## ASSESSING THE TRANSFERABILITY OF RANDOM FOREST AND TIME-WEIGHTED DYNAMIC TIME WARPING FOR AGRICULTURE MAPPING

MWANAIDI MOHAMEDI DADI March 2019

SUPERVISORS: dr. M.Belgiu dr. M.Marshall

# **ASSESSING THE TRANSFERABILITY OF RANDOM** FOREST AND TIME-WEIGHTED DYNAMIC TIME WARPING FOR **AGRICULTURE MAPPING**

MWANAIDI MOHAMEDI DADI Enschede, The Netherlands, March 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: [Name course (e.g. Applied Earth Sciences)]

SUPERVISORS: dr. M.Belgiu dr. M.Marshall

THESIS ASSESSMENT BOARD: prof.dr.ir. A. Stein dr. D. Arvor (External Examiner, Université Rennes 2, France)



#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

### ABSTRACT

Automated methodologies for mapping cropland over large geographical areas enable frequent and updated agriculture statistics within a short period. Unavailability of such automated methodologies makes cropland mapping a tedious and time-consuming task. Therefore, most of the existing solutions dedicated to cropland mapping are mainly focused on a small geographic area. The aim of this research is to assess the transferability of Random Forest (RF) and Time-Weighted Dynamic Time Warping (TWDTW) across different years in the same study area and across different agroecosystems for mapping crop type using Sentinel-2 (S2) time series. RF and TWDTW classification models were tested in two different agroecosystems, namely Mississippi Alluvial Plain and Central California Valley. The classification models trained in one agroecosystem were applied across years and different agroecosystems to map similar crops. Furthermore, the performance of the two classification models in a selected study area was evaluated. The classification output indicated that RF performed better than TWDTW in all study areas with an overall accuracy of more than 92% except for the neighboring agroecosystems in California where the overall accuracy was 87.59%. The transferability results concluded that RF and TWDTW classification models are transferable with accuracy more than 84% across years and agroecosystems having the same growing conditions. However, unbalanced class proportions in two test areas under investigation may restrict their transferability. The classification models are not transferable across different agroecosystems having different growing conditions and dissimilar agricultural management practices. In light of these results, automated methodologies for mapping cropland remain essential to ensure significant contribution to the food security world.

Keywords; Satellite image time series, RF, TWDTW, transferability, red-edge bands, agroecosystems

## ACKNOWLEDGEMENTS

First, I thank the Almighty God for his protection, guidance, and care throughout my stay and study in the Netherlands.

I would like to express my sincere deepest gratitude to my supervisors Dr. Mariana Belgiu and Dr. Michael Marshall for their constant support, guidance, constructive criticism and encouragement throughout the research period. I couldn't have wished for friendlier supervisors than you. Thank you so much.

I am also extending my gratitude to the Government of the Netherlands for granting me NFP Scholarship to pursue my M.Sc. course at ITC.

Thanks to GFM staff members for their lectures, lessons and extra relevant materials they provide which helped me to absorb knowledge in the best ways.

Many thanks to my GFM classmates who I interacted with on a daily basis and new friends I made from other courses for making my stay interesting and enjoyable.

Special thanks go to my beloved husband for his patience and encouragement throughout the course period. Last but not least I would like to express my heartfelt thanks to my parents, brothers, and sisters for their moral support.

## TABLE OF CONTENTS

1.	INTF	ODUCTION	1
	1.1.	Motivation and problem statement	1
	1.2.	Research objectives	2
2.	LITE	RATURE REVIEW	4
	2.1.	Transferability of supervised clasifiers in remote sensing	4
	2.2.	Satellite image time series (SITS) for agriculture mapping using RF and TWDTW	4
	2.3.	Sentinel-2 Vegetation Indices (VI)	5
3.	STUI	DY AREAS, DATA PRE-PROCESSING AND SOFTWARE	7
	3.1.	Study areas	7
	3.2.	Reference data	8
	3.3.	Sentinel-2 (S2) data	10
	3.4.	Computation of vegetation indices	11
	3.5.	Generating training and validation samples	12
	3.6.	Software used	13
4.	MET	HODOLOGY	14
	4.1.	Adopted methodology	14
	4.2.	Processing	15
5.	RESU	JLTS	21
	5.1.	Random Forest Classification results	21
	5.2.	Time-Weighted Dynamic Time Warping classfication results	24
	5.3.	Producer's (PA) and user's accuracies (UA) of RF and TWDTW	25
	5.4.	Post classification analysis	27
	5.5.	Assessment of the interclass similarity and intra-class variability	29
	5.6.	Transferability assessment of the RF and TWDTW classification models	29
6.	DISC	USSION	35
	6.1.	Phenological cycles of crops	35
	6.2.	Potential of time series Sentinel 2 data for mapping crop type	35
	6.3.	RF and TWDTW classification models	36
	6.4.	RF and TWDTW model transferability across different agroecosystems	36
	6.5.	RF and TWDTW model transferability across same agroecosystems	37
	6.6.	RF and TWDTW model transferability across years	37
7.	CON	CLUSION AND RECOMMENDATIONS	38
	7.1.	Conclusion	38
	7.2.	Recommendations and future works	39

## LIST OF FIGURES

Figure 3-1. Study areas in Central California Valley and Mississippi Alluvial Plains
Figure 3-2. Reference maps of California and Mississippi (https://nassgeodata.gmu.edu/CropScape/)10
Figure 4-1. Flowchart of the proposed methods
Figure 4-2. Logistic weight with steepness $\alpha = -0.1$ and midpoint $\beta = 100$
Figure 5-1. Classified images of California in 2017 (a), (c), and (e), Mississippi 2017 (b), California 2016 (d),
and neighboring area within California 2017 (f)
Figure 5-2. The variable importance measures and rankings using mean decreases in accuracy (MDA) of
California 2017 (a), Mississippi 2017 (b), California 2016 (c) and an area within California 2017(d)23
Figure 5-3. Classified maps of California in 2017 (a), (c), and (e), California 2016 (d), Mississippi 2017 (b)
and another study area within California 2017 (f)25
Figure 5-4. Producer's and user's accuracy obtained by RF and TWDTW for California and Mississippi
study areas in 2017
Figure 5-5. Producer's and user's accuracy obtained by RF and TWDTW for California in 2016 and 2017
Figure 5-6. Producer's and user's accuracy obtained by RF and TWDTW for California and its neighbor
region in 2017
Figure 5-11. Area uncertainties of the mapped and adjusted areas extracted from error matrix (at 95%
confidence interval)
Figure 5-7. Interclass similarities of cotton and tomatoes, and intra-class variabilities of others class29
Figure 5-8. Transferability of RF and TWDTW classification models across different agroecosystem31
Figure 5-9. Transferability of RF and TWDTW classification models across years in the same
agroecosystem
Figure 5-10. Transferability of RF and TWDTW classification models across the same agroecosystem34

## LIST OF TABLES

Table 3-1. Description of the study areas	7
Table 3-2. CDL accuracy assessment of some of the crops in 2016 and 2017 in Mississippi and Califo	rnia
(CropScape and Cropland Data Layers, 2018)	9
Table 3-3. Wavelength range of each S2 red-edge band (ESA, 2019)	10
Table 3-4. Central California Valley 2016 and 2017 image information and Mississippi Alluvial Plain	11
Table 3-5. The formulas used to compute spectral indices of NDVIre	12
Table 3-6. Training and validation samples for California 2016 and 2017	12
Table 3-7. Training and validation samples for California and Mississippi 2017	13
Table 5-1. Overall accuracy and Kappa obtained by RF in all each study areas	22
Table 5-2. Overall accuracy and Kappa index obtained by TWDTW in all investigated study areas	24

## 1. INTRODUCTION

#### 1.1. Motivation and problem statement

The United Nations (UN) established 17 interrelated Sustainable Development Goals (SDG) each with its unique objectives. No poverty and zero hunger are two of the goals with the intention to raise the poverty line as high to \$5 per day and to end hunger and all sorts of malnutrition by 2030 (Hickel, 2015). One way to fulfil these two goals is to utilize remote sensing data and techniques in providing valuable information in the agricultural sector (Kumar, 2015). For example, remote sensing provides platforms like airborne and satellites for acquiring data in a large area within a short period which is helpful in facilitating agriculture monitoring. Also, remote sensing provides techniques such as machine learning for processing and analysing large volumes of data. Machine learning algorithms are proved to be very fast and accurate in various studies conducted in the agriculture domain (Duro et al., 2012, Li et al., 2016, Watts et al., 2008, Xiaomu et al., 2005).

Machine learning classifiers such as RF (Breiman, 2001), and TWDTW (Victor Maus et al, 2016), are efficient supervised classifiers in the agricultural domain. RF predicts through constructing many decision trees and the output is based on the majority vote of the classification output (Zafari, Zurita-Milla, & Izquierdo-Verdiguier, 2017). TWDTW is an adaptation of Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978) used to classify land use and land cover classification from time series data. DTW is the technique useful in finding the best alignment between two dependent time series and in providing similarity measures between them (Victor Maus et al., 2015). RF, DTW and TWDTW classifiers are detailed explained under section 4.2.1 and 4.2.4 of this research.

RF and TWDTW classifiers have recently gained popularity in crop type mapping using time series approach due to their ability of handling the main challenges encountered when mapping crop type based on time series analysis (Maus et al., 2017, Maus et al., 2016, Belgiu & Drăguţ, 2016, Lebourgeois et al., 2017). These challenges include insufficient samples for training the supervised classifiers, missing time series data due to cloud coverage and changes of phenological cycles annually due to weather or variability of agricultural practices (Petitjean, Inglada, & Gançarski, 2012). These classifiers achieved promising results in mapping agriculture using satellite image time series (SITS), but they are typically calibrated and validated for a single area. Their transferability across geographical regions and years have been rarely studied. There are few studies reported on the transferability assessment of RF. For example, Jin et al. (2018) investigated the transferability of RF in canopy height estimation from multi-source remote sensing data. Wenger & Olden. (2012) assessed the transferability of ecological models using RF and Artificial Neural Networks (ANN) algorithms, and Juel et al. (2015) tested the transferability of RF classification models to map vegetation structures from aerial images and Digital Elevation Model (DEM). All authors

concluded that RF classification models cannot be transferable from one location to another. So far, no studies have been reported on assessing the transferability of TWDTW.

Most of the remote sensing classifiers such as RF and TWDTW are proved to achieve excellent results when exploited in a small area (local scale) (Belgiu & Csillik, 2018, Müller et al., 2015, Torbick et al., 2017, Victor Maus et al., 2015). However, transferability of these classifiers to regional and national scale was not thoroughly investigated. This research aims to test the transferability of RF and TWDTW machine learning across different agroecosystems and years.

Transferability of classification-based models is an important issue to be addressed since training these supervised classifiers is a time-consuming and expensive task. Therefore, reusing the classification model trained on a specific study area to another study area reduces the computational burden and could facilitate operationalization of SITS classification over large areas. In addition, a model trained in one area can be applied in a new test area for which survey data have not been obtained as long as the two areas have similar contexts (Sanko & Morikawa, 2010). Therefore, this research will focus on assessing the transferability of the RF and TWDTW supervised classifiers across different agroecosystems and years by utilizing various spectral indices computed from Sentinel-2 SITS data (Fernández-Manso, Fernández-Manso, & Quintano, 2016). The developed automated solution to classify crops in this research might contribute to the realization of the SDGs, especially in food security.

