMC values. RMSE is calculated for all estimations mentioned above.

After RMSE of the final fusion is determined ($RMSE_{fused}$), it is compared with the lowest obtained RMSE value from all of the previously mentioned estimations ($RMSE_{min}$). If $RMSE_{fused}$ is lower than $RMSE_{min}$, then $RMSE_{fused}$ becomes the new $RMSE_{min.}$

Finally, after repeating this process for all of the possible combinations of weights satisfying the same condition, the weights corresponding to the lowest RMSE values ($RMSE_{min}$) are the ones that get chosen as the optimal weights for the final fusion, and those weights are saved in case of absence of ground truth data in future estimations. Any future inclusions of more SMC ground truth measurements will only improve the accuracy of SMC estimation system, the addition of different study areas with different weather conditions, soil compositions, vegetation cover distributions, surface roughness profiles and surface topography, will allow the proposed SMC estimation to create a configuration of weights suitable for each future scenario.

The experimental results of all of the aforementioned estimations are available in chapter 5, Testing and Evaluation.

3.5 Conclusion

Chapter 3 has explored the different issues and limitations posed by each individual method (*PDI, TVDI, IEM* inversion) and proposed a data fusion scheme to maximise the accuracy of SMC estimation. Chapter 3 also provided a description of the elements of novelty pertinent to this research. The first presented element was the updated empirically adapted IEM. The addition of a semi-empirical calibration parameter *Lopt* was proposed to increase the validity range of surface roughness value of the empirically adapted IEM, which was yet to be performed in literature. The second element of novelty presented in chapter 3, is this specific configurations of features vectors, more specifically FLF3. These specific set of features (*s*, l,θ_i,σ_{VV}^0,*TVDI*,*PDI*), and especially the inclusion of σ_{VV}^0 feature (which is the direct output of the updated empirically adapted IEM) have never been investigated for the same EO based SMC estimation system. The argument behind this proposition was to increase the dimensionality of the feature space which would potentially result in lower inaccuracies in terms of SMC estimation. Chapter 3 also introduced the third element of novelty which is the estimation level fusion scheme of the soil estimation system. The concept of the fusion centre was explained, where it was argued that assigning weights to each estimation (including the 3 succinct feature level fusions), would ameliorate the overall accuracy of estimation. Assigning these weights to this specific combination of estimations ($ANN_{TVDI,}$ ANN_{PDI}, $ANN_{EA\text{-}IEM}$, ANN_{FLF1}, ANN_{FLF2}, ANN_{FLF3}), which are the outputs of ANNs, and using loops in the fusion centre have never been proposed in the literature. The third element constitutes the most significant addition to the state of the art that this research has to offer.

In order to validate this system, ground truth measurements collected from study areas are required. Chapter 4: "Generating Test Datasets" will supply a description of the used study areas.

4. Generating Test Datasets

4.1 Introduction

Chapter 4 is dedicated to all the technical details about study areas or areas of interest for this research, in terms of geographical location, coordinates, size, and intensity of vegetation cover as well as the mineralogical composition of the soil or soil type.

Chapter 4 will also supply a detailed description of all the earth observation data adapted (whether be it datasets from Sentinel-1 or Landsat-8) and all *in situ* sensing instruments (such as SMC levels probes and laser and needle profilometres).

4.2 Study Area Details

The validation of all the methodologies used in this research requires direct measurements from suitable areas of interest or study areas. Three different study areas have been selected for this research.

The first and earliest dataset is one of the agricultural fields in Blackwell farms, located in Guildford, the county town of Surrey in South East England. The size of the farm is around 295 m x 308 m, and at the time of collection of measurement, the field was visibly spatially homogenous in terms of vegetation cover intensity. The field is also characterised by a non-flat surface topography (it has a small slope in the middle of it).

The second study area is an agricultural field in Sidi Rached, Tipasa, Algeria, and this study area is by far the largest (540 m x 180 m), with the agricultural field had a spatially heterogeneous in terms of vegetation cover intensity, some areas of the field contained almost bare soil, and others had intense vegetation cover. This field, however, is characterised by relatively flatter surface topography.

The third study area is another agricultural field in a different location at Sidi Rached, Tipasa. This field was the smallest of all of the datasets (180 m x 180 m) due to restriction made by the field owner. This field was also visibly heterogeneous in terms of vegetation cover, and similarly to the second study, the field has a flat surface topography.

Table 6 contains relevant information for all study areas, such as the location, coordinates, size, and respective NDVI means for the datasets as an indicator of the intensity of vegetation cover, as well as the soil type, or rather the mineralogical composition of each soil (which will be important for the IEM inversion application):

Table 6. **Details of study areas.**

Study area	**Location**	**Coordinates (Latitude, longitude)**	**Size (m x m)**	**NDVI (Mean)**	**Soil type**
Blackwell farms	Guildford, Surrey, United Kingdom	51° 14’ 10” N, 000°37’ 32” W	295 x 308	0.26	Clay loam
Sidi Rached 1	Tipasa, Algeria	36° 33’ 18” N, 002° 31’ 28” E	540 x 180	0.43	Sandy loam
Sidi Rached 2	Tipasa, Algeria	36° 31’ 30” N, 002° 32’ 38” E	180 x 180	0.33	Sandy loam

The selection of suitable study areas was one of the most difficult challenges that faced this study due to the scarcity of concurrent acquisitions from different satellites. Normally, satellite acquisitions (Sentinel-1, Landsat-8) only quite coincide in the same day only twice or three times a month, and that was made worse by the frequent poor weather condition and cloud presence in the United Kingdom. At one point of this research, there was no possibility for ground truth collection campaign from the period of November 2017 to March 2018. This complication has prompted the author of this research to pursue different study areas elsewhere, and the chosen study areas were two agricultural fields in Sidi Rached, Tipasa, Algeria. There was a fourth study area in Mezriaa, Biskra, which is a county in the south of Algeria (latitude: 34° 40′ 01″ N, longitude: 06° 16′ 05″ E), but all SMC and roughness parameters measurements were disregarded due to a significant temporal gap between the corresponding satellites acquisitions times (Sentinel-1: 16/04/2018 at 05:38 a.m., Landsat-8: 18/04/2018 at 10:07 a.m., approximately 52 hours) not to mention that the sandy texture of the soil (weak water retention capabilities) and extreme weather conditions in terms of temperature have altered soil moisture content levels drastically between acquisitions. This alteration was manifested when the same data points of SMC measurements had substantially different values between the beginning and end of the campaign (differences in the order of 10 to 20 %).

Figures 18, 19, 20, 21, 22, and 23 depict satellites views of the concerned study areas:

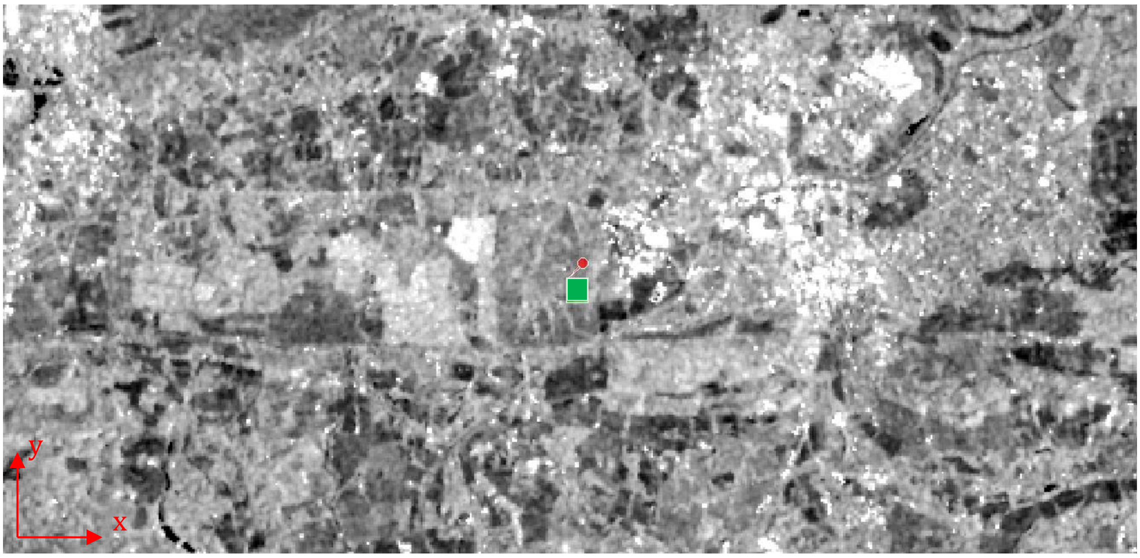

Figure 18. Image from Sentinel-1 dataset of the study area Blackwell farms represented by the green rectangle (10x11 pixels).

Figure 19. Image from Landsat-8 dataset of the study area Blackwell farms represented by the green rectangle (10x11 pixels).

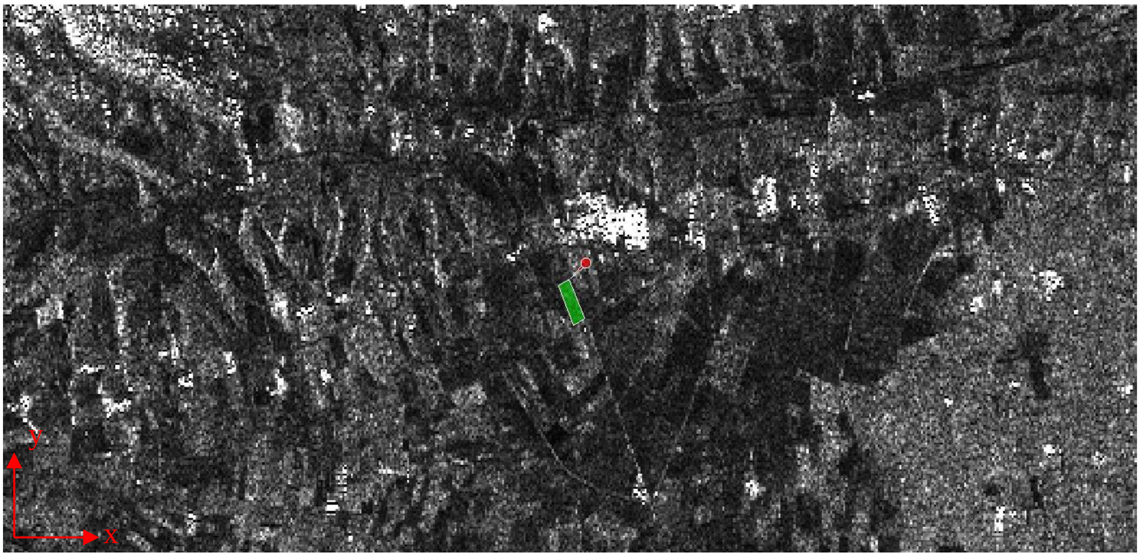

Figure 20. **Image from Sentinel-1 dataset of the study area Sidi Rached 1 represented by the green rectangle (6 x 19 pixels).**

Figure 21. **Image from Landsat-8 dataset of the study area Sidi Rached 1 represented by the green rectangle (6 x 19 pixels).**

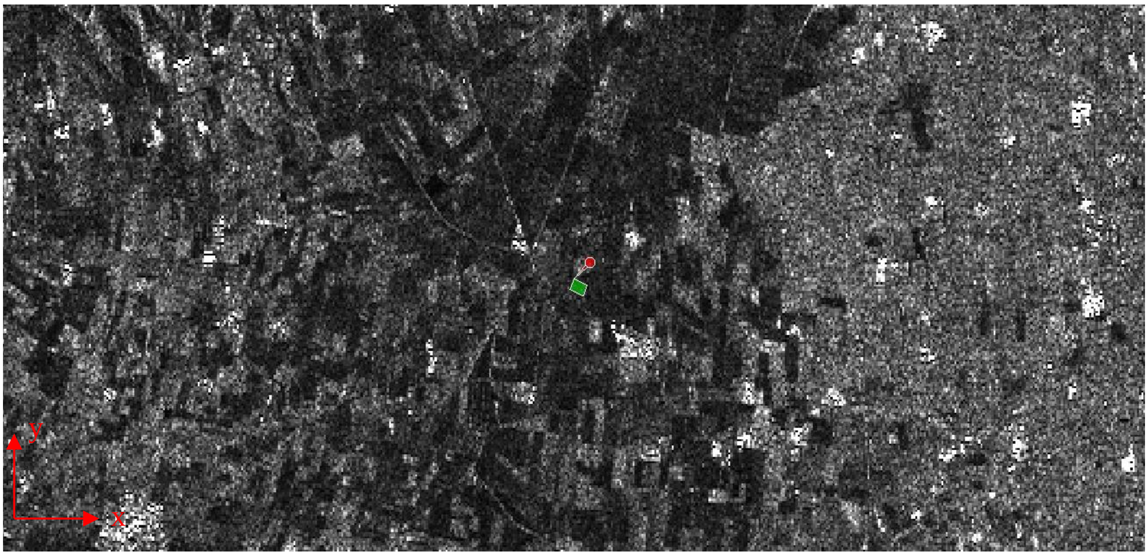

Figure 22. Image from Sentinel-1 dataset of the study area Sidi Rached 2 represented by the green rectangle (6 x 6 pixels).

