Crop type identification and spatial mapping using Sentinel-2 satellite data with focus on field-level information

Murali Krishna Gumma , Kimeera Tummala , Sreenath Dixit , Francesco Collivignarelli , Francesco Holecz , Rao N. Kolli & Anthony M. Whitbread

Check for updates

Crop type identification and spatial mapping using Sentinel-2 satellite data with focus on field-level information

Murali Krishna Gumma[a] (iD), Kimeera Tummala[a], Sreenath Dixit[a] (iD), Francesco Collivignarelli[b] (iD), Francesco Holecz[b], Rao N. Kolli[c] (iD) and Anthony M. Whitbread[d] (iD)

[a]RS/GIS Lab, Innovation Systems for the Drylands, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India; [b]sarmap, Caslano, Switzerland; [c]International Reinsurance and Insurance Consultancy and Broking Services Pvt. Ltd, Mumbai, India; [d]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Tanzania

ABSTRACT

Accurate monitoring of croplands helps in making decisions (for insurance claims, crop management and contingency plans) at the macro-level, especially in drylands where variability in cropping is very high owing to erratic weather conditions. Dryland cereals and grain legumes are key to ensuring the food and nutritional security of a large number of vulnerable populations living in the drylands. Reliable information on area cultivated to such crops forms part of the national accounting of food production and supply in many Asian countries, many of which are employing remote sensing tools to improve the accuracy of assessments of cultivated areas. This paper assesses the capabilities and limitations of mapping cultivated areas in the Rabi (winter) season and corresponding cropping patterns in three districts characterized by small-plot agriculture. The study used Sentinel-2 Normalized Difference Vegetation Index (NDVI) 15-day time-series at 10 m resolution by employing a Spectral Matching Technique (SMT) approach. The use of SMT is based on the well-studied relationship between temporal NDVI signatures and crop phenology. The rabi season in India, dominated by non-rainy days, is best suited for the application of this method, as persistent cloud cover will hamper the availability of images necessary to generate clearly differentiating temporal signatures. Our study showed that the temporal signatures of wheat, chickpea and mustard are easily distinguishable, enabling an overall accuracy of 84%, with wheat and mustard achieving 86% and 94% accuracies, respectively. The most significant misclassifications were in irrigated areas for mustard and wheat, in small-plot mustard fields covered by trees and in fragmented chickpea areas. A comparison of district-wise national crop statistics and those obtained from this study revealed a correlation of 96%.

CONTACT Murali Krishna Gumma ✉ m.gumma@cgiar.org; gummamk@gmail.com

1. Introduction

Dryland agriculture covers about 68% of India's total cropland area, characterized by lower mean annual rainfall compared to potential evaporation. Dryland areas are highly vulnerable because of high variability in rainfall. Forty percent of the rural population and 60% of the livestock population directly depend on dryland agriculture (Singh et al. 2004; Misra et al. 2010). The changing character and extent of dryland agriculture over the years underlines the importance of monitoring croplands continuously to ensure sustainable food production. Information on district-level cropland extent and statistics are important for decision making, not only with regard to issues concerning sustainable agriculture and food security, but also on crop insurance.

Crop area statistics collected by different government organizations (departments of district revenue, agriculture and irrigation) generally show deviations of up to 30% and don't truly correspond to the effective seasonal cultivated area (Van Genderen et al. 1978; Stehman et al. 2003; Gallego 2004; Biggs et al. 2006). Satellite images provide a valid alternative, particularly when the area needs to be estimated at the state or national level (Thiruvengadachari and Sakthivadivel 1997; Lobell et al. 2003; Thenkabail et al. 2010). Studies carried out in the past decades, based mainly on coarse resolution data such as Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS), demonstrated that cropland extent derived from time-series data has higher accuracy than a single-date image (Gómez et al. 2016; Teluguntla et al. 2017; Xiong, Thenkabail, Gumma, et al. 2017; Xu et al. 2018). Studies based on Landsat-8 and Sentinel-2 single-date images have shown the advantages of moderate resolution images for crop mapping, particularly by exploiting object-based methods (Walker and Blaschke 2008; Peña-Barragán et al. 2011; Petitjean et al. 2012; Long et al. 2013; Matton et al. 2015; Belgiu and Csillik 2018) and the region merging approach (Moscheni et al. 1998).

The utility of Sentinel-2 data for agricultural applications, and specifically crop type mapping, has been evaluated by many researchers. European Space Agency's sen2agri (Sentinel-2 for agriculture) project generates a crop type map with 4-5 main crops of the region at 10 m spatial resolution using the Random Forest-based methodology proposed by Inglada J et al (Inglada et al. 2015). Vijayasekaran (Vijayasekaran 2019) assessed the sen2agri products for the Indian agricultural scenario and reported an overall classification accuracy of 70% on an average for various crops.

In recent times, machine-learning algorithms like Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) have been used by researchers to classify satellite data and its derivatives into crop types. Saini et al. (Saini and Ghosh 2018) performed crop classification on a single-date Sentinel-2 image using RF and SVM algorithms. Sun C et al. used a combination of spectral bands, derived textural measures, and vegetation indices from Sentinel-1 SAR, Sentinel-2 optical and Landsat-8 optical datasets to perform crop classification using machine-learning methods SVM, ANN, and RF (Sun et al. 2019). More similarly, Feng (Feng et al. 2019)used Sentinel-2 time-series data alone for mapping crop types. They successfully used spectral bands, texture parameters, vegetation indices, and phenological parameters derived from Sentinel-2 time-series data as inputs to RF and SVM classification algorithms. They examined the use of short-wave infra-red (SWIR) and water vapour bands of Sentinel-2 data for differentiating between crops.

