Machine Learning Approaches and Sentinel-2 Data in Crop Type Mapping

Pranay Panjala, Murali Krishna Gumma, and Pardhasaradhi Teluguntla

Abstract Crop monitoring becomes essential in attaining food security for implementation of various agricultural serving programs. So, fast and reliable crop monitoring is must. Using traditional methods, crop monitoring maps need high amount of satellite data downloading and processing time. Google Earth Engine (GEE) cloud platform enables us to save time in downloading and processing of time series satellite data, the every satellite imagery is converted into Normalized Difference Vegetation Index (NDVI) image and stacked monthly wise maximum images. The stacked image was used for conducting supervised classification. The main objective of this study is to evaluate the performance of different supervised machine learning (ML) classifiers in GEE platform and Spectral Matching Technique (SMT) using Sentinel-2 10 m satellite imagery in specific crop type classification. The crop classification for the year 2018–19 (*rabi* season) was carried for Jhansi District using supervised classifiers like Random Forest (RF), Support Vector Machine (SVM) and Classification and Regression Trees (CART) in GEE platform and also with SMT with the help of ground data. It was attained nearly 81.8% accuracy for RF, 68.8% for SVM, 64.9% for CART and 88% for SMT. The results obtained using RF classifier were nearly relative to SMT classification map. The study indicates that classifier's performance depends on the quality of ground data used, RF can reduce the error samples in ground samples and produce satisfactory results. This study compared results obtained from all the above classifiers with agricultural statistics and also compared crop-wise accuracies. In the study, it was observed that RF classification is outperformed when compared with other classifiers considered in the study.

P. Panjala · M. K. Gumma (✉)
Geospatial Sciences and Big Data, Global Research Program on Resilient Farm and Food Systems, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), 502 324 Patancheru, India
e-mail: M.Gumma@cgiar.org

P. Teluguntla
Western Geographic Science Center, U.S. Geological Survey (USGS), AZ 86001 Flagstaff, USA

Bay Area Environmental Research Institute (BAERI), NASA Research Park, Moffett Field, San Jose, CA 94035, USA

G. P. O. Reddy et al. (eds.), *Data Science in Agriculture and Natural Resource Management*, Studies in Big Data 96, https://doi.org/10.1007/978-981-16-5847-1_8

Keywords Crop type mapping · Sentinel-2 · Dryland agriculture · Agriculture · Google earth engine

1 Introduction

Monitoring agricultural changes, ecosystem dynamics, climate change scenarios and other human-induced changes on earth surface require information on land use and land cover (LULC). In agriculture, monitoring of crop changes, land use changes and other physical parameters helps in estimating agriculture statistics, crop yield and drought risk, which in turn helps solving problems faced by farming community. The above estimation requires satellite imagery for LULC mapping. The coarse resolution satellite data like Moderate Resolution Imaging Spectroradiometer (MODIS) 250 m has been used to generate LULC and changes maps, also used for global cropland extent (Gumma et al. 2016). The entry of high-resolution imagery like Sentinel-1, Sentinel-2 and Landsat-8 allows to generate and monitor LULC changes efficiently. But analyzing at large scale, high-resolution data requires high volume of data to download and it is time consuming and cumbersome process. These problems can be overcome by using GEE cloud platform and it contains updated time series satellite imagery data.

Earth observations and spatial sciences have high potential for sustainable agriculture and environmental development (Im 2020). However, cropland spatial maps combined with weather information and social factors to identify areas where productivity is affected and areas is supported by insurance companies. Furthermore, spatial maps could be used as important input parameter for crop yield assessment. One example is the development of short duration or stress-tolerant varieties that reduce the risk of drought exposure. High spatial-temporal imagery serves rapid, large-scale, cost-efficient and continuous monitoring of agriculture crops lands (Brown et al. 2007; Anderson et al. 2012; Thenkabail et al. 2012; McCarthy et al. 2018). Numerous studies have been conducted on LULC monitoring using various resolutions of satellite imagery with various methods at field to Global scale. Remote sensing is a powerful tool to monitor spatial distribution of agriculture croplands along with LULC classes (Pittman et al. 2010). Several land cover products at coarse and high resolution, such as Global Land Cover 2000, Global Land Cover by National Mapping Organizations (GLCNMO) (Tateishi et al. 2014) and Glob Cover (Bicheron et al. 2008). Mapped irrigated areas across the globe (GIAM) (Thenkabail et al. 2009). Several methods and techniques were used for crop classification, mainly temporal vegetation-based algorithms (Jeganathan et al. 2014; Dong et al. 2015; Pan et al. 2015; Xiong et al. 2017a), classification decision tree and regression tree algorithms (Friedl and Brodley 1997; Ozdogan and Gutman 2008; Deng and Wu 2013; Egorov et al. 2015), discrete Fourier harmonic analysis (Harris, 1978; Platas-Garza and de la O Serna 2010; Boche and Mönich, 2020), SMT (Thenkabail et al. 2007; Gumma et al. 2014; Gumma et al. 2018), RF algorithms (Millard and Richardson, 2015; Teluguntla et al. 2018; Oliphant et al. 2019; Xie et al. 2019). These studies were conducted

regional to global scale, which have some limitations for field scale management. Hence, monitoring croplands at field scale using high resolution is important for decision making and improving agriculture production. Many studies were conducted with supervised and unsupervised classification algorithms for mapping agriculture croplands through time series remote sensing imagery (Xiong et al. 2017a; Feyisa et al. 2020; Gumma et al. 2020). Croplands mapping using time series analysis were proven to map precise crop extent maps rather than single date imagery (Long et al. 2013; Gumma et al. 2014; Gómez et al. 2016; Belgiu and Csillik, 2018).

Above studies were conducted on classification of multi-temporal satellite images for cropland monitoring including LULC mapping. Monitoring major crop types using Earth observation (EO) data is very complex and has many challenges (Griffiths et al. 2019), such as crop changes within and between years according to local demand or water availability. In dryland areas, field size is very small and dynamically changes the cropping pattern and many areas have an intercropping pattern, especially in the present study area. Several studies have been demonstrated to map croplands using Sentinel-2 EO data (Veloso et al. 2017; Xiong et al. 2017b; Belgiu and Csillik, 2018; Lambert et al. 2018). Lambert et al. developed an approach for improved synthesis over fine temporal intervals and mapped national crop and land cover map. Xiong et al. (2017b) mapped 30 m crop extent for entire Africa using pixel and object-based approaches. Belgiu and Csillik (2018) was used pixel and object-based time-weighted dynamic time warping (DTW) analysis to map croplands. Veloso et al. (2017) were used Sentinel-1 time series data with Sentinel-2 NDVI data for monitoring winter and summer crops and this study was conducted in southwest France, where field size is very large. Given the above background, the present study tested various methods of classification in study area, where there is complex cropping pattern and average field size is one hectare.

