The Identification of Irrigated Crop Types Using Support Vector Machine, Random Forest and Maximum Likelihood Classification Methods with Sentinel-2 Data in 2018: Tashkent Province, Uzbekistan

Erdanaev, E.,* Kappas, M. and Wyss, D.
Cartography, GIS and Remote Sensing Department, Institute of Geography, University of Göttingen, Goldschmidt Street 5, 37077 Göttingen, Germany
E-mail: elbek.erdanaev@geo.uni-goettingen.de,* mkappas@uni-goettingen.de, daniel.wyss@uni-goettingen.de
Corresponding Author

DOI: https://doi.org/10.52939/ijg.v18i2.2151

Abstract

Accurately mapping land use and land cover including agricultural use and the state of crops at various stages is important to address specific agro-ecological challenges, to implement sustainable agricultural practices, and monitor crops periodically. This study aims to provide a timely and accurate main irrigated crop types mapping at 10m resolution for Tashkent province based on multi-temporal Sentinel-2 data acquired for the growing season in 2018. This paper shows the potential use of multitemporal Sentinel-2 satellite data to derive an up-to-date irrigated crop types classification map of the study area. As single-date satellite imagery does not allow proper cropland classification, multitemporal and high-resolution Sentinel-2 data was used to capture small cropland fields and specific crop types for the vegetation period (April to October 2018). NDVI monthly profiles of crop types as well as additional 10 m resolution bands 2 and 3 were used as input data to perform and assess three classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Maximum Likelihood Classification (MLC). Accuracy assessment results showed that SVM showed the highest Overall Accuracy (OA) and Kappa Accuracy (KA). KA of classified images for SVM were 0.90 and 0.89 for the RF algorithm. Both performed well with close values. But MLC showed a lower result of KA 0.60. The paper also compares the area of derived irrigated cropland area with data from the State Committee for Statistics of Uzbekistan for selected crop types. Values for the crops "cotton" and "wheat" derived by SVM and RF methods show a high correlation with the provided statistical data. Based on the results, the SVM classification method is recommended for further mapping and monitoring of irrigated crop types in the region when Sentinel-2 data is used.

1. Introduction

Detailed and appropriate land use and land cover (LULC) maps are of high importance for a variety of sectors in the developing world including food security, land use planning, water, and natural resources management decisions (Saah et al., 2019). Nowadays remote sensing (RS) imagery together with geoinformation systems (GIS) tools is widely used for LULC mapping. Medium and coarse resolution RS imagery provide general LULC classifications but cannot distinguish specific LUs i.e. on crop level due to similar reflectance values (Jia et al., 2018). Therefore, freely available high spatial resolution RS imagery such as Sentinel-2 is necessary to derive timely and accurate information on irrigated croplands and to better understand the impact of spectral-temporal properties of various crop types allowing control and monitoring of irrigated croplands over larger areas.

Uzbekistan is one of the Central Asian (CA) countries and is located at the heart of the Eurasian continent (Liu, 2011). Climate change in CA countries is expected to strongly impact agricultural productivity and food security in the region. Temperatures are predicted to increase by 1.7 °C and 2.6 °C and precipitation to rise by 9 % and 12 % under global warming scenarios of 1.5 °C and 2.0 °C, respectively (Li et al., 2020).

Besides, population increase in CA is leading to higher consumption of natural resources, expansion of irrigated croplands, and intensive land-use practices (Erdanaev et al., 2015). Agriculture is one of the main sectors of Uzbekistan's economy with agriculture, forestry, and fisheries contributing 32,40 percent of the country's total GDP in 2018. 53,20% of agricultural products belong to crop production and 46,80% to livestock. Eighty percent of the country is classified as desert or semi-desert. Agricultural land totals 20,26 million hectares including only 3.7 million hectares of irrigated agricultural land in 2018. The major crops are cotton and wheat. The share of Tashkent province towards the total volume of agricultural products, forestry, and fisheries amounts to 15.7 percent, the highest in the country (Zvi Lerman, 2019).

In Uzbekistan, accurate accounting and monitoring of land use in agriculture are very important. Unfortunately, several problems have accumulated in the field of geodesy, cartography, and cadastre in the country. The last land survey was conducted 40 years ago. The procedure for allocating agricultural land was adopted 20 years ago and does not meet modern requirements. This is evidenced by the fact that 150 thousand hectares of arable land were arbitrarily occupied in 66 districts of the state (Mirziyoyev, 2020). Several studies on LULC mapping and change detection have been carried out in Uzbekistan using RS imagery and GIS tools. (Chen et al., 2013) studied land use and land cover change and variations of ecosystem services including net primary productivity (NPP), evapotranspiration (ET), and grain production in CA between 1990 and 2009. They found out that most significant changes were triggered through farmland abandonment and reclamation showing a significant increase of farmland between 2000 and 2009. Farmland NPP was higher than natural vegetation and NPP increased with the rise of temperature in 2000 despite a decline in precipitation. The actual ET in the central area was lower than in the northern and eastern parts of Central Asia.

Irrigated lands in Uzbekistan increased from 2.2 million ha in 1953 to 4.21 million ha in 2013, thus the long-term growth rate was about 1.5 % annually (Lerman, 2018). Expansion and densification of irrigated cropland in Kashkadarya Province of Uzbekistan were studied using classification tree methods based on Landsat MSS and TM data from 1972/73, 1977, 1987, 1998, and 2000. Cropland extent developed from 134.800 ha to 477.000 ha between 1972/73 and 2009 and winter wheat harvesting doubled to approximately 211.000 ha from 1987 to 1998 (Edlinger et al., 2012). Irrigated cropland mapping was performed using pixel-based (PB) and field-based (FB) robust non-parametric machine learning algorithms such as RF, SVM, and a common parametric MLC using multi-temporal Landsat 8 images in Khorezm Province of Uzbekistan. Accuracy assessment results showed higher OA and kappa index (KI) for FB-RF and FB-SVM algorithms over the PB-RF, PB-SVM, and PB-MLC algorithms. The parametric FB-MLC showed the lowest OA and KI (Basukala et al., 2017). Alikhanov et al., (2020) studied LULC change of Tashkent Province between 1992-2018 using Landsat images. They found that grassland, shrubland, meadow, as well as agricultural lands, decreased from 7000km2 in 1992 to 3000km2 in 2008 then increased again in 2018. (Juliev et al., 2019) carried out LULC change detection analysis between 1989-2017 in the Bostanlik District of Tashkent Province, which is located in the mountain area. They used Maximum Likelihood Classification using Landsat data with the resulting classes: snow cover, bare soil-rock, forest, waterbody, built-up areas, and agriculture. The results showed that for the last 28 years, significant changes occurred within classes of the forest, built-up areas, bare soil, and snow cover.

LULC change in the mountain areas of Tashkent Province using Landsat NDVI values as an indicator of land degradation was studied from 1989 to 2018 (Pulatov et al., 2020). Their research results showed that overgrazing had a significant effect on the mountain ecosystem of Tashkent Province with a decrease of 29 thousand ha of pastureland from 1989 until1998, which increased again between 2008 and 2018. Gerts et al., (2020) studied agricultural land-use change in Urta Chirchik district of Tashkent province between 1994 and 2017. The study was based on Spectral Correlation Mapper classification using Landsat and Sentinel-2 NDVI profile analysis. The results showed that the combination of both Landsat and Sentinel-1 radar data for Spectral Correlation Mapper classification increases classification accuracy. We believe that long-term change detection analysis should be performed using similar image acquisition dates or months for each year to consider vegetation phenology and specific climatic conditions. This approach is missing in the above-mentioned research and can result in misclassification and errors in LULC change detection analysis (Alikhanov et al., 2020).

Most of the published data on agricultural land use and land cover change analysis of local experts in Uzbekistan are based on the reported statistics through state organization bodies that are usually outdated and not very reliable and there is little information available on the extent of agricultural LUs. Up to now, no study has provided a deeper understanding of spectral-temporal information derived from Sentinel-2 satellite imagery for crop classification on the provincial level in the study area. Thus, RS approach of irrigated crop types mapping using 10 m resolution bands of Sentinel-2 data can play an important role for an effective land use planning and management which are essential to assure the sustenance of its population while fostering environmental sustainability. To achieve these approach relatively modern and robust classification algorithms such as SVM, RF and MLC can be utilized. The objective of this study aims to provide a timely and accurate irrigated crop types map for Tashkent province based on multi-temporal Sentinel-2 data acquired for the growing season in 2018. For this purpose, three supervised classification algorithms are tested and cross-evaluated. Classification results are compared with available state statistics data on crop types for future applicability and can thus provide an important basis for more sustainable development, future monitoring, planning, and decision-making.