S2 satellite images provide multi-spectral data with 13 spectral bands at different spatial resolutions in the visible, near infrared (NIR) and short-wave infrared (SWIR) of the electromagnetic spectrum. Visible and NIR bands are at 10m resolution, the red-edge, narrow NIR, and SWIR at 20m resolution and at 60m resolution is represented by the coastal aerosols, water vapor and SWIR (cirrus). In the red-edge region, there are three bands that other multispectral sensors such as Landsat-8 lacked but which have been important to agricultural monitoring. It is presumed that the red-edge bands have great potential for mapping diverse vegetation characteristics (Sibanda, Mutanga, & Rouget, 2015). For example, Qiu et al. (2017) discovered that classification accuracy of various land cover types including agriculture areas, was improved after adding red-edge bands in the image classification procedures. Forkuor et al. (2018) found that crop classification of S2 red-edge bands achieved better results compared to other S2 bands and Landsat-8 bands. Consequently, these bands could be helpful in assessing the transferability of RF and TWDTW classification models across different agroecosystems.

#### 1.2. Research objectives

The overall aim of this research is to assess to what extent the TWDTW and RF-based classification models are transferable across different years in the same study area and across different agroecosystems for agriculture mapping.

The specific objectives and their respective research questions are;

1. To evaluate the relevance of various S2 based vegetation indices (VI) for mapping crops in various agroecosystems.

**Research question:** What are the most important S-2 based vegetation indices used to map target crops in the selected study areas?

 To develop the TWDTW and RF classification models for mapping different crops based on VI identified in specific objective 1.

**Research question:** Which of the developed classification models performs better in the selected study areas?

**Research question:** Is there any difference in the performance of the two evaluated classifiers in case of high intra-class variability and/or interclass similarity?

3. To test the transferability of the developed classification models across different years and different agroecosystems with no calibration.

**Research question:** How transferable are the developed classification models across different years and different agroecosystems without calibration?

4. To investigate the uncertainties of the obtained classification results.

## 2. LITERATURE REVIEW

This chapter provides the theoretical background of this research. Section 2.1 explains remote sensingbased model transferability, Normalized Difference Vegetation Index (NDVI) calculated using red-edge bands is described under section 2.2, section 2.3 highlights the agricultural mapping from SITS.

#### 2.1. Transferability of supervised clasifiers in remote sensing

Transferability is defined as "the fitness of the transferred model, information, or theory in the new context" (Wilmot, 1982). Model transferability is the process of applying the trained model in a context other than the one which was originally estimated (Koppelman, 1986). The trained model can be transferred across different geographical regions or in the same geographical regions at different periods. Werkowska et al. (2017) explained that transferring the trained model to other periods of time is known as temporal transferability. In contrast, spatial transferability refers to the reusing of the trained model in different geographical locations. It reduces the cost of data gathering and computation burden (Sanko & Morikawa, 2010).

Juel et al. (2015) investigated the transferability of RF classification models for mapping vegetation structures using aerial images and DEM. The authors concluded that the RF classification model was not transferable to other study areas. Jin et al. (2018) tested the transferability of RF from multi-source remote sensing data using LIDAR-derived canopy height as a reference to train the RF model. The authors concluded that the RF classification model trained in one location or built for a specific vegetation type cannot be transferred to another location or other vegetation types. Wenger & Olden. (2012) tested the transferability of ecological models developed using RF and ANN into two study areas in the Western United States. This study revealed that RF and ANN algorithms provided outstanding performance when applied within one location, but they achieved poor transferability in new locations. Wang et al. (2019) developed and examined automatic methods for generating crop type mapping without field-level data by transferring trained RF model across geographic locations and year and utilizing unsupervised clustering procedures. The authors discovered that harmonic coefficients can successfully separate crop types when used in either supervised or unsupervised techniques. This study revealed that RF are transferable with high accuracy across geographical regions where the growing conditions and regional crop composition are similar.

#### 2.2. Satellite image time series (SITS) for agriculture mapping using RF and TWDTW

Recently launched S2 satellite missions by the European Space Agency (ESA) in 2015 and other global Landsat-class missions provides time series of moderate spatial resolution imagery. This time series offers high possibilities to conduct near real-time monitoring and enhance our understanding of complex earth

structure (Zhu, Diao, & Deng, 2018). S2 time series provides a wide range of land cover applications (Julea, Lasserre, Jolivet, Ledo, & Lodge, 2011). Its characteristics offer the opportunity to study the seasonal evolution of crops.

There are a number of studies reported using different algorithms for mapping crop type using time series or single-data satellite image (Ozdarici-Ok., 2015, Petitjean et al., 2012, Rhemtulla and Kalacska, 2012). Methods for mapping crop type from time series images have been confirmed to perform better than the single date methods (Gómez, White, & Wulder, 2016). For example, Simonneaux et al. (2010) successfully classified bare soil, annual crops, trees in the middle of an agricultural field, and trees located on bare soil. Furthermore, the authors estimated evapotranspiration over irrigation fed area in Morocco using high-resolution Landsat TM images. Kontgis et al. (2017) succeeded to illustrate the unique phenology of paddy rice for generating a map of paddy rice extent and classifying them into single, double or triple cropped during the growing season.

Also, there are a number of studies reported on mapping agriculture from time series using the RF algorithm. Belgiu & Csillik. (2018) examined the performance of RF and TWDTW from S2 time series when applied in object-based (OB) and pixel-based (PB) classification of diverse types of crops in Romania, Italia, and the USA. The results show OB-TWDTW performed better than PB-TWDTW and RF achieved high accuracy in the USA than other areas. Müller et al. (2015) investigated the spectral-temporal classification approach for separating between cropland, pasture, natural savanna vegetation, and other relevant land cover classes using Landsat time series. The RF classifier was used for this classification study. The overall classification result achieved was 93%. Pelletier et al. (2016) examined the robustness of RF classifier compared to Support Vector Machine in mapping land cover using high-resolution time series, namely Landsat-8 and SPOT-4. RF yielded the best results compared to SVM in terms of classification performance and computation times.

There are also studies dedicated to mapping agriculture from time series using TWDTW. For example, Maus et al. (2015) improved the proposed TWDTW version of DTW for land use and land cover classification from time series data. The algorithm proved to be flexible in mapping different crop types providing overall good results for land cover classification. High accuracy was achieved when mapping the classes of pasture in a tropical forest area, single cropping, double cropping, and forest. Belgiu & Csillik, (2017) applied OB-TWDTW for mapping rice, sunflower, maize and wheat from S2 time series. The method achieved high accuracies, namely an overall accuracy of 93.43%, and kappa 92%.

#### 2.3. Sentinel-2 Vegetation Indices (VI)

The availability of multispectral data provided by S2 with 13 bands in the visible, near infrared (NIR) and Short-wave infrared (SWIR) allows the calculation of new spectral vegetation indices that proved to be useful to trace vegetation dynamics. For example, NDVI red edges (1,2, and 3) and that of NDVI red-

edges (1,2, and 3) narrow computed from (NIR B8 vs bands 5,6,7) and (NIR B8A vs band 5,6,7) respectively were introduced by Fernández-Manso et al. (2016). The red-edge spectral region is sensitive to chlorophyll contents of the plants (Watson & Lee., 2005). The spectral range in the red-edge region of the electromagnetic spectrum lies between 675nm to 760nm.

There are few studies reported on utilizing red-edge to improve the performance of the RF algorithm. Yet no study has been reported for TWDTW. Forkuor et al. (2018), for example, examined the added value of the red-edge bands of S2 for mapping land use and land cover in rural Burkina Faso using RF and SVM machine learning classifiers. In all five experiments conducted, the RF classification results showed that red-edge bands of S2 outperformed Landsat8 and other S2 bands. Qiu et al. (2017) investigated the capability of S2 vegetation red-edge spectral bands for improving land cover classification using RF classifier. The authors found out that the red-edge spectral bands improved the overall accuracy of the target land cover classes.

There are also studies dedicated to using supervised and unsupervised algorithms for crop type mapping using red-edge spectral index. For example, Schuster et al. (2012) tested the red-edge spectral bands for improving land use classification based on RapidEye high-resolution multispectral satellite data. The results of this study showed that the incorporation of red-edge information can increase classification accuracy. Also, red-edge measurements were found useful for assessing vegetation chlorophyll and Leaf Area Index and are suitable for early stress detection (Filella & Penuelas, 1994). Fernández-Manso et al. (2016) used the red-edge index for discriminating burn severity. Delegido et al. (2011) utilized red-edge for estimation of Leaf Area Index and chlorophyll content.

# 3. STUDY AREAS, DATA PRE-PROCESSING AND SOFTWARE

This chapter describes the study areas, the dataset used and its pre-processing, reference data and software used for data processing.

#### 3.1. Study areas

The research was conducted in two study areas in the United States, California Great Plains, and the Mississippi Alluvial Plains. The agricultural fields under investigation in California are located approximately between 36°25'49.98" to 36°21'01.71"N and 120°06'10.80" to 120°01'23.41"W. The area is very hot and dry in summer periods but cool and damp in winter seasons. During the summer, the temperature at daytime reaches 38°C and common heat waves raise the temperature to 46°C. There is a rainy season in mid-autumn to mid-spring (Western Regional Climate Center 2009). Annual rainfall receives in the area is 292.1mm. The agricultural activities in the area depend on irrigation from surface water diversions and pumped water from the ground. The major crops cultivated in the area are cotton, tomatoes, grapes, almonds, winter wheat, and alfalfa.

Mississippi is one of the largest agriculture areas in the United States covering an area larger than 4 million acres. The agroecosystem in this region is very productive under proper management. Agriculture has been the major contributor to the Mississippi economy and the whole nation (Watson & Lee, 2005). The agricultural fields focused in this research are located approximately between 34° 18' 51.83" to 34° 09' 37.38"N and 90° 32' 12.86" to 90° 21' 57.26"W. The area receives annual rainfall from approximately 1143mm to 1524mm in the northern and southern part, respectively (Watson & Lee, 2005). Average temperature per year is 16.63°C. Snowfall is around 3.3cm per year (National Weather Service, n.d.). There is sufficient water from perennial streams and lakes which serve as water resources for irrigation (Watson & Lee, 2005). The area receives plentiful rainfall but it occurs during the months when the major crops are not produced (Watson & Lee, 2005). Given the rainfall distribution in the study areas irrigation is required to prevent crop loss during drought. The major crops cultivated are cotton, small grain, forage, maize, soybean, and rice.

Table 3:1 below shows the description of the study areas. The information was extracted from (Western Region Climate Center, 2009, National Weather Service, n.d., Watson & Lee, 2005).