GEE allows users to run different machine learning algorithms using java and python languages. GEE collects a multi-petabyte catalog of various satellite imagery and geospatial datasets (Gorelick et al. 2017) for cloud-level analysis. Using GEE, it is easy to monitor decadal changes without downloading data, using and comparing different datasets. GEE also allows to import shape files, images for any cloud analysis and also to export data for offline analysis.

GEE has been widely employed in multi-temporal analyses for large areas and many studies were carried out like sprawl mapping using Landsat 5 and 7 images (Patel et al. 2015), map rice-growing areas in Northeast Asia using Landsat 8 (Dong et al. 2016), automated mapping of cropland of continental Africa (Xiong et al. 2017a), cropland extent and areas of South Asia (Gumma et al. 2020), crop yield map (Lobell et al. 2015), flood mapping (Coltin et al. 2016), land cover change detection (Sidhu et al. 2018) and global non-croplands like global surface water changes (Pekel et al. 2016), global urban land use mapping (Liu et al. 2018), forest cover changes (Hansen et al. 2013).

In this study, we demonstrated different ML approaches and SMT techniques to map major crop types using Sentinel-2 time series data and intensive ground data. The main goal of the study is to test different methods to estimate crop areas at village level and identify the gaps in heterogeneous small holder farming systems.

We applied supervised classification with intensive training data collected in winter season 2018–19. GEE is able to run various machine learning algorithms like decision trees algorithms like CART and RF and optimization algorithms like SVM, which helps in classification.

2 Materials and Methods

2.1 Study Area

Jhansi district lies in 25°6″ to 25°56″ N latitude and 78°18″ to 79°25″ E longitude (Fig. 1), part of Bundelkhand agro-climatic region at an average elevation of 285 m and situation between Pahuj and Betwa rivers. This area experiences about 45–47 °C maximum temperatures during summer and 1–3 °C minimum temperatures during winter. This region receives annual rainfall about 884.6 mm out of which 90% is received during southwest monsoon. This area is commonly affected by droughts and long dry spells that affect both *kharif* and *rabi* seasons. Major soil texture available in this area are loamy, clay and loamy skeletal (Fig. 1). The major crops of Jhansi during *kharif* are paddy, sorghum, sesame, groundnut and pigeon pea, etc., while wheat, chickpea, mustard and pea are the crops in *rabi* season.

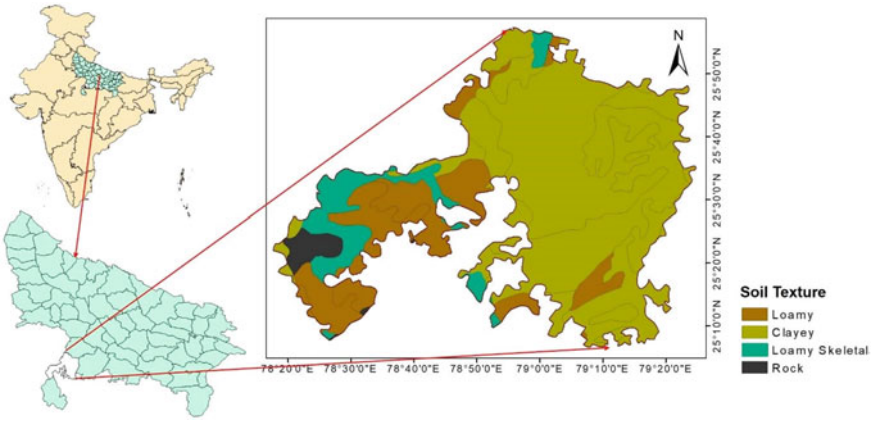

Fig. 1 Location of Jhansi District of Uttar Pradesh state, India with soil textural classes (*Source* NBSS&LUP)

Table 1 Satellite data and their bands with vegetation indices

Satellite imagery/Indices	Bands	Significance
SENTINEL-1	VH	Helps in identification of temporal changes
SENTINEL-2	B2, B4 and B8	
NDVI	(B8 – B4)/(B8 + B4)	Vegetation identification

NDVI—*Normalized Difference Vegetation Index was calculated using band B8 and B4 ranges from −1 to 1, which shows the vegetation level and sensitive to changes;* ***VH****—Vertical–Horizontal polarized median value was extracted from Sentinel-1 imagery*

2.2 Image Acquisition

The present study uses Sentinel-2 and Sentinel-1 satellite imagery for classification (Table 1). The median Sentinel-1 and Sentinel-2 images were collected for November 2018 to March 2019.

2.3 Image Classification

- ***Support Vector Machines (SVM)***

SVM is supervised learning algorithm that analyzes data in classification and regression, which also has an important property of optimization. SVM (Cortes and Vladimir 1995) mainly works on the concept of margin, calculates the minimal distance between the decision boundary and points, followed by creating a margin, which is perpendicular distance between decision boundary and nearby points. The decision boundary is determined by the support vectors (subset of data points). SVM performs both linear and non-linear classification. In linear classification, a line separates the space data points. Non-linear classification is performed using kernel tricks that maps the inputs into high-dimensional feature spaces.

- **Classification and Regression Tree (CART)**

CART is mainly a decision tree supervised classifier that takes the inputs of training data using the theory of information entropy. At every node of the tree, the attribute of training data splits into its subsets and is then enriched into suitable class. The major benefit of using tree models is capturing the non-linearity in the dataset and there is no need for data standardization because they don't calculate distances between the data. The main disadvantage of CART is mostly depending upon training data, small error in training data cause significant change in classification, i.e. high sensitivity.

