A Multi-Sensor Approach to Classify Tillage Practices in Mexico

by

Haoyu Wang

A thesis submitted

in partial fulfillment of the requirements

for the degree of Master of Science

(School for Environment and Sustainability)

in the University of Michigan

May 2020

Thesis Committee:

Assistant Professor Meha Jain, Chair

Associate Research Scientist Kathleen Bergen

Abstract

To minimize soil disturbance, there has been an increased adoption of reduced or zero tillage (ZT) technologies among farmers in different regions across the globe. Yet, to date, the scale of adoption remains unclear because it is difficult to collect adoption data on-the-ground at large spatial and temporal scales. Remote sensing can offer a way to map such technology adoption at large scales and at low cost. This study uses Sentinel-2, Landsat 7 & 8, and Sentinel-1 satellites to map tillage practices in Guanajuato, Mexico, a region where the use of zero-tillage has been promoted by national and international agencies over the last decade. We specifically compared accuracy scores of different sensors and sensor combinations, and different timing of imagery in a random forest classification. The results indicate that when considering the accuracy of a single sensor, Sentinel-2 has the highest classification accuracy. However, using a combination of all three sensors dramatically outperformed all single sensor analyses, with an overall classification accuracy of 85.96%. Considering image timing, we find that using imagery from only the sowing season performs almost as well as using imagery throughout the growing season. We conclude that using freely-available satellite images is effective in classifying tillage practices in Mexico at large spatio-temporal scales.

Keywords: zero tillage; Mexico; optical and SAR sensors; Google Earth Engine; random forest

Acknowledgement

I wish to express my sincere gratitude to my thesis committee chair Dr. Meha Jain for providing the template for this study, and guiding me through the whole process. I would like to thank Dr. Preeti Rao and Weiqi Zhou for the helpful advice about software and coding. It is a great honor for me to work with and learn from all the members of the Jain lab. I want to thank Kai Sonder for the field data support. I would like to also thank my thesis committee member Dr. Kathleen Bergen for leading me into the world of GIS and remote sensing, and helping me build essential geospatial concepts. I also give special thanks to my family and Duo Xu for all the support and inspiration.

Table of Contents

1. Introduction

Over the last several decades, the use of reduced or zero tillage (ZT), a technique where seeds are sown and fertilizer is applied in fields that have limited or no disturbance and have crop residues, is becoming more prevalent [1,2]. ZT offers the advantage of reducing soil erosion, soil water loss, and greenhouse gas emissions compared to conventional tillage (CT) [1,2,3]. ZT can also improve soil condition, improving water use efficiency, decreasing water evaporation, and increasing the infiltration of precipitation [2,4]. While ZT is increasingly adopted globally, there is limited understanding of how widely adopted ZT practices are and how consistently farmers use these techniques through time. This is because traditional methods of recording and tracking the use of ZT, such as social surveys, are usually resource-consuming [5]. Remote sensing can offer an alternative way to efficiently and accurately detect ZT adoption at large spatio-temporal scales.

Although numerous studies have examined how satellite remote sensing can be used to classify ZT practices, few studies have examined how combining information from multiple sensors can improve classification accuracy. Yet, doing so is important given that different sensors can offer different benefits to classification. Specifically, Sentinel-2, which has a 10 m spatial resolution and a 5-day temporal resolution, can provide fine spatial resolution and time series data to classify ZT on smaller agricultural fields. Previous studies have found that Sentinel-2 can improve classification of forest and crop types, yet to date no studies have used Sentinel-2 to classify ZT versus CT fields [6,7]. Sentinel-1, which has a 10 m spatial resolution and a 12-day revisit time, is a radar satellite where microwave signals can penetrate clouds and are less affected by weather conditions [8,9,10]. This sensor can, therefore, be beneficial during periods of high cloud cover and haze. Previous studies have used Sentinel-1 to measure surface roughness with the aim of monitoring land cover and mapping vegetation. Finally, Landsat satellite data, which has a 30 m resolution and a 16 day revisit time, is beneficial because it can produce data over long time periods, since Landsat data are available from the 1970s onwards; Sentinel data, on the other hand, are only available starting in 2014. Previous studies have used Landsat satellite data to quantify the crop residue cover of fields in order to identify tillage intensity [11].

The previous studies that do exist that have used multiple sensors to map agricultural characteristics, including ZT, have found that using multiple sensors can lead to higher accuracies. Beeson et al. [13] suggests that using combined data from Landsat-like multispectral sensors (e.g. Landsat-8 and Sentinel-2) improves the temporal resolution of data, improving accuracies; however, no studies to our knowledge have combined Landsat and Sentinel-2 to map ZT. In addition, few studies combine optical and SAR sensors for agricultural land use classification, including ZT [5, 12, 14]. Azzari et al. [5] used image composites from a combination of optical (Landsat 5, 7, & 8) and synthetic-aperture radar (SAR) Sentinel-1 sensors to map ZT across the United States; the authors found that the combination of the two sensors led to higher accuracies, with an average accuracy of around 75% across models.

In this study, we use single and combined datasets from Sentinel-2, Landsat 7 ETM+, Landsat 8 OLI, and Sentinel-1 to examine which sensor and sensor combinations lead to the highest classification accuracies. We use ground data on zero tillage adoption from 2585 fields in the year 2017 in Guanajuato, Mexico. We specifically focus this study in Mexico because zero tillage adoption has increased since the 1990s due to the promotion of such technologies by the International Maize and Wheat Improvement Center (CIMMYT), which is headquartered in Mexico [15]. In addition, a sharp increase in population in the subtropical highlands in Mexico has also resulted in many sustainability issues, which accelerated the process of ZT adoption in this

region [16]. Several previous studies exist that have used satellite data to map ZT in Guanajuato, Mexico, however, these studies were conducted prior to 2015 and thus did not consider Sentinel data [17,18]. The specific objectives of this study are to 1) examine which sensor or sensor combinations lead to the greatest accuracy in mapping ZT, 2) assess which time periods in the growing season are most critical for mapping ZT, and 3) identify which bands and indices are the most important for mapping ZT accurately.