Li Q et al. applied a maximum likelihood classifier on a combination of Landsat TM 30 m resolution data and features related to crop growth extracted from MODIS NDVI time-series data (Li et al. 2014). Belgiu M and Csilik O performed both object- and pixel-

based classification using time-weighted dynamic time warping and RF analyses on Sentinel-2 time series NDVI images (Belgiu and Csillik 2018). Griffiths P et al. integrated Sentinel-2 and Landsat-8 data to generate an intra-annual time-series composite of ten days to generate a national level crop-type map of Germany using Random Forest classification algorithm (Griffiths et al. 2019).

Many studies have applied SMTs on optical data to obtain information on different land uses and land covers (LULC) (Thenkabail et al. 2007, Gumma et al. 2016, Gumma, Thenkabail, Teluguntla et al. 2018). Thenkabail et al. used SMTs to map historical LULC and irrigated areas using time-series AVHRR pathfinder datasets for the Krishna River Basin, India. Gumma et al. (2011) applied SMTs to map rice ecosystems in South Asia for 2000-01 using MODIS NDVI monthly maximum value composite (MVC) data; to monitor the spatial expansion of chickpea (Gumma et al. 2016) and to differentiate between canal and groundwater irrigation areas in the Krishna River basin using MODIS 250 m data (Gumma et al. 2011). Recently, studies using MODIS NDVI time-series have been reported on mapping rice-fallows for crop intensification using short duration legumes employing NDVI time- series data (Gray et al. 2014, Gumma, Thenkabail, Deevi, et al. 2018), and mapping drought prone areas across India (Gumma et al. 2019). Coarse resolution data were found to be effective in estimating the extent of homogeneous land cover areas, but failed to accurately estimate the area, particularly in small-plot agriculture (Gumma et al. 2014).

In the present study, an attempt was made to apply the SMT using high-resolution (10 m) satellite data at district-level. The focus was on crop type mapping with crop insurance and crop yield modeling as the major consumption areas. With rising emphasis on crop insurance in India, it has become imperative for the stakeholders to have access to the spatial information of crop type, crop health and stress at field level. This information also needs to be available for large areas (districts or states), and include information about all crops grown in those areas. The presence of a wide variety of crops in a small area has always posed difficulties in capturing the complete spatial variability of crops in the drylands of India. In the present study, these drawbacks have been focused on, and improvements have been made to field data collection, data processing including quality check, and spectral matching techniques for classification, in order to obtain the best results for the particular concerns mentioned above. The classification results for *rabi* season of 2018-19 in three dryland districts of India – Jhansi, Chitrakoot and Panna – were studied. The results were validated against ground validation points and area statistics provided by state agriculture departments. A part of a larger study, this study discusses the advantages and shortcomings of using optical data (Sentinel-2; 10 m spatial resolution) for crop type mapping in particular.

2. Study area and data

2.1. Study area

The three study districts of Jhansi, Chitrakoot and Panna are located in Bundelkhand region in India (Figure 1) that lies between $23^050'$ N and $26^000'$ N latitudes and $78^000'$ E and $82^000'$ E longitudes. Bundelkhand, spread across the states of Uttar Pradesh and Madhya Pradesh, is subject to a dry sub-humid and semi-arid climate. Jhansi district is located in western Uttar Pradesh that spans 5,024 km^2, with a population of about 1.998 million with a density of about 400/km^2. The district lies in the plains region of Bundelkhand, and thus, has a larger area (70% of the total area) under cultivation. This
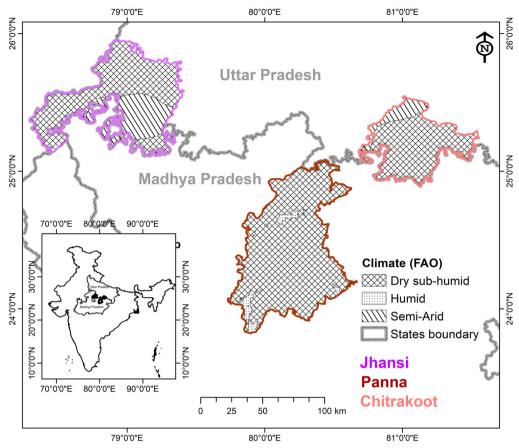
Figure 1. Study districts with location map showing climate zones. (www.FAO.org).

region is distinguished for its deep black soil, known as 'mar', and is very well suited for cotton cultivation. Chitrakoot district covers $3,216 \, km^2$ and has a population density of about $310/km^2$. The district lies in the intermediate region of Bundelkhand, which has a higher percentage of wasteland compared to the plains region. Consequently, only about 50% of the total area of the district is under cultivation. Panna district covers an area of $7,135 \, km^2$, and has a population density of $140/km^2$. Of the three districts, Panna has the highest percentage of wasteland and consequently, only 35% of its total area is under cultivation.