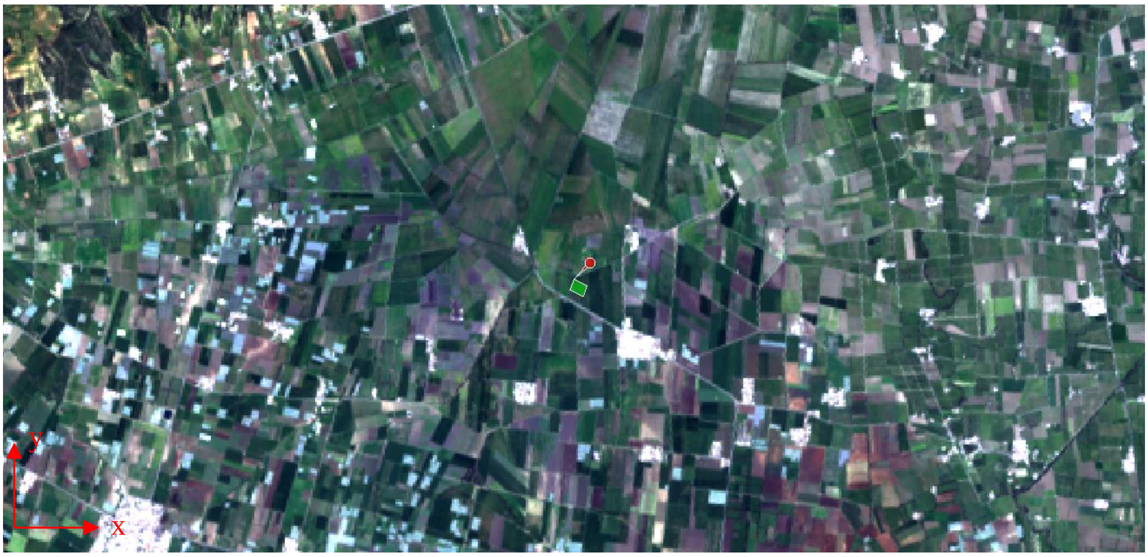

Figure 23. Image from Landsat-8 dataset of the study area Sidi Rached 2 represented by the green rectangle (6 x 6 pixels).

Figure 24. Field photographs of the landscapes of Blackwell farms (a), Sidi Rached 1 (b) and Sidi Rached 2 (c).

The author of this research is well aware that the respective sizes of all study areas in terms of data samples are relatively small, that was due to a lack of availability of ground truth measurements for SMC for high spatial resolution satellite data, and the amount of time and efforts required to collect such measurements on a larger scale.

4.3 Earth Observation Data

Earth observation datasets were collected from two different satellites, Sentinel-1 as the active microwave sensor for the integral equation inversion model, and Landsat-8 as the data source to calculate both *PDI* and *TVDI*.

4.3.1 Sentinel-1

The reason Sentinel-1 was chosen for this system is its all-weather imaging capabilities, high spatial resolution as well its reasonable temporal resolution, (short revisit cycles of 12 days), not to mention the availability of datasets. Sentinel-1 is a mission of twin satellites

Sentinal-1A and Sentinel-1B, both equipped with a C-SAR on board (5.405 GHz frequency) (Huang, S., Ding et al. 2019).

All products used in this study are Ground Range Detected (GRD), the acquisition mode is Interferometric Wide Swath (IWS), and the available polarization in this acquisition mode in the concerned datasets were VV, VH polarizations. However, only acquisitions with VV polarization were considered since the IEM inversion for SMC estimations performs its best in co-polarized configurations (Kornelsen, Coulibaly 2013b). The spatial resolution for this particular acquisition mode is 20 m x 23 m, and the swath is 250 km. Table 7 provides further acquisition details:

Table 7. **Sentinel-1 acquisition details**

Study area	**Spatial resolution (range, azimuth)**	**Incidence angle (°) (min-max)**	**Acquisition date**	**Acquisition time**
Blackwell farms	20.4 m x 22.5 m	38.2-41.52	18/11/2017	06:21
Sidi Rached 1		44.98-45	07/04/2018	17:51
Sidi Rached 2		34.4-34.41	09/05/2018	5:45

All of the used Sentinel-1 data products were downloaded from https://scihub.copernicus.eu/dhus/#/home.

4.3.2 Landsat-8

Both thermal infrared and multispectral images are collected from the Landsat-8 satellite. This satellite was selected for its high spatial resolution in comparison with other satellites used in literature for similar purposes (such as MODIS) (Chen, J., Wang et al. 2011). Another reason is the guaranteed availability of datasets in the study area backed by an acquisition calendar. Landsat-8 satellite was launched on the 11th of February 2013, it has two sensors on-board, Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), with a total of 10 spectral bands and a panchromatic band. Landsat-8 data are acquired at 185 km swaths with a revisit time of 16 days (Roy, Wulder et al. 2014). Table 8 elucidates its technical specifications and the acquisition details of interest for this study:

Table 8. **Technical specifications of Landsat-8**

Sensor	Band number	Wavelength (μm)	Spatial resolution (m)	Swath (km)	Acquisition date	Acquisition time
Operational Land Imager (OLI)	B1- Ultra Blue	0.43-0.451	30	183	17/11/2017 07/04/2018 09/05/2018	10:52 10:25 10:25
	B2- Blue	0.452-0.51	30			
	B3- Green	0.53-0.59	30			
	B4- Red	0.63-0.67	30			
	B5- Near infrared	0.85-0.87	30			
	B6-Shortwave Infrared (SWIR) 1	1.56-1.65	30			
	B7-Shortwave Infrared (SWIR) 2	2.56-2.29	30			
	B8 - Panchromatic	0.50-0.67	15			
	B9 - Cirrus	1.36-1.38	30			
Thermal Infrared Sensor	B10- Thermal infrared 1	10.60-11.19	100			
	B11- Thermal infrared 2	11.50-12.51	100			

All of the used Landsat-8 datasets were Level-1 data products downloaded from https://earthexplorer.usgs.gov/.

It is quite apparent from observing Tables 7 and 8 that there are temporal gaps between the times of acquisitions. It was unfortunately unavoidable due to the revisit cycles of the different used satellites in this research. The longest temporal gap occurred in the Blackwell farms datasets (approximately 20 hours), while the other two datasets had a relatively shorter temporal gap (approximately 5 hours). Certain measures were taken to attempt to lower the temporal gap and the author of this research has scheduled ground truth measurements collection between the times of acquisitions. For Blackwell farms the start time of collection of SMC measurements was from 14:00 to 17:00, for Sidi Rached 1 the start time of collection was from 11:30 p.m. to 15:00, and for Sidi Rached 2, the start time of collection was from 09:00 (a.m.) to 10:30 (a.m.). Furthermore, there were no significant weather changes in terms of temperature, precipitation, and humidity between the times of acquisitions.

4.4 **Ground Truth Measurements**

After the selection of suitable study areas for this research, the next step was the collection of ground measurements. Two types of measurements were required for the proposed system: SMC levels and soil surface roughness. Three distinct instruments were used to gather such measurements, and the ML3 theta soil moisture probe was used for SMC level measurements while needle and laser profilometres were used to measure surface heights.

4.4.1 ML3 Theta Soil Moisture Probe

The instrument used for SMC level measurements is the ML3 Theta probe soil moisture sensor.

The ML3 probe, when powered on, applies a 100 MHz waveform to an array of stainless-steel rods, which transmits an electromagnetic field to the soil. Any water content present in the soil surrounding those rods will dominate the permittivity (ε) of soil (ε of water ≈ 81, while ε of soil is ≈ 4). The ML3 detects the influence of the permittivity on the transmitted electromagnetic field in terms of stable voltage output which represents a sensitive measure of SMC levels. The device has a measurement range of 0-100% with a 1% error for SMC values from 0 to 50%, and it is compatible with several data loggers such GP1, GP2, DL6, DL2. In this research, the HH2 meter was used. The latter provides general calibrations of the ML3 readings for mineral and organic soils and the storage of those readings (ML3 ThetaProbe user Manual).

The condition of the ML3 theta probe condition was brand new at the first time of usage, and all soils used in this research were organic soils.

Given that point based SMC measurements are far from being an accurate representation of SMC and its variability, four measurements were collected every 30 metres. Then the mean of those measurements was assigned to its corresponding pixel of the earth observation data. The use of 30 metres distance between the points of measurements, in particular, is to match the spatial resolution of Landsat-8 in the multispectral bands which is a good compromise between the spatial resolutions of thermal bands (100 m) and Sentinel-1 (20.4 m x 22,5 m). Figure 25 represents the process of SMC measurements:

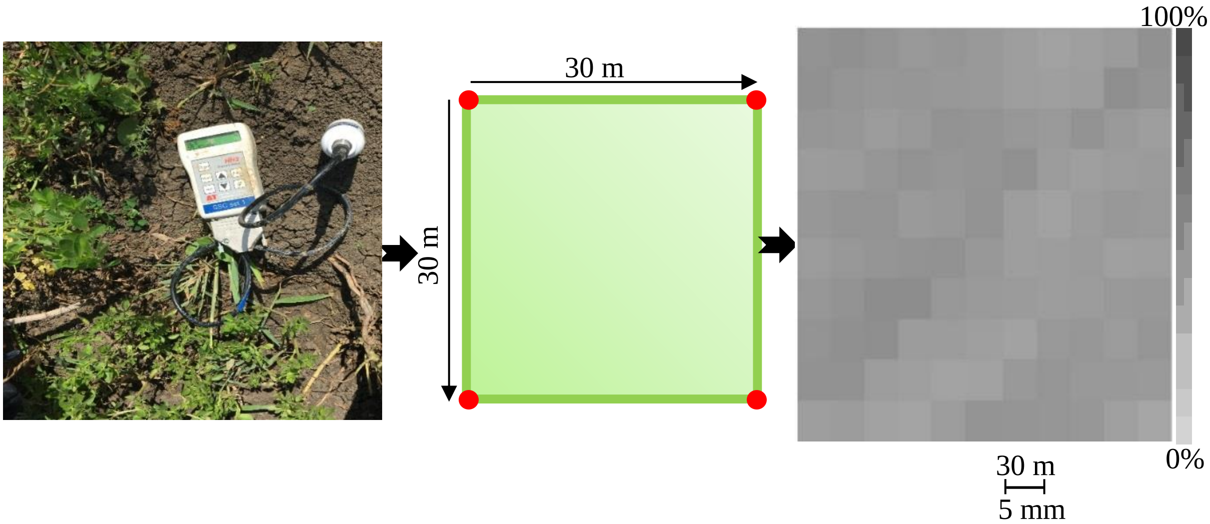

Figure 25. Process of measurements of Soil moisture content level using ML3 Theta probe soil moisture sensor.

The process of SMC measurement was a demanding task; it took a tremendous deal of logistical planning and scheduling to collect such measurement in a timely fashion, an average of 2 hours to collect around 100 data samples, not to mention the amount of time and effort necessary to render the relevant data samples in the same order as their corresponding earth observation geocoded pixels. Consequently, the ground truth measurement campaigns did not yield a large number of data samples to use for the validation of this research.

4.4.2 Profilometre

Secondly, the surface roughness profile (RMS height and correlation length) of the said agricultural fields needed to be determined. Two different profilometres were used to recover such information at different stages of this research.

Initially, a needle profilometer was solely used in the Blackwell farms field; this profilometer was built by technicians from the University of Surrey. To derive the measurements, the Profilometer is positioned in the point of interest, then, 78 needles (1*cm* distance from each other) are inserted into this structure, and their height in the main structure is regulated accordingly to represent the level of the soil in each point. Once this process is done to all needle, the profilometer is superimposed on an A0 paper where a curve of points representing the soil profile can be drawn, as elucidated in Figure 26:

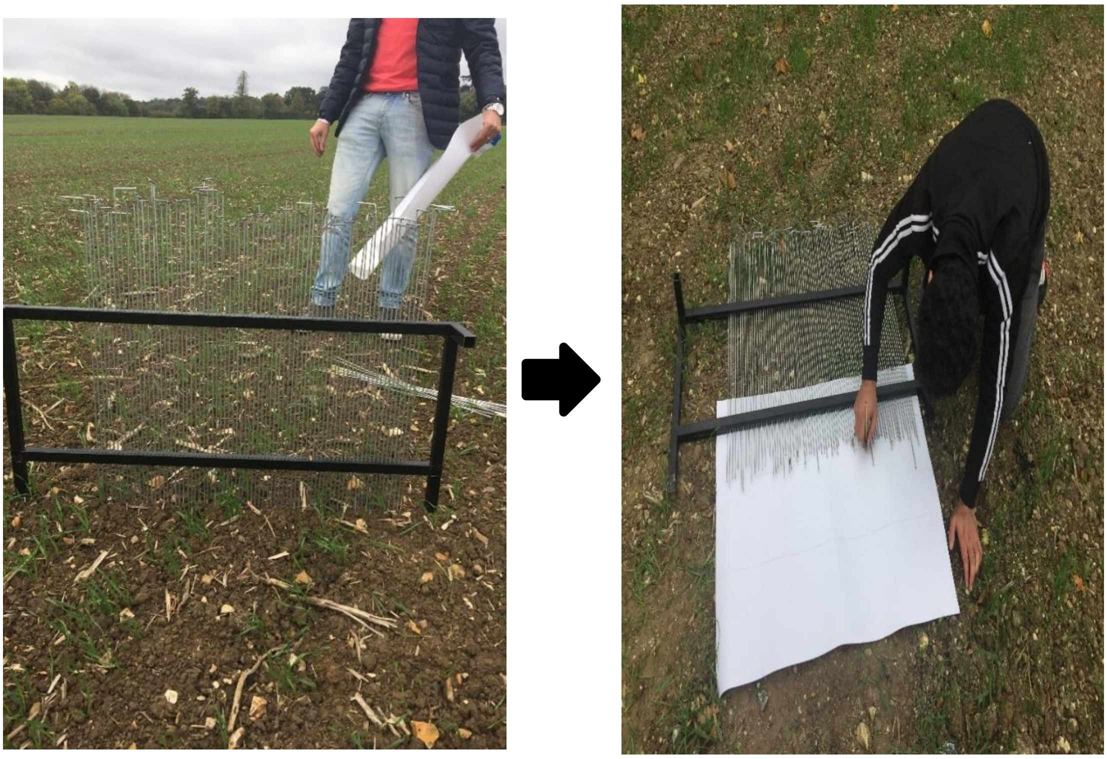

Figure 26. The process of surface roughness parameters measurement using a needle profilometer.