2. Study Area

This study is focused on the state of Guanajuato, Mexico. In the high peaks of Guanajuato, there are sub-humid and semi-cold climates, while warm sub humid conditions occur in the eastern and western ends of the sub-province. The warmest month is in May and the coldest is in January, and the rainy season spans from June to September [2,15]. Crops are planted during two seasons in Guanajuato (Figure 1) [17], the autumn/winter and the spring/summer. For the autumn-winter season, crops are planted between October and February, with the majority of crops being sown in December. The main crops grown during this season are wheat, barley, and vegetables. For the spring-summer season, crops are planted between March and September, with the majority of crops being sown in May. The main crops grown during this season are corn and sorghum. In this study, we focused on maize crops planted during the spring-summer season.

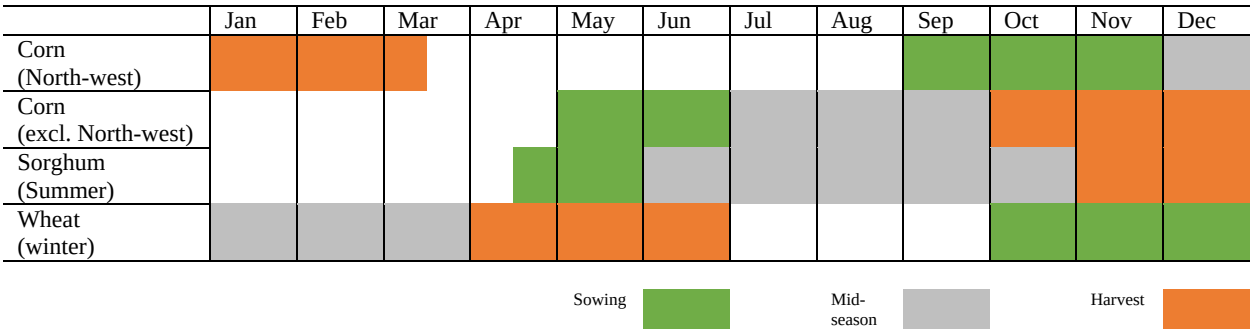

Figure 1. Crop calendar of the major crops in Mexico. Green indicates sowing/planting seasons; grey indicates growing/peak seasons; orange indicates harvest seasons.
Source: USDA FAS/IPAD.

3. Data and Image Processing

3.1. Field data

Ground truth data were collected as field shapefiles by CIMMYT in 2017 [19]. Each field was labeled as either zero tillage (ZT) or conventional tillage (CT). Given that the data collection by CIMMYT was focused on ZT, there were more ZT fields than CT fields. Therefore, in order to generate an unbiased model for classification, we randomly selected an equal number of ZT and CT polygons from the full dataset. We also checked each polygon and deleted those that did not include apparent agricultural features by using visual inspection in Google Earth. This resulted in 818 field polygons selected for each tillage class for a total of 1636 field polygons.

Table 1. Field polygon count before and after processing.

Field Label	Original Count	Processed Count
Zero Tillage (ZT)	1734	818
Conventional Tillage (CT)	851	818
Total	2585	1636

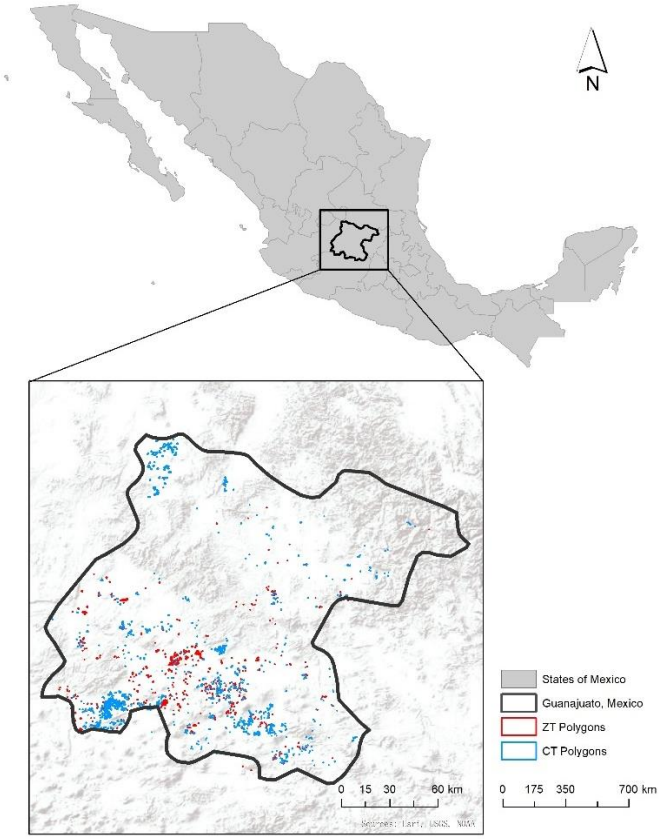

Figure 2. Map of the selected ground truth polygons in the area of interest – the state of Guanajuato, Mexico.

3.2. Sentinel-2

We accessed the European Space Agency (ESA) Copernicus Sentinel-2 satellite data through the Google Earth Engine (GEE) platform [20]. Since no surface reflectance (SR) corrected Level-2A data were available for the area of interest in 2017, we instead used the top-of-atmosphere (TOA) orthoimage products, specifically the Multi-Spectral Instrument (MSI) Level-1C image collection. We corrected the TOA product to SR using the radiative transfer emulator *Second Simulation of the Satellite Signal in the Solar Spectrum* (6S) with Py6S interface. The emulator uses built-in interpolated look-up tables to quantify the atmospheric variables, such as solar zenith, ozone, and surface altitude, which are then used to calculate atmospheric correction coefficients. SR was calculated from at-sensor radiance using the correction coefficients (credit to Sam W. Murphy).