The size of landholdings not only affects agricultural production, but also determines the accuracy of spatial maps generated from satellite data. High variability in crops and small landholdings make it necessary to use satellite data of high spatial resolution that can capture the variability. In Jhansi and Chitrakoot districts, about 20% of the total cultivated area is under marginal landholdings (less than 1 ha) and in Panna district it is 11%. Small holdings (1-2 ha) constitute 22% of the cultivated area in Jhansi district, 21% in Chitrakoot district and 22% in Panna district. Most of the landholdings in the three districts are semi-medium (2-4 ha) and medium (4-10 ha) in size. Fifty-four percent of the cultivated area in Jhansi district, 46% in Chitrakoot district and 58% in Panna district are comprised of landholdings that are 2-10 ha in size (Agriculture Census Input Survey 2001-02). Sentinel-2 data with 10 m spatial resolution was chosen to map the small to

medium size agricultural plots, which form the majority of the crop area in the three study districts.

In Jhansi and Chitrakoot districts, the percentage of land under irrigation is lower than the state average for Uttar Pradesh. About 48% of sown area in Jhansi and about 28% in Chitrakoot is irrigated. In Panna, about 28% of sown area is irrigated (District-wise Land Use Statistics, Ministry of Agriculture, Government of India, May 2008). In all the three districts, more land is cultivated during rabi season than kharif season, owing to the availability of residual moisture and climatic conditions that are conducive to crop growth. In Jhansi district, the major rabi crop is wheat, followed by chickpea, peas and beans, and lentil. Mustard and barley are also cultivated in some areas. In Chitrakoot district, wheat and chickpea are cultivated in most areas during rabi season, followed by lentil. Barley, mustard, linseed, and peas and beans are found in small areas. In Panna district, chickpea is the major crop, followed by wheat, peas and beans. This district also has lentil, and small areas with mustard and linseed.

2.2. Sentinel-2 data

Sentinel-2 10 m resolution with 6-day surface reflectance from the EU Copernicus Programme is ideal for monitoring vegetation at a small scale (Xiong, Thenkabail, Tilton, et al. 2017). In this study, we used Sentinel-2 products, which provide 6-day composite images at 10 m spatial resolution (bands 2, 3, 4 and 8). Sentinel-2 products include blue, green, red, near infrared and mid-infrared bands. Four tiles covering the required region were downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). MAPScape was used to pre-process and mosaic the tiles of the study area, and then stack them as a single composite. Cloud contamination can be severe spanning several consecutive 6-day composites, especially during early rabi seasons (October to January). We chose to address this by generating fortnightly (FN) composites and for use in conjunction with the 6-day composites. Each pixel in the Sentinel-2 dataset contains the best observation during the 15-day period that it covers. The data are described in greater detail in the Scientific Data set documentation for Sentinel-2 (Xiong, Thenkabail, Tilton, et al. 2017).

3. Methods and approaches

3.1. Ground survey data

Ground data were collected during 7–31 January 2019 for 732 sample points covering about 4000 km of road travel in the study districts (Figure 2, Table 1). Ground data were collected based on pre-classified output and Google Earth imagery, and a tracking GPS attached to image-processing software captured ground survey information while moving on the road. Detailed information certain locations were collected for training, i.e., class identification and labelling. Point-specific information was collected from 90 m × 90 m plots and consisted of GPS locations, land use categories, land cover percentages, cropping pattern during different seasons (through farmer interviews), crop types and watering method (irrigated, rainfed). For validation, information on crop type and point coordinates (latitude and longitude values) were only collected at point locations. Samples were obtained within large contiguous areas of a particular LULC. Sentinel-2 False colour Composites were used as additional ground survey information in class identification. A stratified-systematic sample design was adopted based on road network or footpath access. Where possible, sample sites were systematically located every 5 or 10 kilometres along
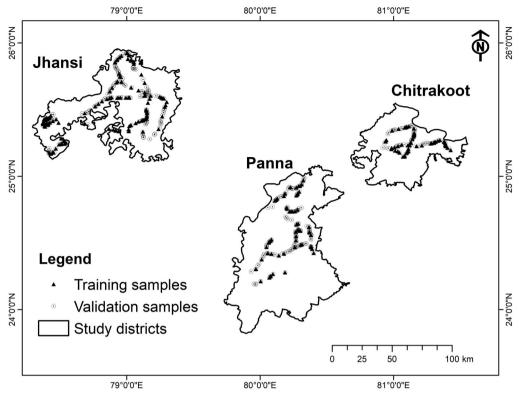
Figure 2. Ground survey data for different crops and spatial distribution of samples for training and validation.

Table 1. Number of training and validation samples collected for training data of each class identification and validation.

Jhansi ground truth information		
Sample class	Training	Validation
01. Wheat	68	79
02. Chickpea	16	11
03. Mustard/ beans	13	11
04. Wheat/mustard/chickpea	19	10
05. Pea beans	9	2
06. Other LULC	15	4
Total	140	117
Chitrakoot ground truth information		
Sample class	Training	Validation
01. Wheat	25	20
02. Chickpea	8	4
03. Mustard/lentil	11	31
04. Wheat/mustard/chickpea/lentil	5	43
05. Other crops	0	2
06. Fallow/ mixed crops	12	9
07. Other LULC	15	7
Total	76	116
Panna ground truth information		
Sample class	Training	Validation
01. Wheat	53	99
02. Chickpea	7	19
03. Wheat/ mustard/ lentil	19	33
04. Fallow/ other crops	9	7
05. Other LULC	16	11
Total	104	169

the road network by vehicle or on foot (Thenkabail et al. 2004; 2005), which is detailed in a description of the ground survey methodological approach.

We collected two independent data sets, one for training and another for validation. Table 1 shows sample size of training data for classification and validation data for accuracy assessment in each district.