2. Materials and Methods

2.1 Study Area

Tashkent Province was formed in 1938 as part of the Uzbek Soviet Socialist Republic (SSR) and is located in the northeast of Uzbekistan between 40.18 and 42.29 N and 68.64 and 71.27 E of the Greenwich meridian or between the western part of Tien Shan mountains and Syrdarya river. The Province borders with Kazakhstan in the north and north-west, with Kyrgyzstan in the north-east, with Namangan Province in the east, with Tajikistan in the south, and with Syrdarya Province in the south-western part (Figure 1).

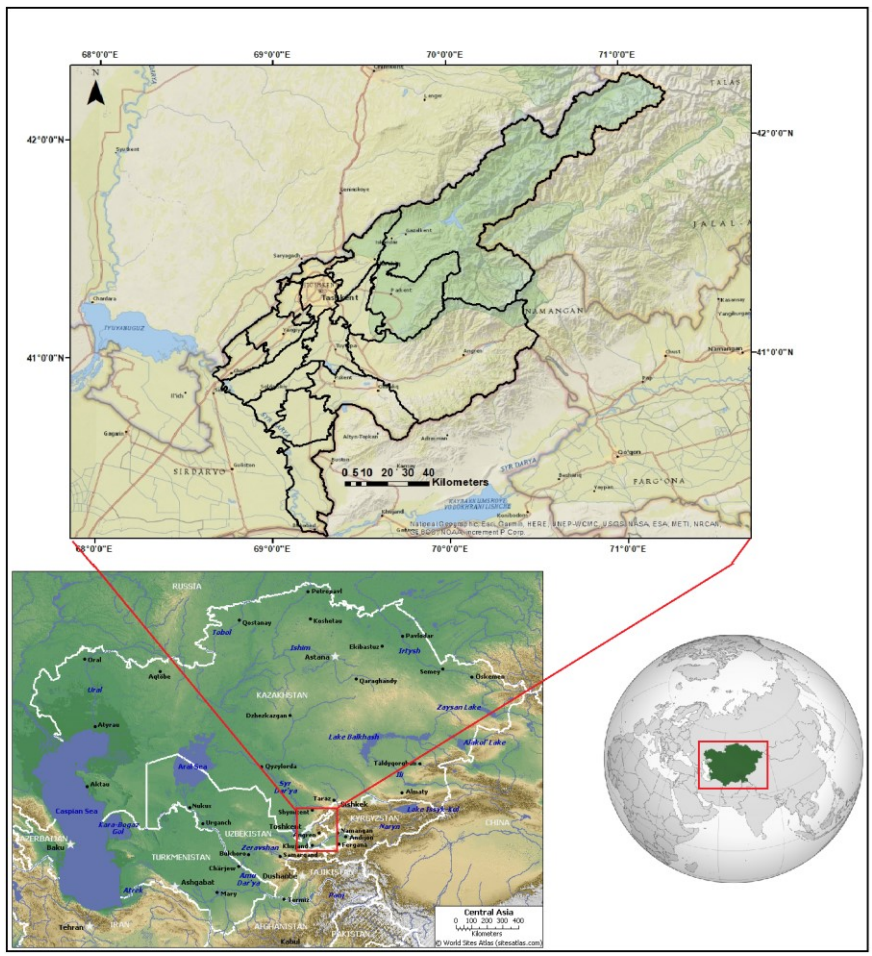

Figure 1: Location of the study area and the administrative districts

Tashkent Province consists of 15 administrative districts and 7 cities with a population of 2.81 million in 2018 (State Statistics, 2019). The climate is a typically continental climate with humid, relatively mild wet winters and long, hot, and dry summers. Mean January temperature is -1°C to -2°C and the mean July temperature is 26.8°C. The average annual precipitation is 300 mm in the plains region, 300-400 mm in the piedmont region, and 500-600 mm in the mountains. Precipitation mostly occurs in the early spring and permanent snow cover is located in the higher mountains. The main river Syrdarya and its tributaries Chirchik and Akhangaron Rivers basins are fed by snow and glaciers and they are used for irrigation and hydroelectric power (Erdanaev et al., 2015). According to statistical data, 50,70% of the Tashkent province's population live in rural areas where most agriculture is practiced. These areas are located in low elevated lands.

2.2 Data Collection and Pre-Processing

The total number of 28 images during crop growing season in 2018 which is freely available Sentinel-2 (Level-1C) multi-spectral satellite images of the study area were downloaded from the Earth Explorer website of the United States Geological Survey (USGS). Band parameters of each spectral band are given in Table 1. These products underwent radiometric and geometric corrections but were not corrected atmospherically (Drusch et al., 2012). Therefore, the images were pre-processed using the Semi-Automatic Classification plugin and the DOS1 (Dark Object Subtraction) correction tool in QGIS (Congedo, 2021) allowing the transformation of Top-Of-Atmosphere (TOA) reflectance to land surface reflectance values which is a basic requirement for multi-temporal image analysis. Generally, Sentinel-2 has temporal resolution of approximately 5 days. But we used only the best cloud-free multi-temporal images for each month with acquisition dates covering the growing season (April to October) were selected for the study area.

Table 1: Band parameter of each spectral band of Sentinel – 2

Band number	Description	Wavelengths (nm)	Spatial Resolution (m)
1	Coastal aerosol	433-453	60
2	Blue	458–523	10
3	Green	543–578	10
4	Red	650–680	10
5	Vegetation Red Edge (RE) 1	698–713	20
6	Vegetation RE 2	733–748	20
7	Vegetation RE 3	773–793	20
8	Near-Infrared (NIR)	785–900	10
8a	Narrow NIR	855–875	20
9	Water vapor	935–955	60
10	Shortwave infrared (SWIR)- Cirrus	1360–1390	60
11	SWIR 1	1565–1655	20
12	SWIR 2	2100–2280	20

Table 2: Sentinel-2 data used for image classification

Scene Path/Row	April	May	June	July	August	September	October
T42TVL	*20*	*25*	*24*	*09*	*03*	*02*	*02*
T42TWK	*12*	*07*	*06*	*01*	*05*	*04*	*04*
T42TWL	*12*	*07*	*06*	*01*	*05*	*04*	*04*
T42TWM	*12*	*07*	*06*	*01*	*05*	*04*	*04*

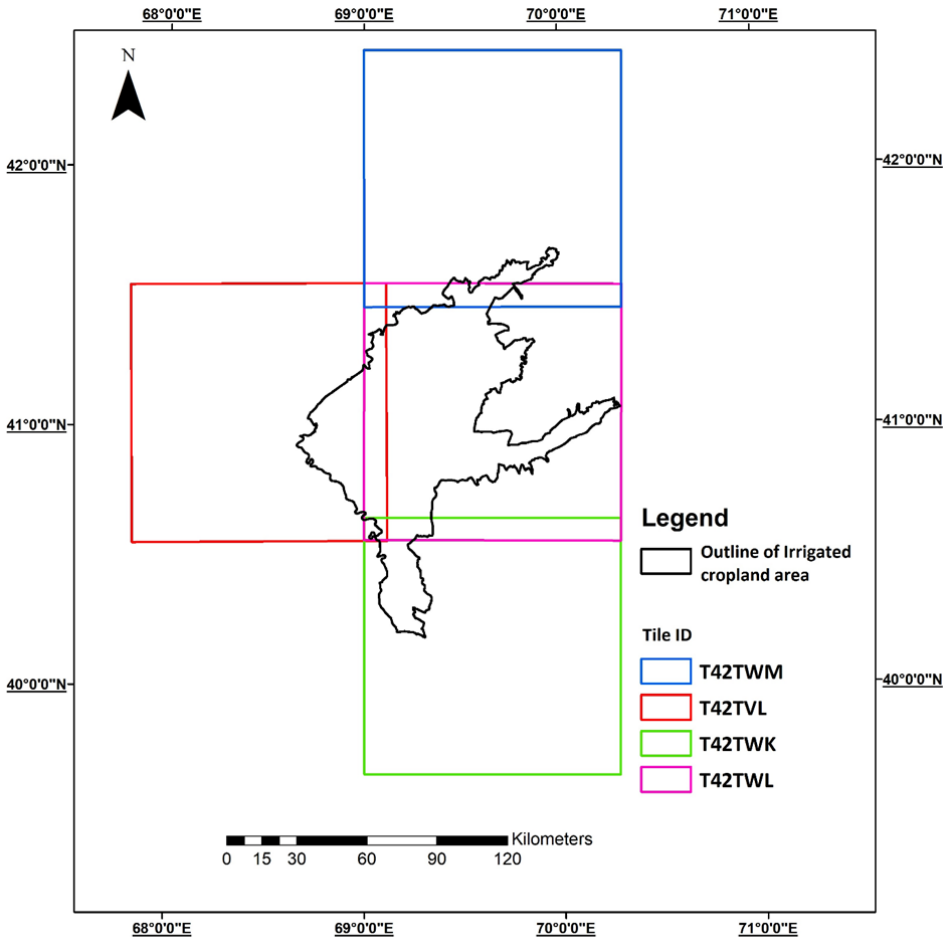

Figure 2: Outline of irrigated cropland area

Irrigated croplands are located in low elevated areas where irrigation canals exist. Therefore, irrigated croplands were separated from Mountain areas using high-resolution base maps in ArcMap. A total of four Sentinel -2 tiles covered the study area (Figure 2). All tiles were processed separately then mosaiced and clipped using the study area shapefile. Only 10 m resolution visible bands: 2, 3, 4, and 8 were used for image classification. Bands 4 and 8 were used for the calculation of monthly NDVI time-series data then composited creating multi-temporal NDVI images for all seven months. Visible bands 2 and 3 were also composited in the same way and combined with the NDVI composites before image classification. The images were projected to WGS 1984 Universal Transverse Mercator (UTM), Zone 42 coordinate reference system. All classified images were smoothed in order to remove the noises and improve the quality of classified output.