Since our study focused on the rainy season in Mexico, there was a significant amount of cloud cover during certain time periods considered in our study. We thus used a cloud mask to mask out cloudy pixels. The cloud mask works by considering two factors. First, it considers the quality assessment band (QA60), which indicates whether there is strong confidence that a cloud/cirrus is present in each pixel. The second considers the reflectance features of clouds and cirrus and uses different spectral bands, such as band 1 (aerosols), band 10 (cirrus), and bands in the visible part (band 2, 3, and 4) to quantify a "cloud score", which indicates the cloudiness of each pixel. Users can manually define the most applicable threshold of the cloud score depending on the study area under consideration. We set the threshold to 25 for this study, which we identified using visual inspection of the cloud mask overlaid over the original Sentinel-2 true color composite image on GEE. Every pixel that has a cloud score larger than 25 was masked out (credit to Ian W. Housman, GEE; relevant code can be found on the website of the GEE help forum: https://groups.google.com/d/msg/google-earth-engine-developers/).

The full band and index information is shown in Table 2. Among all available bands of Sentinel-2 images, we removed bands B1, B9, and B10 from our models. This is because these bands represent aerosols, water vapor, and cirrus, and are developed mainly for the purposes of tracking or estimating atmospheric features and do not measure the actual surface reflectance of land features. We used the built-in function *ee.Image.sampleRegions()* to sample the original pixels and output them with a new projection of 10 m scale for each band. Apart from the selected original bands, two spectral indices were calculated as supplementary measurements for identifying ZT fields according to previous studies, which are the normalized difference tillage index (NDTI) [21] and the crop residue cover index (CRC) [22]. The normalized difference vegetation index (NDVI) was also added for identifying phenology features [5,11,23].

Table 2. Bands and indices information of Sentinel-2 (S2).

	S2	**Description (Wavelength: μm)**
Band	~~*B1*~~	~~*Aerosols*~~
	B2	Blue (0.490)
	B3	Green (0.560)
	B4	Red (0.665)
	B5	Red Edge 1 (0.705)
	B6	Red Edge 2 (0.740)
	B7	Red Edge 3 (0.783)
	B8	NIR (0.842)
	B8A	Red Edge 4 (0.865)
	~~*B9*~~	~~*Water vapor*~~
	~~*B10*~~	~~*cirrus*~~
	B11	SWIR 1 (1.610)
	B12	SWIR 2 (2.190)
Index	NDVI	(NIR – Red) / (NIR + Red)
	NDTI	(SWIR1 – SWIR2) / (SWIR1 + SWIR2)
	CRC	(SWIR1 − Green) / (SWIR1 + Green)

3.3. Landsat

The United States Geological Survey (USGS) Landsat SR Tier 1 product was acquired on the GEE platform in the form of an image collection. We used both the 30m Landsat 8 Operational Land Imager (OLI) and the Landsat 7 ETM. Roy et al. [24] states that, apart from the improvements

of Landsat 8 OLI on the signal-to-noise characteristics and slightly narrower wavelengths of bands compared to its predecessors, the data of all Landsat products are expected to be consistent. Thus, we merged the image collections of Landsat 8 OLI and Landsat 7 ETM+ into one collection in GEE using the built-in function *ee.ImageCollection.merge()*.

To remove cloud cover, we applied a cloud mask that uses the quality assessment band (pixel_qa) in Landsat products. This quality band stores information about the probability that a given pixel is cloud or a cloud shadow and can be used to remove contaminated pixels [7]. Relevant model code can be found on the website of the GEE developer's guide (https://developers.google.com/earth-engine/).

Band and index information is shown in Table 3. We did not include in our final model bands that represent Brightness Temperature (B6 of Landsat 7; B10 & B12 of Landsat 8) and Ultra blue (Landsat 8 B1), as these bands store atmospheric features. We used the built-in function *ee.Image.sampleRegions()* to convert the image to a projection of 10 m for each band in order to use this data along with Sentinel-2 bands in our final combined model. We also calculated the same three spectral indices that were calculated for Sentinel-2 – NDTI, CRC, and NDVI.

Table 3. Bands and indices information of Landsat 7 (L7) and Landsat 8 (L8).

	L7	**Description (Wavelength: µm)**	**L8**	**Description (Wavelength: µm)**
			~~*B1*~~	~~*ultra blue*~~
	B1	Blue (0.45 - 0.52)	B2	Blue (0.45 - 0.51)
	B2	Green (0.52 - 0.60)	B3	Green (0.53 - 0.59)
	B3	Red (0.63 - 0.69)	B4	Red (0.64 - 0.67)
Band	B4	NIR (0.77 - 0.90)	B5	NIR (0.85 - 0.88)
	B5	SWIR 1 (1.55 - 1.75)	B6	SWIR 1 (1.57 - 1.65)
	B7	SWIR 2 (2.08 - 2.35)	B7	SWIR 2 (2.11 - 2.29)
	B6	~~*Brightness Temperature*~~	~~*B10*~~	~~*Brightness Temperature*~~
			~~*B11*~~	~~*Brightness Temperature*~~
	NDVI			(NIR – Red) / (NIR + Red)
Index	NDTI			(SWIR1 – SWIR2) / (SWIR1 + SWIR2)
	CRC			(SWIR1 − Green) / (SWIR1 + Green)

3.4. Sentinel-1

We also used dual-polarization C-band (operating at 5.4 cm wavelength) SAR Sentinel-1 images, which are available and pre-processed by GEE using these steps: 1) thermal noise removal; 2) radiometric calibration; 3) and terrain correction [25]. The backscatter values (σ^0) were converted to decibels (dB) with logarithmic transformation:

$$\sigma^0{}_{dB} = 10 \times \log_{10}(\sigma^0)$$

We filtered the Interferometric Wide (IW) swath mode, which provides polarizations VV (vertical transmit/vertical receive) and VH (vertical transmit/horizontal receive) from both the ascending and descending orbits. The ascending or descending orbit images have different local incidence angles, which leads to differences in backscatter intensity. Considering the topography of the central Mexico highland, we extracted images from both orbits and used them in separate image collections [26]. A filter for Sentinel-1 tile edge was performed for the removal of the "gap" between images [27]. In addition, we used the incident angle correction filter [28]. We finally applied a simple speckle filter with a 3 × 3 pixel kernel [29]. All above filters were applied after

logarithmic transformation from $\sigma^0{}_{dB}$ to the original σ^0, and were then transformed back to $\sigma^0{}_{dB}$ (credit to Guido Lemoine, GEE; relevant code can be found on the website of the GEE help forum: https://groups.google.com/d/msg/google-earth-engine-developers/).