3.2. Ideal NDVI temporal signatures

A prerequisite for employing SMT is to identify ideal NDVI temporal signatures for each class of interest. In this case, Sentinel-2 time-series and ground survey data collected from homogeneous patches along with in-depth information about the cropping system and irrigation methods were used to generate ideal temporal NDVI signatures. The ideal spectra for four major crops of Jhansi district – wheat, chickpea, peas/beans and mustard – are shown in Figure 3. In the spectra, the beginning of the rising spectral profile marks the germination. From Figure 3, it is evident that the germination time for wheat is mid-December, for chickpea it is mid-November, for peas/beans mid-November and for mustard early-mid November. The peak biomass period (peak of season) occurs in mid-late March in wheat, late December in chickpea, late January-early February in peas/beans and late January in mustard. Leaf senescence (end of season) can be observed in April in wheat, February in chickpea and March in peas/beans and mustard. These differentiating spectral characteristics were taken advantage of for separating various crops.

3.3. Seasonal cultivated area

Figure 4 provides an overview of the methodology used for mapping the rabi cropping pattern.

MAPscape-Basic software, which supports the processing of Landsat, Sentinel-2 and MODIS data and spaceborne Synthetic Aperture Radar data, was used to automatically download and generate time-series of various indices related to vegetation, bare soil, built-up, and water at a resolution of either 10 m or 20 m.

Sentinel-2 data for rabi cropping season, i.e., for the months of January, February and March, for the year 2019, was considered for the present analysis. The 15-day MVCs of NDVI were generated and stacked using functions supported by MAPscape-Basic. The K-means classifier was applied on the NDVI MVC time-series stack to generate clusters of similar pixels belonging to 70 classes. The NDVI MVC time-series spectra of these classes, known as class spectra, were plotted and compared. Classes with similar spectra were grouped using spectral similarity values (SSVs). The lower the SSV, the higher the similarity between the two classes. The classes were then identified and labelled using SMT (Gumma et al. 2014).

Intensive ground data were used for class identification. The ideal NDVI MVC spectra for each crop type were established based on 113 unique samples selected from the ground survey data. These samples were selected to effectively represent all major crop types, and in homogeneous areas that are large enough to avoid confusion while determining the crop type using the time-series spectra. For each class, the ideal spectra were taken to be the average of the spectra of all locations belonging to that class. The class spectra were compared with the ideal spectra and labelled with the land use of the matching spectra. For this step, ground reference data and Google Earth images were used to improve class identification.
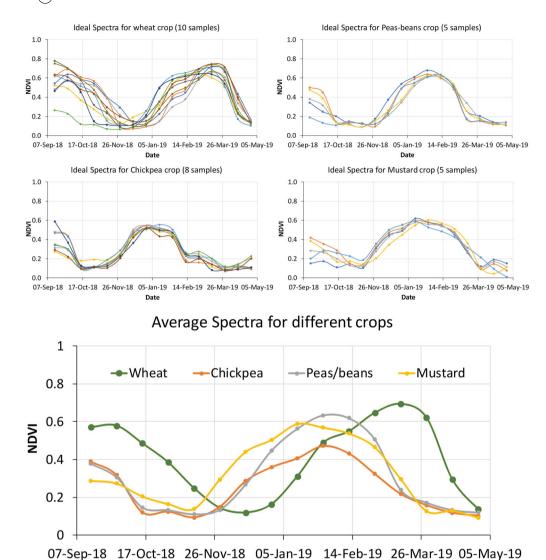
Figure 3. Ideal temporal profiles of NDVI for various crops (top); and temporal profiles of NDVI values from random locations.

3.4. Seasonal cultivated area validation

Accuracy assessment was performed based on the validation data. A total of 402 ground survey samples (Panna = 169, Chitrakoot = 116 and Jhansi = 117) were used to assess the accuracy of the classification results, based on a theoretical description given by Jensen (Jensen 2004), to generate a confusion matrix. The columns of a confusion matrix contain the field-plot data points, and the rows represent the results of the classified crop maps (Congalton 1991). Accuracy measure is Kappa, producer, user and overall accuracy (Cohen 1960), representing the agreement among user and reference ground survey data.
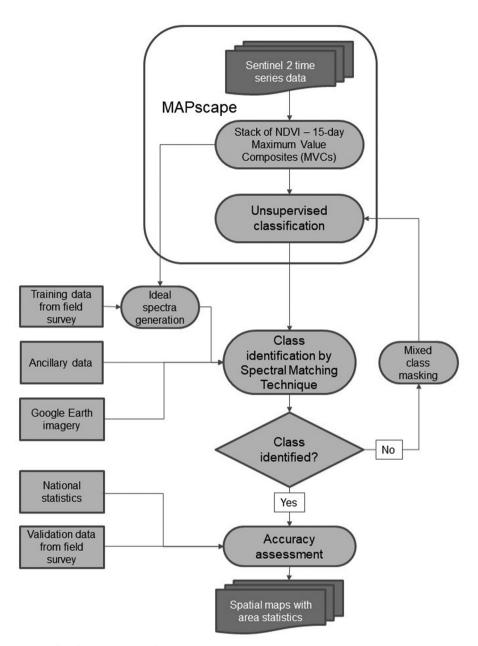
Figure 4. Workflow for rabi crop classification.