Table 2 Pros and Cons of SVM, RF and CART classifiers

Classifiers	Pros	Cons
SVM	1. Solo Classes can be easily classified	1. High Processing time
	2. Solving higher dimensions	2. Poor in classifying mixed classes
RF	1. Better correlation of trees, i.e. classes	1. Need predictive power for features
	2. Handling errors of trees	2. Moderate in classifying mixed classes
CART	1. Easy to perform and visualization	1. Likely to be overfitted
	2. Automatic class selection	2. High sensitive to training data

- **Random Forests (RFs)**

RF classifier (Breiman 2001) is also a decision tree supervised classifier, which rectifies the disadvantage of CART classifier, i.e. sensitivity with construction of group of Decision trees. RF assembles the DT at training time and results into classification of the individual trees, which corrects the overfitting of training data.

- **Pros and Cons of Machine Learning Algorithms**

See (Table 2).

- **Spectral Matching Techniques (SMT)**

SMT is one of the well-known techniques in crop classification using NDVI time series data. This technique mainly generates ideal spectral signatures using ground data and correlates with other spectral signatures to label the classes. First, the process starts with downloading the sentinel-2 satellite data (Fig. 2) for November 2018 to March 2019, i.e. *rabi* season and generating maximum NDVI images for every month and stacking in order. Then, run the unsupervised classification on NDVI stacked image with 50 classes, which were grouped based on spectral similarity.

Ideal spectra signatures were prepared using ground data by extraction of spectral values, i.e. NDVI values. 50 classes were compared with ideal spectral signatures, ground data and classes with alike time series were combined into a single class and labelled accordingly. If any disparities in classes, then that class was masked out and the above process is repeated. Accuracy assessment was conducted using independent validation data.

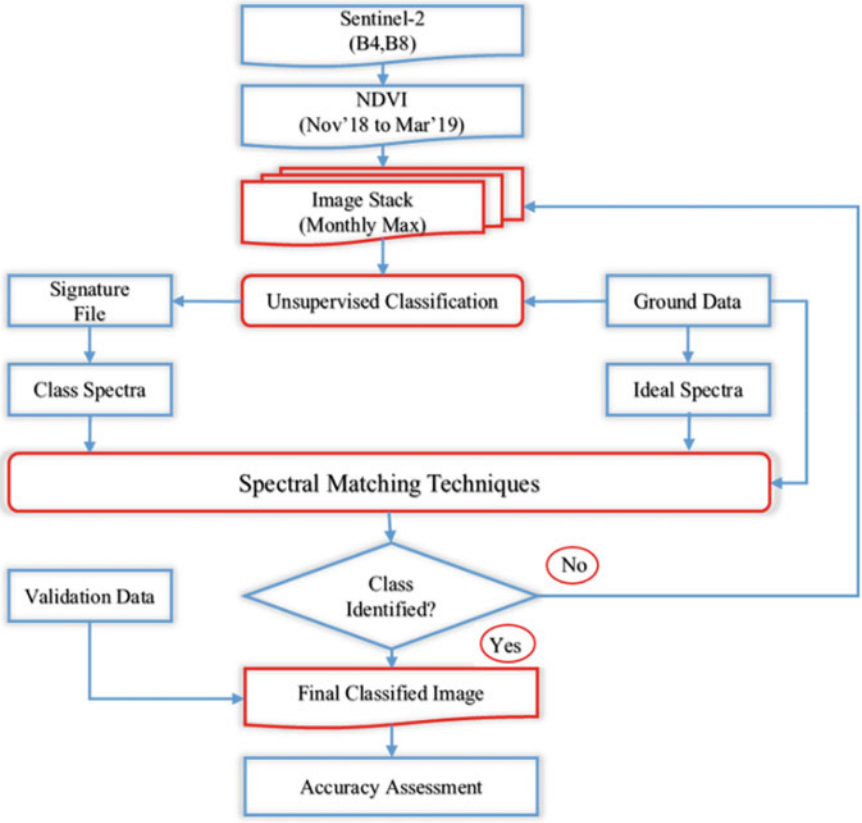

Fig. 2 Flow chart showing methodology of SMT for crop classification and accuracy assessment

2.4 Classification Procedure in GEE Platform

The classification procedure started with preparing image stack from Nov'2018 to Mar'2019 taking median values of NDVI using Sentinel 2 and VH from Sentinel 1 image (Fig. 3). Crop Mask obtained from SMT image was applied on stacked image to extract croplands. Ground data was divided into training and validation data in the ratio of 3:2 randomly and classified the stacked image using ground data with three well-known classification algorithms, i.e. CART, RF and SVM. Accuracy assessment was carried out using independent validation data for every algorithm.

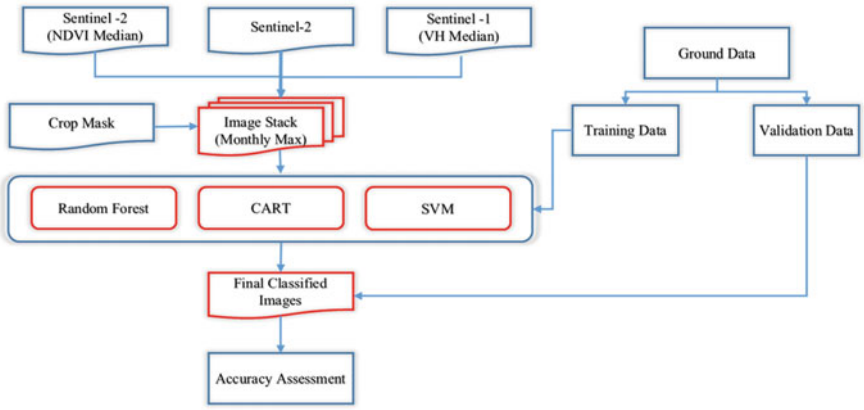

Fig. 3 Flow chart showing methodology used in GEE for crop classification and accuracy assessment

2.5 Ground Data

Ground data collection was carried out during the month of December 2018 through Jhansi districts covering major croplands and other LULC. Overall, 195 samples were collected out of which 77 samples were used for accuracy assessment (Fig. 4).

During ground data collection, minimum size of 90 × 90 m homogenous croplands were considered for sampling that were recorded using handheld Garmin GPS unit. Parameters like GPS, Crop parameters like crop intensity, crop pattern and crop growth, and irrigation techniques were recorded along with field photographs. The distribution of crop-wise ground samples is shown in above Table 3.

3 Results and Discussion

3.1 Spatial Distribution of Crop Classification

The spatial distribution of crops in Jhansi district using different ML algorithms showed in Fig. 5. The classification maps were obtained from RF algorithm, CART algorithm, SVM algorithm and SMT with classes Wheat, Chickpea, Mustard and Mixed crops.