Study area	California	Mississippi
Lat/Long (between)	• 36° 25' 49.98" to 36° 21' 01.71"N	• 34° 18' 51.83" to 34° 09' 37.38"N
	and	and

Table 3-1. Description of the study areas

	• 120° 06' 10.80" to 120° 01' 23.41"W	• 90 <sup>0</sup> 32' 12.86" to 90 <sup>0</sup> 21' 57.26"W
Annual Temperature	• Min 10.72 °C	• Min 11 °C
(°C)	• Max 24.94 °C	• Max 22.28 °C
	• Average 17.83 °C	• Average 16.64 °C
Annual Rainfall (inch)	• 292 mm mid-autumn to mid-spring	• From 1143mm to 1524mm
	• Very hot and dry	
Crops cultivated	• cotton, tomatoes, grapes, almonds, winter wheat, double cropping, pistachios, and alfalfa	• cotton, maize, soybean, and rice
Irrigation	• surface water diversions and pumped water from the ground	<ul><li>perennial streams and lakes</li><li>Rainfed</li></ul>



Figure 3-1. Study areas in Central California Valley and Mississippi Alluvial Plains

#### 3.2. Reference data

The ground truth data were downloaded from the Cropscape website provided by the United States Department of Agriculture (USDA). The Cropland Data Layer (CDL) of the USDA are raster-based, crop specific and already geo-referenced layers (Guyet & Nicolas, 2016). These reference layers have a ground resolution of 30 meters. The CDL layer is created from Landsat 8 OLI/TIRS satellite imagery sensor. The

reference maps for each area of interest are shown in figure 3:2 below. The CDL accuracy assessment produced for tilled crops (fallow/idle cropland), principal crops (cotton, corn, and winter wheat), orchards (almonds) and vegetables (onions, dry beans, and tomatoes) in California and Mississippi are presented in summary in table 3:4 below.

In this research, the resolution of S2 images was 10m which is different from these reference images. The differences in resolution might affect the results if the validation and training samples are situated towards the edges of the agriculture parcels. According to the CDL accuracy assessment for agriculture field edges, the accuracy indicates that most of the crops situated towards the edges of agriculture fields have lower classification accuracy. To avoid that all training and validation samples were located towards the middle of the agricultural fields.

Table 3-2. CDL accuracy assessment of some of the crops in 2016 and 2017 in Mississippi and California (CropScape and Cropland Data Layers, 2018)

	California 2016		California 2017		Mississippi 2017	
	Overall	Kappa	Overall	Kappa	Overall	Kappa
Principal crops	86.7%	0.844	74.7%	0.677	85.8%	0.789
Tilled crops	86.4%	0.848	73.5%	0.683	85.5%	0.787
vegetables	78.0%	0.679	36.5%	0.313	55.4%	0.462
Orchards	87.0%	0.817	96.4%	0.834	58.7%	0.450



![](_page_18_Figure_1.jpeg)

Figure 3-2. Reference maps of California and Mississippi (https://nassgeodata.gmu.edu/CropScape/)

#### 3.3. Sentinel-2 (S2) data

The Level-1C S2 satellite images were downloaded from the United States of Geological Survey (USGS) on Earth Explorer, at 20m spatial resolution. Twelve images of California (six for 2016 and six for 2017) and 6 images of Mississippi were used to fulfil the requirement for this research. The atmospheric correction of Level-1C S2 raw images was performed to remove the scattering and absorption effects of the atmosphere. The Sen2cor plugins incorporated in SNAP were used to convert the Top-Of-Atmosphere (TOA) Level-1C S2 to surface reflectance (Level-2A). Meanwhile, all spectral bands of level-2A products were resampled to 10m resolution using the Nearest neighbor method. Only the red-edge, NIR and narrow NIR bands were selected for further analysis from all 13 spectral bands in each year. The range of wavelengths covered by each S2 red-edge, NIR and narrow NIR spectral bands which were selected for further analysis are shown in table 3:5 below.

Sentinel-2 bands	Central wavelength (nm)
Red-edge 1 (Band 5)	704.1
Red-edge 2 (Band 6)	740.5
Red-edge 3 (Band 7)	782.8
Near-infrared (Band 8)	832.8
Narrow Near-infrared (Band 8A)	864.7

Table 3-3. Wavelength range of each S2 red-edge band (ESA, 2019)

The spatial extent of the area focused for investigation for California in 2016 and 2017 was 16486.97 acres and 14944.81 acres for the neighboring region situated in California. The test area focused in Mississippi

was 16230.55 acres. The temporal resolution of all images is as shown in table 3:4 below. The images were taken starting from early May to late November to match the growing cycle of the crops of interest. The information on the standard planting and harvesting dates were retrieved from the Agricultural handbook of the Department of agriculture in the United States released in 2010.

The crops of interest were selected based on the requirement of transferability test which requires similarities in terms of crops between the two-test area under investigation. For example, in Mississippi and California, the common crop was cotton, whereas in California in 2016 and 2017 the crops of interest were cotton, tomatoes and winter wheat. The crops classified for the test areas located in California were similar to those found in California 2016 and 2017.

The agroecosystem regions in the United States are organized in hierarchical systems (i.e. domains, divisions, provinces and sections) and more detailed levels. In this research we used the lowest level, namely section which is more detailed compared to other levels.

California 2016		California 2017		Mississippi 2017		
Acquisition	Cloud cover in	Acquisition	Cloud cover	Acquisition	Cloud cover in	
date	%	date	in %	date	%	
22/05	0.37	17/05	25.33	01/05	0.1	
21/07	0.01	06/06	0	10/06	13.36	
10/08	0.04	16/07	31.2	20/07	6.78	
19/09	0	14/09	15.59	18/09	17.89	
09/10	5.19	24/10	0	28/10	7.74	
8/11	0	28/11	0	22/11	0	

Table 3-4. Central California Valley 2016 and 2017 image information and Mississippi Alluvial Plain

#### 3.4. Computation of vegetation indices

This research was focused on utilizing the red-edge indices. The Normalized Difference Vegetation Index red-edge (NDVIre) is similar to the common NDVI, however NDVIre uses the fraction of Near-Infrared (NIR) and the edge of the red region (Harris, 2018a).

To obtain the temporal phenology of the crops of interest six red-edge spectral indices NDVIre were computed. The NDVI red-edge 1 (NDVIre1) was introduced by Gitelson and Merzlyak (1994), the authors used the Near-Infrared (NIR) band 8 and the red-edge Band 5. Fernández-Manso & Quintano. (2016) introduced the NDVIre (2 & 3) and the NDVI red-edge narrow (NDVIren) using the NIR with red-edge (6 & 7) and the Narrow-NIR band 8A with red-edge (5,6,7) bands, respectively. The equations to compute the red-edge spectral indices are shown in table 4:2 below

Description	Equation	Reference
NDVI red-edge 1	$NDVI_B8B5 = \frac{NIR B8 - Red Edge B5}{NIR B8 + Red Edge B5}$	Gitelson and Merzlyak (1994)
NDVI red-edge 2	$NDVI_B8B6 = \frac{NIR B8 - Red Edge B6}{NIR B8 + Red Edge B6}$	Fernández-Manso et al. (2016)
NDVI red-edge 3	$NDVI_B8B7 = \frac{NIR B8 - Red Edge B7}{NIR B8 + Red Edge B7}$	Fernández-Manso et al. (2016)
NDVI red-edge 1 narrow	$NDVI_B8AB5 = \frac{NIR B8A - Red Edge B5}{NIR B8A + Red Edge B5}$	Fernández-Manso et al. (2016)
NDVI red-edge 2 narrow	$NDVI_B8AB6 = \frac{NIR B8A - Red Edge B6}{NIR B8A + Red Edge B6}$	Fernández-Manso et al. (2016)
NDVI red-edge 3 narrow	$NDVI_B8AB7 = \frac{NIR B8A - Red Edge B7}{NIR B8A + Red Edge B7}$	Fernández-Manso et al. (2016)

Table 3-5. The formulas used to compute spectral indices of NDVIre

#### 3.5. Generating training and validation samples

In all test areas CropScape – CDL was used for generating stratified random samples. Table 3:6 below show the number of samples generated for each class. For each crop of interest i.e. cotton, winter wheat, and tomatoes 50 samples were generated for training the classifiers and 80 samples for validation. The rest of the crops were grouped into one "others" class in each test area. For each class composing the "others" class, we generated 25 samples for training and 50 for validation. The validation samples were independently generated to minimize spatial autocorrelation with training samples (Millard & Richardson, 2015). Since only a few classes were available and the test areas were small, we checked visually to see the spatial distribution of the samples were not situated in the same parcels.

Table 3-6.	Training and	d validation	samples	for	California	2016	and 2	2017
1 4010 0 0.	i ranning and	a vancaation	Jumpies	TOT	Gamonna	2010	and 2	1011

Study area	California					
Year	Year 2016		2017		The Neighbor region within California 2017	
Class	Training	Validation	Training	Validation	Training	Validation
Winter	50	80	50	80	50	80
wheat						
tomatoes	50	80	50	80	50	80
cotton	50	80	50	80	50	80
Others	50	100	100	200	50	100

Year	2017		2017		
Class	Training	Validation	Training	Validation	
cotton	50	80	50	80	
Others	75	150	50	100	

Table 3-7. Training and validation samples for California and Mississippi 2017

#### 3.6. Software used

The third-party plugins 'Sen2Cor' incorporated with **Sentinel Application Platform** (SNAP) software v6.0.2 developed by the European Space Agency (ESA) was used to perform atmospheric corrections of Level-1C products to obtain Level-2A images.

QGIS software v3.2.0 was used in the computation of the Normalized Difference Vegetation Indices of the red-edge (NDVIre)

ArcGIS software v10.6.1 with a tool developed by Ken Buja in 2016, senior GIS Analyst for NOAA's National Centers for Coastal Ocean Science was used to generate training samples on the reference images.

Rstudio with *RandomForest* package was used for assessing the variable importance (VI) and developing the Random Forest model. TWDTW model was also developed in Rstudio using *dtwsat* package. RF and TWDTW model was implemented by adapting the codes developed by Millard et al., (2015) of Carleton University, Canada and Victor Maus et al., (2016), respectively.

## 4. METHODOLOGY

This chapter provides a detailed description of the methods used to meet the main objective of assessing the transferability of RF And TWDTW models across different agroecosystems and across different periods (but the same areas).

#### 4.1. Adopted methodology

The flowchart shows a general overview of the tasks and methods implemented in each step to meet the final outputs.

![](_page_22_Figure_5.jpeg)

#### 4.2. Processing

#### 4.2.1. Random Forest (RF) classification

RF uses a bagging approach whereby trees are generated by a randomly drawn subset of the training samples through replacement. Some samples may be selected many times whereas others may not be selected at all (Belgiu & Drăguț, 2016). RF splits the samples into training (2/3) called in-bag-samples and validation (1/3) referred to Out-Of-Bag (OOB). These validation samples are not used for building the trees (Breiman, 2001). They are used instead for internal cross-validation during the run to estimate how well the classification results performs by the model (Breiman, 2001).