Sentinel-1 bands and the index are shown in Table 4. The intensity cross-ratio (CR) VV/VH was added to improve classification accuracies, since Vreugdenhil et al. [30] found that CR is helpful for differentiating vegetation types and other ground features. We processed the data at 10 m resolution so that it could be used in the combined sensor model.

Table 4. Bands and indices information of Sentinel-1 (S1).

	S1	**Description**
Band	VV	Single co-polarization, vertical transmit/vertical receive
	VH	Dual-band cross-polarization, vertical transmit/horizontal receive
CR	VV/VH	Log ratio of (VV/VH)

4. Methodology

4.1. Image Compositing

The study area has limited image availability due to high cloud cover during the rainy season in central Mexico. Thus, even if we create monthly (30-day) image composites, there are many gaps in data availability due to cloud cover, mostly from February 1st to March 1st and from May 1st to Sep 1st. Thus, setting reasonable time ranges for image compositing is critical for the study to ensure that there are limited missing pixels due to cloud cover. Since it is likely that CT and ZT are most distinguishable during the sowing period, we created one mosaic during the sowing period. To identify the ideal sowing date range, a phenology curve was made using averaged values from 400 ZT pixels and 400 CT pixels that were randomly selected from our full polygon dataset (Figure 3). This image composite spanned the dates April 1st to July 31st 2017, and captures the widest possible trough periods on the phenology curve, including the intensive sowing period in Guanajuato around May. In addition, given that previous studies have shown that full season composites can classify ZT versus CT fields, we also included an image composite for the second half of the season, which we call the "Peak" season (Figure 3). This second 3-month composite spanned from August 1st to October 31st 2017 and captured the peak and the end of the growing season.

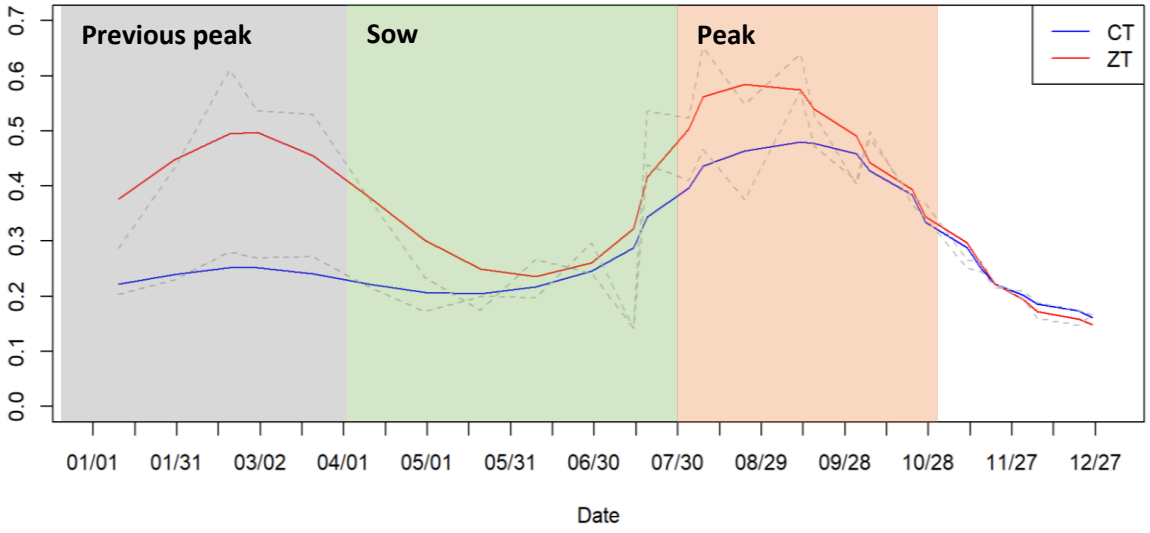

Figure 3. Phenology curves generated from averaged NDVI values of 400 random points in 2017. Timing of the sow image composite is highlighted in green and of the peak image composite is highlighted in orange.

4.1.1. Metric compositing

In order to create the image composites, we must define how pixels for each band will be selected throughout the time period. Following methods from Azzari et al. [5], we chose to apply a percentile compositing method and selected 5 metrics [0, 25, 50, 75, 100] % using the percentile reducer function in GEE for all bands and indices from every sensor. For example, consider the sowing season image composite for Sentinel-2. For each band and index, we took the exact 0%, 25%, 50%, 75%, and 100% values from a histogram of all values across all images throughout the image compositing time period (April 1 to July 31). The percentile metric can be generally applied to any set of images through time, regardless of whether or not there are data gaps due to cloud cover on certain dates as occurred in this study. As shown in Table 5, the use of the percentile compositing method expands the number of features considered in our model [5].

Table 5. Statistics of bands and indices. Showing double composites only (model-1 & 4).

Sensor	Original No. of bands & indices	Feature No. after compositing	Feature Components
Sentinel 2	13	130	(10 bands + 3 indices) × 2 seasonal composites × 5 percentiles
Landsat 7 & 8	9	90	(6 bands + 3 indices) × 2 seasonal composites × 5 percentiles
Sentinel 1	3	60	(2 bands + 1 index) × 2 orbits × 2 seasonal composites× 5 percentiles

4.2. Pixel-Based Random Forest Classification

For the classification done in this study, we used the machine learning algorithm random forest (RF), which is an ensemble supervised classifier based on the classification and regression

tree (CART). With this method, different decision trees come from bootstrap samples, and each tree node is assigned a limited subset of variables. The final class is then selected by majority voting [31]. For classification models, 70% of the field polygons were randomly selected for training, and the remaining 30% of polygons were kept for validation, with balanced ZT and CT labels in both the datasets [5,31]. We found that only 11 polygons contain less than 20 10 x 10 m pixels; thus, for each single polygon, we randomly selected no more than 20 pixels at random. Such an approach both reduces the computational time needed to run our models and also leads to a more balanced number of ZT and CT pixels being used for classification and training. Note that the final total pixel numbers in the models could be slightly different after the data fusion of multi-source sensors/sensor combinations.