3.5. Comparison with regional statistics

This study generated crop type maps for the three study districts. Area statistics were extracted from crop type maps. District-wise maps were tested for accuracy using field-plot data collected by the research team, and national statistical data obtained from government agencies. Areas generated from the present study were compared with the area statistics available to us from the National Statistics report (http://aps.dac.gov.in/APY/Public_Report1.aspx).
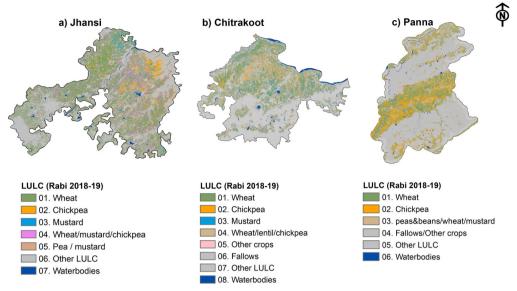
Figure 5. Spatial distribution of rabi crops along with other LULC for the three districts (a) Jhansi, (b) Chitrakoot, and (c) Panna.

4. Results and discussion

4.1. Crop patterns

Wheat occupied the largest cropped area followed by chickpea in all three districts. In Jhansi district, wheat dominated almost the entire area except the north-eastern part, which was dominated by chickpea. Mustard was mainly concentrated in the northern part of Jhansi district (Figure 5a). Out of the 280,000 ha sown (56% of the total geographical area of the district) in Jhansi district, wheat occupied about 64.2%, chickpea around 13.5% and mustard about 6.6% (Table 2a).

The southern part of Chitrakoot district is dominated by forest, shrub and grasslands, with a few small pockets of cultivated land. The northern part of the district, which is mainly cultivated land, is dominated by wheat, followed by chickpea and mustard. All the three crops are distributed throughout the northern part of the district (Figure 5b). Table 2b shows the crop area statistics for Chitrakoot district. A total of 130,000 ha (40% of the total geographical area of the district) was cropped in the district in rabi 2018-19, out of which about 38% was wheat, 12.6% chickpea and 11% mustard. Lentil is another crop that was found in small areas in Chitrakoot district.

Panna district, like Chitrakoot district, has large areas under forest, shrubs and grasslands. The central and northern patches have major agricultural lands, which were dominated by wheat and chickpea. Peas/beans and lentil were also cultivated in small areas (Figure 5c). Crop area statistics for Panna (Table 2c) showed that about 300,000 ha (41% of the total geographic area of the district) was under cultivation during the season, out of which about 40% was under wheat and 27.4% under chickpea.

The exploitation of SMT on Sentinel-2 NDVI time-series allowed the provision of field-level information with clear field boundaries, which is important in the agricultural scenario of Bundelkhand, where about 60% of the agricultural plots are less than 2 ha in size. As illustrated in Figure 6, the Sentinel-2 False Colour Composite (FCC) image and the corresponding classified image are shown for a part of Jhansi district. Temporal

Table 2. Rabi crop area in hectares.

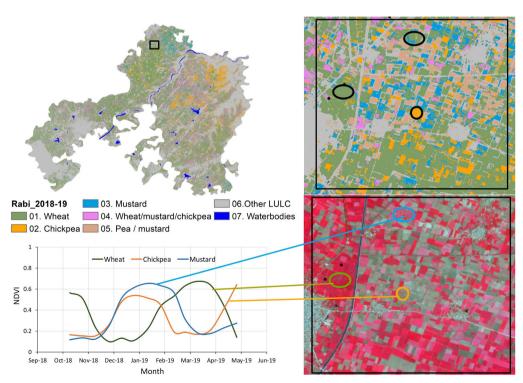| Jhansi	
Crop categories	Area (ha)
01. Wheat	180306
02. Chickpea	37992
03. Mustard	18412
04. Wheat/mustard/chickpea	38442
05. Peas/mustard	5751
06. Other classes	220559
07. Water bodies	5728
Chitrakoot	
Crop categories	Area (ha)
01. Wheat	49145
02. Chickpea	16150
03. Mustard	14543
04. Wheat/lentil/chickpea	42265
05. Other crops	6326
07. Fallow/mixed crops	17528
07. Other classes	152314
06. Water bodies	10095
Panna	
Crop categories	Area (ha)
01. Wheat	117001
02. Chickpea	80772
03. Peas & beans/lentil/wheat	28758
04. Fallow/other crops	68304
05. Other classes	410824
06. Water bodies /wetlands/fallow	5542

Figure 6. Spatial distribution of rabi crops at plot level (top) and the plot-wise spectral profiles for wheat, chickpea and mustard (bottom).

Table 3. Overall accuracy with producer's accuracy and Kappa coefficient of three districts for rabi crop area maps.

		Jhansi	
LULC Class	User's accuracy	Producer's accuracy	Kappa Coefficient
01. Wheat	0.85	0.95	0.64
02. Chickpea	0.71	0.45	
03.Mustard	0.83	0.45	
04. Wheat/mustard/chickpea	0.78	0.70	
05. Peas/mustard	1.00	0.50	
06. Other LULC	0.67	1.00	
		Chitrakoot	
LULC Class	User's accuracy	Producer's accuracy	Kappa Coefficient
01. Wheat	0.61	0.95	0.77
02. Chickpea	0.40	1.00	
03. Mustard	1.00	0.35	
04. Wheat/lentil/chickpea/mustard	1.00	1.00	
05. Other crops	0.67	1.00	
06. Fallow/mixed crops	1.00	1.00	
07. Other LULC	0.78	1.00	
		Panna	
LULC Class	User's accuracy	Producer's accuracy	Kappa Coefficient
01. Wheat	0.95	0.85	0.74
02. Chickpea	0.38	0.67	
03. Peas-Beans/wheat/mustard	0.97	0.94	
04. Fallows/other crops	0.92	0.65	
05. Others LULC	0.40	0.86	

NDVI profiles for the three fields in the area, one each of wheat, chickpea and mustard, showed how the differences in the profiles of different crops were used to classify cropped areas into crop types.