Using Sentinel-2 (10 m) data, we can easily extract croplands and can easily identify the crop patterns up to field level. It was observed that majority part of crop class was wheat and it spreads across the district followed by chickpea crop. Majority of the mustard is sowed in northern part and some parts of southeast of Jhansi district. In most of the area, it was observed chickpea as mono-crop especially in eastern and central part of Jhansi district. During ground data collection it was observed that

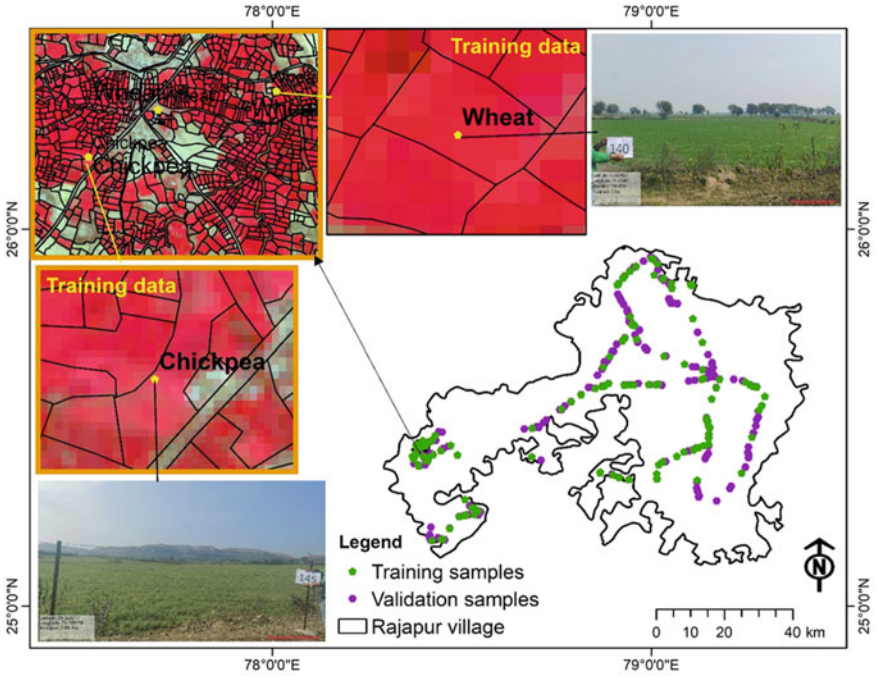

Fig. 4 Ground data collection for Jhansi district; training and validation data used for crop classification

Table 3 Distribution of samples of crops in both training and validation ground Data

Ground data		
Class	Training points	Validation points
Wheat	45	31
Chickpea	39	22
Mustard	20	14
Mixed crops	18	10
Total	122	77

majority of crops in Jhansi district have intermixing crops like chickpea-mustard and wheat-pea.

Figure 6 shows the field level crop indicated red mark in Fig. 5 classification for all classified images and compared with Google Earth High Resolution Image (GEHRI) and nominal Landsat 30 m cropland product. Landsat 30 m cropland contains some other LULC but we successfully classified and eliminated other LULC like built-ups, shrubs and barren, etc., using Sentinel-2 10 m data.

The field level distribution of crops shows that SMT classified mono and mixed crops with good accuracy. But ML algorithms got some gaps in classifying mono

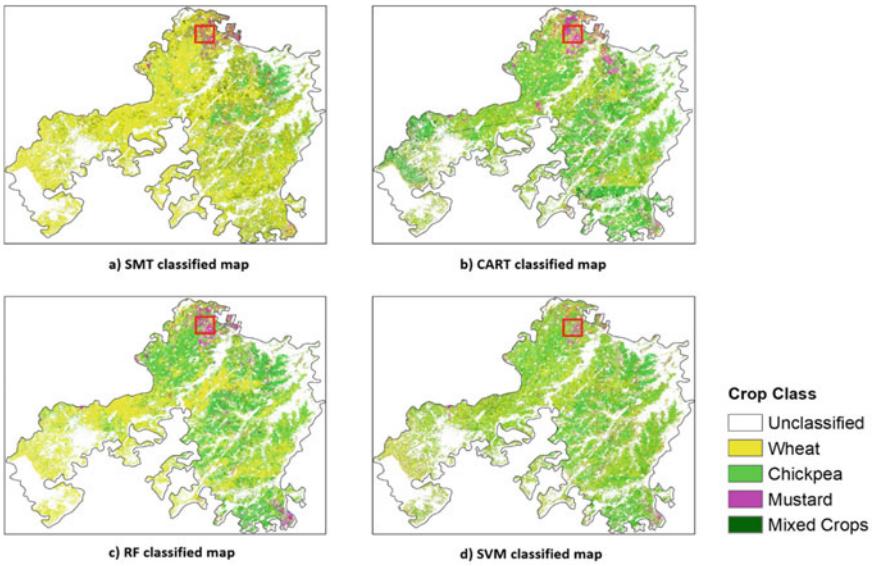

Fig. 5 Spatial Distribution of crops in Jhansi District—SMT classified image (**a**), CART classified image (**b**), RF classified image (**c**) and SVM classified image (**d**) along with selected area for field observation

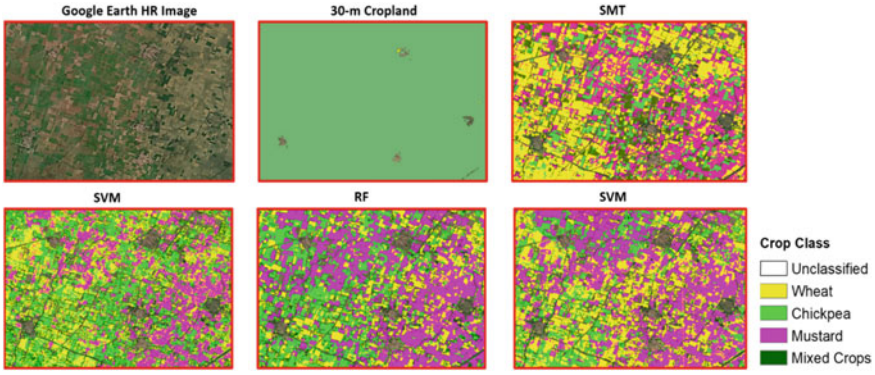

Fig. 6 Field level crop distribution in Jhansi District comparing Google Earth Image, 30 m cropland, SMT classified image, SVM classified image, RF classified image, CART classified image

and mixed crops. RF algorithm classified mustard and chickpea correctly but got some errors in classifying wheat, whereas, CART and SVM algorithms classified all classes with some errors. These errors were mainly due to misperception in deciding the class between mono and mixed crops as both consist of nearly same signatures. These errors can be reduced by using extensive and quality ground data.