The second device used to measure surface roughness parameters was the laser profilometer. The device was built by the author of this research at the Surrey Space Centre laboratory. It consists of a metallic frame with a new BOSCH PLR 15 Laser rangefinder fitted to it. The laser device has a range of 0.5 to 15 metres, with a measuring error of 3 mm (User

Manual of, BOSCH PLR 15). The distance between the fitted laser device and a flat surface is 0.33 m (which is well within its range). The laser points towards the soil surface where the metallic structure is fixed. Then, a measurement is taken and recorded, and the laser device is moved along a rail one centimetre at a time, with 54 possible increments. Figure 27 illustrates the laser profilometer mentioned above:

Figure 27. The process of surface roughness parameters measurement using a laser profilometer.

Table 9 showcases the measurements of different surface roughness parameters (*s*, *l*) expressed in cm, as collected for each study area:

Table 9. **The Values of the measured surface roughness parameters.**

Study area	**Profilometer type**	***s (cm)***	***l (cm)***
Blackwell Farms	Needle	1.57	1.67
Sidi Rached 1	Laser	2.85	5.08
Sidi Rached 2		1.76	8.66

By inspecting the values of *s* and *l* values obtained in Table 9, it is noticeable that Blackwell farms have a short correlation length compared to the other datasets, it can be explained by the fact that the fields were recently ploughed and planted in the time of collection of measurements.

4.5 Conclusion

Chapter 4 offered a detailed account of the different characteristics of the chosen study areas in terms of their locations, vegetation state and the rationale behind their selection, not to mention the initially encountered difficulties in the process of their selection. This chapter also offered a description of all used earth observation datasets; it presented Sentinel-1 as the active microwave sensor to be used for the IEM inversion, and it introduced Landsat-8 sensors OLI and TIRS as data sources to calculate *PDI* and *TVDI*. Then, chapter 4 went through the technical specifications all of the instruments used for proximal sensing, with said instruments being ML3 Theta probe and laser and needle profilometres.

Chapter 4 also featured a presentation of another set of elements of novelty. The first element is that these areas of interest have never been the subject of an EO-related SMC estimation study, especially the regions in Algeria. The SMC ground truth data samples collected using this specific equipment (ML3-Theta Probe) from the areas of interests discussed above have never been used for the validation of any EO-based SMC estimation system before. Another element of novelty is the design and implementation of a laser profilometre and using its completely novel measurements to derive soil surface roughness profiles. Chapter 5: “Testing and Evaluation”, will review all the achieved results across the time of this research as well as a full analysis to interpret those results.

5. TESTING AND EVALUATION

5.1 Introduction

Initially, this chapter showcases the results achieved by the proposed SMC estimation system and offers a deep analysis of each of its corresponding methods, the results and analysis are organised into three groups according to their corresponding study area. Finally, this chapter provides discussion and remarks about the results achieved by the proposed SMC estimation system as well as the performance issues it faces.

5.2 Results and Analysis

The proposed SMC estimation system was developed and implemented in the Matlab environment. In order to analyse and evaluate the accuracy of the proposed system, the author of this research suggests the consideration of the followings metrics of evaluation:

- RMSE, as it is one of the most reliable metrics for model performance evaluations (Willmott 1982).
- The minimum value of the estimated and measured SMC (Min).
- The maximum value of the estimated and measured SMC (Max).
- The mean of the estimated and measured SMC values (Mean).
- The standard deviation of the estimated and measured SMC values (SD).

The results and analysis of the proposed SMC estimation system are organized into three distinct groups according to their study area; Blackwell farms, Sidi Rached 1, Sidi Rached 2. The analysed estimations are named according to their methods, as it is visible in Figures 23, 24 and 25, where the measured SMC (%) is plotted as a function of estimated SMC (%), and the concerned methods of estimation are:

(a). *TVDI.*
(b). *PDI.*
(c). *EA-IEM* inversion.
(d). Feature Level Fusion 1 (FLF1) which is the output of ANN_{FLF1}.
(e). Feature Level Fusion 2 (FLF2) which is the output of ANN_{FLF2}.
(f). Feature Level Fusion 3 (FLF3) which is the output of ANN_{FLF3}.
(g). Weight-Based Fusion (WBF) which the output of the fusion centre.

5.2.1 Blackwell Farms

The first group of results are those corresponding to the Blackwell farms dataset, and Figure 28 and Table 10 provide summaries of the results achieved by each estimation method:

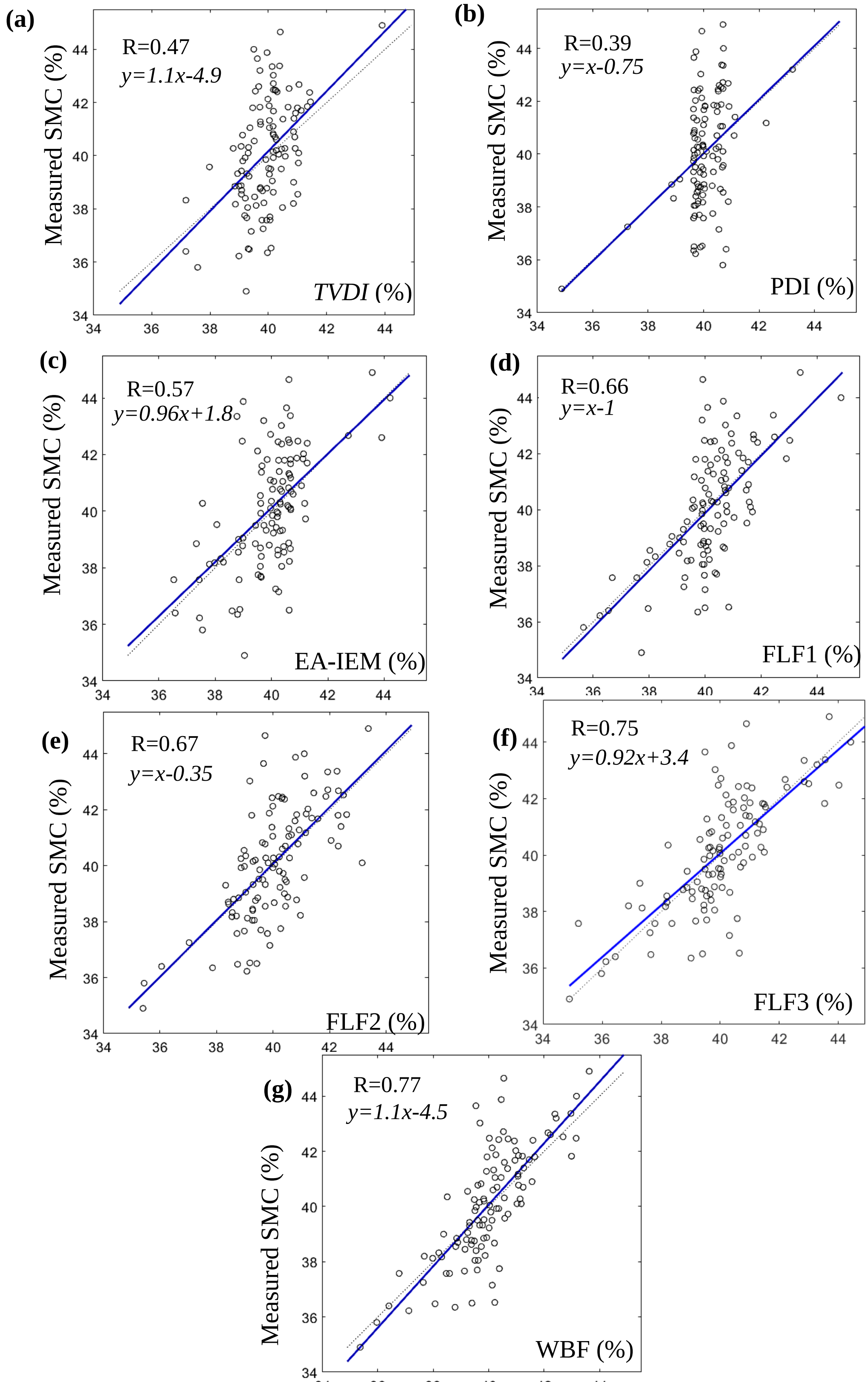

Figure 28. Results of each of the used SMC estimation methods in Blackwell farms datasets.

Table 10. **The results of each of the used estimation methods in the Blackwell farms dataset.**

Estimation methods		Blackwell farms (*n*=110)								
		RMSE (%)	Estimated SMC				Measured SMC			
			Min (%)	Max (%)	Mean (%)	SD (%)	Min (%)	Max (%)	Mean (%)	SD (%)
Estimations using a single method	*TVDI*	1.82	37.16	43.9	39.91	0.87	34.9	44.9	40.07	2.07
	PDI	1.9	34.88	43.2	40.04	0.8				
	EA-IEM	1.7	36.52	44.21	39.96	1.23				
Estimations using feature-level fusion	*FLF1*	1.54	35.65	44.85	40.18	1.34				
	FLF2	1.53	35.39	43.36	40.01	1.37				
	FLF3	1.37	34.88	44.42	40.02	1.69				
Estimation using decision level fusion	*WBF*	**1.32**	35.36	43.62	40.01	1.43				

In this dataset, although soil moisture content values in this field were greater than 35% (minimum measured SMC was 34.9%), the best estimation using a single method in terms of RMSE and degree of correlation was achieved using the *EA-IEM* inversion (RMSE=1.7% and R=0.57), which can be explained by the fact that the field had a non-flat surface, and it was composed mainly of bare soil and sparse vegetation cover, which puts *PDI* in a disadvantage due to the nature of the surface topography, and limits the performance of *TVDI* because of the absence of the full range from bare soil to full vegetation.

It is also quite clear that the estimations using feature-level fusion outperform EA-IEM inversion in terms of RMSE, and that those methods have a stronger correlation as well.

For FLF1, the addition of the synergetic feature *TVDI* to the feature vector of EA-IEM inversion causes immediate improvements on the overall accuracy of estimation (RMSE=1.54%) and the correlation (R=0.66), that addition seems to balance out some of the inaccuracies that could be caused by the high SMC values and surface roughness of the site. For the FLF2, the elimination of radar features, and the addition of a feature (*PDI*) resistant to the limited range of vegetation cover of the study area have increased the accuracy and the correlation slightly further (RMSE=1.53% and R=0.67).

As for FLF3, the inclusion of all of the available features produced the best feature level fusion estimation in terms of accuracy and correlation (RMSE=1.37%, R=0.75), which is quite the improvement. Finally, the estimation produced by WBF has the best accuracy and the strongest correlation out of all of the used methods (RMSE=1.32%, and R=0.77), and the weights achieved for this study area were:

$$\boldsymbol{SMC_{WBF} = 0.08SMC_{TVDI} + 0.26SMC_{FLF2} + 0.66SMC_{FLF3}} \tag{28}$$

The WBF method disregards the *PDI*, *EA-IEM* inversion, and FLF1 estimations completely, and assigns the largest importance to FLF3 (weight= 0.66) which makes sense since it is the most accurate estimation, then the second-largest importance to FLF2 (weight = 0.26) which is the also the second-best estimation in terms of RMSE values, however, it was quite interesting to find that it assigns importance to *TVDI* (weight =0.08).

For this study area, the WBF method produced the best results in terms RMSE and R, with FLF3 a very close second. However, the latter does possess a slightly more accurate range of values in terms of minimum, maximum, and mean and standard deviation.

5.2.2 Sidi Rached 1

The second group of results are those corresponding to the Sidi Rached 1 dataset, and Figure 29 and Table 11 provide summaries of the results achieved by each estimation method:

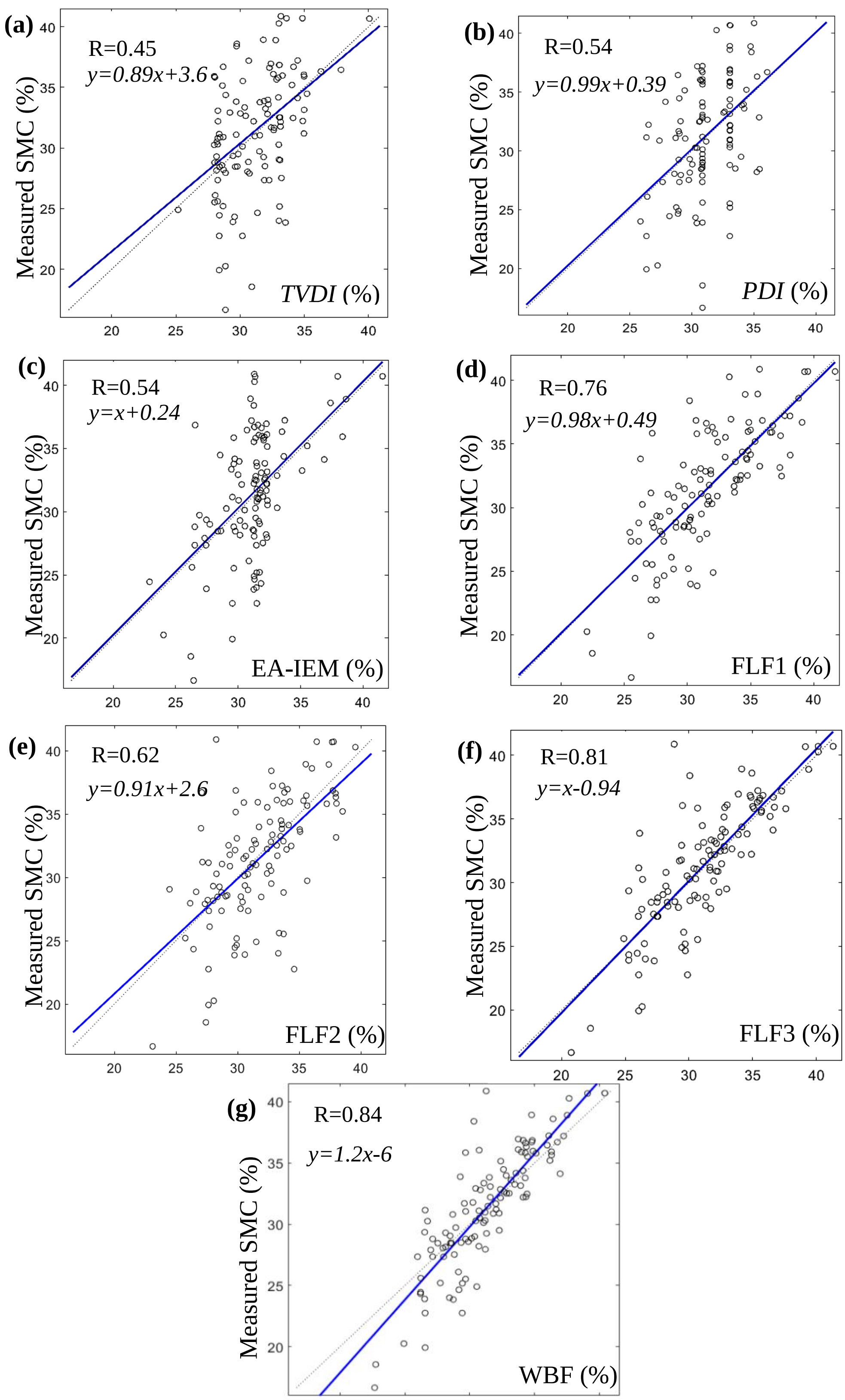

Figure 29. Results each of the used SMC estimation methods in the Sidi Rached 1 datasets.