2.3 Methodology

The methodological workflow flowchart is given in Figure 3 consisting of the following steps: 1) Preparation of input data including the download of cloud-free images and band selection; 2) data processing, including atmospheric correction (TOA to LSR conversion), creation of NDVI time series and temporal spectral profiles for specific crop types as well as delineating multitemporal training samples; 3) irrigated cropland classification using MLC, SVM and RF algorithms; 4) evaluation of classification accuracy and comparison of classified irrigated cropland area with the State Statistics data. The identification of specific crop types can be achieved using a variety of vegetation indices (Sonobe et al., 2018). NDVI-based spectral-temporal profiles taken during the vegetation period have a high potential to properly classify heterogeneous crops on farmland.

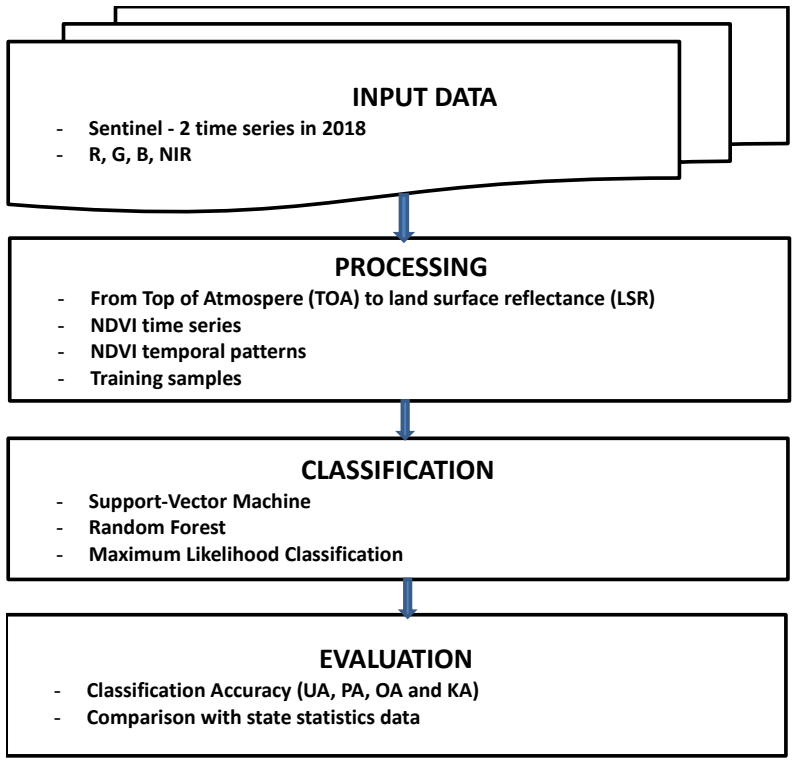

Figure 3: Flowchart of the study methods

More than 85% overall accuracy can be achieved using five or more acquisition dates covering different phenological phases of vegetation, for example, the period before winter-wheat harvest and summer crops in Uzbekistan (Conrad et al., 2014). Thus, monthly NDVI profile values for the growing season between April to October were used for the delineation of training samples. In addition, multitemporal NDVI image composites in combination with composites of Band 2 and Band 3 were used as input data for supervised image classification due to their high spatial resolution of 10 m. The high resolution of all input bands including the NDVI calculations was an important factor to derive crop-specific land use maps.

2.4 Training and Ground Truth data

Archived images from high-resolution satellites for ground-truthing and accuracy assessment are costly for scientists in developing countries. But freely available Google Earth (GE) images with high spatial resolution can be utilized as training samples and ground truth for mapping of LULC on a regional scale (Ahmed Ibrahim Ramz, 2015). Available GE images were also used as training and testing samples by Thanh Noi and Kappas (2017). It is recommended that the training sample size should represent approximately 0.25 percent of the study area for land cover classification of satellite images (Thanh Noi and Kappas, 2017). When the training sample sizes are large enough the performance of classifiers achieves better results (Thanh Noi and Kappas, 2017).

High accuracies are not achieved by having a large number of training samples. The main factors for ensuring higher accuracies are clarity of samples used in training, diversity of training samples, spatially equal distributed samples, and several samples (Gumma et al., 2020). To keep spatially well-distributed samples per class, 135 field samples were taken for each class except “wheat” due to crop rotation practices, requiring more training samples. Some wheat fields are given to farmworkers as a subsidy after harvest where they can use the land to grow necessary crops for themselves until the end of the harvesting season. Overall 8 classes were chosen for image classification (see Table 3): “Cotton”, “Wheat”, “Rice”, “Other Crops”, “Fruits/Trees”, “Bare land”, “Others” and “Water”. “Other crops” include forage, melons, and vegetable crops. “Others” include roads, river sands, and non-vegetated areas. The class “wheat” includes winter wheat and “secondary crops” after winter wheat.

Table 3: Training and ground truthing samples

Class Name	Number of Pixels		Class Color
	Training Samples	Ground Truthing	
Cotton	2528	250	
Wheat	5750	500	
Rice	3038	250	
Other Crops	2248	250	
Fruits/Trees	3642	250	
Bare land	2560	250	
Others	2601	250	
Water	2899	250	

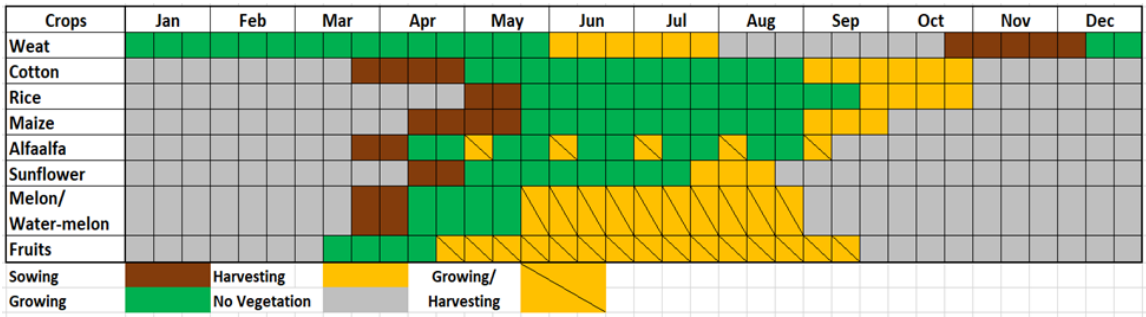

Figure 4: Idealized cropping calendar of the main crop types grown in the study area based on the guidance of regional agricultural land management authority and local experts

Usually, crop rotation is practiced after winter wheat is harvested such as maize, beans, and sunflowers. State statistics do not give any reports on secondary crops after winter wheat. The crops planted after winter wheat are given as a subsidy for farmworkers or as additional income sources for farmers. To check the performance of classifiers 250 ground-truthing pixel points were taken for accuracy assessment. Freely available GE images were used for ground-truthing based on the knowledge of cropping calendar shown in Figure 4.

2.5 Classification Methods

Nowadays, there are efficient and inexpensive GIS tools and methods available for the monitoring of LUs. In general, image classification can be grouped into three categories: a) supervised and unsupervised according to the way of learning, b) parametric and non-parametric based on data distribution assumptions, c) hard and soft based on the number of outputs for each spatial unit. These classifiers are either based on object-oriented classification (OOC) or pixel-based classification (PBC) methods (Jawak et al., 2015). OOC has an advantage over PBC when very high and hyperspatial resolution data are used for image classification (Quynh Trang et al., 2016).