3.2 Accuracy Assessment

Accuracy assessment was carried out for classified image obtained using validation data with the help of error matrices. The error matrix of RF classified image obtained overall accuracy of nearly 82% with kappa coefficient 0.735 and obtained good accuracy in both User's and Producers accuracy (Table 4). In RF, some of the chickpea is misclassified as other crops and also mixed crops were misclassified as wheat and chickpea. It indicates that the mixed crops show nearly same signatures of mono-crops in some areas.

SVM algorithm achieved overall accuracy of nearly 69% with kappa coefficient 0.558 and obtained nearly 70% in both user and producers accuracy (Table 5). In SVM, some of the Wheat class was misclassified as mixed crops, whereas, chickpea is misclassified as wheat. Mustard class is misclassified as chickpea class as it was a mixed crop with mustard.

CART algorithm achieved overall accuracy of nearly 65% with kappa coefficient 0.510 and obtained nearly 65% in both user and producers accuracy in every class (Table 6). In CART classification, some of the wheat class was misclassified as chickpea and chickpea crop was misclassified as wheat crop. Some of the mustard crop was misclassified as wheat and chickpea.

Table 4 Accuracy assessment for the crop type map of RF classified image. Map classes are rows and reference classes are columns

Random forest	Wheat	Chickpea	Mustard	Mixed crops	Total	Users
Wheat	30	1	0	0	31	0.968
Chickpea	2	17	1	2	22	0.773
Mustard	2	1	11	0	14	0.786
Mixed crops	3	2	0	5	10	0.500
Total	37	21	12	7	77	
Producers	0.811	0.810	0.917	0.714	**Overall**	**0.818**
					Kappa	**0.735**

Table 5 Accuracy assessment for the crop type map of SVM classified image. Map classes are rows which reference classes are the columns

SVM	Wheat	Chickpea	Mustard	Mixed crops	Total	Users
Wheat	23	2	1	5	31	0.742
Chickpea	5	15	1	1	22	0.682
Mustard	3	2	8	1	14	0.571
Mixed crops	2	1	0	7	10	0.700
Total	33	20	10	14	77	
Producers	0.697	0.750	0.800	0.500	**Overall**	**0.688**
					Kappa	**0.558**

Table 6 Accuracy assessment for the crop type map of CART classified image. Map classes are rows which reference classes are the columns

CART	Wheat	Chickpea	Mustard	Mixed crops	Total	Users
Wheat	20	6	2	3	31	0.645
Chickpea	4	15	2	1	22	0.682
Mustard	1	1	10	2	14	0.714
Mixed crops	1	4	0	5	10	0.500
Total	26	26	14	11	77	
Producers	0.769	0.577	0.714	0.455	**Overall**	**0.649**
					Kappa	**0.510**

Table 7 Accuracy assessment for the crop type map of SMT classified image. Map classes are rows which reference classes are the columns

SMT	Wheat	Chickpea	Mustard	Mixed crops	Total	Users
Wheat	30	1	0	0	31	0.968
Chickpea	2	18	1	1	22	0.818
Mustard	1	1	12	0	14	0.857
Mixed crops	0	1	1	8	10	0.800
Total	33	21	14	9	77	
Producers	0.909	0.857	0.857	0.889	**Overall**	**0.883**
					Kappa	**0.833**

SMT classified image achieved overall accuracy of nearly 88% with kappa coefficient 0.833 and obtained greater than 80% in both user and producers accuracy in every class (Table 7). Every class was classified with good accuracy. Some of the chickpea crop was misclassified as wheat class and also mixed crop into chickpea and mustard.

3.3 Class Wise Assessment in Terms of User's and Producer's Accuracy

The Fig. 7 shows the comparison of user's accuracy for every individual classified class. It was observed that in wheat class, SMT and RF achieved more than 90% and in chickpea and mustard class, it was near 80%. Other classifiers achieved nearby 70% in all classes.

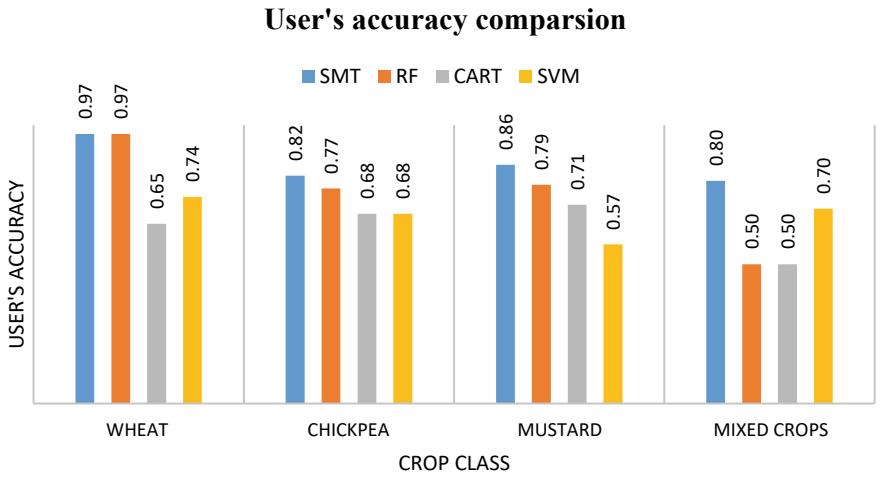

Fig. 7 Comparison of user's accuracy of classified classes, i.e. Wheat, Chickpea, Mustard and Mixed Crops

The Fig. 8 shows the comparison of producer's accuracy for every individual class. We can observe in wheat, chickpea and mustard classes, SMT and RF achieved more than 80%, whereas, other classifiers near to 60% and less.

By looking at the above statistics, it was noticed that most of the classes were classified with good accuracy using SMT and RF algorithm.

