



Wheat Acreage Mapping and Yield Prediction Using Landsat-8 OLI Satellite Data: a Case Study in Sahibganj Province, Jharkhand (India)

Bikash Ranjan Parida¹ · Avinash Kumar Ranjan¹

Received: 25 September 2018 / Revised: 24 December 2018 / Accepted: 18 April 2019 / Published online: 7 May 2019
© Springer Nature Switzerland AG 2019

Abstract

The emergence of remote sensing technologies and availability of satellite data over three decades have facilitated to monitor and understand the agricultural systems in many intensive agricultural regions. Here, we performed a comprehensive study on utilization of multi-temporal satellite data (i.e., Landsat-8 and MODIS) for wheat acreage and yield estimation during winter season (2016–2017) over the Sahibganj District in Jharkhand (India). Phenological variables were derived using the time-series normalized difference vegetation index (NDVI), which helps to understand the phenological transitions of wheat. The NDVI profile was used to derive rules for decision tree classifier to map the acreage of wheat. The key findings indicate that the acreage of wheat was estimated as ~3870 ha. Further, the long-term wheat statistics data were used to derive a yield model. Based on this model, wheat production was predicted as ~4523 t for the winter season 2016–2017, while, the mean was 3482 t. Predicted wheat yield was as ~1.17 t/ha, which was underestimated by 0.07 t/ha. Thus, it can be concluded that the accuracy of yield prediction depends on the precision of wheat acreage map derived from remote sensing data. A significant challenge for accurate acreage mapping could be the coarser spatial resolution of satellite data as the average plot sizes of Indian farmers can be far smaller than pixel sizes of the satellite data. Nevertheless, this comprehensive case study inferred that satellite-derived wheat acreage can be preferred to predict yield instead of traditional-based survey estimates.

Keywords Phenology · Decision tree · Wheat acreage · Yield prediction · Remote sensing

1 Introduction

Agriculture plays a key role in India's economy by contributing 17.3% of India's gross domestic product (GDP) in 2016–2017 and over 55% of the population is engaged in agricultural activity. Given the importance of agricultural sector, India becomes a leading producer of rice and wheat over the years. India is the second biggest producer of wheat in the world with ~30 million hectares of cultivation land. As per the Directorate of Economics & Statistics, DAC&FW [6], Ministry of Agriculture, Government of India, wheat production was ~96.50 million tons in 2016–2017. There are two key agricultural seasons, namely, Kharif (monsoon season) and Rabi (winter season). Wheat is the Rabi season's main crop that sown in November and harvested in March. During Rabi season

(winter), wheat is grown in several states of India, namely, Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Punjab, Chandigarh-Haryana-Delhi, Uttar Pradesh, Bihar, Jharkhand, and Rajasthan. These aforementioned states contributed ~99.5% of total wheat production in the country. Remaining states such as Karnataka, Assam, Chhattisgarh, and other North Eastern states have contributed only ~0.5% of the total wheat production in the country (DES 2017). In the state of Jharkhand (JH), wheat crop plays an imperative role for livelihood of farming communities that supports food security for ~80% of total population of the state [17].

The information on crop acreage and yield is very essential for planning and sustainable development of agriculture system. Reliable and timely detailed information on crop acreage provides valuable information to the planner and policy makers for making decision with respect to procurement, storage, public distribution, export-import, and finally the food security of nation [28]. Many developing countries including India have been used the traditional method (i.e., survey-based) of data collection for crop monitoring and yield estimation, wherein Crop Cutting Experiment (CCE) typically conducted under a complex sampling design (i.e., stratified

✉ Bikash Ranjan Parida
bikashrp@gmail.com

¹ Department of Geoinformatics, School of Natural Resource Management, Central University of Jharkhand, Ranchi 835205, India

multi-stage random sampling). So, the traditional methods are subjective, cumbersome, and expensive as it needs large number of manpower to collect the field data (Singh et al. 2002). By contrast, space-based technologies and crop growth models were used widely in the recent decades for crop monitoring and yield predictions at various scales.