Two parameters needed to bet set when implementing RF, the number of features to best split at the tree node (*Mty*) and the number of decision trees (*Ntree*). The research conducted by Ghosh et al. (2014) indicated that the classification accuracy is less sensitive to *Ntree* than to the *Mtry*. Belgiu & Drăguţ. (2016) suggested the acceptable number of trees is 500 when applying RF classifier on remotely sensed data. Other studies reported also the use of 500 number of trees (Lawrence et al., 2006; Galvan-Tejada et al., 2018) because the errors become stable before reaching this value. This is also the default value in "*randomForest*" R package. The *Mtry* parameter is normally set to the square root of the total number of input variables (Gislason, Benediktsson, & Sveinsson, 2006).

In this study, the *Mtry* was considered as the square root of the number of input variables (Gislason et al., 2006). When *Mtry* is set as equal to the total number of input variables the computation time increases because the algorithm needs to calculate the information gain contributed by all the variables used to split the nodes (Gislason et al., 2006). The number of decision trees, *Ntree* was set equal to 1000 throughout all the experiments conducted because the errors become stable before reaching this classification trees number (Lawrence et al., 2006). Therefore, it can be as large as possible since the RF classifier is computationally efficient (Guan et al., 2013).

The computation time required to develop the RF model is given by (Breiman, 2001)

#### $T\sqrt{MNlog(N)}$

Where; T is the number of decision trees

M is the number of variables for splitting in the node

N is the number of training samples

#### 4.2.2. Assessing variable importance

The variable importance (VI) plots indicate the importance of each variable in classifying the data (Dinsdale & Edwards, 2008.).

Equation (1)

The Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) are among the measures of VI available in *RandomForest* package in R. The variable selection is important because the cost of data storage and collection will decrease enabling computation speed to increase because of a reduced number of variables to process (Han, Guo, & Yu, 2017). In this research, MDA measure was used to assess the importance of variables in classifying the data.

The MDA uses the permuting OOB samples to calculate the importance of the variables (Breiman, 2001). As already mentioned in the previous section the OOB samples are not used to train the tree in RF but are used for self-assessment of the classifier internally. Therefore, MDA is determined during the computation of OOB error. The permutation of the VI in MDA is straightforward based on prediction accuracy instead of splitting criterion (Boulesteix, Janitza, Kruppa, & König, 2012). If the accuracy of the RF model decreases because of the exclusion of a certain variable, it means that those particular variables are very important in correctly classifying the data. In this case, the variables with large MDA values are the most important variable (Han et al., 2017). For the variables which are not relevant for a particular classification purpose, it is expected to have MDA=0 since permutation will not increase or decrease misclassification (Goldstein, Polley, & Briggs, 2011).

Most studies use MDA for VI measure because of its broad applicability and unbiasedness (Schmidt et al., 2014, Hapfelmeier et al., 2014). However, MDG can also be used when the predictive quality of the trees is low i.e. when OOB error is approximately 50% (Goldstein et al., 2011), given that MDA is calculated based on the increase of the misclassification rate after permuting a certain variable. MDA is calculated off of the OOB sample and, therefore, it can be viewed as the predictive quality of that variable (Goldstein et al., 2011).

RF classification under section 4.2.1 was implemented using iterative classification approach in order to assess the stability of predicted class extends and VI (Millard & Richardson, 2015). When running iterative classification approach variable importance measures and rankings varied with few iterations, with more iterations the VI rankings become stable

#### 4.2.3. Temporal patterns

A temporal pattern is a set of time series containing land cover information (Victor Maus et al., 2016). From the package *dtmSat* developed by Victor Maus et al. (2016), the temporal patterns were defined using the function dtwSat::createPatterns which utilize a Generalized Additive Model (GAM)(Wood, 2010) to create a smoothed temporal patterns based on the training samples. To obtain the temporal phenological patterns using the function mentioned above, one needs to set the sampling frequency of the output patterns in days in order to create a smoothed line and GAM smoothing formula (Victor Maus et al., 2016). In this research, the sampling frequency of the output patterns was set to 8 days in order to generate a smoothed line. The temporal patterns generated in this step were used as input in the TWDTW classification.

#### 4.2.4. Time-weighted Dynamic Time Warping (TWDTW) classification

Dynamic Time Warping (DTW) is useful in finding the best alignment between the two-time series and provide the similarity between them (Victor Maus et al., 2015). The sequences of the two-time series are matched to each other in a piece-wise fashion. The original DTW method was designed for speech recognition (Sakoe et al., 1978), and had been widely used for a number of applications such as human motion animation (Hsu et al., 2005), recognition of human activity (Kulkarni et al., 2014), and classification of time series data (Vasimalla, Challa, & Manohar Naik, 2016). However, DTW is challenging when applied to SITS analysis (Victor Maus et al., 2016). For example due to the variability of atmospheric conditions, such as clouds, the time series becomes non-uniform and of unequal length in terms of temporal sampling (Petitjean, Inglada, et al., 2012).

TWDTW is the method used in this study for land cover classification using satellite image time series. The method was implemented using *dtwSat* package developed by Victor Maus et al. (2016).

TWDTW method has the ability to classify crops with various vegetation dynamics. The TWDTW classifier can handle challenges arising due to irregular sampling and out-of-phase time series. It is sensitive to seasonal changes in natural and cultivated crops.

There are two in-built functions provided by the package *dtwSat*, namely the linearWeight and logisticWeight. Victor Maus et al. (2016) found that logistic-weight provides better results than the linear-weight for land cover classification. This is because logistic TWDTW has a low penalty for small-time warps and significant costs for bigger time warps compared to linear-weight which has a large cost for small time differences (Victor Maus et al., 2016). By using the function twdtwApply users can define different weight function (Victor Maus et al., 2016). In Maus et al. (2016) the authors used logistic-weights with two parameters  $\alpha$  and  $\beta$  which represents the steepness and the midpoint respectively. Belgiu & Csillik. (2018) used different values of  $\alpha$  (-0.2, -0.1, 0.1 or 0.2) and  $\beta$  (50 or100), the best classification results were acquired with  $\alpha = -0.1$  and  $\beta = 50$  meaning the authors applied a low penalty for time warps smaller than 50 days and a higher penalty for larger time weights. Figure 1:1 below is an example of the logistic-weight function with steepness  $\alpha = -0.1$  and midpoint  $\beta = 100$  developed by Victor Maus et al. (2016). The x and y-axis represent the absolute difference between two dates and time weight respectively.

![](_page_26_Figure_1.jpeg)

Figure 4-2. Logistic weight with steepness  $\alpha = -0.1$  and midpoint  $\beta = 100$ 

The TWDTW algorithm requires two inputs for image classification, namely (a) phenological cycles of land cover classes (b) long-term satellite image time series images of the same spatial extent with their respective dates (Victor Maus et al., 2016). According to the description of the *dtwSat* package, the TWDTW method consists of three main steps; (a) generating temporal phenological cycles of land cover classes from training samples based on red-edge VI time series (b) applying weighting function (logistic or linear) and (c) Classifying the raster time series (Victor Maus et al., 2016). Steps (a) and (b) are detailed explained under section 4.2.3 and 4:2:4, respectively.

The temporal phenological patterns of the land cover classes were extracted from the red-edge NDVI's calculated under section 3.4. Logistic weighting function was used with different values of the parameters steepness ( $\alpha$ ) and midpoint ( $\beta$ ) to obtain the best classification results.

#### 4.2.5. Accuracy Assessment of Classification results

Overall accuracy is the measure of mapping accuracy which tells us about what proportion out of the reference data that was correctly classified (*The core of GIScience: a systems-based approach*, 2012). Kappa shows the level of agreement between the classified and reference images corrected by chance. It evaluates if the classification of the categories did better than randomly assigning values.

Producer's accuracy (Error of omission) includes samples points which were excluded in the interpretation results (*The core of GIScience: a systems-based approach*, 2012). They are the measure of false negatives which show the fraction of values that belong to a certain class but were predicted in a different class (Harris, 2018b). Since omission error are related to the reference data and therefore, they are shown in the columns of the confusion matrix.

User's accuracy (error of commission) refers to the fraction of values that were predicted in a particular class but do not belong to it (Harris, 2018b). These are incorrectly classified samples (false positives). Errors of commission are indicated in the rows of the confusion matrix except for diagonal values.

The accuracy of the RF and TWDTW classification results was assessed to evaluate how good the maps are in terms of Overall accuracy, producer's and user's accuracy metrics (Congalton, 1991) and kappa coefficient (Cohen, 1990). The formulas to calculate the metrics are as follows;

Overall Accuracy (OA) = 
$$\frac{\text{Number of correctly classified cells}}{\text{Total number of pixels checked}}$$
Equation (2)Kappa (K) =  $\frac{\text{Observed accuracy - chance agreement}}{1 - \text{chance agreement}}$ Equation (3)whereChance agreement= Sum (product of row and column totals per class)Producer accuracy =  $\frac{\text{Total number of correct pixels in a category}}{\text{Total number of pixels of that category (column total)}}$ Equation (4)User accuracy =  $\frac{\text{Total number of correct pixels in a category}}{\text{Total number of classified pixels in that category (row total)}}$ Equation (5)

#### 4.2.6. Assessing the uncertainty of the RF and TWDTW classification results

From the error matrix, the post-classification analysis was performed to assess the uncertainty of the classification results using the information obtained from the map accuracy assessment. The information is useful in estimating the area of a land cover class and constructs confidence intervals that reflect the uncertainty of the area estimated. After performing accuracy assessment and create error matrix, estimating the error-adjusted of land cover area and confidence interval yields valuable information on land cover compared to results obtained from the map itself (Olofsson, Foody, Stehman, & Woodcock, 2013).

The error-adjusted estimator of an area includes the area of map omission error and leaves out the area of map commission error was calculated using the formula below (Olofsson et al., 2013)

$$\widehat{\mathbf{A}}_{j} = \mathbf{A}_{tot} \sum_{i} W_{i} \frac{n_{ij}}{n_{i}}$$

Equation (6)

Where;

 $\widehat{A}_j$  = Error-adjusted estimator of area  $A_{tot}$  = Total area of the map i = Map categories (i = 1,2, ...., q) of the error matrix j = Reference categories (j = 1,2, ...., q) of the error matrix n = cell of the error matrix  $W_i$  = Mapped area proportion given by ( $A_{m,i} \div A_{tot}$ )

 $A_{m,i}$  = The mapped area of category *i* 

Also, an approximate 95% confidence interval for  $A_i$  was estimated using the formula below

$$\widehat{A}_j \pm 2 \times S(\widehat{A}_j)$$
 Equation (7)

Where;

 $S(\widehat{A}_{l})$  = The standard error of the error-adjusted estimated area given by  $(A_{tot} \times S(\widehat{P}_{l}))$ 

 $\widehat{P}_I$  = Reference category of column totals in error matrix

 $S(\hat{P}_{I})$  = The estimated standard error of the estimated area proportion (Cochran, 1977),

given by 
$$\sqrt{\sum_{i=1}^{q} W_i^2 \frac{\frac{n_{ij}}{n_i}(1-\frac{n_{ij}}{n_i})}{n_i-1}}$$
 Equation (8)

#### 4.2.7. Transferability assessment

The transferability assessment of the best models identified under sections 4.2.1 and 4.2.4 was conducted across agroecosystems and years based on the common crops available in both study areas. The best RF and TWDTW classification models trained in California using 2017 datasets was transferred to three study areas, namely California in 2016, a neighboring region situated in California in 2017 and Mississippi in 2017 to map similar crops.