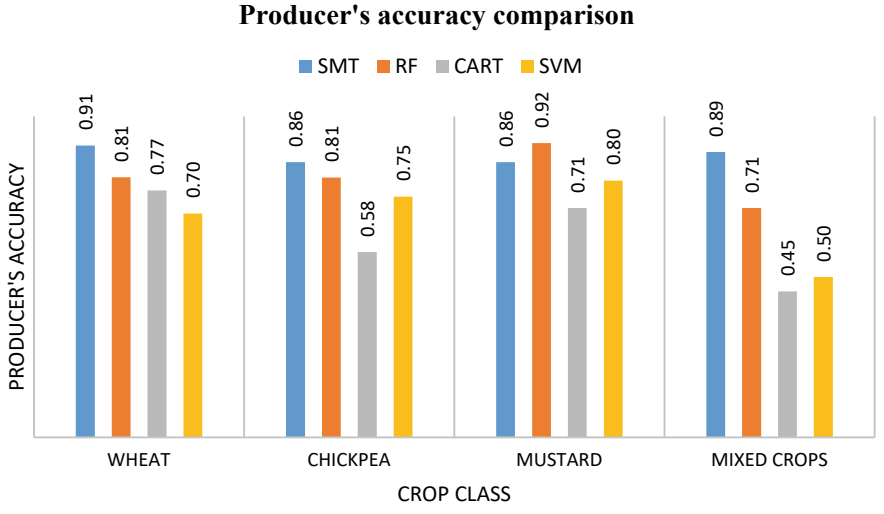

Fig. 8 Comparison of producer's accuracy of classified classes, i.e. Wheat, Chickpea, Mustard and Mixed Crops

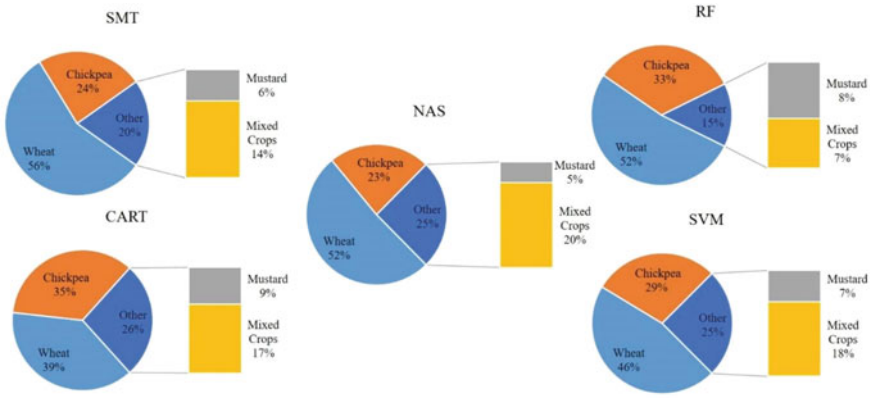

Fig. 9 Comparison of spatial extracted areas using SMT, RF, CART and SVM with National Agriculture Statistics (Meier et al. 2018)

3.4 Comparing Extracted Areas with NAS Statistics

We have compared spatial extracted crop areas with National Agriculture Statistics (Meier et al. 2018.). We have noticed that the areas obtained from SMT classified image were highly correlated with NAS areas in terms of all crops (Fig. 9).

In ML algorithms, RF classified image shows good relation with NAS in terms of wheat crop and mustard crop. But in terms of chickpea, it shows higher area compared to NAS because most of the chickpea crop in Jhansi district is intermixing crop. Other algorithms like CART and SVM show less % of areas compared to NAS. SVM classified image shows good correlation with NAS in terms of all crops but very less area in terms of wheat crop.

4 Discussion

The importance of variables in decision tree classifications, i.e. RF and CART can be identified using GEE code whereas it is not possible in identifying variable importance in linear kernel type SVM classifier as it is difficult to extract features.

The selection of variables plays important role in image classification based on theme of map. In this paper, the main theme was to map specific crop classification so, the variables monitoring the vegetation changes were important. Hence, Sentinel-1 VH and Sentinel-2 NDVI bands were considered for crop classification. Considering variable importance in RF Algorithm, VH band plays a vital role, i.e. about 92% importance in classifying image. Bands like B9, B1, B11, B8 and B2 have more than 80% importance in classification and NDVI with 76% (Fig. 10). It is important to select the small number of variables that work well for our theme than selecting large number of variables. Sometimes, selecting large number of variables might reduce

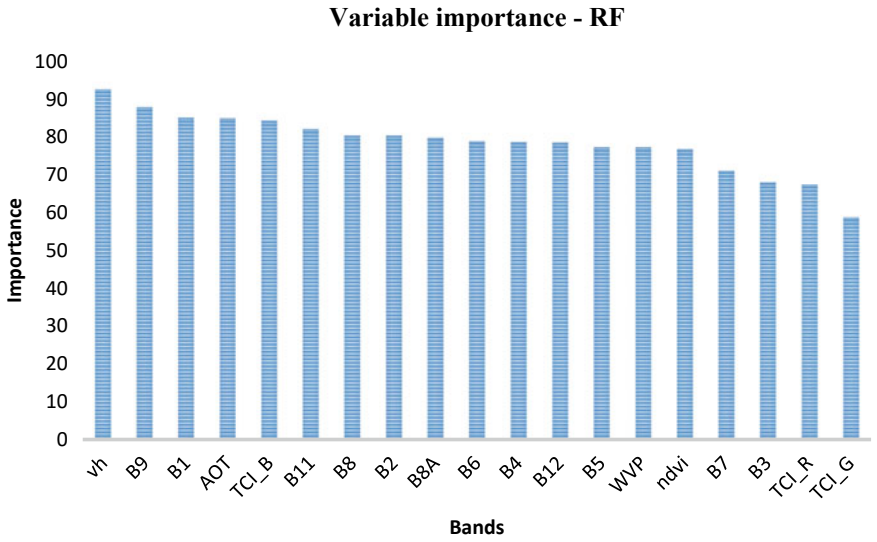

Fig. 10 Bar chart showing variables (Bands) Importance in RF algorithm

accuracy level. The variables with high permutations can train the RF algorithm efficiently results into good image classification.

CART classifier variables and their importance can be observed in Fig. 11. This shows that all the bands were of least importance in classification. B1 with nearly 3.5%% importance and other with less than 1.0%%. This shows that the classification using CART is purely depended on ground data.