4.2. Accuracy assessment

Table 3 shows user's and producer's accuracy for each crop in the three study districts. It was performed independently with 404 validation sample points (Jhansi = 121, Chitrakoot = 116 and Panna = 167). Overall accuracy for all three districts was 84% and the Kappa coefficient was 0.72 (accuracy for Jhansi was 83%, for Chitrakoot 82% and for Panna 86%). Producer's accuracy for Jhansi district cropland map was 68% and user's accuracy was 81%. Accuracies varied based on crop homogeneity, with wheat having the highest spatial extent of 388,717 ha (55% of total cropped area), and correspondingly, the highest accuracy among all the classes. Producer's accuracies for wheat class were 95% for Jhansi and Chitrakoot districts and 85% for Panna district. Total accuracy for wheat crop over all three districts was 92% (8% error of omission). Chickpea occupied the highest area after wheat, with an area of 134,914 ha (19% of total cropped area). Overall accuracy for chickpea crop was 70% (with 30% error of omission), which is lower than that for wheat, because chickpea crop is heterogeneous (fragmented fields). Overall accuracy for mustard crop was 58%, which also has much less spatial extent with 61,713 ha (9% of total cropped areas). Table 3 shows the overall Kappa coefficients for the three study districts.

4.3. Significance of Sentinel-2 time-series for crop type mapping

Crop classification based on Sentinel-2 NDVI 15-day time-series exploiting SMT was successful in differentiating cropping patterns in wheat, chickpea, mustard and beans, which
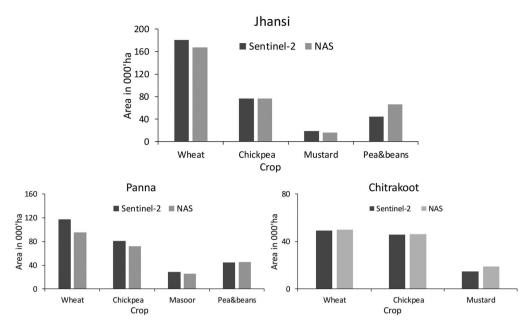
Figure 7. Crop area statistics derived from Sentinel-2 compared with NAS (national agriculture statistics) data for major crops in Jhansi, Panna and Chitrakoot districts.

had not been done in earlier methods. This methodology could correctly classify wheat, chickpea and mustard crops in the dryland areas, where a clear difference between temporal signatures of different crops was identified for the rabi season (Figure 4). However, some areas were mixed due to intercropping pattern in the eastern part of Jhansi (Figure 6). Also, other minor cropped areas with high tree cover percentage were often misclassified as shrub lands and minor fragmented cropped areas mainly in the high altitudes where beans and mustard were the major crops contributing to most of the inaccuracy.

Existing studies show that Sentinel-2 data is sufficient for crop-type mapping provided regular temporal availability during the crop season. They also show that machine-learning techniques like RF, SVM, and ANN are highly efficient in crop type classification. However, these methods have not been successful in mapping crop types in small, fragmented fields and fields with inter-cropping. This paper proposes a scalable classification methodology that is efficient for crop type mapping in areas with small, fragmented and multiple crop types, which dominate the drylands of South Asia and Africa.

The results obtained were compared with national statistics district-wise and there was very good correlation (Figure 7).

To improve accuracy, further work is needed with Hyperspectral images (ex: Hyperion) and Sentinel-1 (SAR data) in fusion with Sentinel-2 time-series datasets (10 m spatial resolution). This will minimize the spectral and spatio-temporal resolution, and provide an ideal platform for mapping specific crops such as wheat, chickpea, and other major crops with high intensity (i.e., grown during more than one season in one year).

4.4. Utilization of high resolution maps for yield assessment and insurance

High spatial resolution crop type maps generated using Sentinel-2 satellite data have numerous important applications in food security planning, temporal change analysis to
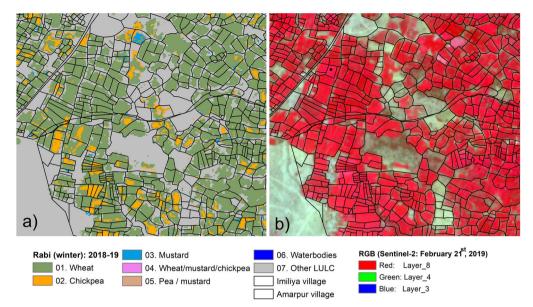
Figure 8. (a) Spatial distribution of rabi crops at village level near Imiliya and Amarpur villages of Jhansi district. (b) False colour composite of the region during rabi season.

identify shifts in cropping patterns of an area with time, impact assessment studies, and many more. Crop insurance is an area that extensively uses crop type maps at various levels. Crop type, along with crop health maps generated using remote sensing technology, are used to optimize the sampling of Crop Cutting Experiment (CCE) locations. CCEs are the basis for crop yield estimation that is used to determine crop loss. Crop acreage estimation at village/sub-district level (Figure 8) helps in making informed decisions on claims of failed sowing and helps prevent the false insuring of more land than what is planted under a particular crop. The satellite-derived crop area statistics can be used for assessing national statistics and augment the decision-making and planning process by providing accurate information of even inaccessible areas. In addition to national level assessments, the crop area and extent maps obtained from this classification technique can also be used for village-level crop assessment for micro-level crop management and advisory. This study has also proven that SMT approach can be applied to high spatial resolution data with good accuracy, and hence, can be used to study the within-field crop variability, which is high in the dryland regions of India.