In unsupervised classifications, pixels are assigned to groups based on each pixel's similarity to other pixels (no truth, or observed, data are required). In supervised classifications, the user instructs the image processing software to specify the land cover classes of interest. For each land cover type of interest, the user creates "training samples" - places on the map that are known to be representative of that land cover type. The main problem for unsupervised classification is that spectral data will not always correspond to spectral classes and that final grouping of clusters thus needs to decided by the human operator. Supervised classification on the other hand, provides more accurate classifications and also allows for more control over the classification process. Especially for large areas and diverse conditions (e.g. differing seasonality of phenology at different sea-levels), extensive and thus expensive training is required (Chuvieco, 2020). Therefore, the following three supervised classification algorithms were used:

Support-Vector Machine Classification: Support Vector Machine method is an excellent classification method if the number of input variables is high (Sonobe et al., 2018). The strength of SVM is to produce good results using small training samples as well as a high level of class

separability. The separation of two classes in the SVMs is done by a hyperplane. A hyperplane is a minimum distance (also called margin) between training samples of two classes. Support vectors are the nearest spectra, which are used for identifying the hyperplane (Baghdadi and Zribi, 2016). The SVMs have four basic kernels: linear; polynomial, radial basis function, and sigmoid. In our research, we used a standard linear basic kernel which is available in ArcMap tools.

Random Forest Classification: The random Forest method is a non-parametric supervised machine learning classification method and it trains a model according to known values given by training samples. RF classifier tool in ArcMap creates models and generates predictions based on Leo Breiman's RF algorithm (Breiman, 2001). As an attribute selection measure, the RF classifier uses the Gini Index. The Index measures the impurity of an attribute related to the classes. For a given training set T, selecting one case (pixel) at random and saying that it belongs to some class Ci, the Gini index can be written as:

$$\sum \sum_{j \neq i} \left(\frac{f(Ci,T)}{|T|}\right)\left(\frac{f(Ci,T)}{|T|}\right)$$

Equation 1

Where, $(\frac{f(Ci,T)}{|T|})$ is the probability that the selected case belongs to class C_i. Each time a tree is grown to the maximum depth on new training data using a combination of features (Pal, 2005). In our research, the maximum number of trees was selected as 50, the maximum tree depth was 30 and the maximum number of samples per class was 1000 by default.

*Maximum-Likelihood Classification:*The Maximum Likelihood Classification algorithm is based on Bayes' theorem and calculates the likelihood distributions for the classes. The likelihood distributions are assumed based on multivariate normal distribution models (Richards and Jia, 2006). There should be a sufficient number of training pixels to allow the calculation of the covariance matrix used in this algorithm. The likelihood function, described by Richards and Jia (2006), is calculated for each pixel as:

$$g_k(x) = lnp(C_k) - \frac{1}{2} ln \left|\sum_k\right| - \frac{1}{2}(x - y_k)^t \sum_k^{-1} (x - y_k)$$

Equation 2

where, C_k = land cover class k; x = spectral signature vector of a image pixel; $p(C_k)$ = probability that the correct class is C_k ; $|\sum_k|$ = determinant of the covariance matrix of the data in class C_k; $\sum_k^{-1}$ = inverse of the covariance matrix; y_k = spectral signature vector of class k.

2.6 Cropping Calendar

In Uzbekistan cotton and winter wheat are the two major crops cultivated. Cotton is mostly exported crop of the country, while wheat is a main grain crop of the country. Other irrigated crops are rice, watermelons, alfalfa, maize, sunflower and fodder crops. The cropping times may differ slightly according to climate conditions. The cropping calendar was created based on the guidance of regional agricultural land management authority and local experts of Tashkent Province, Uzbekistan (Figure 4). The similar cropping calendar was created for another region of Uzbekistan (Conrad et al., 2014). As shown most of the crops have almost similar cropping season in the study area. The crops cotton, rice, maize, alfa-alfa, sunflower and melons are planted between end of March and beginning of May and harvested in July to October. Only winter wheat is sown in October and November and harvested in June and July. Fruit trees and gardens are scattered throughout the study area and it is impossible to make comparison of classified area of Fruits and Trees with state statistics data. Cropping calendar and vegetation growth data play an important role to take training samples for supervised classifiers.

3. Results

In 2018 main crops in Tashkent Province were wheat and cotton. 85 % of crops are cultivated on farmland. Cotton is planted between the end of March and during April. Thus, NDVI values in May show high deviation error bars which are higher compared to June due to the growth of weeds (see Figure 5). After 40 to 50 days, at the end of June and July, flower buds start to form and cotton bolls begin to fill from July until mid-August. As a result, NDVI reaches its maximum value in August. End of August and in mid-September cotton bolls are fully open and are ready for picking. At this point, NDVI values start declining. Cotton is picked in September until mid-October. Wheat is planted end of October until the beginning of December. The vegetation period of wheat continues until June. Unfortunately, cloud-free satellite images are difficult to acquire during wintertime. After harvesting wheat, the land will be used for secondary crops or stay as furrows until the next cropping season.

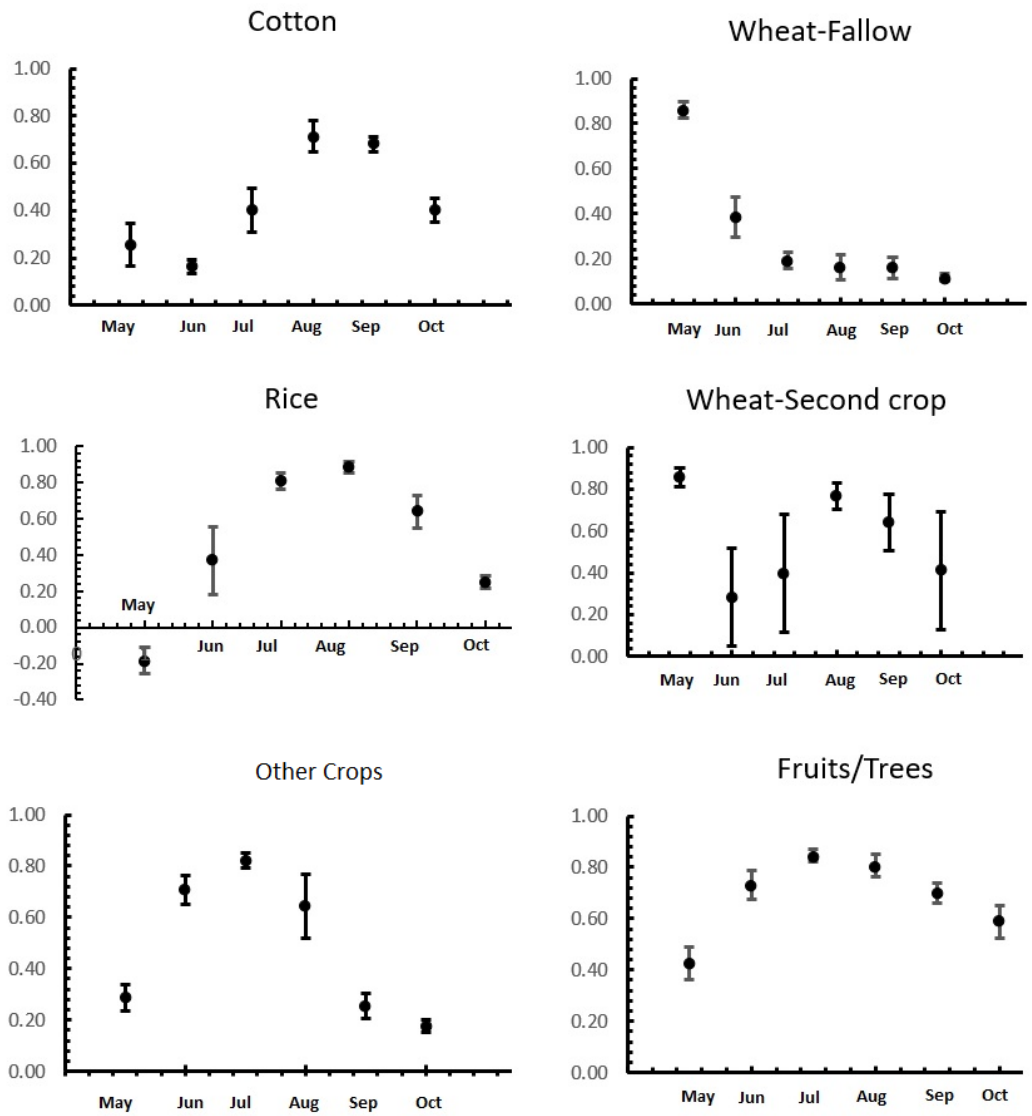

Figure 5: Monthly NDVI profiles during the vegetation period in 2018 at field level retrieved from multitemporal Sentinel-2 data. Mean NDVI and its standard deviation error bars are also shown

Thus, NDVI values fluctuate highly during crop rotation after winter wheat. Secondary crops such as vegetables, cucurbits, potatoes, and others are sown to fulfil the food requirements of the growing population. The NDVI profiles of Rice are clearly distinguishable because of negative values in May and peak values in July and August. Other crops are forage crops and vegetables with vegetation periods from May to August. Fruits and trees grow throughout the whole vegetation period resulting in constantly high NDVI values. These can easily be distinguished by all classification algorithms. A comparison of this class with state statistics was not possible as the class includes fruits and trees of private households and private landowners.