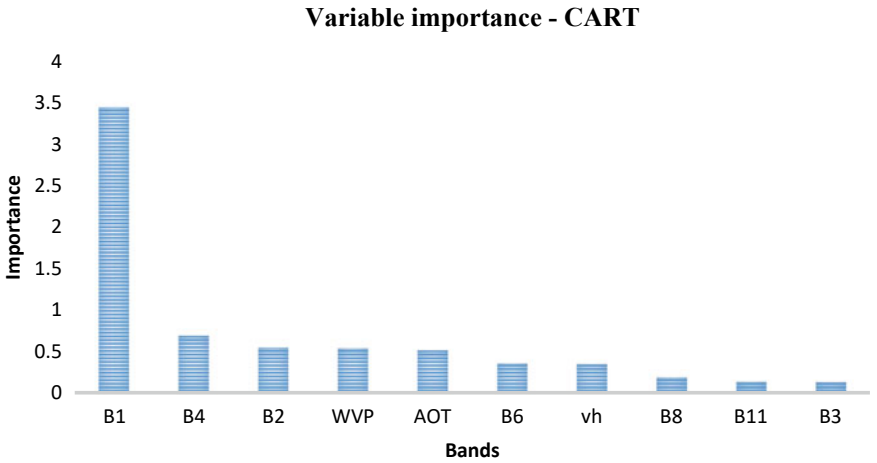

Fig. 11 Bar chart showing variables (Bands) Importance in CART algorithm

Considering accuracy assessment, SMT is better than RF classification. With SMT, initially satellite images were classified into high number of classes using unsupervised classes and later which were simplified into major classes using ground data spectra values but in RF, only classified image based on ground data training points using supervised classification. So, the accuracy of image using RF mostly depends upon equal distribution of training points. SMT needs satellite data to download whereas, using RF worked on GEE cloud and this reduces the time consumption for downloading and processing time. Using RF, it is possible to work for large areas but using SMT, it is a hectic task as it takes a lot of time and reaches computation limit.

In this study, major crop types were mapped using different ML techniques and semi-automated algorithms. SMT technique was used for mapping crop types using Sentinel-2 time series imagery with ground data which was used in ML algorithms. The process starts with unsupervised classification followed by creation of NDVI MVCs. In the process of class identification, the main advantage of this approach is the usage of monthly cloud-free or near-cloud-free images with creation of maximum value composites. In addition, ground data to identified different cropping systems across season. SMT's were predominantly successful in differentiating cropping patterns.

The same training data was used for all three ML algorithms and SMT. Since, ML algorithms work on different statistical approaches, which helps in identifying classes. Here, SVM algorithm calculates the minimum distance between decision boundary and samples, this approach increases the chance of mixing different crop classes. So, the high amount of training data is required to make efficient decision boundary whereas, CART algorithm works on DT where quality of training data plays a role in deciding classes using DT, i.e. error in training sample effect the class. These possibilities make SVM and CART cause less accuracy but in RF algorithm, rectifying the disadvantage of CART in deciding the class. So, RF achieved good accuracy as compared to other algorithms. With SMT, the decision of class was based on spectral signatures comparing with ideal spectral signatures. These help in understanding the signatures and helps in deciding the class. So, high accuracy was achieved using SMT.

5 Conclusions

The present study shows the classification of crop types at small holding farms using different ML and SMT techniques. Each classifier has its advantages in classifying various LULC. RF algorithm classified mono-crops correctly, however, it has its limits in classifying mixed crops. SMT classifies every class and attained good accuracy as it was done by reclassifying class disparities. RF is classified nearly comparable to SMT. Remaining classifiers like CART, SVM were classified satisfactorily and algorithm mistakenly classified other crops. It was observed that ground data plays a key role in classification of crop types. RF classification can

run efficiently using some ground points whereas CART requires high amount of ground data. RF classification can neutralize error points using decision trees, but in SVM, error points affect the classification. It is concluded that every classifier with its limits of ground data can efficiently classifies the map. Mapping crop types at high resolution is very important for macro-level planning, yield predictions and crop stresses to support small holder farmers. High-resolution specific crop maps are important inputs for assessing yield predictions at village level and national food security studies. The classification limitation arises when we used satellite data during *kharif* season because of high cloud cover. This limitation can be rectified using Synthetic Aperture Radar data, i.e. Sentinel-1 data. This work was carried out in dryland area in which most of the croplands were small land patches, i.e. small holding farmers and having complex cropping patterns. So, it enables us to generate specific crop maps using four different ML methods and assessed the performance of those methods.

Acknowledgements This research was supported by the RIICE III project funded by the Swiss Agency for Development Cooperation (SDC), the Government of Uttar Pradesh, India, through its 'Doubling Farmers Income' in Bundelkhand, and the CGIAR Research Program on Water, Land and Ecosystems (WLE). The authors are grateful to Divya Kashyap, Arindom Baidya, and other team members for their suggestions and recommendations. We also thank Jameeruddin A and Ismail for their support during ground data collection.

References

Anderson MC, Allen RG, Morse A, Kustas WP (2012) Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens Environ 122:50–65

Belgiu M, Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens Environ 204:509–523

Bicheron P, Huc M, Henry C, Bontemps S, Lacaux J (2008) GlobCover products description manual, Issue 2, Revision 2. European Space Agency (ESA)

Boche H, Mönich UJ (2020) Turing computability of Fourier transforms of bandlimited and discrete signals. IEEE Trans Signal Process 68:532–547

Brown JC. Jepson WE. Kastens JH. Wardlow BD. Lomas JM. Price K. (2007) Multitemporal, moderate-spatial-resolution remote sensing of modern agricultural production and land modification in the Brazilian Amazon. GISci. Remote Sens. 44: 117–148

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Coltin B, McMichael S., Smith T, Fong T (2016) Automatic boosted flood mapping from satellite data. Int J Remote Sens 37:993–1015

Cortes C, Vladimir V (1995) Support-vector networks. Mach Learn 20(3):273–297

Deng C, Wu C (2013) The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. ISPRS J Photogramm Remote Sens 86:100–110

Dong J, Xiao X, Kou W, Qin Y, Zhang G, Li L, Jin C, Zhou Y, Wang J, Biradar C (2015) Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. Remote Sens Environ 160:99–113

Dong J, Xiao X, Menarguez MA, Zhang G, Qin Y, Thau D, Biradar C, Moore III B (2016) Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens Environ 185:142–154

Egorov A, Hansen M, Roy DP, Kommareddy A, Potapov P (2015) Image interpretation-guided supervised classification using nested segmentation. Remote Sens Environ 165:135–147