Crop yield can be estimated using multiple approaches such as (i) semiphysical RS-based model, (ii) satellite index-based crop yield models, (iii) correlation-weighted agrometeorological-based model, (iv) crop simulation model, and (v) data obtained from CCE. Number of previous studies have used crop growth models, namely, WTGROWS, DSSAT, and InfoCrop, for crop monitoring and yield prediction at various scale [2, 19, 23, 41, 44]. In particular, remote sensing (RS) satellite data have revolutionized the agriculture system by providing wide range of data to monitor crop growth and to predict crop yields. These RS data were mainly used to enhance the accuracy of crop yield estimation and forecasting yield [3, 5, 38]. So the RS data are very important for the benefit of the society. Besides wide range of RS application in agriculture, RS data provide a way to quantify accurate crop acreage and corresponding yield [1, 32–34]. Satellite-derived normalized difference vegetation index (NDVI) has been extensively used for crop yield estimation, wherein it utilizes the red (R) and near-infrared (NIR) bands of electromagnetic spectrum. Many previous studies have showed a well correlation between satellite-derived NDVI and yield [4, 16, 47], which further helps to extrapolate yield at regional level and national level. Crop phenology provides the first-hand information on timing and stages of the crop growth cycles and was very crucial for acreage estimation [4, 40]. The similarity in the spectral reflectance curve of various crops may pose various challenges to identify different crops using phenology metrics [9, 31, 48]. Hence, reliability of RS data for estimating acreages of crops was important to achieve accurate yield as well as to know the state of food security of nation [36].

Despite large volume of RS satellite data, there was no functional acreage estimation methods developed in India including many other Asian countries. On the contrary, a large volume of RS-based approaches has produced, which are important steps forward in the field of RS and agriculture. Notably, there were literature gaps on the status of wheat crop mapping using satellite data and its acreage for the Sahibganj District of Jharkhand. So, the overarching objectives of this comprehensive study were (i) to derive the phenology of wheat crop based on multi-temporal MODIS and Landsat-8-based NDVI data during Rabi (winter) season and (ii) to estimate the wheat acreage and yield using Landsat-8 data over the Sahibganj District of Jharkhand.

2 Study Area

The Sahibganj District of Jharkhand state in India lies between 24° 42' 30" to 25° 21' 27" N latitude and 87° 25' 10" to 87° 54'

35" E longitude (Fig. 1). Based on satellite imagery, the total geographical area of the district is ~2201 km². In the Sahibganj District, there are three distinct seasons such as (i) summer season (March to May), (ii) rainy Season (June to October), and (iii) winter season (November to February). The average air temperature is ~15 °C during winter season, while it can increase to 30 to 40 °C during summer season. The district receives an annual rainfall of ~1500 mm and > 75% of annual rainfall typically occurred during rainy season [26]. The district covers ~1030 km² of cultivable land, wherein only ~84 km² (~5%) of cultivable area has been covered under irrigation. The prime source of irrigation is ground water harvest, deep well, canals, and rivers. The dominant soil orders were alfisol, inceptisol, and entisol [14]. The predominant Rabi (winter) crops were wheat, rice, maize, gram, khesari, pea, etc. These crops are usually sown in last week of November and harvested in the month of March.

The district comprises a series of Rajmahal Hills in the middle portion of the district. The Rajmahal Hills alone comprise ~70% of forest area of the district and were marked by dark red tone in the false color composite (FCC) map (Fig. 1). The FCC map was created using Landsat-8 OLI (Operational Land Imager) sensor data (bands used: 543). The low-lying areas along the Rajmahal Hills were covered with cultivable land. The rivers such as Ganga, Gumani, and Bansloi flow through some blocks of the district (north-east to south-east and some of the part of south-west). These rivers have developed fertile alluvium soils over the cultivable land, and therefore, it helps for crop production. The cultivable lands were marked by light red tone in the FCC map, and they were scattered over the north, north-east, south-east, and south-west of the study area (Fig. 1).

3 Materials and Methods

Satellite data used for this study were Landsat-8 (OLI sensor), MODIS-based land use and land cover (LU/LC) product (MCD12Q1), and NDVI product (i.e., MOD13Q1). These data were obtained at bimonthly to monthly time scale for the period 2016 to 2017, which corresponds to Rabi (winter) season crops. The wheat statistics (area, production, yield) of the district over the period 2002–2016 has been obtained from DES, DAC&FW (2017). The detail descriptions of data usages were provided in Table 1.