Besides, many mulberry trees are grown around croplands for sericulture. The results of high-resolution irrigated cropland classifications based on SVM, RF, and MLC methods are presented in Figure 6 and their subsets in Figure 7. It can be seen the crops "Cotton" and "Wheat" are evenly distributed across the entire province. Rice fields are located along Chrichik and Syrdarya rivers as well as main irrigated canals. The class "Other Crops" was highly misclassified using the MLC method. Visually only small differences can be seen when comparing SVM and RF classification results (Figure 7). Overall, the SVM-derived classification shows higher accuracy compared to RF and MLC.

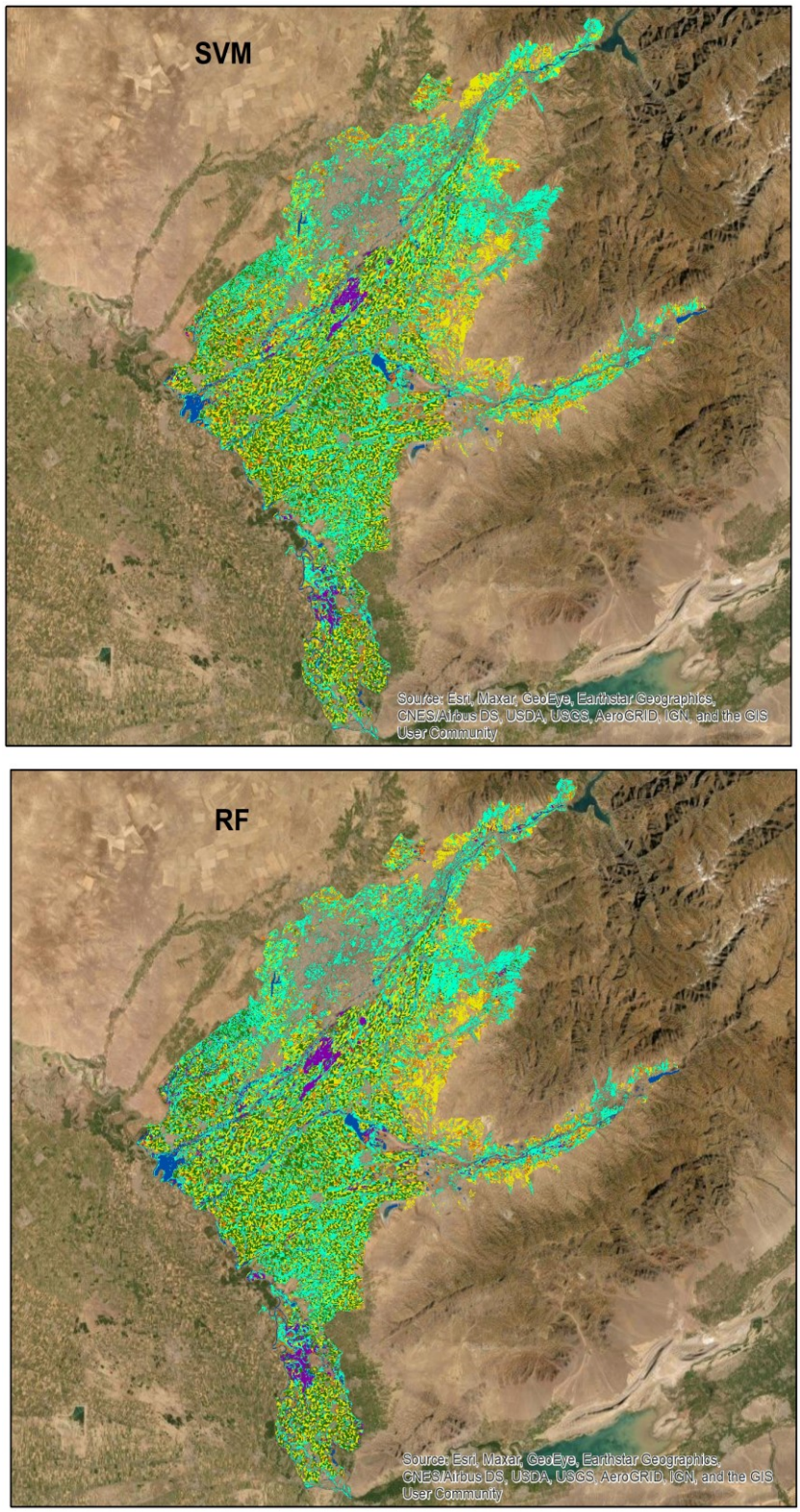

Figure 6: Supervised classification methods output and the subset of Agricultural LU maps in 2018: SVM classification; RF classification; and MLC classification (Continue next page)

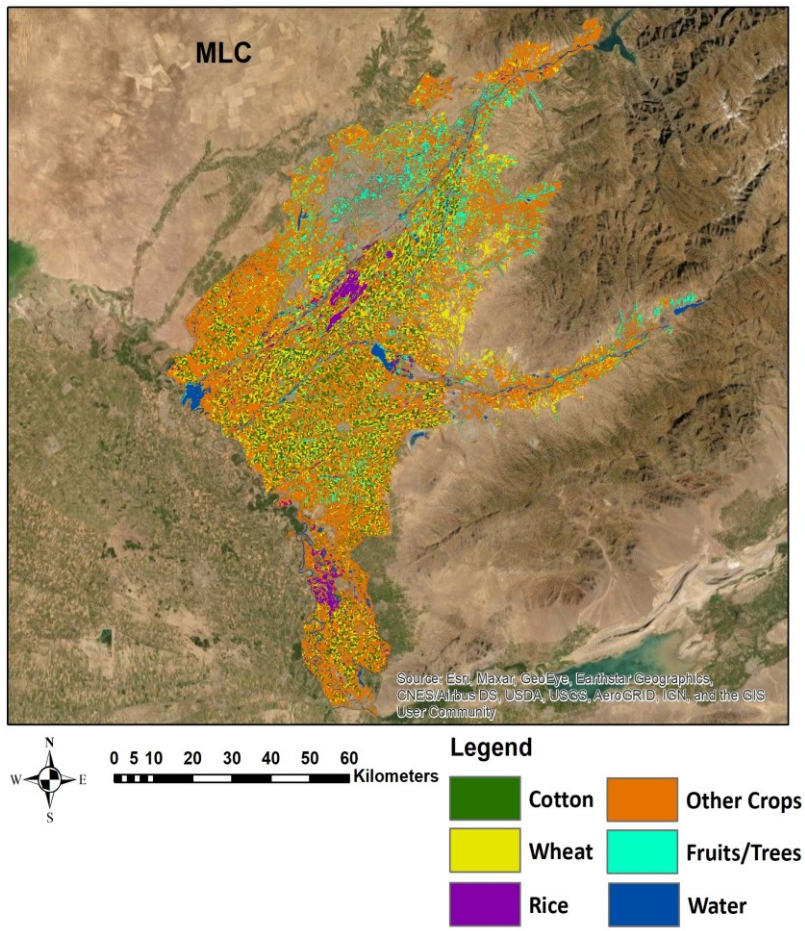

Figure 6: Supervised classification methods output and the subset of Agricultural LU maps in 2018: SVM classification; RF classification; and MLC classification

3.1 Accuracy Assessment of Image Classification

To evaluate the derived irrigated cropland maps, accuracy assessment was performed for each classification method. Detailed confusion matrices showing classification accuracies are given in Table 4. A total of 2250 well-distributed ground-truthing samples were taken for the accuracy assessment in Tashkent Province. The results show that the overall accuracies (OA) of SVM and RF classification algorithms are good and similarly high, whereas the MLC algorithm shows a lower OA. In the SVM confusion matrix OA was 91.3%, in RF 90.5% and in MLC 85,5%. User's accuracy (UA) and Producer's accuracy (PA) of all classes in SVM and RF algorithms were higher than 80%, but MLC showed a moderate PA of 44.8% for the class "Fruit/Trees" and 51.7% UA for the class "other crops". All other cropland classes PA and UA in MLC were higher than 78%. The highest UA for the cropland class "cotton" was achieved using MLC, for the class "wheat" and "rice" SVM showed better performances. The class "other crops" performed best using the RF algorithm. SVM shows the highest PA for the cropland classes "cotton", "wheat" and "fruit/trees" and MLC for the classes "rice" and "other crops", respectively. Kappa accuracies (KA) of classified images for SVM were 0.90 and 0.89 for the RF algorithm. Both performed well with similar values. MLC showed a lower result of KA 0.60.

3.2 Comparison Sentinel Derived Cropland Products with National Statistical Data

In this study, we calculated areas of crop types for the entire Tashkent Province and compared them with the area given by statistics available from the State Statistics Committee of Uzbekistan. Figure 8 compares Sentinel-2 derived irrigated croplands areas with different supervised classification methods.