Feyisa GL, Palao LK, Nelson A, Gumma MK, Paliwal A, Win KT, Nge KH, Johnson DE (2020) Characterizing and mapping cropping patterns in a complex agro-ecosystem: an iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. Comput Electron Agr 175:105595

Friedl MA, Brodley CE (1997) Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61:399–409

Gómez C, White JC, Wulder MA (2016) Optical remotely sensed time series data for land cover classification: A review. ISPRS J Photogramm Remote Sens 116:55–72

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27

Griffiths P, Nendel C, Hostert P (2019) Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sens Environ 220:135–151

Gumma MK, Thenkabail PS, Deevi KC, Mohammed IA, Teluguntla P, Oliphant A, Xiong J, Aye T, Whitbread AM (2018) Mapping cropland fallow areas in myanmar to scale up sustainable intensification of pulse crops in the farming system. GISci Remote Sens 55:926–949

Gumma MK, Thenkabail PS, Maunahan A, Islam S, Nelson A (2014) Mapping seasonal rice cropland extent and area in the high cropping intensity environment of Bangladesh using MODIS 500 m data for the year 2010. ISPRS J Photogramm Remote Sens 91:98–113

Gumma MK, Thenkabail PS, Teluguntla PG, Oliphant A, Xiong J, Giri C, Pyla V, Dixit S, Whitbread AM (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. GISci Remote Sens 57:302–322

Gumma MK, Thenkabail PS, Teluguntla P, Rao MN, Mohammed IA, Whitbread AM (2016) Mapping rice-fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data. Int J Digital Earth 9(10):981–1003

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova S, Tyukavina A, Thau D, Stehman S, Goetz SJ, Loveland TR (2013) High-resolution global maps of 21st-century forest cover change. Science 342:850–853

Harris FJ (1978) On the use of windows for harmonic analysis with the discrete Fourier transform. Proc IEEE 66:51–83

Im J (2020) Earth observations and geographic information science for sustainable development goals. GISci Remote Sens 57:591–592

Jeganathan C, Dash J, Atkinson P (2014) Remotely sensed trends in the phenology of northern high latitude terrestrial vegetation, controlling for land cover change and vegetation type. Remote Sens Environ 143:154–170

Lambert M-J, Traoré PCS, Blaes X, Baret P, Defourny P (2018) Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sens Environ 216:647–657

Liu X, Hu G, Chen Y, Li X, Xu X, Li S, Pei F, Wang S (2018) High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. Remote Sens Environ 209:227–239

Lobell DB, Thau D, Seifert C, Engle E, Little B (2015) A scalable satellite-based crop yield mapper. Remote Sens Environ 164:324–333

Long JA, Lawrence RL, Greenwood MC, Marshall L, Miller PR (2013) Object-oriented crop classification using multitemporal ETM+SLC-off imagery and random forest. GISci Remote Sens 50:418–436

McCarthy MJ, Radabaugh KR, Moyer RP, Muller-Karger FE (2018) Enabling efficient, large-scale high-spatial resolution wetland mapping using satellites. Remote Sens Environ 208:189–201

Meier J, Zabel F, Mauser W (2018) A global approach to estimate irrigated areas–a comparison between different data and statistics. Hydrol Earth Syst Sci 22:1119

Millard K, Richardson M (2015) On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sens 7:8489–8515

Oliphant AJ, Thenkabail PS, Teluguntla P, Xiong J, Gumma MK, Congalton RG, Yadav K (2019) Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud Int J Appl Earth Obser Geoinf 81:110–124

Ozdogan M, Gutman G (2008) A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: an application example in the continental US. Remote Sens Environ 112:3520–3537

Pan Z, Huang J, Zhou Q, Wang L, Cheng Y, Zhang H, Blackburn GA, Yan J, Liu J (2015) Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. Int J Appl Earth Obser Geoinf 34:188–197

Patel NN, Angiuli E, Gamba P, Gaughan A, Lisini G, Stevens FR, Tatem AJ, Trianni G (2015) Multitemporal settlement and population mapping from Landsat using Google Earth Engine. Int J Appl Earth Obser Geoinf 35:199–208

Pekel J-F, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. Nature 540:418–422

Pittman K, Hansen MC, Becker-Reshef I, Potapov PV, Justice CO (2010) Estimating global cropland extent with multi-year MODIS data. Remote Sens 2:1844–1863

Platas-Garza MA, de la O Serna JA (2010) Dynamic harmonic analysis through Taylor–Fourier transform. IEEE Trans Instrum Meas 60:804–813

Sidhu N, Pebesma E, Câmara G (2018) Using Google Earth Engine to detect land cover change: Singapore as a use case. Eur J Remote Sens 51:486–500

Tateishi R, Hoan NT, Kobayashi T, Alsaaideh B, Tana G, Phong DX (2014) Production of global land cover data-GLCNMO2008. J Geog Geol 6:99

Teluguntla P, Thenkabail PS, Oliphant A, Xiong J, Gumma MK, Congalton RG, Yadav K, Huete A (2018) A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J Photogramm Remote Sens 144:325–340

Thenkabail P, GangadharaRao P, Biggs T, Krishna M, Turral H (2007) Spectral matching techniques to determine historical land-use/land-cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR pathfinder datasets. Photogramm Eng Remote Sens 73:1029–1040

Thenkabail PS, Biradar CM, Noojipady P, Dheeravath V, Li Y, Velpuri M, Gumma M, Gangalakunta ORP, Turral H, Cai X (2009). Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int J Remote Sens 30:3679–3733

Thenkabail PS, Knox JW, Ozdogan M, Gumma MK, Congalton RG, Wu Z, Milesi C, Finkral A, Marshall M, Mariotto I (2012) Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help? PE&RS, Photogramm Eng Remote Sens 78:773–782

Veloso A, Mermoz S, Bouvet A, Le Toan T, Planells M, Dejoux J-F, Ceschia E (2017) Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sens Environ 199:415–426

Xie Y, Lark TJ, Brown JF, Gibbs HK (2019) Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine. ISPRS J Photogramm Remote Sens 155:136–149

Xiong J, Thenkabail PS, Gumma MK, Teluguntla P, Poehnelt J, Congalton RG, Yadav K, Thau D (2017a) Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J Photogramm Remote Sens 126:225–244

Xiong J, Thenkabail PS, Tilton JC, Gumma MK, Teluguntla P, Oliphant A, Congalton RG, Yadav K, Gorelick N (2017b). Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. Remote Sens 9:1065