3.1 MODIS-Based MCD12Q1 and MOD13Q1

The MODIS-based MCD12Q1 provides LU/LC data at 500-m spatial resolution at annual interval. We have used the International Geosphere-Biosphere Programme (IGBP) classification scheme to show the different LU/LC of the district in the year 2013. The MOD13Q1 product provides 16-day composite NDVI data with 250 m of spatial resolution. These data were originally having sinusoidal projection, and

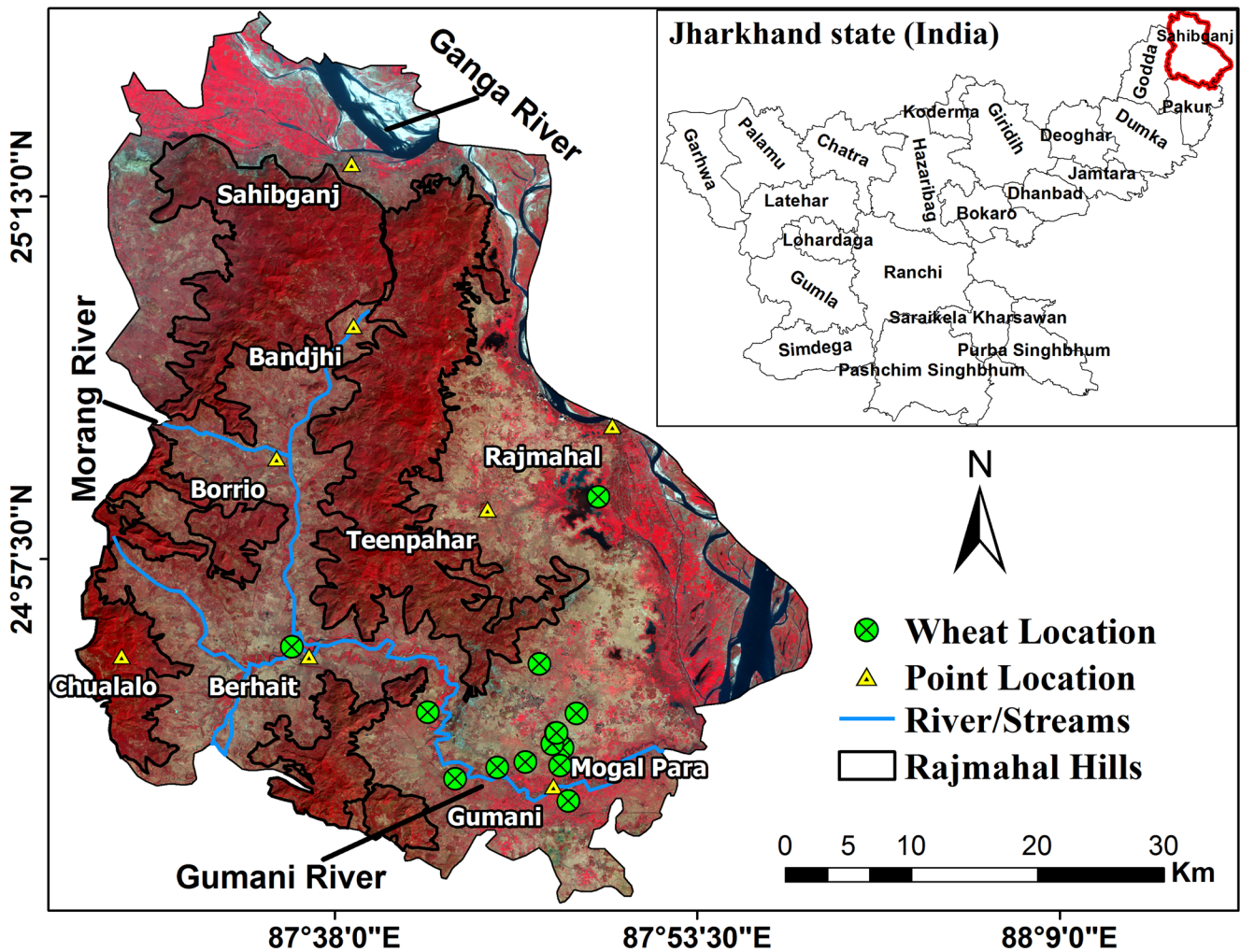


Fig. 1 Location map of Sahibganj District in the state of Jharkhand, India. The demarcated boundary on the FCC map (Landsat-8 acquired on 15 January 2017) showed the Rajmahal Hills that comprises forests

area. The GPS points (green symbols) represent locations of wheat crops collected during field data collection

tile number used for this study was *h25v07* (*h* as horizontal and *v* as vertical). These data were reprojected to geographical latitude and longitude projection and the datum was assigned as WGS 1984. The MODIS-based time-series NDVI data (December 2016–April 2017) were used to capture the phenology of wheat crop in the district of Sahibganj, Jharkhand. These MODIS products were downloaded from United States Geological Survey (USGS) (<https://earthexplorer.usgs.gov/>).