Table 11. **The results of each of the used estimation methods in the Sidi Rached 1 dataset.**

<table>
<tr><th colspan="2" rowspan="3">Estimation methods</th><th colspan="9">Sidi Rached 1 (n=114)</th></tr>
<tr><th rowspan="2">RMSE (%)</th><th colspan="4">Estimated SMC</th><th colspan="4">Measured SMC</th></tr>
<tr><th>Min (%)</th><th>Max (%)</th><th>Mean (%)</th><th>SD (%)</th><th>Min (%)</th><th>Max (%)</th><th>Mean (%)</th><th>SD (%)</th></tr>
<tr><td rowspan="3">Estimations using a single method</td><td>TVDI</td><td>4.41</td><td>25.16</td><td>40.07</td><td>31.14</td><td>2.52</td><td rowspan="7">16.65</td><td rowspan="7">40.87</td><td rowspan="7">31.4</td><td rowspan="7">4.94</td></tr>
<tr><td>PDI</td><td>4.4</td><td>25.85</td><td>36.08</td><td>31.25</td><td>2.26</td></tr>
<tr><td>EA-IEM</td><td>4.1</td><td>22.91</td><td>41.55</td><td>31.12</td><td>2.67</td></tr>
<tr><td rowspan="3">Estimations using feature-level fusion</td><td>FLF1</td><td>3.19</td><td>22.05</td><td>41.68</td><td>31.48</td><td>3.84</td></tr>
<tr><td>FLF2</td><td>3.89</td><td>23.1</td><td>39.55</td><td>31.67</td><td>3.38</td></tr>
<tr><td>FLF3</td><td>2.9</td><td>20.75</td><td>41.32</td><td>31.22</td><td>3.87</td></tr>
<tr><td>Estimation using decision level fusion</td><td>WBF</td><td><u>2.7</u></td><td>22.67</td><td>40.4</td><td>31.35</td><td>3.51</td></tr>
</table>

In this dataset, the RMSE error values were higher due to the spatial variability in terms of vegetation cover intensity and soil moisture content (SD=4.94%) in this particular agricultural field, and the best estimation using a single method in terms of RMSE and degree of correlation was again achieved using the *EA-IEM* inversion (RMSE=4.1%, R=0.54). The relatively lower performance of *TVDI* could be explained by the fact that the number of pixels representing bare soil and low vegetation intensity was low, only 24% of all pixels had *NDVI* values less than 0.3, which caused a less accurate determination of the dry edge, and the opposite could be said about the performance of *PDI*, due to the high number of pixels associated with dense vegetation, the soil line is not accurately represented given that 76% of the pixels had *NDVI* values more than 0.3.

It is also apparent that the estimations using feature-level fusion outperform *EA-IEM* inversion in terms of RMSE and degree of correlation in this study area as well. For FLF1, the addition of the *TVDI* to the feature vector of EA-IEM inversion has yet again, made a positive impact on the overall accuracy of estimation (RMSE=3.17%) and the correlation (R=0.76), that addition seems to lower the impact of the presence of dense vegetation cover on the accuracy of EA-IEM inversion estimation, not to mention increasing the accuracy of estimation for values that range from 28% to 33% (as visible in plot (d) in Figure 24). For the FLF2, unlike the results achieved in the first study area, the elimination of radar features did not produce a more accurate estimation (RMSE=3.89% and R=0.62), which could be for the same reasons affecting the estimations of *PDI* and *TVDI* individually. As for FLF3, the inclusion of all of the available features produced the best feature level fusion estimation in terms of accuracy and correlation (RMSE=2.9%, R=0.81) once again.

Finally, the estimation produced by WBF has again yielded the best accuracy and the strongest correlation out of all of the used methods (RMSE=2.7%, and R=0.84), and the weights achieved for this study area were:

$$\boldsymbol{SMC_{WBF} = 0.36SMC_{FLF1} + 0.08SMC_{FLF2} + 0.56SMC_{FLF3}} \tag{29}$$

The WBF method disregards the *TVDI*, *PDI*, *EA-IEM* inversion, and FLF1 estimations completely, indeed the WBF in this study area seems to assign weights according to the accuracy of each estimation, it assigns the largest weight to the most accurate estimation FLF3 (weight= 0.56), then the second-largest weight to weight FLF1 (weight = 0.36) and the smallest weight to *FLF2* (weight =0.08).

For this study area, the WBF method produced the best results in terms RMSE and R, and range of values in terms of minimum, maximum, and mean and standard deviation.

It is noticeable that the correlation of WBF estimation in this study area (R=0.84) is stronger than that of the Blackwell farms counterpart (R=0.77), which could be attributed to the longer temporal gap between the time of acquisitions of the Blackwell farms ground truth data.

5.2.3 Sidi Rached 2

The third group of results are those corresponding to the Sidi Rached 2 dataset, This particular dataset is by far the smallest in terms of the number of samples (*n*=36), and Figure 30 and Table 12 offer summaries of the results achieved by each estimation method:

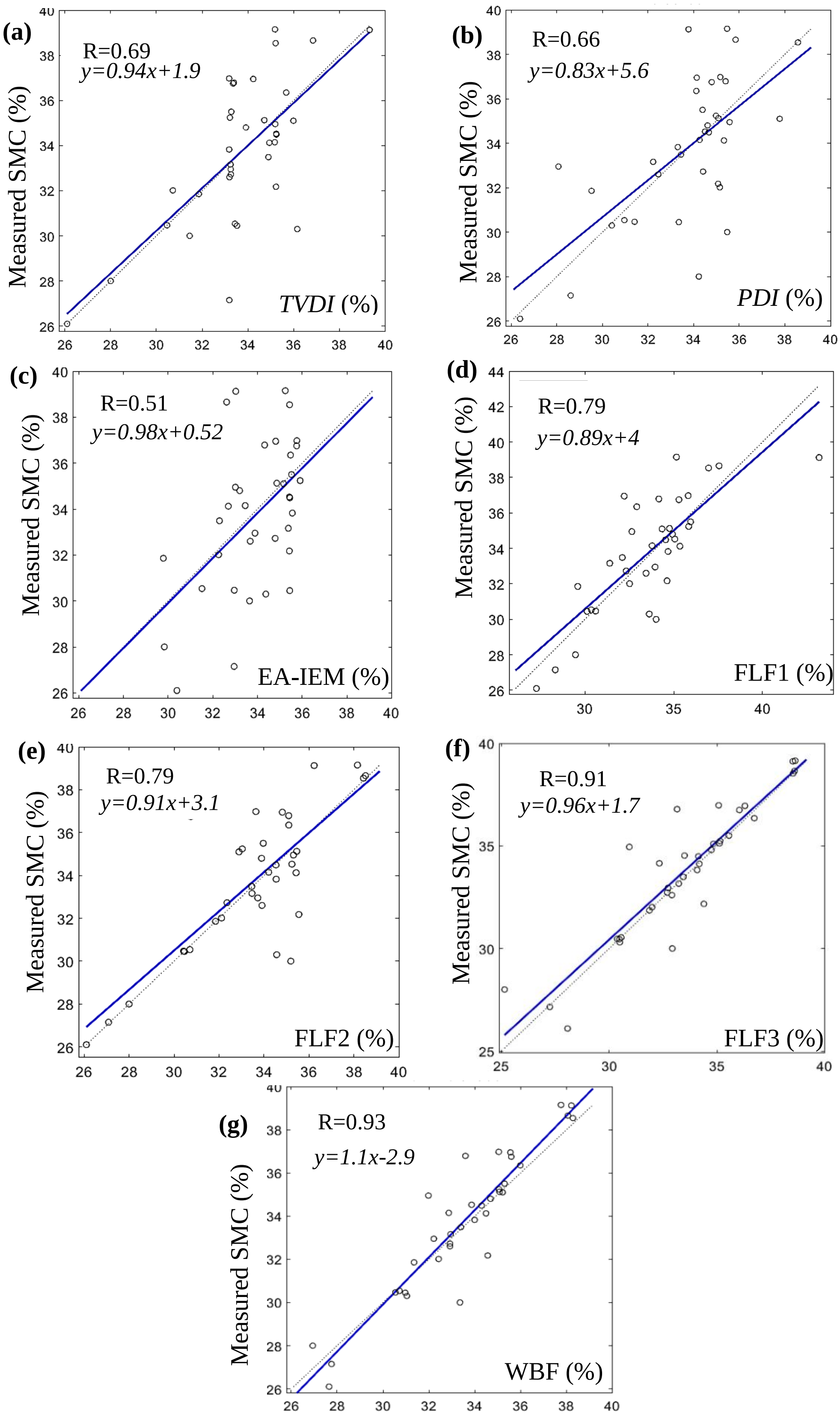

Figure 30. Results of each of the used SMC estimation methods in the Sidi Rached 2 datasets.

Table 12. **The results of each of the used estimation methods in the Sidi Rached 2 dataset.**

Estimation methods	**Sidi Rached 2 (*n*=36)**								
	RMSE (%)	Estimated SMC				Measured SMC			
		Min (%)	Max (%)	Mean (%)	SD (%)	Min (%)	Max (%)	Mean (%)	SD (%)
SMC_{TVDI}	2.32	26.09	39.32	33.73	2.34	26.1	39.15	33.75	3.2
SMC_{PDI}	2.44	36.83	38.57	33.69	2.53				
$SMC_{EA\text{-}IEM}$	2.74	29.78	35.9	33.92	1.69				
SMC_{FLF1}	1.97	27.26	43.19	33.58	2.87				
SMC_{FLF2}	1.98	26.09	38.51	33.56	2.77				
SMC_{FLF3}	1.34	25.14	38.62	33.46	3.05				
SMC_{WB}	**1.23**	26.93	38.27	33.51	2.72				

The best estimation using a single method in terms of RMSE and degree of correlation was achieved using the *TVDI* method this time (RMSE=2.32%, R=0.69), the potential reason for that is that the full range of vegetation cover intensity is present, with 41% of the overall pixels having *NDVI* values below 0.3. The performance of *PDI* was slightly poorer, which is understandable due to the presence of pixels with dense vegetation, which can have a negative impact on the determination of the soil line, another possible reason for *TVDI* and *PDI* outperforming EA-IEM inversion; is that the time of ground truth collection (09:00 a.m. to 10:30 a.m.) was closer to the Landsat-8 acquisition time (10:25 a.m.) than that of Sentinel-1 (05:45 a.m.).

The estimations using feature-level fusion outperform *TVDI* estimation in terms of RMSE and degree of correlation in this study area too. FLF1 and FLF2, seem to perform almost identically with the same degree of correlation (R=0.79), with slightly different RMSE values (1.97% and 1.98% respectively). As for FLF3, the inclusion of all of the available features has generated the best feature level fusion estimation in terms of accuracy and correlation (RMSE=1.34%, R=0.91) for this dataset too.

The estimation produced by WBF has again produced the best accuracy and the strongest correlation out of all of the used methods (RMSE=1.27%, and R=0.93), and the weights achieved for this study area were:

$$\boldsymbol{SMC_{WBF} = 0.04SMC_{TVDI} + 0.14\,SMC_{PDI} + 0.08SMC_{FLF1} + 0.02SMC_{FLF2} + 0.72SMC_{FLF3}} \tag{30}$$

The WBF method disregards the *EA-IEM* inversion estimations completely, the WBF in this study area generates an interesting configuration of weights, it assigns the largest weight to the most accurate estimation FLF3 (weight= 0.72). However, the second-largest weight was the one assigned to *PDI* estimation (weight = 0.14), then smaller weights to the *FLF1* estimation (weight =0.08), the *TVDI* estimation (weight=0.04) and the smallest weight was assigned to the FLF2 estimation (weight=0.02).

In this study area, the WBF method produced the best results in terms RMSE and R, and the closest to the measured SMC in terms of the range of the values of minimum, maximum, and mean.