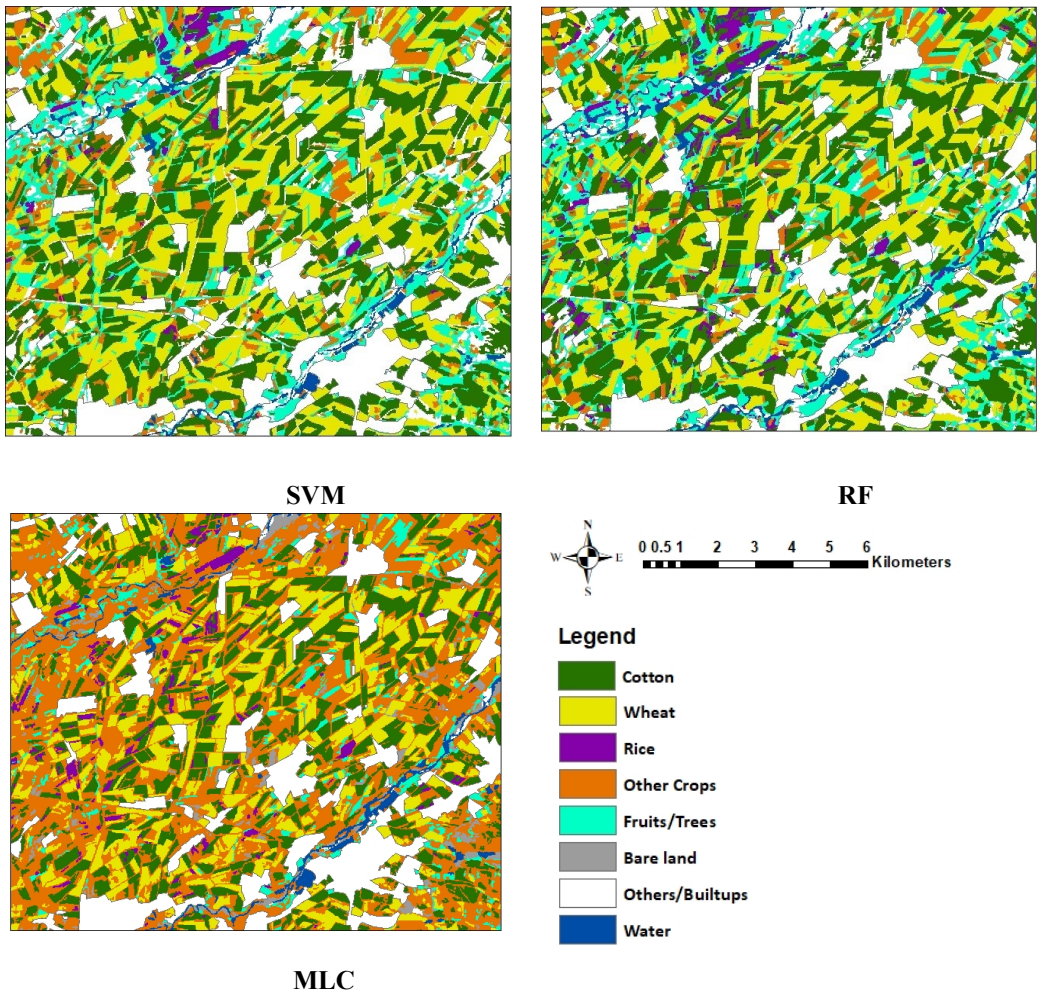

Figure 7: LU map subsets from different classification methods

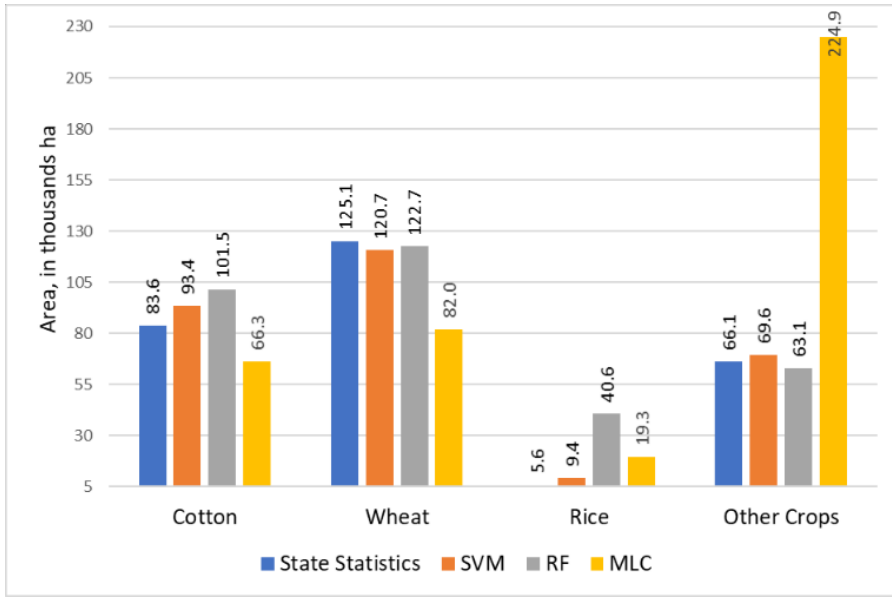

Figure 8: Comparison of Sentinel 2 -derived cropland product with national statistical data

Table 4: Confusion matrices and classification methods accuracy based on ground-truthing data

Support Vector Machine Classification										
	Ground Truth									
Classes	Cotton	Wheat	Rice	Other Crops	Fruits/ Trees	Bare land	Others	Water	Grand Total	UA, %
Cotton	244		14		4	4			266	91.7
Wheat		462	1	2		23			488	94.7
Rice	1		225	5		2			233	96.6
Other Crops	3	18	6	238	14	2	14	2	297	80.1
Fruits/Trees	1	20		4	222	4		2	253	87.7
Bare land	1		1	1	10	204	22		239	85.4
Others						11	214	1	226	94.7
Water			3					245	248	98.8
Grand Total	250	500	250	250	250	250	250	250	2250	
PA, %	97.6	92.4	90	95.2	88.8	81.6	85.6	98		
Overall Accuracy	91.3%									
Kappa	0.90									
Random Forest Classification										
Cotton	241	1	8	5	6	5	2		268	89.9
Wheat		450	2	9		22			483	92.2
Rice	6	1	235	11	13	4			270	87.0
Other Crops	2	28	3	213	10	5		1	262	81.3
Fruits/Trees	1	20		10	220	2		2	255	86.3
Bare land				2	1	205	20	1	229	89.5
Others						7	228	2	237	96.2
Water			2					244	246	99.2
Grand Total	250	500	250	250	250	250	250	250	2250	
PA, %	96.4	90.0	94	85.2	88	82	91.2	97.6		
Overall Accuracy	90.5%									
Kappa	0.89									
Maximum Likelihood Classification										
Cotton	226		3		2	1			232	97.4
Wheat		414	1	2		16			433	84.8
Rice	9		237		5	1		1	253	93.7
Other Crops	14	57	4	247	125	14	13	4	478	51.7
Fruits/Trees	1	29		1	112				143	78.3
Bare land					6	212	4		222	95.5
Others						6	233	2	241	96.7
Water			5					243	248	98.0
Grand Total	250	500	250	250	250	250	250	250	2250	
PA, %	90.4	82.8	94.8	98.8	44.8	84.8	93.2	97.2		
Overall Accuracy	85.5%									
Kappa	0.84									

The estimated area of irrigated crop types using SVM and RF fit well to State Statistics data. The MLC classification results show the highest differences compared to state statistics and are significantly lower than state statistics data. This especially accounts for the class “other crops” due to the complexity of phenological stages during the vegetation period as shown in the NDVI monthly profiles. Other cropland classes in MLC do not significantly differ from state statistics data

In the SVM classification method, Sentinel-2 derived irrigated cropland area was 93.4 thousand ha for cotton, 120.7 thousand ha for wheat, 9.4 thousand ha for rice, and 69.6 thousand ha for other crops compared to 83.6, 125.1, 5.6, and 66.1 thousand ha of state statistics data respectively. Whereas in RF, it was 101.5 thousand ha for cotton, 122.7 thousand ha for wheat, 40.6 thousand ha for rice, and 63.1 thousand ha for other crops. MLC classified an area of 66.3 thousand ha for cotton, 82 thousand ha for wheat, 19.3 thousand ha for rice, and 224.9 thousand ha for other crops. Overall, the "cotton" and "wheat" classes show reliable results in all classification methods as both are state-order crops that can be correctly reported. Other crops vegetation period starts in early spring in April. At the same time the precipitation amount is high and it can result to grow grasses on the bare lands during this time of period. Usually these areas are located nearby irrigation channels, rivers and high slope areas. Besides, after harvesting winter wheat, the second crop is cultivated by farmers which challenges the classification of this class. Therefore, MLC method misclassified the class "Other Crops". But SVM and RF classification methods have privilege to distinguish this class and classified accurately.