First, the models were trained to identify cotton crops and transferred to map similar crop of 2017 in Mississippi. Second, the models were trained again based on three crops, namely cotton, winter wheat, and tomatoes and it was transferred to two study areas, California 2016 and neighboring area within California 2017.

In a new study area where the models were applied, there was no model calibration. The models were applied with the same parameters tuned in order to obtain the best classification models before transferability.

In another case, the best RF and TWDTW models developed and trained in three study areas, namely California 2016, a neighboring region in California 2017 and Mississippi 2017 were applied in California 2017 to map similar crops same as in case one. The purpose of this test was to determine whether unequal class proportion between the two areas under test has an influence on the transferability test results.

The performance evaluation of the models was assessed based on the overall accuracy, producer's and user's accuracy metrics and kappa coefficient.

## 5. RESULTS

This chapter shows the results of the implemented methods. The chapter begins with presenting the results of the RF model, followed by TWDTW results in section 5.2. Section 5.3 presents a discussion of the Producer's and User's accuracies for both RF and TWDTW classification models. Section 5.4 shows the results of interclass similarity and intra-class variability assessment. Transferability assessment is presented and evaluated in section 5.5. The last section 5.6 present uncertainties of the obtained classification.

#### 5.1. Random Forest Classification results

Figure 5:1 (a) to (f) represents the classification results obtained by RF algorithm. Cotton was the common crop across different agroecosystem (California and Mississippi). To assess the transferability across different years and across neighboring areas, we classified the following classes: cotton, winter wheat, and tomatoes. The final implementation of the RF model was at 300 iterations which produced the best classification results. The computation time per iteration was approximately 2 minutes depending on the number of classes to be classified and the number of layers in the NDVI stack. In this research, 36 layers of NDVI red-edge were used as input, because we computed six spectral indices for six months.

#### 5.1.1. Overall accuracies of the classification results yielded by Random Forest

The best classification results obtained in California, in 2017 had an overall accuracy of 99.12% and kappa 98.82% compared to other areas as shown in Table 5:1. The accuracy of the classification yielded for the same area, but in a different year, namely 2016 was lower, namely we obtained an overall accuracy of 92.08% and kappa 89.38%. This was due to numerous changes that occurred from one year to another in the study area under investigation. For example, most of the tomatoes and winter wheat crops cultivated in 2016 were replaced by cotton in 2017. According to the CDL accuracy assessments which reflect the accuracy of each crop, tomatoes and winter wheat classes have lower accuracy compared to cotton in both years. Therefore, large coverage of tomatoes and winter wheat that were misclassified caused the lower accuracy obtained in 2016 as compared to those obtained in 2017. The accuracy obtained in different areas which belong to the same agroecosystem, was lower compared to other areas due to high confusion between winter wheat and "others" class. In California overall accuracy and kappa were 87.89% and 83.40% and its neighbor region were 87.59% and 83.15%, respectively.

Study area	Overall Accuracy	Карра
	in %	in %
California 2016	92.08	89.38
California 2017	99.12	98.82
California 2017	97.92	95.24
Mississippi 2017	98.33	96.63
California 2017	87.89	83.40
California 2017 (another	87.59	83.15
location)		

Table 5-1. Overall accuracy and Kappa obtained by RF in all each study areas

![](_page_30_Figure_3.jpeg)

Figure 5-1. Classified images of California in 2017 (a), (c), and (e), Mississippi 2017 (b), California 2016 (d), and neighboring area within California 2017 (f)

#### 5.1.2. Evaluation of the variable's importance

According to the variable assessment carried out in this work the best suitable time to classify crops in Mississippi is October when the crops reached their maturity, and the harvesting period starts. The NDVI-red edge values at the moment are lower and their temporal profiles indicate the end of the season. In Mississippi, the harvesting season starts on mid-September. This is influenced by the planting dates which are not fixed, and which normally range from April to mid or late May. In California 2016, California 2017 and the neighbor region, September was identified as the most relevant month to classify the target crops. This is the moment when the NDVI-red edge values are at the maturity stage and the harvesting season is approaching (from October until November).

Therefore, classifying target crops when they have reached the maturity and senescence stages increases the probabilities of the crops to be correctly classified. Figure 5:2 shows the first 30 important variables ranked from the most important (top) to the least ones (bottom).

![](_page_31_Figure_4.jpeg)

Figure 5-2. The variable importance measures and rankings using mean decreases in accuracy (MDA) of California 2017 (a), Mississippi 2017 (b), California 2016 (c) and an area within California 2017(d).

Note; ndvi\_b8b5-ndvi\_b8b7 and ndvi\_b8ab5-ndvi\_b8ab7 are the six calculated indices in all study areas. The last digits 1 to 6 represents months from May-November, respectively except August.

#### 5.2. Time-Weighted Dynamic Time Warping classification results

To implement the TWDTW classification model, different values of  $\alpha$  ( $\alpha = -0.2$ , -0.1) and  $\beta$  ( $\beta = 50$  or 100) parameters were used to run the classification. The best classification results were achieved when  $\alpha$  and  $\beta$  were set to -0.1 and 50, respectively. The processing time in this study varied between 9 to 12 hours depending on the number of crops classified. For the test areas with four crops, the processing time was between 11 and 12 hours, whereas for those areas where only two crops were classified the computation time was about 10 hours (Computer processor; Intel(R) Core (TM) i7-7700HQ CPU @2.80GHZ). For the purpose of transferability assessment, we considered again only those crops that were similar across evaluated study areas. For example, cotton was the common crops for California and Mississippi 2017. For testing transferability across years and across different study areas, but the same agroecosystem region, we considered the following crops: winter wheat, cotton, and tomatoes. Figure 5:2 show the classification results obtained by applying TWDTW.

#### 5.2.1. Overall accuracies of the classification results obtained by Time-Weighted Dynamic Time Warping

California 2017 yielded the best classification results with an overall accuracy of 96.67% and kappa 92.45% compared to all test areas under investigation. These accurate classification results were obtained due to distinctive temporal profile of cotton and "others" classes. The performance of TWDTW across different years but the same study area (e.g. California 2016 and 2017) were different due to the high overlap between the temporal profiles of tomatoes and "others" classes in 2016. Therefore, the classification results for 2016 were lower as compared to those obtained for 2017. Overall accuracy was 94.46% and kappa 92.55, in 2017 and 2016 the overall accuracy was 80.41% and kappa 73.69%. The test areas situated within the same agroecosystems obtained the lowest accuracy compared to other areas because of the high confusion between cotton and "others" class. Table 5:2 shows the results of TWDTW classification model.

Study area	Overall Accuracy	Карра
	in %	in %
California 2016	80.41	73.69
California 2017	94.46	92.55
California 2017	96.67	92.45
Mississippi 2017	91.11	82.35

Table 5-2. Overall accuracy and Kappa index obtained by TWDTW in all investigated study areas

California 2017	81.31	74.59
California 2017	89.66	85.98
(another location)		

![](_page_33_Figure_2.jpeg)

Figure 5-3. Classified maps of California in 2017 (a), (c), and (e), California 2016 (d), Mississippi 2017 (b) and another study area within California 2017 (f)

#### 5.3. Producer's (PA) and user's accuracies (UA) of RF and TWDTW

Figures 5:3 shows the PA and UA obtained by RF and TWDTW models in the investigated areas situated across different agroecosystems. The PA and UA for cotton and "others" classes yielded highest values ranging from 96% to 100% for both RF and TWDTW in California, whereas the classification results obtained for Mississippi were lower than 90%. This was due to small overlap of the temporal profiles of the cotton class with "others" from the start of the growing phase before maturity stage.

![](_page_34_Figure_1.jpeg)

Figure 5-4. Producer's and user's accuracy obtained by RF and TWDTW for California and Mississippi study areas in 2017

PA for cotton, tomatoes and winter wheat obtained by RF and TWDTW in 2017 achieved best classification results (higher than 90%). In the case of UA, all crops acquired high values except for those obtained by both classifiers for the tomatoes, namely 43.75% and "others" classes 76.47%, in 2016. Generally, the results obtained for California in 2017 by both RF and TWDTW were excellent compared to 2016. This is because in 2016 there was a high confusion between "others" class overlapped with those of tomatoes. In California 2017 the area covered by tomatoes and "others" classes were smaller compared to 2016. Therefore, the confusion between the two classes had a lower impact on the classification results obtained for 2017. Figure 5:4 presented below shows the PA and UA across years, for RF and TWDTW.

![](_page_34_Figure_4.jpeg)

Figure 5-5. Producer's and user's accuracy obtained by RF and TWDTW for California in 2016 and 2017

In the other study area situated in California the PA and UA obtained by RF and TWDTW for tomatoes and UA for cotton, for both RF and TWDTW were high from 97% to 100%, as shown in figure 5:5. In

California, the PA for cotton and "others" were 69.56% and 43.75%, respectively obtained by TWDTW and UA for others and winter wheat were 28% and 30% acquired by TWDTW and RF, respectively. This was due to the confusion between the classes of interest (cotton and winter wheat) and "others" class. The temporal patterns of "others" class showed high intra-variability and high overlap with winter wheat in both test areas. Furthermore, the winter wheat class was underrepresented in California, i.e. about 3 times smaller area compared to neighbor region situated in California. Therefore, the confusion between winter wheat and "others" classes and small proportion of winter wheat in California, contributed to the lower UA obtained by RF. Also, cotton and "others" class showed overlap of their temporal profiles from the start until the mid of the growing season. In its neighbor region, the temporal patterns of winter wheat and "others" were the same in the whole season. This confusion explains the poor classification results obtained by RF and TWDTW.

![](_page_35_Figure_2.jpeg)

Figure 5-6. Producer's and user's accuracy obtained by RF and TWDTW for California and its neighbor region in 2017

#### 5.4. Post classification analysis

By using the approach described by Olofsson et al. (2013), the error-adjusted estimate of the area (at 95% confidence interval) was computed. In California 2017, the adjusted and mapped area of cotton and tomatoes was the same (Figure 5.5(a)). There were no significant differences, because of their distinct temporal patterns which result in a good classification result. RF mapped a large area of cotton and "others" classes about 2948.19 ha and 1514.22 respectively, compared to the TWDTW. TWDTW mapped a large area of tomatoes about 2974.44 ha compared to RF. The area mapped by the winter wheat for both RF and TWDTW is smaller compared to other crops whereby the margin error of TWDTW is greater than that of RF.

In California 2017, where two classes were mapped, the area coverage of "others" class was larger than the cotton for both TWDTW and RF (Figure 5.5(b)). The differences between the mapped and adjusted areas for cotton and "others" class was 166.56 ha and 82.27 ha for RF and TWDTW, respectively. Therefore, the margin error of the mapped and adjusted area for cotton and "others" was the same for RF and TWDTW.