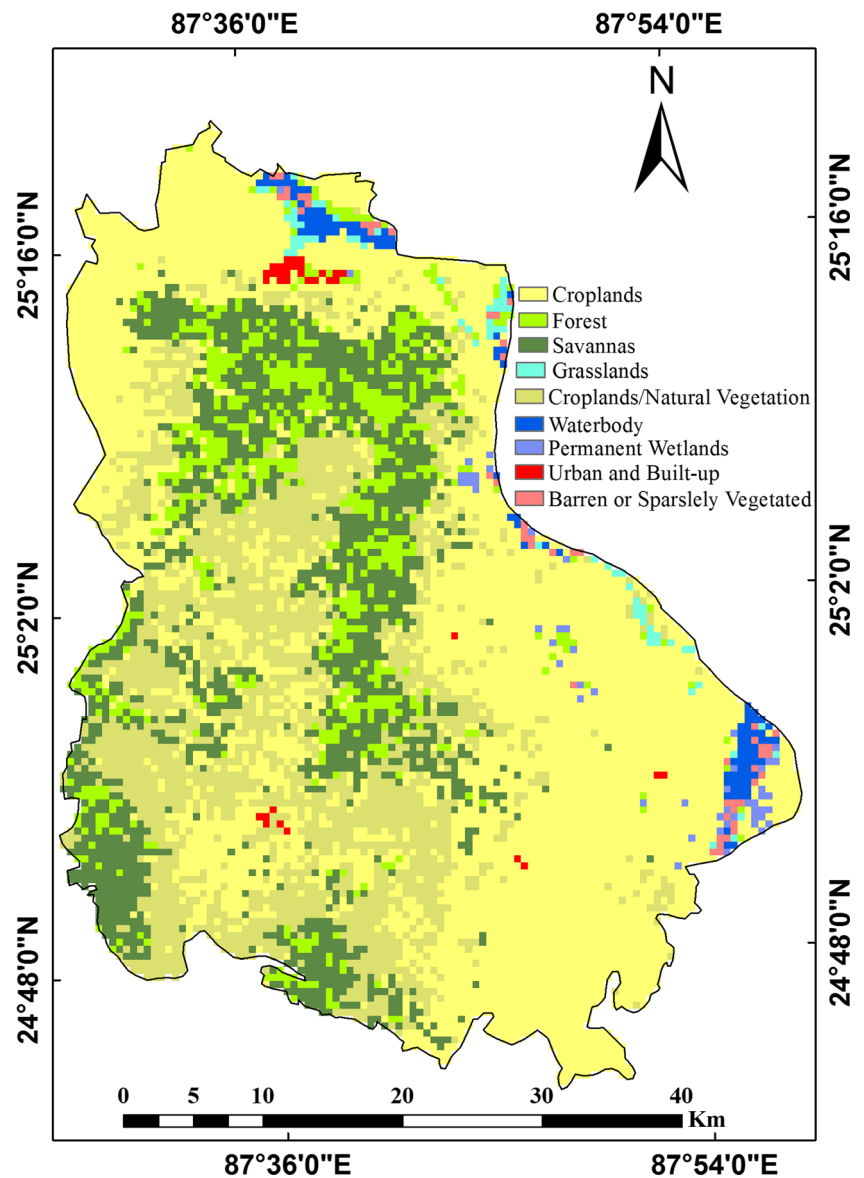
The LU/LC map of the Sahibganj has been obtained from the MODIS data and the product used was MCD12Q1 (Fig. 2). The dominant land use classes in the district were croplands followed by woody savannas and forest. The croplands account for ~50% area, while forest covers including savannas account for ~25%, and croplands/natural vegetation covers account for ~22% of total area. The spatial extent of individual features was

Table 1 Satellite and ancillary data used for this study and their characteristics

Data	Acquisition date	Spatial resolution	Temporal resolution	Purpose	Source
MODIS (MCD12Q1)	2013	500 m	Yearly	LU/LC	USGS
MODIS (MOD13Q1)	2 December 2016 to 7 April 2017	250 m	16-day composite	Wheat phenology	USGS
Landsat-8 (OLI)	15 January 2017	30 m	16 days	Wheat acreage	USGS
	16 February 2017				
	20 March 2017				
Wheat APY statistics	2002–2016	–	–	Statistical wheat yield model	DES

APY area, production, and yield

Fig. 2 MODIS-based LU/LC map of Sahibganj District for the year 2013. The LU/LC map was prepared based on International Geosphere Biosphere Programme (IGBP) classification scheme



tabulated with their corresponding percentage of coverage of total area (Table 2). The MODIS-based LU/LC has been included to display various LU/LC classes of the study area.