The rabi season in India, dominated by non-rainy days, is best suited for the application of this method, as persistent cloud cover could hamper the availability of images, to generate clearly differentiating temporal signatures. A clear difference between the temporal signatures of wheat, chickpea and mustard was detected. However, there were some areas where classification was challenging. Irrigated mustard areas were mixed with irrigated wheat areas, particularly in the northern part of Jhansi. Differentiating the spectral signatures of mustard and wheat is still a challenge without auxiliary information on agronomic practices (including source of irrigation). In addition, small areas cropped with mustard with high percentage of tree cover were often misclassified as other LULC. Some discrepancies were also found during the comparison between district-wise national statistics and Sentinel-2 crop area. The mismatch between the Sentinel-2 cropped area and the landscape patterns of cropland will inevitably lead to mixed land cover pixels.

5. Conclusions

This study has led to the development of a method for mapping croplands during rabi season using Sentinel-2 NDVI time-series, and exploiting the SMT approach along with ground survey data. Major crop extents were mapped with higher accuracy during rabi season for the three study districts based on intensive training data. The study showed the potential of high-resolution temporal images and ground data for mapping cropland at field scale. Mapping rabi crop areas is the first step in characterizing important crop-growing environments that help macro-level planning, leading to sustainable use of resources and improvement in drylands. Seasonal crop maps and statistics with a high degree of accuracy such as these are important inputs for assessing the impact of abiotic stresses, such as droughts and heat stress, which regularly affect the region and are predicted to increase in frequency and intensity in a changing climate scenario.

This methodology, however, cannot be used in kharif (monsoon) season in the Indian sub-continent due to the persistence of cloud cover. Non-availability of enough cloud-free Sentinel-2 data during the kharif season makes it impossible for the generation of multi-temporal NDVI spectra that can satisfactorily differentiate between various crops. This necessitates the use of Synthetic Aperture Radar imagery, like that provided by Sentinel-1, which has all-weather capability.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Murali Krishna Gumma 🔾 http://orcid.org/0000-0002-3760-3935
Sreenath Dixit 🔾 http://orcid.org/0000-0002-3607-8729
Francesco Collivignarelli 🔾 http://orcid.org/0000-0001-7442-9051
Rao N. Kolli 🔾 http://orcid.org/0000-0002-7436-219X
Anthony M. Whitbread 🔾 http://orcid.org/0000-0003-4840-7670

References

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.

Biggs TW, Thenkabail PS, Gumma MK, Scott CA, Parthasaradhi GR, Turral HN. 2006. Irrigated area mapping in heterogeneous landscapes with Modis time series, ground truth and census data, Krishna basin, India. Int J Remote Sens. 27(19):4245–4266. Available from: http://www.informaworld.com/10.1080/01431160600851801.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas. 20(1):37–47.

Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ. 37(1):35–46. http://www.sciencedirect.com/science/article/pii/003442579190048B.

Feng S, Zhao J, Liu T, Zhang H, Zhang Z, Guo X. 2019. Crop type identification and mapping using machine learning algorithms and sentinel-2 time series data. IEEE J Sel Top Appl Earth Observations Remote Sensing. 12(9):3295–3306.

Gallego FJ. 2004. Remote sensing and land cover area estimation. Int J Remote Sens. 25(15):3019–3047.

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.

Gray J, Friedl M, Frolking S, Ramankutty N, Nelson A, Gumma MK. 2014. Mapping Asian cropping intensity with Modis. IEEE J Sel Top Appl Earth Observations Remote Sensing. 7(8):3373–3379.

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, Deevi K, Mohammed I, Varshney R, Gaur P, Whitbread A. 2016. Satellite imagery and household survey for tracking chickpea adoption in andhra pradesh, india. Int J Remote Sens. 37(8):1955–1972. Available from: http://www.tandfonline.com/doi/abs/10.1080/01431161.2016.1165889.

Gumma MK, Nelson A, Yamano T. 2019. Mapping drought-induced changes in rice area in India. Int J Remote Sens. 40(21):8146–8173.

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 Sen. 55(6):926–949. Available from:.

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 500m data for the year 2010. ISPRS J Photogramm Remote Sens. 91:98–113. Available from: http://www.sciencedirect.com/science/article/pii/S0924271614000446.

Gumma MK, Thenkabail PS, Nelson A. 2011. Mapping irrigated areas using Modis 250 meter time-series data: A study on Krishna river basin (India). Water. 3(1):113–131.

Gumma MK, Thenkabail PS, Teluguntla P, Whitbread AM. 2018. Monitoring of spatiotemporal dynamics of Rabi rice fallows in south Asia using remote sensing. Geospatial technologies in land resources mapping, monitoring and management. Cham: Springer; p. 425–449.

Inglada J, Arias M, Tardy B, Hagolle O, Valero S, Morin D, Dedieu G, Sepulcre G, Bontemps S, Defourny P, et al. 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sens. 7(9):12356–12379.

Jensen JR. 2004. Iintroductory digital image processing: A remote sensing perspective, 3rd ed., Upper Saddle River (NJ): Prentice Hall; p. 544.