![](_page_36_Figure_2.jpeg)

Figure 5-7. Area uncertainties of the mapped and adjusted areas extracted from error matrix (at 95% confidence interval)

In California 2016, there was the greatest difference between mapped and adjusted for tomatoes produced by TWDTW, namely 895.70 ha as compared to RF which was 453.88 ha (figure 5:9 (c)). This happened due to the overlapping of its temporal profiles with "others" class. A large discrepancy was also identified for the winter wheat and cotton classes, where the adjusted area was higher with 203.96 and 537.42 ha for TWDTW. In case of RF, the mapped and adjusted areas of winter wheat have a smaller discrepancy, namely 13.27 ha. The area of tomatoes was underestimated for TWDTW because of the confusion between this class and "others" class. This also reflects the lower UA, namely 43.75% for TWDTW and larger area margin of 237.82 ha for RF.

In Mississippi 2017, the area mapped by cotton class was larger with a difference of approximately 3000 ha compared to "others" for both TWDTW and RF (figure 5:9 (d)). The adjusted area for cotton was larger with 276.28 ha for TWDTW. In case of RF-based classification results the mapped and adjusted area have smaller differences, namely 22.15 ha. The amount increased of adjusted area for cotton was the same quantity decreased by "others" class in the adjusted area obtained by TWDTW. The same differences in gain and loss between mapped and adjusted were identified for cotton and "others" for RF. This is the reason why the margin error of cotton and "others" classes are the same in adjusted and mapped for TWDTW and RF.

#### 5.5. Assessment of the interclass similarity and intra-class variability

Figure 5:6 (a) below show interclass similarities between the two crops i.e. cotton and tomatoes. Cotton growing phase is between May and July followed by the maturity stage from July to September and was harvested in October. In case of tomatoes, the green up phase usually starts on February (Urban Farmer LLC, 2019), the maturity stage is in June and its harvest takes place on September. Therefore, there is a clear distinction between the temporal profiles of the two crops. Figure 5.3 (b) indicate the inter-class variability of "others" class. As this class is composed of more than one class, namely idle cropland and dry beans there is a high inter-class variability and their profiles show different patterns. In this class the temporal profiles which overlap indicate similar class. The high variability of this class is directly reflected in the low classification accuracy obtained in our study. Also, some overlap on the temporal profile of this class with the other classes of interest influenced the classification accuracy obtained by RF and TWDTW.

![](_page_37_Figure_4.jpeg)

Figure 5-8. Interclass similarities of cotton and tomatoes, and intra-class variabilities of others class.

#### 5.6. Transferability assessment of the RF and TWDTW classification models

The best RF and TWDTW classification models were those obtained for California in 2017 as already mentioned in section 5.1.1 and 5.2.1. Therefore, these models were used to test their transferability for the

selected crops. No model calibrations were performed when applied on a new test area, the models were applied with the same parameters which were originally trained with. For RF, the best models were obtained by setting 1000 number of trees (*Ntree*) and 6 number of variables (*Mtry*). In the case of TWDTW, the best models were achieved when steepness ( $\alpha$ ) and midpoint ( $\beta$ ) parameters were set to -0.1 and 50, respectively.

We assessed the transferability by applying the best RF and TWDTW classification models (i.e. those obtained for California 2017) to Mississippi 2017, California 2016, and neighboring region in California. Also, another transferability test was conducted by applying the best RF and TWDTW classification models obtained for California 2016, neighboring region situated in California and Mississippi 2017 to California 2017.

#### 5.6.1. Model transferability across different agroecosystems

By transferring the RF and TWDTW classification models from California to Mississippi, the classification results dropped significantly. The overall accuracy decreased by 54.58% for RF and 49.45% for TWDTW, as shown in Figure 5:6 (a) and (b). The kappa index for RF indicated a clear disagreement between the two study areas and for TWDTW it was reduced by 88.41%. The PA and UA for all crops were decreased by approximately 50% for RF. TWDTW obtained a slightly better UA for cotton which was increased by 0.59%.

The models trained in Mississippi and applied in California also acquired extremely poor results, as shown in figure 5:6 (c) and (d). The overall accuracy was reduced by 42% and 76.28% for RF and TWDTW, respectively. Their kappa indicates no effective agreement between the two study areas. PA and UA accuracies for cotton was 1.8% and 2.5%, respectively, for TWDTW. These indices were 0 for RF because cotton was not identified in a new test area.

In general, model's transferability from Mississippi to California yielded poor results. The temporal profile of the crop of interest, i.e. cotton, between the two areas was different. Therefore, it was challenging to identify this crop in a new test area, especially for TWDTW. The class proportion of cotton in Mississippi was 4841.66 ha, approximately two times larger compared to its proportion in California. Transferring models with a large proportion of cotton from Mississippi to California yielded poor results. This indicates that unbalanced class proportion influenced the classification results obtained, by RF and TWDTW. "other" class was identified in both test areas. However, its composition was different between the two areas.

![](_page_39_Figure_1.jpeg)

Figure 5-9. Transferability of RF and TWDTW classification models across different agroecosystem

#### 5.6.2. Transferability of classification models across years

Figures 5:7 (a) and (b) below show the transferability results obtained in California in 2017 and 2016. There was no increase in accuracy after model transferability from 2017 to 2016, the overall accuracy decreased by 15.25% and kappa by 20.56% for RF. The same for TWDTW the overall accuracy and kappa were decreased by 17.83% and 23.51%, respectively. Therefore, no improvement was observed that exceeded the accuracies before transferability, for both RF and TWDTW.

However, when the classification models from 2016 were applied to 2017, TWDTW shows improvement since overall accuracy increased by 3.67% and kappa by 4.37%. PA and UA for all classes increased except for the PA obtained for the winter wheat and UA obtained for "others" class. This happened because of the high spectral confusion between these two classes. Most of the "others" pixels class were not identified in the new test area and were mapped as winter wheat since their temporal patterns overlapped.

Overall accuracy and kappa of the RF classification model showed no improvement. The results are as shown in figures 5:7 (c) and (d).

The results obtained indicate that, correctly classified classes in the test area where a model is applied have an influence on the final results of transferability. Transferability of RF and TWDTW models from California 2017 to 2016 yielded poor results because in 2016 the classification accuracy was low compared to 2017, therefore, the accuracy of the new predicted classes also became low. Transferability from California 2016 to 2017 achieved good results for TWDTW because the classification accuracy of 2017 classes was higher compared to 2016. PA and UA for all classes increased except for the PA for winter wheat and UA for other class. This is because of the high spectral confusion between these two classes. Most of the pixels belonging to the "others" class were not correctly identified and were mapped as winter wheat since their temporal patterns overlapped.

In case of RF classification model, its transferability from 2017 to 2016 indicated no improvement because the classification accuracy in 2016 was lower compared to 2017 where the model was trained. RF model transferability from 2016 to 2017 achieved low accuracy because of the unbalanced proportions of the classes of interests especially winter wheat and because of the high confusion between "others" class and tomatoes class in the new test area. PA and UA for cotton were increased to 100% since its accuracy in 2017 was good. UA for tomatoes was also increased from 76% to 100%. Winter wheat showed no improvement because of unequal proportional between the two areas. In California 2017 winter wheat area was three times smaller compared to its coverage in 2016.

![](_page_40_Figure_4.jpeg)

![](_page_41_Figure_1.jpeg)

Figure 5-10. Transferability of RF and TWDTW classification models across years in the same agroecosystem

#### 5.6.3. Model transferability across the same agroecosystems

After applying RF and TWDTW classification models trained in California 2017 to its neighbor region, the overall accuracy of the RF model decreased by 12.37% and kappa 16.81%, respectively as shown in figures 5:7 (a) and (b). There was no increase also for the TWDTW, the overall accuracy was reduced by 21.65% and kappa by 28.09%. PA of winter wheat and UA of tomatoes, was the same 100%, for RF. Also, the UA for cotton for both TWDTW and RF models its accuracy remained 100%. Its PA was reduced by 12.86% for RF and 2.33% for TWDTW because of the underrepresentation of this class in its neighbor region which was approximately three times smaller compared to its coverage in California 2017. Winter wheat was not identified by the TWDTW in the neighbor region due to non-overlapping temporal patterns between the two areas throughout the season. Therefore, the poor performance of RF and TWDTW classification models when transferred to a neighbor region is caused by the poor classification accuracy of this test area before transferability.

Model transferability from a neighbor region situated within California to California, TWDTW and RF classification models improved their accuracy. RF accuracy increased by 18.51% outperforming TWDTW which was improved by 4.45%. Kappa for TWDTW was 6.04% higher than RF which scored 1.93%. PA and UA of all crops for RF and TWDTW scored higher compared to those obtained before transferability, as shown in figures 5:8 (b) and (c).

This indicates, after applying the model to the new area where classification accuracy is high, the model performance for mapping the same crops will improve.

![](_page_42_Figure_1.jpeg)

Figure 5-11. Transferability of RF and TWDTW classification models across the same agroecosystem

## 6. **DISCUSSION**

The chapter discusses some of the key findings in this research. The discussion starts by highlighting crops phenology and reflecting on the results acquired. It continues by presenting the potential of S2 for mapping time series data, followed by the performance of RF and TWDTW. The last three sections focused on RF and TWDTW transferability results achieved in our study.

#### 6.1. Phenological cycles of crops

According to the crop calendar, the cotton is cultivated starting from April-November. Given the fact that cloud coverage is one of the challenges in mapping crop type based on time series (Petitjean, Inglada, et al., 2012), two images of April and August were missed because of the cloud coverage. Therefore, the focus was on six images from May-November with exception of August. Within this temporal interval (May-November) there was no overlap in the phenology represented between cotton and "others" classes in all study areas. Therefore, cotton was relatively good classified.

The high heterogeneity of the "others" class caused the relatively low UA and PA obtained in our study. The crop calendar indicates that tomatoes crop is normally cultivated starting from February-September. Within this temporal range, there was an overlap with "others" class in California 2016 which resulted in low UA. The rest of the test areas obtained good classification results.

The planting dates for winter wheat are from October to December and the harvest period ranges between June-July. The temporal range represented in this study was during the harvest period. In most cases, this crop was confused with "others" class because one of the components of the "others" class was idle cropland, which has a temporal profile similar to the winter wheat class profile.

#### 6.2. Potential of time series Sentinel 2 data for mapping crop type

This research was motivated by the freely available high temporal resolution of S2 for mapping crops. S2 adds a new image every 5 days (with both satellites in orbit) and makes it very appealing for crop monitoring and tracing vegetation dynamics in time (Escolà, Badia, Arnó, & Martínez-Casasnovas, 2017). The increase of temporal resolution of the Multispectral Instrument carried onboard Sentinel platform allowed us to capture growing cycles of the target crops and detect their inter-annual processes. However, to obtain the cloud-freely images covering the whole season for the crops of interest is a challenging task. The presence of clouds in an image covers the true surface reflectance. In this study, several cloud-free images were available as shown in table 3:2. These data were sufficient for achieving satisfactory results and meet the main objective of our research. There are a number of studies indicating the potential of S2 time series for crop type mapping using RF and TWDTW (Belgiu & Csillik, 2018), for identifying and

updating crop field boundaries using proposed automated approach (Solano-Correa, Bovolo, Bruzzone, & Fernandez-Prieto, 2017), or for constructing a dynamic cropland mask to map cultivated areas using RF (Valero et al., 2015).