Processing, tuning, and accuracy assessments of the models were performed in R statistical software v 3.6.1 [33]. We used the default values for RF parameters: $n_{tree} = 500$, since the increasing of trees has negligible effect on the classification; $m_{try} = \sqrt{p}$, where p equals the number of features in the target dataset. All RF classifier operations were done using the package "randomForest" [34]. Although the RF classifier is insensitive to multicollinearity, we still tested and performed a feature selection process [31,32]. We created a correlation matrix, and a cut-off correlation coefficient of higher than 0.9 was used to remove highly correlated features resulting in better performance and a reduction of feature dimensionality. This process was done using the package "caret" [35]. RF also produces measures of feature importance. The importance of a feature is recorded in the RF classifier through the measurement of mean decrease in the model accuracy with the removal of that feature.

4.3. Models

We created 14 different random forest models to identify which sensor and/or sensor combinations and which composite timing resulted in the highest accuracies. Specifically, we created models for each individual sensor, Sentinel 1, Sentinel 2, and Landsat, and we also created models that included all sensor combinations. In addition, for each individual sensor and sensor combination, we created models that only used data from the sowing season image composite (model 1), and one that used data from both the sowing and peak season composites (model 2). We did this to better understand how effective using only early season imagery could be in detecting zero tillage compared to using the full growing season.

5. Results

5.1. Accuracy Assessments

As shown in Table 6, model 2 had the highest classification accuracies when using single sensors (72.95% for Sentinel 2, 69.27% for Landsat, and 68.25% for Sentinel 1). Accuracies were 3% higher for Sentinel-2, 5% higher for Sentinel-1, and 0.5% higher for Landsat for model-2 than model-1, suggesting that the model that considered both the sowing and peak seasons performed better than only considering the sowing season. Although the classification model using Sentinel-2 data performed similarly to the model using Landsat data for model 1, the accuracy of the Sentinel-2 model is 3% higher than the models using Landsat or Sentinel-1 for model-2.

Table 6. Accuracies from single sensors and two composite models.

Model No.	Composite	Single Sensor	Overall Accuracy (%)
1	Sow (4-month single composite)	S2	69.36
		S1	63.50
		L8_7	68.71
2	Sow + Peak (4- + 3-month double composites)	S2	72.95
		S1	68.25
		L8_7	69.27

Note: S2 = Sentinel-2; S1 = Sentinel-1; L8_7 = Merged data of Landsat 8 & Landsat 7

Table 7. Accuracies from sensor combination scenarios with two composite models.

Model No.	Composite	Sensor Combination	Overall Accuracy (%)
1	Sow (4-month single composite)	S2 + S1	77.04
		S1 + L8_7	76.15
		S2 + L8_7	77.52
		S2 + S1 + L8_7	81.36
2	Sow + Peak (4- + 3-month double composites)	S2 + S1	79.34
		S1 + L8_7	81.19
		S2 + L8_7	82.82
		S2 + S1 + L8_7	85.96

Note: S2 = Sentinel-2; S1 = Sentinel-1; L8_7 = Merged data of Landsat 8 & Landsat 7

Table 7 indicates that the use of sensor combinations is generally more accurate than the use of a single sensor among models. The overall accuracies of model-2 are still higher than model-1 across sensor combinations, similar as what we showed in Table 6. Improvement among models, compared to the best performing single sensor model (Sentinel 2) can be found in Table 8.

The improved accuracies when combining Sentinel-2 with Sentinel-1, Landsat 8 & 7, and Sentinel-1 + Landsat 8 & 7 are similar, with an increase of around 7%, 9%, and 13%, respectively. From these figures we can also see that the improvements from combining different sensors are different – the combination of Sentinel-2 with Landsat seems to enhance the model performance more than adding Sentinel-1 data.

The highest classification accuracy came from the model that used all three sensors and model-2 (multi-month double composites), with an overall accuracy of 85.96%. More detailed pixel-wise confusion matrices and producer/user accuracies can be found in Table 9. The producer/user accuracies of Sentinel-2 and its combinations ranged from 77.58% to 84.72%, and 70.81% to 86.90% respectively. The sensor combination of Sentinel-2 and Sentinel-1 had the lowest producer accuracy of 77.58%, which indicates the highest omission error.

Table 8. Accuracy improvements among models compared to the single sensor Sentinel-2 model

Basis Sensor	vs.	Sensor combination	Improved accuracy (%) among models	
			Model-1	Model-2
S2	vs.	S2 + S1	+ 7.68	+ 6.39
S2	vs.	S2 + L8_7	+ 8.16	+ 9.87
S2	vs.	S2 + S1 + L8_7	+ 12.00	+ 13.51

Note: S2 = Sentinel-2; S1 = Sentinel-1; L8_7 = Merged data of Landsat 8 & Landsat

Table 9. Confusion matrices with producer/user accuracies using model-2 with Sentinel-2, Sentinel-1, and Landsat 7 & 8 sensor combinations.

Sensors/Sensor Combinations	**Classification (RF)**	**Reference (Ground Truth Tillage Types)**			
		CT	ZT	Row	User Accuracy (%)
	CT	3204	1423	4627	69.25
S1	ZT	1675	3454	5129	67.34
	Column	4879	4877	9756	
	Producer Accuracy (%)	65.67	70.82		***68.25***
		CT	ZT	Row	User Accuracy (%)
	CT	3231	1044	4275	75.58
S2	ZT	1540	3736	5276	*70.81*
	Column	4771	4780	9551	
	Producer Accuracy (%)	67.72	*78.16*		**72.95**
		CT	ZT	Row	User Accuracy (%)
	CT	3200	1320	4520	70.80
L8_7	ZT	1687	3577	5264	67.95
	Column	4887	4897	9784	
	Producer Accuracy (%)	65.48	73.04		**69.27**
		CT	ZT	Row	User Accuracy (%)
	CT	3650	1009	4659	78.34
S2 + S1	ZT	850	3491	4341	*80.42*
	Column	4500	4500	9000	
	Producer Accuracy (%)	81.11	*77.58*		**79.34**
		CT	ZT	Row	User Accuracy (%)
	CT	3764	879	4643	81.07
S1 + L8_7	ZT	864	3761	4625	81.32
	Column	4628	4640	9268	
	Producer Accuracy (%)	81.33	81.06		**81.19**
		CT	ZT	Row	User Accuracy (%)
	CT	3681	717	4398	83.70
S2 + L8_7	ZT	835	3800	4635	*81.98*
	Column	4516	4517	9033	
	Producer Accuracy (%)	81.51	*84.13*		**82.82**
		CT	ZT	Row	User Accuracy (%)
	CT	3793	666	4459	85.06
S2 + S1 + L8_7	ZT	557	3694	4251	*86.90*
	Column	4350	4360	8710	
	Producer Accuracy (%)	87.20	*84.72*		**85.96**