The strong correlation of WBF estimation in this study area (R=0.93) reinforces the assumption that the correlation could be attributed to the length of the temporal gap between the times of acquisitions.

5.3 Remarks and Discussions

The achieved results come in good agreement with the findings of (Yahia, Guida et al. 2018a, Yahia, Guida et al. 2018b), the proposed SMC estimation system systematically improves the accuracy of SMC estimation in terms of lowered RMSE values in the orders of at least 0.38%, 1.4%, 1.09% for the datasets of Blackwell farms, Sidi Rached 1, and Sidi Rached 2 respectively, which is a great improvement on the accuracy of estimation achieved by a single method. However, more investigations are required to understand the signification of weight assignments.

The time of execution of the proposed SMC estimation system is somewhat reasonable (Blackwell farms=895.79 seconds, Sidi Rached 1=880.11 seconds, Sidi Rached 2= 891.27 seconds), and any future weights determinations will only be required if a new ground truth data is presented to the system (which would require further training).

However, the performance of the proposed SMC estimation system is sensitive to the vegetation cover intensity, it seems to perform better in bare soil to low vegetation cover (Blackwell farms), or in datasets with a full range of vegetation cover intensity from bare soil to dense vegetation, as it is the case in Sidi Rached 2, but its performance is relatively poorer in fully vegetated areas (Sidi Rached 1).

In addition to the performance issues described above, the author of this research is concerned that the low number of available samples for ANNs training could cause potential overfitting issues. Overfitting would limit the constraints of the use of the proposed SMC estimation system; it would render the latter specialised to areas with similar characteristics to those of the study areas in terms SMC distribution, the vegetation cover intensity and distribution, as well as soil compositions and surface roughness parameters. The temporal resolution of the used satellites is another concern, as large SMC changes could occur within the temporal gap between the times of acquisitions of the satellites. That will hopefully change by the inclusion of future satellite missions, especially the Landsat-9 mission (will be launched in December 2020), the mission largely replicates its predecessor Landsat-8 and in terms the on-board sensors (Markham, Jenstrom et al. 2016), which will reduce the temporal gap and increase the accuracy and correlation of the collected measurements to the earth observation data.

5.4 Comparison of the fusion centre to methodologies in Literature

In order to correctly gauge the value of this research, it is a good practice to compare its findings to those of the most relevant methodologies literature (available detailed description of said methodologies is section 2.6). However, each of these methodologies used different EO-data, auxiliary data, and ground truth data corresponding to a variety of study areas, at different spatial and temporal scales. It is extremely difficult to replicate these methodologies for the research presented in this thesis due to a variety of constraints. These numerous constraints could be any of these factors: the nature of the used satellites (active/passive), their frequency, their spatial resolution, the topography of the corresponding study, the intensity of vegetation cover, the species of vegetation cover, surface roughness parameters, and the instruments used for the validation. This renders the comparison of the results of these studies to the results of this research rather problematic. For instance, this research concentrates on SMC estimation for the regional scale at high spatial resolution. That rules out any methodologies based on coarse spatial resolution satellites such as SMAP, SMOS, AMSE-R, MODIS, which constitutes the majority of these studies. With that being said, it is still entirely crucial to the identity of the benchmark of accuracy associated with the state of the art EO-based SMC estimation systems in order appreciate qualitatively the value of the proposed system results. The different results of these studies, as well as the mean of all the results achieved by the feature level fusion as well as the fusion centre, are illustrated in Table 13.

Table 13. **Comparison between the results attained by relevant studies and the proposed SMC estimation system.**

Study	**RMSE (%)**	**R**	**Additional information**
(Kurucu, Sanli et al. 2009)	/	0.72	Image fusion of SPOT-2 and Radarsat-1.
(Bai, L., Long et al. 2019)	3.8	/	Image fusion of MODIS and Landsat-8.
(Bai, L., Long et al. 2019)	3.8	0.67	Image fusion in the form of downscaled and blended MODIS-LANDSAT-8 data to use in the trapezoidal method.
(Moran, Hymer et al. 2000)	(Absolute error) 2.51	0.96	Feature level fusion of data from Landsat-TM and ERS-2 C-band.
(Notarnicola, Posa 2001)	(Absolute Error) 10%	/	Feature level fusion using passive and active microwave data from RASAM truck-mounted radiometer-scatterometer at a frequency of 4.6 GHz.
(Posa, Notarnicola et al. 2004)	BAY1: 1.94 ANN1: 2.21 BAY2: 5.08 ANN2:3.28	BAY1: 0.8 ANN1: 0.78 BAY2: 0.68 ANN2:0.84	Feature level fusion using passive and active microwave data from Truck-mounted radiometer and scatterometer, and C-band scatterometer data.
(Park, Im et al. 2017)	/	AMSR-E=0.53 In-situ=0.77	Feature level fusion using data from AMSR-E, MODIS, and TRMM.

(Van der Schalie, De Jeu et al. 2018)	LS-reg=3 LPRM=2 ANN= 1.9	0.7	Feature level fusion using passive and active microwave data from AMSR-E validated by SMOS data.
(Portal, Vall-Llosscra et al. 2018)	7	0.8	Feature level fusion using microwave data from SMOS and optical and thermal data from MODIS.
(Xu, Yuan et al. 2019)	(ubRMSE) 7.1	0.74	Feature level fusion using passive microwave data from SMAP, data from MODIS, and data from the GEOS-5 model.
(Yuan, Xu et al. 2020)	(ubRMSE) 5	0.88	Feature level fusion using passive microwave data from SMAP, data from MODIS, and data from the GEOS-5 model.
(Huang, Liang et al. 2019)	13.1	0.89	Feature level fusion of features derived from GNSS data (SNR).
(Ren, Liang et al. 2019)	6	0.94	Feature level fusion of features derived from GNSS data (SNR).
Feature level fusion FLF3 (mean of the results of 3 study areas)	1.87	0.82	/
Fusion centre (mean of the results of 3 study areas)	1.75	0.85	/

Inspecting Table 13 reveals, that the fusion centre outperforms all other studies when it comes the RMSE values (1.75 %), all while still offering a strong correlation (R=0.85). These RMSE values cannot be contested nor compared due to the absence of any other similar studies, especially in terms of estimation/decision level fusion systems in the literature. The same cannot be said about the proposed feature levels fusions, as numerous studies exploit that particular processing level albeit in different spatial scale, and sensors. The proposed feature level fusion in this research (FLF3) still outperforms all the other studies in the reviewed literature in terms of RMSE (1.87%), with relatively weaker correlation (R=0.82). the study authored by (Moran, Hymer et al. 2000) produces the strongest correlation out of all the reviewed literature (R=0.96), however, it is hard to evaluate its accuracy of estimation due to the study using another metric of accuracy (absolute error) instead of RMSE. Studies using GNSS (Huang, Liang et al. 2019, Ren, Liang et al. 2019) data seem to offer strong overall correlations to their respective validation data (R=0.94, R=0.89) while offering less than optimal RMSE values (6% and 13% respectively). The study proposed by (Posa, Notarnicola et al. 2004) seems also to produce a good compromise between the correlation and low RMSE values, but that could be explained by the fact that the used sensors are in fact more proximal than remote (Truck mounted radiometer and scatterometer). The study proposed by (Van der Schalie, De Jeu et al. 2018) produced low RMSE values (one of the methods achieved ANN RMSE=1.9%), however, it is important to note that these values are not consistent throughout the whole validation samples (only at specific NDVI intervals) and all pixels corresponding outside of that interval are not considered for the calculations of RMSE.

In conclusion, it is an accurate statement to declare that the results achieved in this research are up to the standard of state of the art. Correlation wise, it could be argued that the reason that the correlation is not even stronger is due to the temporal resolution, or the heterogeneity of the used EO-data as well as the used methods. Despite that, the results achieved by the proposed methodology still produce the lowest RMSE values in all of the reviewed literature at the time of this research.

5.5 Conclusion

In chapter 5, the results achieved by the proposed SMC estimation system were analysed according to their study area, the proposed system has to achieve the best results in terms of RMSE values in the Sidi Rached 2 dataset (1.27%), followed by the Blackwell farms dataset (1.34%) and finally by Sidi Rached 1 dataset (2.7%), which suggest that the system performs its best in the presence of a full range of vegetation cover. However, it is worth noting that the most significant improvement of this system on estimations using single methods, was achieved in the Sidi Rached 1 data with lowered RMSE values in the order of 1.4%. Chapter 5 also discussed and offered a few remarks on the factors that cause or could cause performance issues of the proposed SMC estimation system.

The following and final chapter will be Chapter 6: Conclusions and Future Works.

6. Conclusions and Future Work

6.1 Summary

The objective of this thesis was to develop a novel multi-sensory data fusion system for soil moisture content estimation.

The designed system is composed of several distinct soil moisture content retrieval methodologies from different remote sensing technologies as described by the state of the art chapter, each of those methods, whether be it the *IEM* inversion, *PDI* or *TVDI,* suffer from several limitations that have an adverse effect on their respective performance in terms of accuracy, which prompted the authors of this research to propose the use of data fusion techniques as a possible solution to minimise those performance-related issues. Furthermore, the single scattering IEM has a limited validity range of surface roughness values, which caused the author to propose the use of the Empirically Adapted IEM with the addition of a semi-empirical calibration (*Lopt*) instead.

The rationale behind the fusion scheme of the proposed SMC estimation system was explained in the methodology chapter. It was maintained that increasing the dimensionality of the feature space would increase the accuracy of SMC estimation. It was also argued that assigning weights to each of the individual estimations, including the feature level fusions, will ameliorate the accuracy of the final output of the proposed SMC estimation system.

The novel SMC estimation system required ground truth measurements from study areas for tests and validations. Those study areas were subject to a detailed description in the study area chapter. That description included the different characteristics in terms of their locations, vegetation state and the reasons behind their selection, not to mention the experienced difficulties in the process of their choice. That description also included a detailed account of Sentinel-1 and Landsat-8 satellite missions as well as the ML3 Theta Probe and laser and needle profilometres.

The novel SMC estimation system was evaluated and analysed according to their study areas. The proposed system produces the best results in terms of RMSE values in the Sidi Rached 2 dataset (1.27%), followed by the Blackwell farms dataset (1.34%) and finally by Sidi Rached 1 dataset (2.7%).

The author concludes that the accuracy of the proposed system is better than any of the individual estimations achieved by *PDI*, *TVDI*, and the updated *EA-IEM* inversion and that the system performs its best in the presence of a full range of vegetation cover. Despite these promising results, there were also some concerns regarding the performance of the proposed system. The low number of collected SMC measurements used for the ANN training phase could potentially cause overfitting. Another concern is the temporal gap between the acquisition time of the satellites that produce EO-data used in the proposed system, which can cause a massive effect on the correlation of EO-data to SMC levels. Another issue to consider is the revisit time of the satellites (the longest is Landsat-8: 16 days), especially, as it is the case in this research, if the idea is to monitor SMC levels for agricultural practices. The impact of the latter issue could be minimised by the inclusion of additional satellites such as Sentinel-2, and the probable addition of Landsat-9 after its launch. Finally, further research is required to address the limitation of the system in the absence of the full range of vegetation cover.

6.2 Main Novelty Contributions

In the time spent conducting this research, the author was able to attain the following achievements:

- Acquired of the necessary backgrounds in remote sensing, especially multispectral, thermal and SAR imaging technologies.
- Conducted a literature survey of the relevant soil moisture retrieval methods in remote sensing.
- Designed a novel SMC estimation system using the knowledge acquired from the literature survey.
- Addressed the limitations of IEM inversion in terms of the range validity of soil roughness parameters, by proposing the use of *Lopt* and a later version of the IEM (EA-IEM).
- Got familiarised with the instrument used for SMC measurement (ML3 ThetaProbe).
- Planned and conducted ground truth measurement campaigns in two different countries (the United Kingdom and Algeria).
- Built a portable laser profilometer to measure surface roughness parameters.
- Processed ground truth SMC measurements and matched them to their corresponding pixels in the earth observation data.

- Developed the novel SMC estimation system and all its relevant methodologies in the Matlab environment.
- Achieved more accurate results in terms of lowered RMSE values across all the datasets.
- Published two conference papers.

The perceived novelty and contribution to the state of the art provided by the research done during this PhD course are the following sorted in order of signification:

- The design and implementation of the fusion centre, which consists of an EO-based SMC estimation through a decision level fusion of estimations from multiple methodologies (TVDI, PDI, and an updated EA-IEM inversion) as well as multiple feature fusions (FLF1, FLF2, and FLF3) by assigning a weight to each estimation.
- Obtaining encouraging results manifested in high accuracy SMC estimation. These results were represented by the low RMSE values (mean of 1.75%), which were lower than any of the individual estimation methods, and slightly lower than those achieved by different variations of feature level fusions. Those results were also compared to the results achieved by the most relevant studies. The results achieved by the proposed system were proven to have lower RMSE values than those found in the relevant literature. However, the aforementioned results exhibited a relatively weaker correlation which could subject to further investigations.
- The design and implementation of multiple feature level fusions using a different and completely novel combination of features extracted from different methods (PDI, TVDI, an updated EA-IEM inversion).
- The semi-empirical calibration the EA-IEM model (Song, Zhou et al. 2009) to include a wider range of surface roughness profiles by the replacement of the correlation length by a semi-empirical parameter *Lopt* (Baghdadi, Holah et al. 2006).
- The alteration of the performance function of the used ANN from absolute error to RMSE.
- The inception and implementation of a new Laser profilometre for surface roughness measurements.
- The conception, planning, and execution of field campaigns for SMC and surface roughness truth measurements for validations and tests in 3 different study areas in Blackwell farms (UK), Sidi Rached (Tipasa, Algeria). These study areas have never been the subject of a research study before.