In this research, we used the NDVI red-edge spectral index which proved to add value to the S2 bands for crop type mapping. Forkuor et al. (2018) proved that the classification accuracy of land use and cover increased when red-edge bands were used in addition to the rest of the S2 bands and Landsat-8. Qiu et al. (2017) assessed and proved that red-edge spectral bands improve land cover classification. Also, the red-edge bands are suitable for discriminating burn severity (Fernández-Manso et al., 2016).

#### 6.3. RF and TWDTW classification models

RF classier outperformed TWDTW in all test areas except in the neighboring area situated in California. This happened because RF proved to better handle the high confusion between winter wheat and "others" classes. The efficiency of this classifier for mapping agriculture was also confirmed in the previous studies (Belgiu & Csillik, 2018, Hao et al., 2015, Pelletier et al., 2016). The overall performance of TWDTW was not good because it was a one-dimensional classification model compared to RF which relied on more input variables, e.g. red-edge based NDVI indices.

The biggest differences between TWDTW and RF classification results were obtained for California in 2016. TWDTW showed lower UA for tomatoes and PA for "others" because of the overlapping of their temporal profiles. The performance of RF for classifying "others" class as compared with the remaining target classes was not good especially when inspecting the classification results obtained in California 2016, 2017 and its neighboring region. This is because training and validation samples for "others" were underrepresented compared to other classes of interest. This study showed that RF cannot easily cope with an unequal proportion of training or validation samples and that this classification method favors the most representative classes (Dalponte et al., 2013, Haibo and Garcia, 2009). This is reflected in the low PA and UA obtained for the 'others' class. Cotton, winter wheat and tomatoes were over-represented classes in the training or validation of these classes whereas the under-represented class may also be under-represented in the classification result.

#### 6.4. RF and TWDTW model transferability across different agroecosystems

RF classification model transferability across different agroecosystem corresponds to the findings obtained by Jin et al. (2018) and Juel et al. (2015) who reported that RF model trained in one location cannot be transferred to new locations. RF and TWDTW models trained in California when applied to Mississippi to map similar crops their accuracy decreased. We performed another test by transferring RF and TWDTW models trained in Mississippi and apply them to California to map similar crops. The classification results yielded were extremely poor compared to the model's transferability from California to Mississippi.

#### 6.5. RF and TWDTW model transferability across same agroecosystems

RF and TWDTW model trained in California 2017 and applied to its neighboring region (transferability across same agroecosystems) achieved less accurate classification results. This is because of the low classification results achieved by RF and TWDTW in the neighboring region compared to California. Most of the pixels belonging to the winter wheat class were not accurately identified by TWDTW.

Another transferability test was carried out from a neighboring region to California. In this case, RF and TWDTW classification accuracy increased. This is due to the fact that classification accuracy of California was higher compared to the neighboring region where the models were trained. RF transferability results resemble with the study conducted by Wang et al. (2019) who concluded that RF is capable to obtain high accuracies of more than 80% when applied to the nearby geographies when regional crop composition and growing degree days (GDD) are similar. However, the authors omitted to check the class proportions of similar classes between the two areas.

#### 6.6. RF and TWDTW model transferability across years

Before transferability, the accuracy of models in California 2017 was excellent compared to the classification results obtained for 2016. When we applied the models trained in California 2017 to 2016 its accuracy decreased for both RF and TWDTW. This is due to the lower classification accuracy in 2016. Again, we tested the model transferability from California 2016 back to 2017 the accuracy of the TWDTW model increased as expected. In the case of RF, it showed no increase in accuracy after transferability. This is because tomatoes and winter wheat were over-represented for California in 2016 compared to 2017, after transferability their UA and PA showed no improvement.

## 7. CONCLUSION AND RECOMMENDATIONS

The chapter focused on the conclusion of this research and recommendation for future studies. Section 7.1 conclude based on findings of the main objective and the last section 7.2 presents a recommendation for future studies.

#### 7.1. Conclusion

The main objective of this research was to assess to what extent the TWDTW and RF-based classification models are transferable across different agroecosystems for agriculture mapping. The data used in this research was a Sentinel-2 time series captured in three study areas, namely Mississippi Alluvial Plain in 2017, California Great Plains in 2016 and 2017.

This study revealed that RF and TWDTW based classification models are not transferable across different agroecosystems with different climatic condition and dissimilar agriculture management practices. The models are transferable across the same agroecosystems having similar growing conditions. However, class proportions of similar classes between the two areas under investigations should be equal or have a small discrepancy. Also, correctly classified classes have an influence on the new predicted classes transferred from another test area.

The answers to the research questions raised in our study are:

a. What are the most relevant vegetation indices used to map target crops in the selected study areas? The most relevant vegetation index to map target crops in Mississippi in 2017 is the NDVI rededge1 index calculated using NIR band 8 and the red-edge band 5. Furthermore, the crop classification should be carried on October when the crops have reached their maturity stage.

In 2016 and 2017 in California, NDVI red-edge1 narrow (NDVI calculated using Narrow-NIR 8A and the red-edge band 5) is the most suitable vegetation index identified. The most relevant month to classify target crops is September when harvesting date is approaching. In a neighboring region located in California, the suitable time for the crop classification is the same as in California in 2016 and 2017, but its relevant vegetation index is NDVI red-edge 1 (NDVI of NIR band 8 and the red-edge Band 5).

b. Which of the developed classification models perform better in the selected study areas?

RF classification model performed better in all study areas in terms of measured overall accuracy metric and computational time, except for the neighboring region in California its overall accuracy was lower by 2% compared to TWDTW.

Low performance of the TWDTW classification model was due to a high intra-class variability presented by some of the crops which led to overlapping temporal patterns of different crops.

c. Is there any difference in the classifier performance in case of high intra-class variability and/or interclass similarity?

RF classifier was less affected by the high intra-class variability and interclass similarity, in all study areas as shown in Table 5:1 and under section 5.1.1. However, in the neighboring region in California, high intra-class variability of "others" created confusion with cotton. This confusion challenged the RF classier.

In most cases, TWDTW was challenged by the intra-class variability of "others" class whereby its temporal profiles were overlapping with some of the crops of interest. This is the reason why TWDTW performed less accurately in our study areas.

d. How transferable are the developed classification models across different agroecosystems without calibration?

RF and TWDTW classification models are transferable across agroecosystems in the same location having similar growing conditions. However, imbalance class proportions between the two areas may hinder their transferability.

The models are also transferable in the same agroecosystems across different years. However, high spectral confusion between classes and unequal proportions of classes between the two areas may limit their transferability.

#### Limitations:

- This study was challenged by the quality of the ground truth data available in Cropscape (CropScape and Cropland Data Layers, 2018). For example, some crops like pistachios, fallow/idle cropland, tomatoes and winter wheat classified in this study have lower accuracy according to the CDL accuracy assessment for 2017 datasets. The classes also indicated confusion among themselves in the selected study area.
- Similarities between the two test areas for transferability was limited. For example, in Mississippi and California cotton was only similar crop in both study area the rest of the crops were different.

#### 7.2. Recommendations and future works

We recommend the following for future work

a. TWDTW proved to be more robust than other classifiers when comes to inconsistent time series. However, we emphasize on a suggestion made in Belgiu & Csillik. (2018), further studies

should be conducted on improving the TWDTW in a situation when the time series available doesn't correspond to the phenological cycles of crop types under investigation.

b. Transferability test of RF and TWDTW classification models performed in this study is at pixel level. We recommend transferability assessment of object-based RF and TWDTW classification models. Object-based TWDTW classification model was proven in Belgiu & Csillik (2018) to achieve best results than pixel-based RF and TWDTW.

#### LIST OF REFERENCES

- Belgiu, M., & Csillik, O. (2017). Cropland mapping from Sentinel-2 time series data using object-based image analysis. Retrieved from https://agile-
- online.org/conference\_paper/cds/agile\_2017/shortpapers/127\_ShortPaper\_in\_PDF.pdf

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based timeweighted dynamic time warping analysis. *Remote Sensing of Environment*, 204(October 2017), 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), 493–507. https://doi.org/10.1002/widm.1072

- Breiman, L. (2001a). Random Forests. Retrieved from https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
- Breiman, L. (2001b). Random Forests (Vol. 45). Retrieved from https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf
- CLIMATE OF CALIFORNIA. (2009). Retrieved January 2, 2019, from
  - https://wrcc.dri.edu/narratives/CALIFORNIA.htm
- Cochran, W. G. (1977). Sampling Techniques, third edition, 428.
- Cohen, J. (1990). A COEFFICIENT OF AGREEMENT FOR NOMINAL SCALES 1. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/001316446002000104
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B
- CropScape and Cropland Data Layers. (2018). USDA National Agricultural Statistics Service Research and Science - CropScape and Cropland Data Layers. Retrieved February 17, 2019, from https://www.nass.usda.gov/Research\_and\_Science/Cropland/sarsfaqs2.php#Section3\_22.0

Dalponte, M., Orka, H. O., Gobakken, T., Gianelle, D., & Naesset, E. (2013). Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5), 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, 11(7), 7063–7081. https://doi.org/10.3390/s110707063

Dinsdale, L., & Edwards, R. (n.d.). Metagenomics. Statistics. Retrieved November 22, 2018, from https://dinsdalelab.sdsu.edu/metag.stats/index.html

Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259–272. https://doi.org/10.1016/J.RSE.2011.11.020

ESA. (2019). MSI Instrument – Sentinel-2 MSI Technical Guide – Sentinel Online. Retrieved February 13, 2019, from https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument

Escolà, A., Badia, N., Arnó, J., & Martínez-Casasnovas, J. A. (2017). Using Sentinel-2 images to implement Precision Agriculture techniques in large arable fields: First results of a case study. Advances in Animal Biosciences, 8(02), 377–382. https://doi.org/10.1017/S2040470017000784

Fernández-Manso, A., Fernández-Manso, O., & Quintano, C. (2016). SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 50, 170–175. https://doi.org/10.1016/j.jag.2016.03.005

FILELLA, I., & PENUELAS, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459– 1470. https://doi.org/10.1080/01431169408954177

Forkuor, G., Dimobe, K., Serme, I., & Tondoh, J. E. (2018). Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience and Remote Sensing*, 55(3), 331–354. https://doi.org/10.1080/15481603.2017.1370169