Note: S2 = Sentinel-2; S1 = Sentinel-1; L8_7 = Merged data of Landsat 8 & Landsat 7

5.2. Feature Importance

Table 10 shows the features that ranked in the top five of all features considered for each sensor and sensor combination with model-1. Interestingly, considering which features were most important in the sensor combination groups, we see that instead of Sentinel-2 features, features from Landsat data ranked higher, with CRC ranked as the most important variable in the full sensor model. Table 11 shows the features that ranked in the top five of all features considered for each sensor and sensor combination with model-2. Sentinel-1 features were the most important when

looking at the full sensor model, with VV polarization at the 100th percentile ranking as the most important variable. In addition, the NDTI from Sentinel-2 and the CRC from both Landsat and Sentinel-2 are also important features in the full sensor combination model.

Table 10. Top five important features of different sensors and sensor combinations with model-1 (sowing composite).

Important features rank	S2	S1	L8_7	S2 + S1	S2 + L8_7	S1 + L8_7	S2 + S1 + L8_7
1	B2_p100 _S2	LogVVVH_p0 _asc_S1	CRC_p100 _L	VV_p100 _desc_S1	SWIR2_p100 _L	VV_p100 _desc_S1	CRC_p0 _L
2	B8A_p0 _S2	VV_p25 _desc_S1	R_p100 _L	B2_p100 _S2	CRC_p0 _L	R_p100 _L	VH_p100 _desc_S1
3	B5_p75 _S2	VV_p100 _asc_S1	NIR_p0 _L	VH_p100 _desc_S1	B2_p100 _S2	VH_p100 _desc_S1	NIR_p100 _L
4	B2_p75 _S2	VH_p100 _asc_S1	CRC_p0 _L	B2_p75 _S2	NDVI_p25 _L	CRC_p100 _L	SWIR2_p100 _L
5	B11_p100 _S2	VH_p0 _asc_S1	G_p75 _L	VH_p75 _desc_S1	B7_p0 _S2	B_p100 _L	VV_p100 _desc_S1

Table 11. Top five important features of different sensors and sensor combinations with model-2 (sowing + peak composites).

Important features rank	S2	S1	L8_7	S2 + S1	S2 + L8_7	S1 + L8_7	S2 + S1 + L8_7
1	B12_p0 _peak_S2	VV_p100 _asc_sow_S1	NIR_p0 _peak_L	VH_p100 _asc_peak_S1	CRC_p50 _ peak_L	VV_p100 _desc_sow_S1	VV_p100 _asc_sow_S1
2	NDVI_p75 _peak_S2	VV_p100 _asc_peak_S1	CRC_p75 _peak_L	VV_p100 _asc_sow_S1	CRC_p25 _ peak_L	VV_p100 _asc_sow_S1	VV_p100 _asc_peak_S1
3	B7_p50 _peak_S2	VH_p100 _asc_peak_S1	NIR_p25 _peak_L	VV_p100 _desc_ sow_S1	B2_p25 _ peak_S2	VV_p75 _desc_sow_S1	NDTI_p75 _sow_S2
4	B6_p100 _peak_S2	LogVVVH_p100 _desc_sow_S1	NDVI_p100 _peak_L	VV_p100 _asc_ peak_S1	G_p0 _sow_L	VV_p50 _desc_sow_S1	B2_p50 _ peak_S2
5	B12_p25 _peak_S2	VV_p50 _desc_sow_S1	B_p100 _sow_L	LogVVVH_p100 _asc_ peak_S1	B8A_p100 _peak_S2	NIR_p0 _peak_L	G_p100 _sow_L

Note: Codes can be interpreted as the band number, percentile, composite, and sensor separated by underscores. For example, "**VV_p100_asc_peak**" means the specific pixel value for VV band was obtained at the 100th percentile in the peak season from the ascending orbit of Sentinel 1. The same format is used with other sensors.

6. Discussion

This study examined which satellite sensors and sensor combinations along with image composite timing led to the greatest classification accuracies for mapping ZT fields in Central Mexico in 2017. We created two different seasonal compositing models, one that focused on only the sowing period and one that focused on the sowing period and the peak growing period, with a percentile metric compositing method. Considering sensors, we used data from two optical satellite sensors, Sentinel-2 and Landsat, and from one SAR sensor, Sentinel-1. Our results illustrate that Sentinel-2 results in the highest classification accuracies if only one sensor is used, however, the highest accuracies are achieved when all three sensors are combined. Considering image timing, we find that using images from only the sowing period have relatively similar accuracies to using images from the entire growing season. Overall, our results show that multi-source remote sensing images can classify ZT and CT agriculture reliably at a large spatial scale.

Considering which sensor and sensor combinations resulted in the highest accuracies, we found that single sensor models have an accuracy of around 70%, with the model that used Sentinel-2 data having the highest classification accuracy (Table 6). There are several reasons why Sentinel-2 may perform better than Landsat. First, the relatively coarse spatial resolution of Landsat data (30 m) compared to that of Sentinel-2 data (10 m) may lead to spectral mixture of classes, which can reduce classification accuracy [37]. Second, Sentinel 2 includes several more bands than Landsat, including bands 5, 6, 7 and 8A in the red-edge, which have been shown to lead to increased classification accuracy of vegetation in other studies [37].