However, the author also recognises the potential improvement of the novel SMC estimation system, which will be discussed in section 6.3.

6.3 Future Work

While the proposed system has produced satisfactory results, the system still has room for improvements, especially in terms of the subjectivity or overfitting of the ANN to the available data samples. That could be remedied by the acquisition of more ground truth measurements of SMC and surface roughness parameters. Another aspect to investigate is the process of weight assignment. The method should assign the weights only for the most accurate estimations (which it does). However, it sometimes assigns small weights to less accurate estimations at the expense of more accurate ones (it is apparent in the Blackwell farms and Sidi Rached 2 dataset).

Another area that needs to be addressed is the temporal gap. While waiting for the launch of Landsat-9 satellite mission, it is worth investigating the effect of using Sentinel-2 data when the Landsat-8 data is unavailable, while that would sacrifice the thermal information that Landsat-8 provides, it would increase the frequency of having the satellites revisit the same area in the same day, which only occurs twice a month for the Landsat-8 and Sentinel-1, and the proposed system is designed to be reconfigurable enough to cope with the absence of Landsat-8 data. Even with the potential inclusion of Sentinel-2, the temporal gap of at least 3 hours is too long, especially if the study area had extreme weather conditions. That could be addressed by the use of data assimilation techniques and the acquisition of better understanding the temporal variability of SMC. The proposed data assimilations methodology by authors in (Zaman, McKee et al. 2012) seems quite interesting, as it explores the possibility of using past and current data to predict soil moisture weather in a different spatial level i.e. root zone SMC, and different temporal intervals. Which could potentially be applicable in future iterations of the proposed system in this research.

More evaluations are also required to judge the performance of the proposed system with different plant species as all study areas were grasslands.

Another possible direction this research could take is the consideration of weather conditions factors like precipitation, air temperature, humidity and wind speed in future updates of the novel system.

Finally, the effect of SMC values in the lower layers of the soil surface on the surface SMC is yet to be investigated, the potential use an L-band SAR if available, as the latter is famed for its ability for surface penetration, it would provide SMC estimations at slightly deeper layers, which would clarify the degree of significance of those values on the performance of the novel SMC.

7. REFERENCES

ABBOTT, M.B., BATHURST, J.C., CUNGE, J.A., O'CONNELL, P.E. and RASMUSSEN, J., 1986. An introduction to the European Hydrological System—Systeme Hydrologique Europeen,"SHE", 2: Structure of a physically-based, distributed modelling system. *Journal of hydrology,* **87**(1-2), pp. 61-77.

AISYAH, A., KUSRATMOKO, E. and WIBOWO, A., 2019The Spread Of Dry Area Based On TVDI Index (Temperature Vegetation Dryness Index), *IOP Conference Series: Earth and Environmental Science* 2019, IOP Publishing, pp. 012006.

ALEXAKIS, D.D., MEXIS, F.K., VOZINAKI, A.K., DALIAKOPOULOS, I.N. and TSANIS, I.K., 2017. Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A hydrological approach. *Sensors,* **17**(6), pp. 1455.

AL-YAARI, A., WIGNERON, J., KERR, Y., RODRIGUEZ-FERNANDEZ, N., O'NEILL, P.E., JACKSON, T.J., DE LANNOY, G., AL BITAR, A., MIALON, A. and RICHAUME, P., 2017. Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sensing of Environment,* **193**, pp. 257-273.

AMANI, M., SALEHI, B., MAHDAVI, S., MASJEDI, A. and DEHNAVI, S., 2017. Temperature-Vegetation-soil Moisture Dryness Index (TVMDI). *Remote Sensing of Environment,* **197**, pp. 1-14.

AMOLINS, K., ZHANG, Y. and DARE, P., 2007. Wavelet based image fusion techniques—An introduction, review and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing,* **62**(4), pp. 249-263.

ASCE TASK COMMITTEE ON APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY, 2000. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering,* **5**(2), pp. 115-123.

BAGHDADI, N., CRESSON, R., EL HAJJ, M., LUDWIG, R. and LA JEUNESSE, I., 2012. Estimation of soil parameters over bare agriculture areas from C-band polarimetric

SAR data using neural networks. *Hydrology and Earth System Sciences,* **16**, pp. p. 1607-p. 1621.

BAGHDADI, N., GAULTIER, S. and KING, C., 2002. Retrieving surface roughness and soil moisture from synthetic aperture radar (SAR) data using neural networks. *Canadian Journal of Remote Sensing,* **28**(5), pp. 701-711.

BAGHDADI, N., CHAAYA, J.A. and ZRIBI, M., 2011. Semiempirical calibration of the integral equation model for SAR data in C-band and cross polarization using radar images and field measurements. *IEEE Geoscience and Remote Sensing Letters,* **8**(1), pp. 14-18.

BAGHDADI, N., EL HAJJ, M., CHOKER, M., ZRIBI, M., BAZZI, H., VAUDOUR, E., GILLIOT, J.M., BOUSBIH, S. and MWAMPONGO, D.E., 2018Potential of Sentinel-1 for Estimating the Soil Roughness Over Agricultural Soils, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* 2018, IEEE, pp. 7516-7519.

BAGHDADI, N., GHERBOUDJ, I., ZRIBI, M., SAHEBI, M., KING, C. and BONN, F., 2004. Semi-empirical calibration of the IEM backscattering model using radar images and moisture and roughness field measurements. *International Journal of Remote Sensing,* **25**(18), pp. 3593-3623.

BAGHDADI, N. and ZRIBI, M., 2006. Evaluation of radar backscatter models IEM, OH and Dubois using experimental observations. *International Journal of Remote Sensing,* **27**(18), pp. 3831-3852.

BAGHDADI, N., ZRIBI, M., LOUMAGNE, C., ANSART, P. and ANGUELA, T.P., 2008. Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields. *Remote Sensing of Environment,* **112**(12), pp. 4370-4379.

BAGHDADI, HOLAH and ZRIBI, 2006. Calibration of the integral equation model for SAR data in C-band and HH and VV polarizations. *International Journal of Remote Sensing,* **27**(4), pp. 805-816.

BAI, L., LONG, D. and YAN, L., 2019. Estimation of surface soil moisture with downscaled land surface temperatures using a data fusion approach for heterogeneous agricultural land. *Water Resources Research,* **55**(2), pp. 1105-1128.

BAI, X., HE, B. and LI, X., 2016. Optimum surface roughness to parameterize advanced integral equation model for soil moisture retrieval in prairie area using Radarsat-2 data. *IEEE Transactions on Geoscience and Remote Sensing*, **54**(4), pp. 2437-2449.

BARRETT, B.W., DWYER, E. and WHELAN, P., 2009. Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote sensing*, **1**(3), pp. 210-242.

BERRY, S.K., KALRA, C.L., SEHGAL, R.C., KULKARNI, S.G., SUKHVIR, K., ARORA, S.K. and SHARMA, B.R., 1988. Quality characteristics of seeds of five okra (Abelmoschus esculentus L.) cultivars. *Journal of Food Science and Technology*, **25**(5), pp. 303-305.

BINDLISH, R. and BARROS, A.P., 2000. Multifrequency soil moisture inversion from SAR measurements with the use of IEM. *Remote Sensing of Environment*, **71**(1), pp. 67-88.

BROWN, M.E., ESCOBAR, V., MORAN, S., ENTEKHABI, D., O'NEILL, P.E., NJOKU, E.G., DOORN, B. and ENTIN, J.K., 2013. NASA's soil moisture active passive (SMAP) mission and opportunities for applications users. *Bulletin of the American Meteorological Society*, **94**(8), pp. 1125-1128.

BRYANT, R., MORAN, M.S., THOMA, D.P., COLLINS, C.H., SKIRVIN, S., RAHMAN, M., SLOCUM, K., STARKS, P., BOSCH, D. and DUGO, M.G., 2007. Measuring surface roughness height to parameterize radar backscatter models for retrieval of surface soil moisture. *IEEE Geoscience and Remote Sensing Letters*, **4**(1), pp. 137-141.

BYUN, K., LIAQAT, U.W. and CHOI, M., 2014. Dual-model approaches for evapotranspiration analyses over homo-and heterogeneous land surface conditions. *Agricultural and Forest Meteorology*, **197**, pp. 169-187.

C. BROUWER, A. GOFFEAU and M. HEIBLOEM, 1985-last update, Irrigation Water Management: Training Manual No. 1 - Introduction to Irrigation. Chapter 2 - Soil and Water . Available: http://www.fao.org/3/r4082e/r4082e03.htm [04/06/, 2020].

CANDIAGO, S., REMONDINO, F., DE GIGLIO, M., DUBBINI, M. and GATTELLI, M., 2015. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote sensing,* **7**(4), pp. 4026-4047.

CASAMITJANA, TORRES-MADROÑERO, BERNAL and VARGA, 2020. Soil Moisture Analysis by Means of Multispectral Images According to Land Use and Spatial Resolution on Andosols in the Colombian Andes. *Applied Sciences*, **10**(16), pp. 5540.

CENCI, L., PULVIRENTI, L., BONI, G. and PIERDICCA, N., 2018. Defining a trade-off between spatial and temporal resolution of a geosynchronous SAR mission for soil moisture monitoring. *Remote Sensing,* **10**(12), pp. 1950.

CHEN, J., WANG, C., JIANG, H., MAO, L. and YU, Z., 2011. Estimating soil moisture using Temperature–Vegetation Dryness Index (TVDI) in the Huang-huai-hai (HHH) plain. *International Journal of Remote Sensing,* **32**(4), pp. 1165-1177.

CHEN, K.S., YEN, S.K. and HUANG, W.P., 1995. A simple model for retrieving bare soil moisture from radar-scattering coefficients. *Remote Sensing of Environment,* **54**(2), pp. 121-126.

CHEN, SUN, LIU and PEI, 2019Drought monitoring using MODIS derived perpendicular drought indexes, *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* 2019, IEEE, pp. 1-4.

CHEN, SUN, WANG and PEI, 2019Application of Sentinel 2 data for drought monitoring in Texas, America, *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* 2019, IEEE, pp. 1-4.

CHEN, WEN, JIANG, ZHAO, ZHANG and CHEN, Y., 2015. Temperature vegetation dryness index estimation of soil moisture under different tree species. *Sustainability,* **7**(9), pp. 11401-11417.

CHO, J., LEE, Y. and LEE, H., 2014. Assessment of the relationship between thermal-infrared-based temperature– vegetation dryness index and microwave satellite-derived soil moisture. *Remote sensing letters,* **5**(7), pp. 627-636.

DAI, X. and KHORRAM, S., 1999. Data fusion using artificial neural networks: a case study on multitemporal change analysis. *Computers, Environment and Urban Systems,* **23**(1), pp. 19-31.

DALIAKOPOULOS, I.N., COULIBALY, P. and TSANIS, I.K., 2005. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology,* **309**(1), pp. 229-240.

DAWSON, M.S., FUNG, A.K. and MANRY, M.T., 1997. A robust statistical-based estimator for soil moisture retrieval from radar measurements. *IEEE Transactions on Geoscience and Remote Sensing,* **35**(1), pp. 57-67.

DITZLER, C., SCHEFFE, K. and MONGER, H.C., 2017. Soil science division staff. *Soil survey manual.USDA Handbook,* **18**, pp. 120-128.

DONG, J., ZHUANG, D., HUANG, Y. and FU, J., 2009. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors,* **9**(10), pp. 7771-7784.

DÖRING, B.J. and SCHWERDT, M., 2013. The radiometric measurement quantity for SAR images. *IEEE Transactions on Geoscience and Remote Sensing,* **51**(12), pp. 5307-5314.

DU, L., SONG, N., LIU, K., HOU, J., HU, Y., ZHU, Y., WANG, X., WANG, L. and GUO, Y., 2017. Comparison of two simulation methods of the temperature vegetation dryness index (TVDI) for drought monitoring in semi-arid regions of China. *Remote Sensing,* **9**(2), pp. 177.

DUBOIS, P.C., VAN ZYL, J. and ENGMAN, T., 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing,* **33**(4), pp. 915-926.

ENTEKHABI, D., NJOKU, E.G., O'NEILL, P.E., KELLOGG, K.H., CROW, W.T., EDELSTEIN, W.N., ENTIN, J.K., GOODMAN, S.D., JACKSON, T.J. and JOHNSON, J., 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE,* **98**(5), pp. 704-716.

ENTEKHABI, D., NJOKU, E., O'NEILL, P., SPENCER, M., JACKSON, T., ENTIN, J., IM, E. and KELLOGG, K., 2008The soil moisture active/passive mission (SMAP), *IGARSS*

2008-2008 IEEE International Geoscience and Remote Sensing Symposium 2008, IEEE, pp. III-4.

ENTEKHABI, D., YUEH, S., O'NEILL, P.E., KELLOGG, K.H., ALLEN, A., BINDLISH, R., BROWN, M., CHAN, S., COLLIANDER, A. and CROW, W.T., 2014. SMAP handbook–soil moisture active passive: Mapping soil moisture and freeze/thaw from space.

FAO, 2011-last update, Agri-environmental indicators (Land). Available: http://faostat3.fao.org/download/E/EL/E [August, 2016].

FENTON, O., VERO, S., IBRAHIM, T.G., MURPHY, P., SHERRIFF, S.C. and Ó HUALLACHÁIN, D., 2015. Consequences of using different soil texture determination methodologies for soil physical quality and unsaturated zone time lag estimates. *Journal of contaminant hydrology,* **182**, pp. 16-24.

FREI, M. and HENKEL, J., 2002Influence of physical surface characteristics on SAR backscatter, *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology* 2002, International Society for Optics and Photonics, pp. 236-242.

FUNG, A.K., LI, Z. and CHEN, K., 1992. Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing,* **30**(2), pp. 356-369.