4. Discussion

In this paper irrigated cropland extent of Tashkent Province was mapped using high-resolution 10-m Sentinel-2 data and a comparison of supervised classification algorithms such as SVM, RF, and MLC performance was carried out. Knowledge of vegetation phenological processes and NDVI time-series data were used during the image classification process illustrating the importance of continuous satellite-based time series, capturing vegetation dynamics and plant growth characteristics to map agricultural crop types. Results show that the use of multi-temporal datasets, in comparison to single-date imagery, provides an improved basis for the discrimination of crop types. The results are comparable to similar research conducted by Gumma et al., (2020) and Xie et al., (2019).

This study results showed that irrigated cropland classification using SVM and RF algorithms provided the highest OA and KA compared to MLC (Figure 9). Comparable studies in other regions have shown similar results of LU mapping (Burai et al., 2015, Deilmai et al., 2014, Thanh Noi and Kappas, 2017 and Xie et al., 2019). SVM and RF perform better than MLC when there is a limited number of training samples (Basukala et al., 2017). The low performance of MLC is because the capability of the MLC algorithm depends on a very accurate estimation of the average vector and the covariance matrix for each spectral class. Therefore, MLC relies on a satisfactory number of training samples per class and it performs unreasonably when a small number of training samples is used (Basukala et al., 2017). For example, when the training samples were reduced from 30 to 10 accuracy assessment was also significantly reduced from 80.78% to 52.56% (Burai et al., 2015). The lowest performance of MLC is partially due to crop types having very heterogeneous raster values in training samples, including several crops with different spatial and spectral preferences. The integration of additional imagery to provide higher temporal resolution would most probably improve classification results (Basukala et al., 2017). Unfortunately, it was impossible to get additional imagery in early spring due to the high percentage of cloud cover. Due to the high OA, KA, UA, and PA results for all irrigated crop types, SVM and RF can be recommended for LU mapping in the region as an effective and accurate classifier.

Accuracy assessment results of the class "rice" showed high values but its comparison with statistics data was not close for all classifiers. It might be the wrong reporting by state statistics because the local government is criticizing that the statistical data were not correctly reported previously in many cases (Shavkat Mirziyoyev, 2021). Even a few years ago rice fields area were two times high than in 2018. Besides, there some farmers plant rice crop as a secondary crop after winter wheat is harvested and it is not included in the Statistics. There should be detailed field observation to prove it. Moreover, some crops are more profitable than other crops and they can be traded on the local market except cotton due to government control (Platonov et al., 2014). As a result, there is a high demand for rice in the local market. For example, in 2020 Uzbekistan imported 13.6 thousand tons of rice but the country's demand is about 335 thousand tons. Uzbekistan is planning to double the rice production by planting new rice sorts and introducing two crop production (Economy Uzbekistan, 2021). But the state-order crop cotton's growing period very long which occupies two crop production periods. Thus, rice in two crop production might be challenging due to water shortage in the region in future. Besides the climate of Tashkent Province is not convenient for two crop production in comparison with other regions of the country.

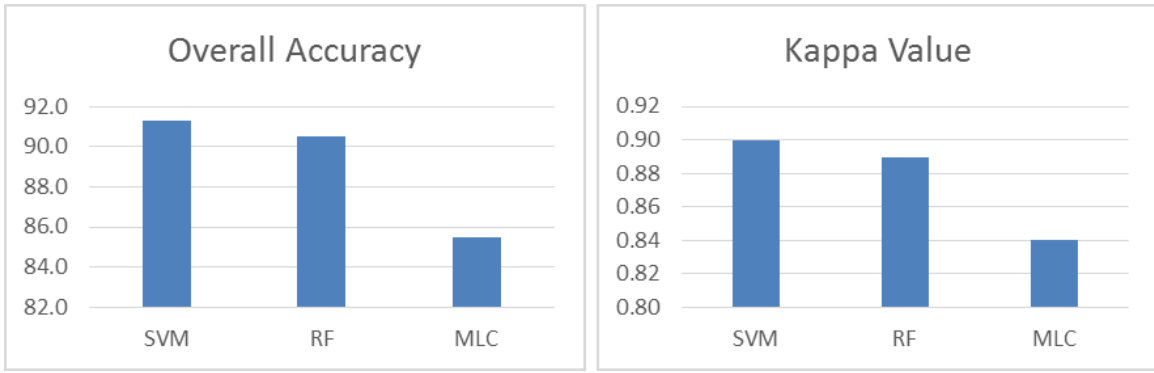

Figure 9: OA and KA value of classifiers

With the introduction of winter wheat, competition comes up between state crops and cash crops. Because state crops are not profitable for the farmers and farmers can earn additional income planting cash crops after winter wheat (Platonov et al., 2014). Cotton is traded only by the government and wheat harvest should be given to the government as a part of the contract with the farmers. Thus, their classification results were satisfied by all classification methods which was believed the official data was correctly reported. The study showed feasible results at the Provincial level for irrigated cropland classification and comparison of classifiers' performance with statistics data were not found in previous studies in the research object. As a first-time approach to this method, it can be a good start and should be developed in the future together with field trips and local experts' knowledge. Besides, the potential of using other indices and spectral bands of Sentinel -2 such as red edge and SWIR for cropland classification raises important questions in the future. In this paper, it is unknown whether using other indices and spectral bands can result in better performance of irrigated cropland classification. And we intended to carry out more studies in the future, to improve the performance of different classification methods by using other indices and spectral bands.

5. Conclusions

The study produced a 10-m high resolution irrigated crop types product (based on Sentinel-2 data with 1 pixel = 0.01 ha). The analysis was applied based on pixels in one agricultural calendar year using SVM, RF, and MLC methods using ArcMap and open-source software QGIS. NDVI monthly profiles during the vegetation period of each class showed the potential of irrigated crop types classification. And the study of irrigated crop types mapping at the provincial level with different supervised classification methods and comparing the performance of the classifiers as well as the comparison of crop types area with official data is the first application in the region.

Among the tested methods the SVM classifier performed the highest OA and KA and produced a visually pleasant irrigated cropland map in Tashkent Province. Therefore, the authors conclude that the SVM classification algorithm was the best-suited classifier for mapping heterogeneous irrigated crop types and it can be used alternatively where digital cadastre LU maps are limited for agro-environmental assessment studies in the region. Although the RF classifier had lower accuracy than SVM, it performed better than MLC. MLC method could not yield high accuracy and handle many pixels incorrectly classified, so we do not recommend this classifier for irrigated crop types classification in the study area.

The comparison of classified crop types area with official State Statistics data showed that the state order crops "cotton", "wheat" and the class "other crops" in SMV and RF methods were close to the official data. The area of class "rice" is higher than official data which requires additional study further or correct statistical data which also includes rice area after winter wheat. And the class "others crops" showed unreliable results in MLC due to its complexity and heterogeneity. Comparing the area of class "fruits/trees" is complicated because private households and landowners have fruits and trees on their property. And also, many cropland fields are surrounded by mulberry trees for sericulture which is not included in the official data. Comparison of remote sensing-based irrigated cropland area with official data showed the applicability of these methods over other areas of the region which has similar agro-ecological conditions but there should also be further research with field data and using other indices or bands that can improve the estimates in rice and other crops classes.

We are also motivated to use and test other bands to derive Sentinel-2 based vegetation indices incorporation the three RE and SWIR bands in a 20 m resolution for image classification in the future.

Acknowledgments

The authors highly appreciate the support of the ERASMUS MUNDUS Action 2 TIMUR (Training of Individuals through Mobility to EU from the Uzbek Republic) project for providing scholarship during this study period.

References

Ahmed Ibrahim Ramzi, V., 2015, Ground Truth and Mapping Capability of Urban Areas in Large Scale Using GE images. *Proc. Earth Resources and Environmental Remote Sensing/GIS Applications VI*, Vol. 9644, https://doi-.org/10.1117/12.2193727.

Alikhanov, B., Alikhanova, Sh., Oymatov, R., Fayzullaev, Z. and Pulatov, A, 2020, Land Cover Change in Tashkent Province During 1992 – 2018. *IOP Conf. Ser.: Mater. Sci. Eng.*, Vol. 883, DOI: 10.1088/1757-899X/883/1/012088.

Baghdadi, N. and Zribi, M., 2016, *Optical Remote Sensing of Land Surfaces. Techniques and Methods (Eds.)*. London, Oxford: ISTE Press Ltd.; Elsevier Ltd (Remote sensing observations of continental surfaces set).