For multiple sensor models, both optical + SAR satellite combinations (S2 + S1; L8_7 + S1) performed similarly. Though Sentinel-1 performed poorly in single sensor analyses, it improved accuracies in multiple sensor analyses, consistent with the result from Denize et al. [14]. Denize et al. [14] finds that SAR images can distinguish between bare soil and crop residues well, which likely explains one reason why Sentinel-1 adds important information to our classification models. In addition, given that the maize crops in this study are planted during the rainy season when there is a significant amount of cloud cover, it is possible that Sentinel-1 improves accuracies by providing information during periods of the growing season covered by clouds, when optical sensors do not provide valuable information.

Considering image timing, the highest accuracies occurred when using the composite of the whole growing season (model-2). Previous work has shown that considering whole-season composites can lead to high classification accuracy of tillage practices [5]. Specifically, Azzari et al. [5] used four-season composites with Landsat and Sentinel-1 data. They found an overall accuracy of 79% for their best model for the classification of low-intensity tillage, which is similar to the accuracy found in our study (81.19%) using a similar full season model with Sentinel-1 and Landsat. Interestingly, although model-1 considered images only during the sowing window, the overall accuracies were only slightly lower than the full growing season model. Despite dropping 50% of the available features, the accuracy only decreased by less than 5.3% across sensors. Although not using image composites, studies from Zheng et al. [36] and Robles et al. [17] focused on only using images during the tillage and sowing time window of their study area, and this approach resulted in high accuracies. Though the models using only data from the sowing season have lower accuracy than those that use data from the whole season, doing so may be desirable in cases where images are only available during the sowing season, or one wants to classify ZT and CT during the middle of the growing season. Also, using data from only one season is more computationally friendly which could be helpful if doing an analysis across large study areas and time-spans (Table 5).

Interestingly, when examining feature importance of model-1, features from Landsat data are generally more important, even though Sentinel-2 is the sensor with the highest single sensor classification accuracy (Table 10). One possible reason for the greater importance of Landsat variables could be there is more Landsat data available during the sowing time window compared to Sentinel-2. Due to the variability in cloud cover across days, different revisit dates and times for Landsat and Sentinel-2 sensors result in possible coverage availability differences [37]. The feature importance table of model-2 suggests that Sentinel-1 bands and indices are helpful in increasing classification accuracies – Sentinel-1 features represent the top five most important features in models with sensor combinations when looking at the multi-sensor combinations of Sentinel-2 and Sentinel-1 and Landsat and Sentinel-1; in particular, the maximum VV polarization variable is most important (Table 11). This is consistent with current studies that use backscattering polarization VV or VH, and also the cross ratio to separate crop residue from other land features

[14]. However, Azzari et al. [5] shows that none of the Sentinel-1 features were considered important by the RF classifier, and concluded Sentinel-1 was not helpful for classification. This may be because of incident angle differences and speckle noise issues that were difficult to control for since the study was conducted across extremely large spatio-temporal scales.

Considering limitations of our study, first, tillage practices were labeled only as ZT and CT and not tillage intensity; other studies which focus on tillage classification usually quantify the intensity of tillage according to the CRC [5,13]. Future work should examine whether our methods could similarly detect tillage intensity in our study region. Second, our study only was conducted in the state of Guanajuato; future work should examine whether we would achieve similar results and accuracies in different states across Mexico. Additionally, our study only focused on a single year, 2017. Future work would benefit from developing multi-year models. Finally, in consideration of the improved accuracies from adding Sentinel-1 data, future work should consider the use of other SAR sensors, such as Cosmo-SkyMed, to further test the effectiveness of SAR data [38].

7. Conclusion

This study used data from multi-source optical satellites, Sentinel-2 and Landsat 8 & 7, and a SAR satellite, Sentinel-1, to identify tillage practices in the state of Guanajuato, Mexico in 2017. Considering individual sensors, Sentinel-2 had the highest classification accuracies (72.95%) across composites. When considering multiple sensor combinations, SAR sensor Sentinel-1 improved classification accuracy when used as a complement to optical sensors. Overall, the full sensor combination, considering Sentinel-2, Landsat, and Sentinel-1 data, across the full growing season performed the best, with an overall accuracy of 85.96%. For image timing, we found that the whole-season composite (model-2) led to the highest accuracies, though the model that used only data from the sowing season (model-1) has relatively high accuracies. These results illustrate that easily and freely-accessible global remote sensing images can be used to classify tillage practices at large spatio-temporal scales.

References

1. Poswal, R. Accelerating adoption of zero tillage technology. Indian Research Journal of Extension Education. **2007**, *7*. 6-10.
2. Dendooven, L.; Patiño-Zúñiga, L.; Verhulst, N.; Luna-Guido, M.; Marsch, R.; Govaerts, B. Global warming potential of agricultural systems with contrasting tillage and residue management in the central highlands of Mexico. *Agriculture, Ecosystems & Environment* **2012**, *152*, 50–58.
3. Six, J.; Ogle, S. M.; Breidt, F. J.; Conant, R. T.; Mosier, A. R.; Paustian, K. The potential to mitigate global warming with no-tillage management is only realized when practised in the long term. *Global Change Biology* **2004**, *10*, 155–160.
4. Honsdorf, N.; Verhulst, N.; Crossa, J.; Vargas, M.; Govaerts, B.; Ammar, K. Durum wheat selection under zero tillage increases early vigor and is neutral to yield. *Field Crops Research* **2020**, *248*, 107675.
5. Azzari, G.; Grassini, P.; Edreira, J. I. R.; Conley, S.; Mourtzinis, S.; Lobell, D. B. Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. *Remote Sensing of Environment* **2019**, *221*, 417–429.
6. Quemada, M.; Hively, W. D.; Daughtry, C.; Lamb, B.; Shermeyer, J. Improved Crop Residue Cover Estimates from Satelite Images by Coupling Residue and Water Spectral Indices. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* **2018**.
7. Tian, F.; Wu, B.; Zeng, H.; Zhang, X.; Xu, J. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth Engine Cloud Platform. *Remote Sensing* **2019**, *11*, 629.
8. Baghdadi, N.; Zribi, M.; Loumagne, C.; Ansart, P.; Anguela, T.P. Analysis of TerraSAR-X Data and Their Sensitivity to Soil Surface Parameters over Bare Agricultural Fields. Remote Sens. Environ. **2008**, *112*, 4370–4379.
9. Zribi, M.; Gorrab, A.; Baghdadi, N. A new soil roughness parameter for the modelling of radar backscattering over bare soil. *2014 IEEE Geoscience and Remote Sensing Symposium* **2014**.
10. Abdikan, S.; Sanli, F.B.; Ustuner, M.; Calò, F. Land Cover Mapping Using Sentinel-1 SAR Data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **2016**, 41, 757.
11. Zheng, B.; Campbell, J. B.; Beurs, K. M. D. Remote sensing of crop residue cover using multi-temporal Landsat imagery. *Remote Sensing of Environment* **2012**, *117*, 177–183.
12. Hadria, R.; Duchemin, B.; Baup, F.; Toan, T. L.; Bouvet, A.; Dedieu, G.; Page, M. L. Combined use of optical and radar satellite data for the detection of tillage and irrigation operations: Case study in Central Morocco. Agricultural Water Management **2009**, *96*, 1120–1127.
13. Beeson, P. C.; Daughtry, C. S. T.; Hunt, E. R.; Akhmedov, B.; Sadeghi, A. M.; Karlen, D. L.; Tomer, M. D. Multispectral satellite mapping of crop residue cover and tillage intensity in Iowa. *Journal of Soil and Water Conservation* **2016**, *71*, 385–395.
14. Denize, J.; Hubert-Moy, L.; Betbeder, J.; Corgne, S.; Baudry, J.; Pottier, E. Evaluation of Using Sentinel-1 and -2 Time-Series to Identify Winter Land Use in Agricultural Landscapes. *Remote Sensing* **2018**, *11*, 37.
15. Produce Guanajuato AC Foundation. Producción de trigo: integradora agropecuaria del centro s.a. de c.v. **2010**.