- Galvan-Tejada, C. E., Lopez-Monteagudo, F. E., Alonso-Gonzalez, O., Galvan-Tejada, J. I., Celaya-Padilla, J. M., Gamboa-Rosales, H., ... Zanella-Calzada, L. A. (2018). A Generalized Model for Indoor Location Estimation Using Environmental Sound from Human Activity Recognition. *Isprs International Journal of Geo-Information*, 7(3). https://doi.org/10.3390/ijgi7030081
- Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observations and Geoinformation*, 26, 49–63. https://doi.org/10.1016/j.jag.2013.05.017
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. In *Pattern Recognition Letters* (Vol. 27, pp. 294–300). https://doi.org/10.1016/j.patrec.2005.08.011
- Goldstein, B. A., Polley, E. C., & Briggs, F. B. S. (2011). Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*. Berkeley Electronic Press. https://doi.org/10.2202/1544-6115.1691
- Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. https://doi.org/10.1016/j.isprsjprs.2016.03.008
- Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z., & Yang, X. (2013). Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *International Journal of Remote Sensing*, 34(14), 5166–5186. https://doi.org/10.1080/01431161.2013.788261
- Guyet, T., & Nicolas, H. (2016). Long term analysis of time series of satellite images. *Pattern Recognition Letters*, 70, 17–23. https://doi.org/10.1016/J.PATREC.2015.11.005
- Haibo He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239
- Han, H., Guo, X., & Yu, H. (2017). Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 219–224. https://doi.org/10.1109/ICSESS.2016.7883053
- Hapfelmeier, A., Hothorn, T., Ulm, K., & Strobl, C. (2014). A new variable importance measure for random forests with missing data. *Statistics and Computing*, 24(1), 21–34. https://doi.org/10.1007/s11222-012-9349-1
- Harris. (2018a). Narrowband Greenness. Retrieved November 21, 2018, from https://www.harrisgeospatial.com/docs/narrowbandgreenness.html#red\_edge\_NDVI
- Harris, G. solution. (2018b). Calculate Confusion Matrices. Retrieved February 13, 2019, from https://www.harrisgeospatial.com/docs/CalculatingConfusionMatrices.html
- Hickel, J. (2015). The Problem with Saving the World. Retrieved June 5, 2018, from https://www.jacobinmag.com/2015/08/global-poverty-climate-change-sdgs/
- Hsu, E., Pulli, K., & Popovi´c, J. P. (2005). *Style Translation for Human Motion*. Retrieved from https://people.csail.mit.edu/jovan/assets/papers/hsu-2005-stf.pdf
- Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., & Guo, Q. (2018). The transferability of Random Forest in canopy height estimation from multi-source remote sensing data. *Remote Sensing*, 10(8), 1–21. https://doi.org/10.3390/rs10081183
- Juel, A., Groom, G. B., Svenning, J. C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 42, 106–114. https://doi.org/10.1016/j.jag.2015.05.008
- Julea, A., Lasserre, C., Jolivet, R., Ledo, F., & Lodge, F. (2011). Unsupervised Spatio-Temporal Mining of Satellite Image Time Series. *IEEE Transaction on Geoscience and Remote Sensing*, 49(4), 1417–1430.
- Kontgis, C., Warren, M. S., Skillman, S. W., Chartrand, R., & Moody, D. I. (2017). Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics. In 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp) (pp. 1–4). IEEE. https://doi.org/10.1109/Multi-Temp.2017.8035199
- Koppelman, F. (1986). MULTIDIMENSIONAL CHOICE MODEL TRANSFERABILITY. Retrieved from https://ac.els-cdn.com/0191261586900469/1-s2.0-0191261586900469main.pdf?\_tid=51971d3d-0c0d-41dd-999f-

6f51361400af&acdnat=1546416263\_c48c529171bcb1b2cbbc3cccc91cbd75

Kulkarni, K., Evangelidis, G., Cech, J., Horaud, R., Kulkarni, K., Evangelidis, G., ... Cech, J. (2014). Continuous Action Recognition Based on Sequence Alignment. Int J Comput Vis, 112, 90–114. https://doi.org/10.1007/s11263-014-0758-9 Kumar, U. (2015). Application of Remote Sensing in Agriculture. Retrieved June 7, 2018, from https://www.slideshare.net/UTTAMKUMAR89/application-of-remote-sensing-in-agriculture

- Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). https://doi.org/10.1016/j.rse.2005.10.014
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., & Bégué, A. (2017). A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sensing*, 9(3). https://doi.org/10.3390/rs9030259
- Li, X., Chen, W., Cheng, X., Wang, L., Li, X., Chen, W., ... Wang, L. (2016). A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. *Remote Sensing*, 8(6), 514. https://doi.org/10.3390/rs8060514
- Maus, V., Appel, M., & Giorgino, T. (2017). Time-Weighted Dynamic Time Warping for satellite image time series analysis (Software), 1–11. Retrieved from http://pure.iiasa.ac.at/id/eprint/14514/
- Maus, V., C<sup>^</sup>, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & Ribeiro, G. Q. (2015). A Time-Weighted Dynamic Time Warping method for land use and land cover mapping. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, XX(X), 1–10. https://doi.org/10.1109/JSTARS.2016.2517118
- Maus, V., Câmara, G., Appel, M., & Pebesma, E. (2016). dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. *Journal of Statistical Software*, VV(Ii). Retrieved from http://www.jstatsoft.org/
- Millard, K., & Richardson, M. (2015). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7(7), 8489–8515. https://doi.org/10.3390/rs70708489
- Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A. J., & Hostert, P. (2015). Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment*, 156, 490–499. https://doi.org/10.1016/j.rse.2014.10.014
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122–131. https://doi.org/10.1016/j.rse.2012.10.031
- Ozdarici-Ok, A., Ok, A. O., & Schindler, K. (2015). Mapping of agricultural crops from single highresolution multispectral images-data-driven smoothing vs. parcel-based smoothing. *Remote Sensing*, 7(5), 5611–5638. https://doi.org/10.3390/rs70505611
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016a). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156–168. https://doi.org/10.1016/j.rse.2016.10.010
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016b). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156–168. https://doi.org/10.1016/j.rse.2016.10.010
- Petitjean, F., Inglada, J., & Gançarski, P. (2012). Satellite image time series analysis under time warping. IEEE Transactions on Geoscience and Remote Sensing, 50(8), 3081–3095. https://doi.org/10.1109/TGRS.2011.2179050
- Petitjean, F., Kurtz, C., Passat, N., & Gançarski, P. (2012). Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, 33, 1805–1815. https://doi.org/10.1016/j.patrec.2012.06.009
- Qiu, S., He, B., Yin, C., & Liao, Z. (2017). Assessments of Sentinel-2 vegetation red-edge spectral bands for improving land cover classification. *International Archives of the Photogrammetry*, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(2W7), 871–874. https://doi.org/10.5194/isprsarchives-XLII-2-W7-871-2017
- Rhemtulla, J. M., & Kalacska, M. (2012). Mapping Crop Types, Irrigated Areas, and Cropping Intensities in Heterogeneous Landscapes of Southern India Using Multi-Temporal Medium-Resolution Imagery Adapting to Climate Change in Urbanising Watersheds (ACCUWa) (2012-2016) View project Evaluating the resiliency of Mexican farming to Climate Change. View project. https://doi.org/10.14358/PERS.78.8.815
- Sakoe, H., & Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055
- Sanko, N., & Morikawa, T. (2010). Temporal transferability of updated alternative-specific constants in

disaggregate mode choice models. Transportation, 37(2), 203–219. https://doi.org/10.1007/s11116-009-9252-6

- Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing*, 33(17), 5583–5599. https://doi.org/10.1080/01431161.2012.666812
- Sibanda, M., Mutanga, O., & Rouget, M. (2015). Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110(March 2018), 55–65. https://doi.org/10.1016/j.isprsjprs.2015.10.005
- Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A., Chehbouni, A. G., ... Chehbouni {, A. G. (2010). International Journal of Remote Sensing The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco The use of high-resolution image time series for crop classification and. *International Journal of Remote Sensing*, 29(1), 95–116. https://doi.org/10.1080/01431160701250390
- Solano-Correa, Y. T., Bovolo, F., Bruzzone, L., & Fernandez-Prieto, D. (2017). Spatio-temporal evolution of crop fields in Sentinel-2 Satellite Image Time Series. In 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp) (pp. 1–4). IEEE. https://doi.org/10.1109/Multi-Temp.2017.8035236
- The core of GIScience: a systems-based approach. (2012). Retrieved from https://blackboard.utwente.nl/bbcswebdav/pid-1059031-dt-content-rid-2609632\_2/courses/U17-EDU-137/CoreBook2012\_2013.pdf
- Torbick, N., Chowdhury, D., Salas, W., Qi, J., Torbick, N., Chowdhury, D., ... Qi, J. (2017). Monitoring Rice Agriculture across Myanmar Using Time Series Sentinel-1 Assisted by Landsat-8 and PALSAR-2. Remote Sensing, 9(2), 119. https://doi.org/10.3390/rs9020119
- Urban Farmer LLC. (2019). California: Vegetable Planting Calendar Urban Farmer Seeds. Retrieved February 2, 2019, from https://www.ufseeds.com/learning/planting-schedules/california-vegetableplanting-calendar/
- US Department of Commerce, N. N. W. S. (n.d.). National Weather Service. Retrieved from https://www.weather.gov/
- Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., ... Defourny, P. (2015). Processing Sentinel-2 image time series for developing a real-time cropland mask. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 2731–2734). IEEE. https://doi.org/10.1109/IGARSS.2015.7326378
- Vasimalla, K., Challa, N., & Manohar Naik, S. (2016). Efficient dynamic time warping for time series classification. *Indian Journal of Science and Technology*, 9(21), 1–7. https://doi.org/10.17485/ijst/2016/v9i21/93886
- Wang, S., Azzari, G., & Lobell, D. B. (2019). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222(December 2018), 303–317. https://doi.org/10.1016/j.rse.2018.12.026
- Watson, V. H., & Lee, J. C. (2005). Mississippi Agricultural & amp; Forestry Experiment Station (MAFES) B1143 Current Agricultural Practices of the Mississippi Delta. Retrieved from http://mafes.msstate.edu/publications/bulletins/b1143.pdf
- Watts, J. D., Lawrence, R. L., & Group, W. (2008). MERGING RANDOM FOREST CLASSIFICATION WITH AN OBJECT-ORIENTED APPROACH FOR ANALYSIS OF AGRICULTURAL LANDS. Retrieved from http://www.isprs.org/proceedings/xxxvii/congress/7\_pdf/4\_WG-VII-4/18.pdf
- Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x
- Werkowska, W., Márquez, A. L., Real, R., & Acevedo, P. (2017). A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25(1), 127–133. https://doi.org/10.1139/er-2016-0045
- Wilmot, C. G. (1982). *Transferability Analysis of Disaggregate Choice* \10dels. Retrieved from http://onlinepubs.trb.org/Onlinepubs/trr/1982/895/895-004.pdf
- Wood, S. N. (2010). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Statist. Soc. B (Vol. 73). Retrieved from https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2010.00749.x

- Xiaomu Song, Guoliang Fan, & Mahesh Rao. (2005). Automatic CRP mapping using nonparametric machine learning approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 888–897. https://doi.org/10.1109/TGRS.2005.844031
- Zafari, A., Zurita-Milla, R., & Izquierdo-Verdiguier, E. (2017). Integrating support vector machines and random forests to classify crops in time series of Worldview-2 images. In L. Bruzzone, F. Bovolo, & J. A. Benediktsson (Eds.), *Image and Signal Processing for Remote Sensing XXIII* (Vol. 10427, p. 34). SPIE. https://doi.org/10.1117/12.2278421
- Zhu, Z., Diao, C., & Deng, C. (2018). Remote Sensing Time Series Analysis with Moderate Spatial Resolution Imagery I | Session Gallery | AAG Annual Meeting 2018. Retrieved December 31, 2018, from https://aag.secure-abstracts.com/AAG Annual Meeting 2018/sessions-gallery/535