Basukala, A. K., Oldenburg, C., Schellberg, J., Sultanov, M. and Dubovyk, O., 2017, Towards Improved Land Use Mapping of Irrigated Croplands: performance assessment of different image classification algorithms and approaches. *European Journal of Remote Sensing*, Vol. 50(1), 187–201. DOI: 10.1080/22797254.20-17.1308235.

Breiman, L., 2001, Random Forests. *Machine Learning*, Vol. 45(1), 5–32. DOI: 10.1023/A:-1010933404324.

Burai, P., Deák, B., Valkó, O. and Tomor, T., 2015, Classification of Herbaceous Vegetation Using Airborne Hyperspectral Imagery. Remote Sensing, Vol. 7(2), 2046–2066. DOI: 10.3390/rs70202046.

Chen, X., Bai, J., Li, X., Luo, G., Li, J. and Li, B. L., 2013, Changes in Land Use/Land Cover and Ecosystem Services in Central Asia during 1990–2009. *Current Opinion in Environmental Sustainability*, Vol. 5(1), 16–127. DOI: 10.1016/j.cosust.2012.12.005.

Chuvieco, E., 2020, *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. CRC press.

Congedo, L., 2021, Semi-Automatic Classification Plugin: A Python Tool for the Download and Processing of Remote Sensing Images in QGIS. *JOSS*, Vol. 6(64), DOI: 10.21105/joss.03172.

Conrad, Ch., Dech, S., Dubovyk, O., Fritsch, S., Klein, D., Löw, F., Schorcht, G. and Zeidler, J., 2014, Derivation of Temporal Windows for Accurate Crop Discrimination in Heterogeneous Croplands of Uzbekistan Using Multitemporal RapidEye Images. *Computers and Electronics in Agriculture*, Vol. 103, 63–74. DOI: 10.1016/j.-compag.2014.02.003.

Deilmai, B. R., Ahmad, B. B. and Zabihi, H., 2014, Comparison of Two Classification Methods (MLC and SVM) to Extract Land Use and Land Cover in Johor Malaysia. IOP Conf. Ser.: Earth Environ. Sci., Vol. 20, DOI: 10.1088/1755-1315/20/1/012052.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C.,Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P., 2012, Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment,* Vol. 120, 25–36, DOI: 10.1016/j.rse.2011.11.026.

Economy Uzbekistan, 2021, Uzbekistan Eyes to Double Rice Production this Year. Available online at https://www.thetribune.com/uzbeki-stan-eyes-to-double-rice-production-this-year/.

Edlinger, J., Conrad, C., Lamers, J., Khasankhanova, G. and Koellner, T., 2012, Reconstructing the Spatio-Temporal Development of Irrigation Systems in Uzbekistan Using Landsat Time Series. Remote Sensing, Vol. 4(12), 3972–3994. DOI: 10.3390/rs4123972.

Erdanaev, E., Kappas, M., Pulatov, A. and Klinge, M., 2015, Short Review of Climate and Land Use change Impact on Land Degradation in Tashkent Province. International Journal of Geoinformatics, Vol. 11(4), 39-48.

Gerts, J., Juliev, M. and Pulatov, A., 2020, Multi-Temporal Monitoring of Cotton Growth Through the Vegetation Profile Classification for Tashkent province, Uzbekistan. GeoScape, Vol. 14(1), 62–69. DOI: 10.2478/geosc-2020-0006.

Gumma, M. K., Thenkabail, P. S., Teluguntla, P. G., Oliphant, A., Xiong, J., Giri, C., Pyla, V., Dixit, V. and Whitbread, A. M., 2020, Agricultural Cropland Extent and Areas of South Asia Derived Using Landsat Satellite 30-m Time-

Series Big-Data Using Random Forest Machine Learning Algorithms on the Google Earth Engine cloud, *GIScience and Remote Sensing*, Vol. 57(3), 302-322, DOI: 10.1080/154816-03.2019.1690780.

Jawak, S. D., Devliyal, P. and Luis, A. J., 2015, A Comprehensive Review on Pixel Oriented and Object Oriented Methods for Information Extraction from Remotely Sensed Satellite Images with a Special Emphasis on Cryospheric Applications. *Advances in Remote Sensing*, Vol. 4(3), 177-195. DOI: 10.4236/ars.2015.43015.

Jia, Yuanxin; Ge, Yong; Ling, Feng; Guo, Xian; Wang, Jianghao; Le Wang; Chen, Yuehong; Li, Xiaodong, 2018, Urban Land Use Mapping by Combining Remote Sensing Imagery and Mobile Phone Positioning Data. *Remote Sensing*, Vol. 10(3), DOI: 10.3390/rs10030446.

John, A. and Richards, X., J., 2006, *Image Classification Methodologies. Remote Sensing Digital Image Analysis*. Berlin/Heidelberg: Springer-Verlag, 295–332.

Juliev, M., Pulatov, A., Fuchs, S. and Hübl, J., 2019, Analysis of Land Use Land Cover Change Detection of Bostanlik District, Uzbekistan. *Pol. J. Environ. Stud.,* Vol. 28(5), 3235–3242. DOI: 10.15244/pjoes/94216.

Li, Z., Fang, G., Chen, Y., Duan, W. and Mukanov, Y., 2020, Agricultural Water Demands in Central Asia under 1.5 °C and 2.0 °C Global Warming. Agricultural Water Management, Vol. 231, DOI: 10.1016/j.agwat.2020.106020.

Liu, M. Y., 2011, Central Asia in the Post–Cold War World. Annu. Rev. Anthropol., Vol. 40(1), 115–131. DOI: 10.1146/annurev-anthro-081309-145906.

Pal, M., 2005, Random Forest Classifier for Remote Sensing Classification. International Journal of Remote Sensing, Vol. 26(1), 217–222. DOI: 10.1080/01431160412331269698.

Platonov, A., Wegerich, K., Kazbekov, J. and Kabilov, F., 2014, Beyond the State Order? Second Crop Production in the Ferghana Valley, Uzbekistan. International Journal of Water Governance, Vol. 2(2), 83–104. DOI: 10.7564/14-IJWG58.

Pulatov, B., Umarova, Sh. and Alikhanov, B., 2020, Assessment of Land Degradation Changes in Mountain Areas in Tashkent Province. IOP Conf. Ser.: Mater. Sci. Eng., Vol. 883, DOI: 10.1088/1757-899X/883/1/012087.

Quynh Trang, N. Thi., Le Toan, Q., Huyen Ai, T. Thi., Vu Giang, N. and Viet Hoa, P., 2016, Object-Based vs. Pixel-Based Classification of Mangrove Forest Mapping in Vien An Dong Commune, Ngoc Hien District, Ca Mau Province Using VNREDSat-1 Images. Advances in Remote Sensing, Vol. 5, 284-295 DOI: 10.4236/ars.2016.54022.

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., Nguyen, Q. H., Patterson, M., Johnson, G., Markert, K., Flores, A., Anderson, E., Weigel, A., Ellenberg, W. L., Bhargava, R., Aekakkararungroj, A., Bhandari, B., Khanal, N., Housman, I. W., Potapov, P., Tyukavina, A., Maus, P., Ganz. D., Clinton. N. and Chishtie, F., 2019, Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. *Front. Environ. Sci*., Vol. 7, doi: 10.3389/fenvs.2019.00150.

Shavkat Mirziyoyev, 2020, Digital Technologies will Be Introduced in the Sphere of Cadastre. Available online at https://president.uz/uz/lists/view/3717.

Shavkat Mirziyoyev, 2021, These Numbers are noting Any Statistics. I declare to answer before my conscience - President Spoke about the Situation with Serious Diseases in Karakalpakstan. Tashkent. Available online at https://kun.uz-/news/2021/09/21/bu-raqamlar-hech-qaysi-statistikada-yoq-vijdonim-oldida-javob-berish-uchun.

Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N. and Mochizuki, K., 2018, Crop Classification from Sentinel-2-derived Vegetation Indices Using Ensemble Learning. J. Appl. Rem. Sens., Vol. 12 (02), DOI: 10.1117/1-.JRS.12.026019.

State Committee for Statistics, 2018, *Statistical Review - Agriculture of Uzbekistan between 2014-2018*. The State Committee for Statistics of the Republic of Uzbekistan, Tashkent.

Thanh Noi, P. and Kappas, M., 2017, Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors (Basel, Switzerland), Vol. 18 (1). DOI: 10.3390/s18010018.

Xie, Z., Chen, Y., Lu, D., Li, G. and Chen, E., 2019, Classification of Land Cover, Forest, and Tree Species Classes with ZiYuan-3 Multispectral and Stereo Data. *Remote Sensing*, Vol. 11 (2), DOI: 10.3390/rs11020164.

Zvi, L., 2018, Agricultural Development in Uzbekistan: The Effect of Ongoing Reforms. Available online at http://departments.agri.huji.-ac.il/economics/indexe.html.