16. Romero-Perezgrovas, R.; Verhulst, N.; Rosa, D. D. L.; Hernández, V.; Maertens, M.; Deckers, J.; Govaerts, B. Effects of Tillage and Crop Residue Management on Maize Yields and Net Returns in the Central Mexican Highlands Under Drought Conditions. *Pedosphere* **2014**, *24*, 476–486.
17. Robles, N. S.; Segoviano, A. Z.; Sixto, J. M. C.; Serwatowski, R.; Barrón, S. G.; Vaca, C. G.; Ortega, A. F.; García, Á. F. Discriminación de la superficie agrícola sembrada bajo labranza de conservación empleando análisis multitemporal. *Revista Mexicana de Ciencias Agrícolas* **2012**, *3*.
18. Robles, N. S.; Juárez, E. M.; Vaca, C. G.; García, Á. F.; Serwatowski, R.; Ortega, A. F. Monitoreo satelital de la adopción de Agricultura de Conservación. *Revista Ciencias Técnicas Agropecuarias* **2013**, *22*.
19. ESRI. ArcGIS Desktop: Release 10. Redlands, CA: *Environmental Systems Research Institute*. **2011**.
20. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27.
21. van Deventer, A. P.; Ward, A. D.; Gowda, P. H.; Lyon, A.J.G.. Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices. 2006 ,pp. 1–7.
22. Sullivan, D. G.; Truman, C. C.; Schomberg, H. H.; Endale, D. M.; Strickland, T. C. Evaluating Techniques for Determining Tillage Regime in the Southeastern Coastal Plain and Piedmont. *Agronomy Journal* **2006**, *98*, 1236–1246.
23. Carlson, T. N.; Ripley, D. A. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment* **1997**, *62*, 241–252.
24. Roy, D.; Kovalskyy, V.; Zhang, H.; Vermote, E.; Yan, L.; Kumar, S.; Egorov, A. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment* **2016**, *185*, 57–70.
25. GEE, 2019. Official Google Earth Engine User Guide. https://developers.google.com/ earth-engine/sentinel1, Accessed: 2019.
26. Sabel, D.; Bartalis, Z.; Wagner, W.; Doubkova, M.; Klein, J.-P. Development of a Global Backscatter Model in support to the Sentinel-1 mission design. *Remote Sensing of Environment* **2012**, *120*, 102–112.
27. D'Andrimont, R.; Lemoine, G.; Velde, M. V. D. Targeted Grassland Monitoring at Parcel Level Using Sentinels, Street-Level Images and Field Observations. *Remote Sensing* **2018**, *10*, 1300.
28. Delancey, E. R.; Kariyeva, J.; Bried, J. T.; Hird, J. N. Large-scale probabilistic identification of boreal peatlands using Google Earth Engine, open-access satellite data, and machine learning. *Plos One* **2019**, *14*.
29. Lee, J.-S.; Lee, J.-S.; Wen, J.-H.; Ainsworth, T.; Chen, K.-S.; Chen, A. Improved Sigma Filter for Speckle Filtering of SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2009**, *47*, 202–213.
30. Vreugdenhil, M.; Wagner, W.; Bauer-Marschallinger, B.; Pfeil, I.; Teubner, I.; Rüdiger, C.; Strauss, P. Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study. *Remote Sensing* **2018**, *10*, 1396.
31. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *114*, 24–31.

32. Duro, D. C.; Franklin, S. E.; Dubé, M. G. 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* **2012**, *118*, 259–272.
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria **2019**.
34. Liaw, A.; Wiener, M. Classification and Regression by randomForest. R News. **2002**.
35. Kuhn, M. The caret package. J. Stat. **2008**.
36. Zheng, B.; Campbell, J. Multitemporal remote sensing of crop residue cover and tillage practices: A validation of the minNDTI strategy in the United States. *Journal of Soil and Water Conservation* **2013**, *68*, 120-131.
37. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest Type Identification with Random Forest Using Sentinel-1A, Sentinel-2A, Multi-Temporal Landsat-8 and DEM Data. *Remote Sensing* **2018**, *10*, 946.
38. Gorrab, A.; Zribi, M.; Baghdadi, N.; Mougenot, B.; Chabaane, Z. Potential of X-Band TerraSAR-X and COSMO-SkyMed SAR Data for the Assessment of Physical Soil Parameters. *Remote Sensing* **2015**, *7*, 747–766.