GAO, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment,* **58**(3), pp. 257-266.

GAO, Z., GAO, W. and CHANG, N., 2011. Integrating temperature vegetation dryness index (TVDI) and regional water stress index (RWSI) for drought assessment with the aid of LANDSAT TM/ETM images. *International Journal of Applied Earth Observation and Geoinformation,* **13**(3), pp. 495-503.

GAO, Z., XU, X., WANG, J., YANG, H., HUANG, W. and FENG, H., 2013. A method of estimating soil moisture based on the linear decomposition of mixture pixels. *Mathematical and Computer Modelling,* **58**(3), pp. 606-613.

GARDNER, M.W. and DORLING, S.R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment,* **32**(14-15), pp. 2627-2636.

GE, S., ZHANG, J., SUN, W., WANG, L., GENG, Y. and FAN, S., 2018. Application of three drought indexes in soil moisture inversion using remote sensing in marsh wetlands of arid area. *Acta Ecologica Sinica,* (7), pp. 6.

GHERBOUDJ, I., MAGAGI, R., BERG, A.A. and TOTH, B., 2011. Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. *Remote Sensing of Environment,* **115**(1), pp. 33-43.

GHULAM, A., QIN, Q., TEYIP, T. and LI, Z., 2007. Modified perpendicular drought index (MPDI): a real-time drought monitoring method. *ISPRS journal of photogrammetry and remote sensing,* **62**(2), pp. 150-164.

GHULAM, A., QIN, Q. and ZHAN, Z., 2007. Designing of the perpendicular drought index. *Environmental Geology,* **52**(6), pp. 1045-1052.

HACHANI, A., OUESSAR, M., PALOSCIA, S., SANTI, E. and PETTINATO, S., 2019. Soil moisture retrieval from Sentinel-1 acquisitions in an arid environment in Tunisia: application of Artificial Neural Networks techniques. *International Journal of Remote Sensing,* , pp. 1-22.

HAJNSEK, I. and PAPATHANASSIOU, K., 2005. Rough surface scattering models. *ESA Earth Online,* .

HALL, D.L. and LLINAS, J., 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE,* **85**(1), pp. 6-23.

HALLIKAINEN, M.T., ULABY, F.T., DOBSON, M.C., EL-RAYES, M.A. and WU, L., 1985. Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing,* (1), pp. 25-34.

HASSAN-ESFAHANI, L., TORRES-RUA, A., TICLAVILCA, A.M., JENSEN, A. and MCKEE, M., 2014Topsoil moisture estimation for precision agriculture using unmmaned aerial vehicle multispectral imagery, *2014 IEEE Geoscience and Remote Sensing Symposium* 2014, IEEE, pp. 3263-3266.

HAUSER, M., ORTH, R. and SENEVIRATNE, S.I., 2016. Role of Soil Moisture vs. Recent Climate Change for the 2010 Heat Wave in Western Russia. *Geophysical Research Letters,* .

HECHT-NIELSEN, R., 1988. Theory of the backpropagation neural network. *Neural Networks,* **1**(Supplement-1), pp. 445-448.

HILLEL, D., 1998. *Environmental soil physics: Fundamentals, applications, and environmental considerations.* Academic press.

HUANG, S., DING, J., ZOU, J., LIU, B., ZHANG, J. and CHEN, W., 2019. Soil Moisture Retrival Based on Sentinel-1 Imagery under Sparse Vegetation Coverage. *Sensors,* **19**(3), pp. 589.

HUANG, LIANG, YUAN, QIU, HUANG and REN, 2019Research on Estimation Method of Surface Soil Moisture Content Based on Multi-star Fusion, *China Satellite Navigation Conference* 2019, Springer, pp. 193-202.

HUETE, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment,* **25**(3), pp. 295-309.

HUNT, E.R., HIVELY, W.D., DAUGHTRY, C.S., MCCARTY, G.W., FUJIKAWA, S.J., NG, T.L., TRANCHITELLA, M., LINDEN, D.S. and YOEL, D.W., 2008Remote sensing of crop leaf area index using unmanned airborne vehicles, *Proceedings of the Pecora 17 Symposium, Denver, CO* 2008.

IERVOLINO, 2015. *Ship detection with SAR:*
modelling, designing and real
data validation, University of Surrey.

JACKSON, T.J., 1993. III. Measuring surface soil moisture using passive microwave remote sensing. *Hydrological Processes,* **7**(2), pp. 139-152.

JACKSON, T.J., SCHMUGGE, J. and ENGMAN, E.T., 1996. Remote sensing applications to hydrology: soil moisture. *Hydrological Sciences Journal,* **41**(4), pp. 517-530.

JIANG, D., YANG, X., CLINTON, N. and WANG, N., 2004a. An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing,* **25**(9), pp. 1723-1732.

JIANG, D., YANG, X., CLINTON, N. and WANG, N., 2004b. An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing,* **25**(9), pp. 1723-1732.

KARTHIKEYAN, L., PAN, M., WANDERS, N., KUMAR, D.N. and WOOD, E.F., 2017. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Advances in Water Resources*, **109**, pp. 106-120.

KELLOGG, C.E., 1993. Soil survey division staff: soil survey manual. *United States Department of Agriculture, Washington,* .

KERR, Y.H., FONT, J., MARTIN-NEIRA, M. and MECKLENBURG, S., 2012. Introduction to the Special Issue on the ESA's Soil Moisture and Ocean Salinity Mission (SMOS)-Instrument Performance and First Results. *IEEE Trans.Geoscience and Remote Sensing,* **50**(5-1), pp. 1351-1353.

KERR, Y.H., WALDTEUFEL, P., WIGNERON, J., DELWART, S., CABOT, F., BOUTIN, J., ESCORIHUELA, M., FONT, J., REUL, N. and GRUHIER, C., 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proceedings of the IEEE,* **98**(5), pp. 666-687.

KERR, Y.H., WALDTEUFEL, P., WIGNERON, J., MARTINUZZI, J., FONT, J. and BERGER, M., 2001a. Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing,* **39**(8), pp. 1729-1735.

KERR, Y.H., WALDTEUFEL, P., WIGNERON, J., MARTINUZZI, J., FONT, J. and BERGER, M., 2001b. Soil moisture retrieval from space: The Soil Moisture and Ocean

Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing,* **39**(8), pp. 1729-1735.

KHABAZAN, S., MOTAGH, M. and HOSSEINI, M., 2013. Evaluation of Radar Backscattering Models IEM, OH, and Dubois using L and C-Bands SAR Data over different vegetation canopy covers and soil depths. *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* , pp. 225-230.

KOGAN, F.N., 1995. Application of vegetation index and brightness temperature for drought detection. *Advances in space research,* **15**(11), pp. 91-100.

KORNELSEN, K.C. and COULIBALY, P., 2013a. *Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications.*

KORNELSEN, K.C. and COULIBALY, P., 2013b. Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. *Journal of Hydrology,* **476**, pp. 460-489.

KOYAMA, C.N., LIU, H., TAKAHASHI, K., SHIMADA, M., WATANABE, M., KHUUT, T. and SATO, M., 2017. In-Situ Measurement of Soil Permittivity at Various Depths for the Calibration and Validation of Low-Frequency SAR Soil Moisture Models by Using GPR. *Remote Sensing,* **9**(6), pp. 580.

KURUCU, Y., SANLI, F.B., ESETLILI, M.T., BOLCA, M. and GOKSEL, C., 2009. Contribution of SAR images to determination of surface moisture on the Menemen Plain, Turkey. *International Journal of Remote Sensing,* **30**(7), pp. 1805-1817.

LAMBIN, E.F. and EHRLICH, D., 1996. The surface temperature-vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing,* **17**(3), pp. 463-487.

LEI, S., BIAN, Z., DANIELS, J.L. and LIU, D., 2014. Improved spatial resolution in soil moisture retrieval at arid mining area using apparent thermal inertia. *Transactions of Nonferrous Metals Society of China,* **24**(6), pp. 1866-1873.

LOBELL, D.B. and ASNER, G.P., 2002. Moisture effects on soil reflectance. *Soil Science Society of America Journal*, **66**(3), pp. 722-727.

LOWDER, S.K., SKOET, J. and RANEY, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Development,* **87**, pp. 16-29.

MANSOURPOUR, M., RAJABI, M.A. and BLAIS, J., 2006Effects and performance of speckle noise reduction filters on active radar and SAR images, *Proc. ISPRS* 2006, pp. W41.

MAO, K.B., TANG, H.J., ZHANG, L.X., LI, M.C., GUO, Y. and ZHAO, D.Z., 2008. A method for retrieving soil moisture in Tibet region by utilizing microwave index from TRMM/TMI data. *International Journal of Remote Sensing,* **29**(10), pp. 2903-2923.

MARKHAM, B.L., JENSTROM, D., MASEK, J.G., DABNEY, P., PEDELTY, J.A., BARSI, J.A. and MONTANARO, M., 2016Landsat 9: status and plans, *Earth Observing Systems XXI* 2016, International Society for Optics and Photonics, pp. 99720G.

MARQUARDT, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics,* **11**(2), pp. 431-441.

MATTIA, F., BALENZANO, A., SATALINO, G., LOVERGINE, F., PENG, J., WEGMULLER, U., CARTUS, O., DAVIDSON, M.J., KIM, S. and JOHNSON, J., 2018Sentinel-1 & Sentinel-2 for soil moisture retrieval at field scale, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* 2018, IEEE, pp. 6143-6146.

MIRMAZLOUMI, S.M. and SAHEBI, M.R., 2016. Assessment of different backscattering models for bare soil surface parameters estimation from SAR data in band C, L and P. *European Journal of Remote Sensing,* **49**(1), pp. 261-278.

MIRSOLEIMANI, H.R., SAHEBI, M.R., BAGHDADI, N. and EL HAJJ, M., 2019. Bare Soil Surface Moisture Retrieval from Sentinel-1 SAR Data Based on the Calibrated IEM and Dubois Models Using Neural Networks. *Sensors,* **19**(14), pp. 3209.

ML3 THETAPROBE USER MANUAL, , ML3 ThetaProbe user Manual. Available: https://www.delta-t.co.uk/wp-content/uploads/2017/02/ML3-user-manual-version-2.1.pdf [22/10/, 2019].

MO, T., SCHMUGGE, T.J. and JACKSON, T.J., 1984. Calculations of radar backscattering coefficient of vegetation-covered soils. *Remote Sensing of Environment,* **15**(2), pp. 119-133.

MORAN, M.S., HYMER, D.C., QI, J. and SANO, E.E., 2000. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. *Agricultural and Forest Meteorology,* **105**(1-3), pp. 69-80.

MORAN, M.S., PETERS-LIDARD, C.D., WATTS, J.M. and MCELROY, S., 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian journal of remote sensing,* **30**(5), pp. 805-826.

MOREIRA, A., PRATS-IRAOLA, P., YOUNIS, M., KRIEGER, G., HAJNSEK, I. and PAPATHANASSIOU, K.P., 2013. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine,* **1**(1), pp. 6-43.

MULLA, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering,* **114**(4), pp. 358-371.

NASR, A.H. and RAMADAN, T.M., 2008. Data fusion using IHS transformations for exploring ore deposits in Northeastern part of the Saharan Metacraton. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* **37**(B7), pp. 1119-1124.

NHEMACHENA, C., MATCHAYA, G., NHEMACHENA, C.R., KARUAIHE, S., MUCHARA, B. and NHLENGETHWA, S., 2018. Measuring baseline agriculture-related sustainable development goals index for southern Africa. *Sustainability,* **10**(3), pp. 849.

NIKOLAKOPOULOS, K.G., 2008. Comparison of nine fusion techniques for very high resolution data. *Photogrammetric Engineering & Remote Sensing,* **74**(5), pp. 647-659.

NOTARNICOLA, C. and POSA, F., 2001Bayesian fusion of active and passive microwave data for estimating bare soil water content, *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)* 2001, IEEE, pp. 1167-1169.

OH, Y., SARABANDI, K. and ULABY, F.T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing,* **30**(2), pp. 370-381.

OLDAK, A., JACKSON, T.J., STARKS, P. and ELLIOTT, R., 2003. Mapping near-surface soil moisture on regional scale using ERS-2 SAR data. *International Journal of Remote Sensing,* **24**(22), pp. 4579-4598.

OLKKONEN, H. and PESOLA, P., 1996. Gaussian pyramid wavelet transform for multiresolution analysis of images. *Graphical Models and Image Processing,* **58**(4), pp. 394-398.

OLTRA-CARRI, R., BAUP, F., FABRE, S., FIEUZAL, R. and BRIOTTET, X., 2015. Improvement of soil moisture retrieval from hyperspectral VNIR-SWIR data using clay content information: From laboratory to field experiments. *Remote Sensing,* **7**(3), pp. 3184-3205.

PALOSCIA, S., PAMPALONI, P., PETTINATO, S. and SANTI, E., 2008. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Transactions on Geoscience and Remote Sensing,* **46**(10), pp. 3274-3284.

PARK, S., IM, J., PARK, S. and RHEE, J., 2017. Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. *Agricultural and Forest Meteorology,* **237**, pp. 257-269.

PENG, LOEW, MERLIN and VERHOEST, 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics,* **55**(2), pp. 341-366.

PETROPOULOS, G., CARLSON, T.N., WOOSTER, M.J. and ISLAM, S., 2009. A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. *Progress in Physical Geography,* **33**(2), pp. 224-250.

PETROPOULOS, G.P., GRIFFITHS, H.M., DORIGO, W., XAVER, A. and GRUBER, A., 2013. Surface soil moisture estimation: Significance, controls and conventional measurement techniques. *Remote sensing of energy fluxes and soil moisture content*, , pp. 29-48.