Li Q, Cao X, Jia K, Zhang M, Dong Q. 2014. Crop type identification by integration of high-spatial resolution multispectral data with features extracted from coarse-resolution time-series vegetation index data. Int J Remote Sens. 35(16):6076–6088.

Lobell DB, Asner GP, Ortiz-Monasterio JI, Benning TL. 2003. Remote sensing of regional crop production in the Yaqui valley, Mexico: Estimates and uncertainties. Agric Ecosyst Environ. 94(2):205–220. Available from: http://www.sciencedirect.com/science/article/B6T3Y-45CW274-1/2/53a983e89193412b917f97e05211c703.

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 (4):418–436. Available from:.

Matton N, Canto GS, Waldner F, Valero S, Morin D, Inglada J, Arias M, Bontemps S, Koetz B, Defourny P. 2015. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. Remote Sens. 7(10):13208–13232. Available from: http://www.mdpi.com/2072-4292/7/10/13208.

Misra A, Rama Rao C, Ravishankar K. 2010. Analysis of potentials and problems of dairy production in rainfed agro-ecosystem of India. Indian J Animal Sci. 80 (11):1126.

Moscheni F, Bhattacharjee S, Kunt M. 1998. Spatio-temporal segmentation based on region merging. IEEE Trans Pattern Anal Machine Intell. 20 (9):897–915.

Peña-Barragán JM, Ngugi MK, Plant RE, Six J. 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. Remote Sens Environ. 115(6):1301–1316.

Petitjean F, Kurtz C, Passat N, Gançarski P. 2012. Spatio-temporal reasoning for the classification of satellite image time series. Pattern Recog Lett. 33(13):1805–1815. Available from: http://www.sciencedirect.com/science/article/pii/S0167865512001973.

Saini R, Ghosh S. 2018. Crop classification on single date sentinel-2 imagery using random forest and support vector machine. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (Vol. XLII-5).

Singh HP, Sharma KD, Subba Reddy G, Sharma KL. 2004. Dryland agriculture in India. Challenges and strategies of dryland agriculture. [accessed on 2019 may 12]. https://dl.Sciencesocieties.Org/publications/books/abstracts/cssaspecialpubl/challengesandst/67. p. 67–92.

Stehman SV, Wickham J, Smith J, Yang L. 2003. Thematic accuracy of the 1992 national land-cover data for the eastern united states: Statistical methodology and regional results. Remote Sens Environ. 86 (4): 500–516.

Sun C, Bian Y, Zhou T, Pan J. 2019. Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. Sensors. 19(10):2401.

Teluguntla P, Thenkabail PS, Xiong J, Gumma MK, Congalton RG, Oliphant A, Poehnelt J, Yadav K, Rao M, Massey R. 2017. Spectral matching techniques (SMTS) and automated cropland classification algorithms (ACCAS) for mapping croplands of Australia using Modis 250-m time-series (2000–2015) data. Int J Digital Earth. 10(9):944–977. Available from:.

Thenkabail PS, 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 AVHRR pathfinder datasets in the Krishna river basin, India. Photogramm Eng Remote Sens. 73(10):1029– 1040.

Thenkabail P, Hanjra M, Dheeravath V, Gumma M. 2010. A holistic view of global croplands and their water use for ensuring global food security in the 21st century through advanced remote sensing and non-remote sensing approaches. Remote Sens. 2(1):211–261. Available from: http://www.mdpi.com/2072-4292/2/1/211/.

Thenkabail PS, Schull M, Turral H. 2005. Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of Modis data. Remote Sens Environ. 95(3):317–341. Available from: http://www.sciencedirect.com/science/article/pii/S0034425705000180.

Thenkabail PS, Stucky N, Griscom BW, Ashton MS, Diels J, Van Der Meer B, Enclona E. 2004. Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using Ikonos data. Int J Remote Sens. 25(23):5447–5472.

Thiruvengadachari S, Sakthivadivel R. 1997. Satellite remote sensing for assessment of irrigation system performance: A case study in India. Research report 9. Colombo, Sri Lanka: International irrigation management institute.

Van Genderen J, Lock B, Vass P. 1978. Remote sensing: Statistical testing of thematic map accuracy. Remote Sens Environ. 7(1):3–14.

Vijayasekaran D. 2019. Sen2-agri-crop type mapping pilot study using sentinel-2 satellite imagery in india. Int Arch Photogramm Remote Sens Spatial Inf Sci. XLII-3/W6:175–180.

Walker JS, Blaschke T. 2008. Object-based land-cover classification for the phoenix metropolitan area: Optimization vs. Transportability. Int J Remote Sens. 29 (7):2021–2040. Available from:.

Xiong J, Thenkabail PS, Gumma MK, Teluguntla P, Poehnelt J, Congalton RG, Yadav K, Thau D. 2017. Automated cropland mapping of continental Africa using google earth engine cloud computing. ISPRS J Photogramm Remote Sens. 126:225–244. Available from: http://www.sciencedirect.com/science/article/pii/S0924271616301575.

Xiong J, Thenkabail P, Tilton J, Gumma M, Teluguntla P, Oliphant A, Congalton R, Yadav K, Gorelick N. 2017. 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(10):1065.

Xu Y, Yu L, Zhao FR, Cai X, Zhao J, Lu H, Gong P. 2018. Tracking annual cropland changes from 1984 to 2016 using time-series Landsat images with a change-detection and post-classification approach: Experiments from three sites in Africa. Remote Sens Environ. 218:13–31. Available from: http://www.sciencedirect.com/science/article/pii/S003442571830419X.