PETROPOULOS, IRELAND and BARRETT, 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C*, **83**, pp. 36-56.

PETTORELLI, N., VIK, J.O., MYSTERUD, A., GAILLARD, J., TUCKER, C.J. and STENSETH, N.C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution*, **20**(9), pp. 503-510.

PORTAL, G., VALL-LLOSSCRA, M., PILES, M., CAMPS, A., CHAPARRO, D., PABLOS, M., ROSSATO, L. and AABOUCH, K., 2018Microwave and Optical Data Fusion for Global Mapping of Soil Moisture at High Resolution, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* 2018, IEEE, pp. 341-344.

POSA, F., NOTARNICOLA, C. and ANGIULLI, M., 2004Soil parameters retrieval from remotely sensed data: efficiency of neural network and Bayesian approaches, *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium* 2004, IEEE, pp. 4682-4685.

PUREVDORJ, T.S., TATEISHI, R., ISHIYAMA, T. and HONDA, Y., 1998. Relationships between percent vegetation cover and vegetation indices. *International Journal of Remote Sensing*, **19**(18), pp. 3519-3535.

RAHIMZADEH-BAJGIRAN, P., BERG, A.A., CHAMPAGNE, C. and OMASA, K., 2013. Estimation of soil moisture using optical/thermal infrared remote sensing in the Canadian Prairies. *ISPRS journal of photogrammetry and remote sensing*, **83**, pp. 94-103.

RAHIMZADEH-BAJGIRAN, P., OMASA, K. and SHIMIZU, Y., 2012. Comparative evaluation of the Vegetation Dryness Index (VDI), the Temperature Vegetation Dryness Index (TVDI) and the improved TVDI (iTVDI) for water stress detection in semi-arid regions of Iran. *ISPRS Journal of Photogrammetry and Remote Sensing*, **68**, pp. 1-12.

RANDALL B. SMITH, 2013. *Hyperspectral Imaging.*

REICHLE, R.H., ARDIZZONE, J.V., KIM, G., LUCCHESI, R.A., SMITH, E.B. and WEISS, B.H., 2018. Soil Moisture Active Passive (SMAP) Mission Level 4 Surface and Root Zone Soil Moisture (L4_SM) Product Specification Document.

REN, C., LIANG, Y., LU, X. and YAN, H., 2019. Research on the soil moisture sliding estimation method using the LS-SVM based on multi-satellite fusion. *International Journal of Remote Sensing,* **40**(5-6), pp. 2104-2119.

RODRÍGUEZ-FERNÁNDEZ, N.J., DE SOUZA, V., KERR, Y.H., RICHAUME, P. and AL BITAR, A., 2017Soil moisture retrieval using SMOS brightness temperatures and a neural network trained on in situ measurements, *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* 2017, IEEE, pp. 1574-1577.

ROSS, A. and GOVINDARAJAN, R., 2005Feature level fusion using hand and face biometrics, *Proceedings of SPIE conference on biometric technology for human identification II* 2005, pp. 196-204.

ROY, D.P., WULDER, M.A., LOVELAND, T.R., WOODCOCK, C.E., ALLEN, R.G., ANDERSON, M.C., HELDER, D., IRONS, J.R., JOHNSON, D.M. and KENNEDY, R., 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment,* **145**, pp. 154-172.

SAHEBI, M.R. and ANGLES, J., 2010. An inversion method based on multi-angular approaches for estimating bare soil surface parameters from RADARSAT-1. *Hydrology and Earth System Sciences,* **14**(11), pp. 2355-2366.

SANDHOLT, I., RASMUSSEN, K. and ANDERSEN, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment,* **79**(2), pp. 213-224.

SANO, E.E., QI, J., HUETE, A.R. and MORAN, M.S., 1998The use of SAR/TM synergy for estimating soil moisture content over a semi-arid rangeland, *Proc. 2nd Latino-American Seminar on Radar Remote Sensing INPE/ESA* 1998, Citeseer.

SCHOWENGERDT, R.A., 2007. Remote sens: models and methods for image processing.

SEMMENS, K.A., ANDERSON, M.C., KUSTAS, W.P., GAO, F., ALFIERI, J.G., MCKEE, L., PRUEGER, J.H., HAIN, C.R., CAMMALLERI, C. and YANG, Y., 2016. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. *Remote Sensing of Environment*, **185**, pp. 155-170.

SENEVIRATNE, S.I., CORTI, T., DAVIN, E.L., HIRSCHI, M., JAEGER, E.B., LEHNER, I., ORLOWSKY, B. and TEULING, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews,* **99**(3–4), pp. 125-161.

SHA, HU, WANG and LI, 2018Preliminary study on the applicability of several remote sensing drought indices to agricultural drought monitoring in Gansu province of China, *Land Surface and Cryosphere Remote Sensing IV* 2018, International Society for Optics and Photonics, pp. 107770H.

SHAHABFAR, A. and EITZINGER, J., 2011. Agricultural drought monitoring in semi-arid and arid areas using MODIS data. *The Journal of Agricultural Science,* **149**(4), pp. 403-414.

SHAO, H., CHU, L., JALEEL, C.A. and ZHAO, C., 2008. Water-deficit stress-induced anatomical changes in higher plants. *Comptes rendus biologies,* **331**(3), pp. 215-225.

SHIPPERT, P., 2003. Introduction to hyperspectral image analysis. *Online Journal of Space Communication,* **3**.

SIKDAR, S.K., GLAVIC, P. and JAIN, R., 2004. *Technological choices for sustainability.* Springer Science & Business Media.

SOMERS, B., GYSELS, V., VERSTRAETEN, W.W., DELALIEUX, S. and COPPIN, P., 2010. Modelling moisture-induced soil reflectance changes in cultivated sandy soils: a case study in citrus orchards. *European Journal of Soil Science,* **61**(6), pp. 1091-1105.

SONG, K., ZHOU, X. and FAN, Y., 2009. Empirically adopted IEM for retrieval of soil moisture from radar backscattering coefficients. *IEEE Transactions on Geoscience and Remote Sensing,* **47**(6), pp. 1662-1672.

STATHAKI, T., 2011. *Image fusion: algorithms and applications.* Academic Press.

STEP ESA, 2015-last update, Sentinel Application Platform. Available: https://step.esa.int/main/toolboxes/snap/ [13/02/, 2020].

THOMA, D.P., MORAN, M.S., BRYANT, R., RAHMAN, M.M., COLLINS, C.H., KEEFER, T.O., NORIEGA, R., OSMAN, I., SKRIVIN, S.M. and TISCHLER, M.A., 2008. Appropriate scale of soil moisture retrieval from high resolution radar imagery for bare and minimally vegetated soils. *Remote Sensing of Environment,* **112**(2), pp. 403-414.

TRENBERTH, K.E., DAI, A., VAN DER SCHRIER, G., JONES, P.D., BARICHIVICH, J., BRIFFA, K.R. and SHEFFIELD, J., 2014. Global warming and changes in drought. *Nature Climate Change,* **4**(1), pp. 17-22.

ULABY, F.T., LONG, D.G., BLACKWELL, W.J., ELACHI, C., FUNG, A.K., RUF, C., SARABANDI, K., ZEBKER, H.A. and VAN ZYL, J., 2014. Miacrowave Radar and Radiometric Remote sensing. The Universisty of Michigan press. *Ann Arbor,* .

ULABY, F.T., 1982. Microwave remote sensing active and passive. *Rader remote sensing and surface scattering and emission theory,* , pp. 848-902.

USER MANUAL OF, BOSCH PLR 15, , BOSCH PLR 15 Laser rangefinder. Available: https://www.tooled-up.com/artwork/ProdPDF/0603672000.pdf [22/10/, 2019].

VAN DER SCHALIE, R., DE JEU, R., PARINUSSA, R., RODRIGUEZ-FERNANDEZ, N., KERR, Y., AL-YAARI, A., WIGNERON, J. and DRUSCH, M., 2018. The effect of three different data fusion approaches on the quality of soil moisture retrievals from multiple passive microwave sensors. *Remote Sensing,* **10**(1), pp. 107.

VERHOEST and LIEVENS, 2013. Soil Moisture Retrieval from Synthetic Aperture Radar: Facing the Soil Roughness Parameterization Problem. *Remote Sensing of Energy Fluxes and Soil Moisture Content,* , pp. 323.

VERHOEST, LIEVENS, WAGNER, LVAREZ-MOZOS, MORAN and MATTIA, 2008. On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar. *Sensors,* **8**(7), pp. 4213-4248.

VERMOTE, E.F., TANRÉ, D., DEUZE, J.L., HERMAN, M. and MORCETTE, J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing,* **35**(3), pp. 675-686.

WANG, C., QI, S., NIU, Z. and WANG, J., 2004. Evaluating soil moisture status in China using the temperature–vegetation dryness index (TVDI). *Canadian Journal of Remote Sensing,* **30**(5), pp. 671-679.

WANG, C., CHEN, J., CHEN, X. and CHEN, J., 2019. Identification of Concealed Faults in a Grassland Area in Inner Mongolia, China, Using the Temperature Vegetation Dryness Index. *Journal of Earth Science,* **30**(4), pp. 853-860.

WANG, J.R. and CHOUDHURY, J., 1995. Passive microwave radiation from soil: examples of. *Passive Microwave Remote Sensing of Land--Atmosphere Interactions,* , pp. 423.

WANG, L. and QU, J.J., 2007. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters,* **34**(20),.

WANG, P., LI, X., GONG, J. and SONG, C., 2001Vegetation temperature condition index and its application for drought monitoring, *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)* 2001, IEEE, pp. 141-143.

WEIYING, C., QIANGUANG, X. and YONGWEI, S., 1994. Application of the anomaly vegetation index to monitoring heavy drought in 1992. *Remote Sensing of Environment,* **9**(2), pp. 106-112.

WILLMOTT, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society,* **63**(11), pp. 1309-1313.

XU, H., YUAN, Q., LI, T., SHEN, H. and ZHANG, L., 2019Estimating Surface Soil Moisture from Satellite Observations Using Machine Learning Trained on In Situ Measurements in the Continental US, *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* 2019, IEEE, pp. 6166-6169.

YAHIA, O., GUIDA, R. and IERVOLINO, P., 2018aSentinel-1 and Landsat-8 feature level fusion for soil moisture content estimation, *Proceedings of the 12th European Conference on Synthetic Aperture Radar (EUSAR 2018)* 2018a, EUSAR.

YAHIA, O., GUIDA, R. and IERVOLINO, P., 2018bWeights Based Decision Level Data Fusion of Landsat-8 and Sentinel-L for Soil Moisture Content Estimation, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* 2018b, IEEE, pp. 8078-8081.

YANG, X., WU, J.J., SHI, P.J. and YAN, F., 2008. Modified triangle method to estimate soil moisture status with MODerate resolution Imaging Spectroradiometer (MODIS) products. *Proc.Int.Arch.Photogramm.Remote Sens.Spat.Inf.Sci*, **37**, pp. 555-560.

YANG, Y., GUAN, H., LONG, D., LIU, B., QIN, G., QIN, J. and BATELAAN, O., 2015. Estimation of surface soil moisture from thermal infrared remote sensing using an improved trapezoid method. *Remote Sensing,* **7**(7), pp. 8250-8270.

YANG, GUAN, LONG, LIU, QIN and BATELAAN, 2015. Estimation of surface soil moisture from thermal infrared remote sensing using an improved trapezoid method. *Remote Sensing*, **7**(7), pp. 8250-8270.

YUAN, Q., XU, H., LI, T., SHEN, H. and ZHANG, L., 2020. Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental US. *Journal of Hydrology,* **580**, pp. 124351.

ZAMAN, B., MCKEE, M. and NEALE, C.M., 2012. Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines. *International Journal of Remote Sensing,* **33**(20), pp. 6516-6552.

ZENG, Y., ZHANG, J. and VAN GENDEREN, J.L., 2006Comparison and analysis of remote sensing data fusion techniques at feature and decision levels, *ISPRS Commission VII Mid-term Symposium" Remote Sensing: From Pixels to Processes* 2006.

ZHANG, D. and ZHOU, G., 2016. Estimation of soil moisture from optical and thermal remote sensing: A review. *Sensors,* **16**(8), pp. 1308.

ZHANG, J., ZHOU, Z., YAO, F., YANG, L. and HAO, C., 2014. Validating the modified perpendicular drought index in the North China region using in situ soil moisture measurement. *IEEE Geoscience and Remote Sensing Letters,* **12**(3), pp. 542-546.

ZHANG, X., ZHAO, H. and YANG, J., 2019. Spatial downscaling of land surface temperature in combination with TVDI and elevation. *International Journal of Remote Sensing,* **40**(5-6), pp. 1875-1886.

ZHU, W., JIA, S. and LV, A., 2017. A time domain solution of the Modified Temperature Vegetation Dryness Index (MTVDI) for continuous soil moisture monitoring. *Remote Sensing of Environment,* **200**, pp. 1-17.

ZOTARELLI, L., DUKES, M.D. and MORGAN, K.T., 2010. Interpretation of soil moisture content to determine soil field capacity and avoid over-irrigating sandy soils using soil moisture sensors. *The Agricultural and Biological Engineering Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida., AE460,* .

ZRIBI, M. and DECHAMBRE, M., 2003. A new empirical model to retrieve soil moisture and roughness from C-band radar data. *Remote Sensing of Environment,* **84**(1), pp. 42-52.

ZRIBI, M., MUDDU, S., BOUSBIH, S., AL BITAR, A., TOMER, S.K., BAGHDADI, N. and BANDYOPADHYAY, S., 2019. Analysis of L-Band SAR Data for Soil Moisture Estimations over Agricultural Areas in the Tropics. *Remote Sensing,* **11**(9), pp. 1122.