3.2 Landsat-8 OLI Sensor Data

The Landsat-8 OLI sensor (level-1, terrain corrected) data have been acquired from USGS site, and data were obtained for the month of January, February, and March during the year 2017. The bands of OLI sensor have 30-m spatial resolution, and the spectral range varies from visible to infrared bands. For this study, we used NIR (band 5) and R (band 4) bands to derive NDVI.

Table 2 MODIS-based major LU/LC classes and the corresponding areal proportion (%) in Sahibganj District

LU/LC classes	Area (km ²)	Area (%)
Croplands	1096	49.7
Forests	181.6	8.2
Savannas	361.1	16.4
Grassland	17.6	0.8
Croplands/natural vegetation	482.11	21.9
Urban	8.5	0.4
Water body	27.64	1.3
Permanent wetlands	13.17	0.7
Barren or sparsely vegetated	13.3	0.6
Total	2201	100

3.3 Historical Wheat Statistics

Historical wheat statistics comprising area, production, and yield (APY) for Sahibganj District have been obtained from the DES, DAC&FW (2017). As per the availability of data, we acquired APY data from the year 2002 to 2016. It can be noted that there was one missing data during the year 2012–2013 in the database of DES (2017). Further, this data were used to develop a simple regression model to predict wheat yield for the district.

3.4 Field Data Collection Using GPS

In the first week of March 2017, a field visit was accomplished to collect geographic location (latitude and longitude) of wheat plots across the district. The location information was collected by handheld Global Positioning System (GPS), manufactured by Garmin (Etrex-30 model). In total, 15 GPS locations were collected (Table 3) and the spatial distribution of locations of wheat plots has been shown in Fig. 1. At the time of field visit, the wheat reached at maturing/ripening stage. These GPS points were further used for accuracy assessment of classified LU/LC map obtained from Landsat-8 (OLI) data using the decision tree (DT) classification technique.

3.5 Methods

The methods include satellite data processing, DT classification, and statistical wheat yield model. The self-explanatory flowchart of methodology has been provided in Fig. 3. The broad categories of data used were satellite, field observations, and historical wheat APY statistics. The satellite data processing involves image pre-processing and conversion of digital number (DN) to surface reflectance. The DN of band 4 (red) and band 5 (NIR) of Landsat-8 OLI was converted to top-of-atmosphere (TOA) reflectance using reflectance rescaling coefficient, which were provided in the product metadata file (MTL file). The equations used to convert the DN values to TOA reflectance and Sun angle correction can be referred to USGS [46]. In image pre-processing, both Landsat-8 OLI and MODIS data were reprojected to a common coordinate system called geographical latitude and longitude with datum of WGS84. The district shapefile has been obtained from DIVA-GIS to demarcate the Sahibganj District of

Jharkhand. Using the Landsat-8 OLI sensor data, the NDVI has been derived. The NDVI index is an indicator to assess the greenness of the vegetation. Its value varies from -1 to $+1$, where positive value shows the vegetation greenness and zero to negative value indicates water bodies.

3.5.1 Phenological Metrics and Interpretation

The phenology has been obtained from the time-series data of NDVI (i.e., MCD13Q1), wherein we have used in total of nine satellite NDVI images covering the period December 2016 to April 2017 (Rabi season). Using the NDVI values, three important phenological indicators such as “maximum growing rate,” “maximum peak value,” and “maximum fading date” were investigated based on the phenological cycle of Rabi (winter) season wheat. The maximum growing rate represents the maximum NDVI during the active growth cycle of Rabi season. The maximum peak value is the date when the derivative of the NDVI composite reaches its maximum value. The maximum fading date corresponds to the maximum NDVI value during the maturing stage of wheat.

3.5.2 DT Classification

The DT classifier offers an effective application for crop mapping based on hierarchical classification concept and non-parametric assumptions. Due to conceptual simplicity and computational ability, the DT classifier has become more important in crop mapping. The DT classifiers can automatically execute the feature selection, which can reduce the complexity. Its tree structure provides easily understandable and interpretable information regarding the logical or generalization ability of the classification [10]. As compared to artificial neural networks (ANNs), DT algorithms have no black box, wherein many hidden layers were concealed. A DT takes as input an object/satellite data described by a set of properties, and outputs were in the form of yes/no decision. DT algorithm therefore represents Boolean functions. In DT algorithms, decisions are introduced by a set of rules in multi-temporal LU/LC classification: if a condition exists, then inference is applied, which is very suitable in distinguishing crop types [22].

Table 3 Wheat plots and corresponding GPS points collected in March 2017 in Sahibganj District

Plot no.	Latitude (N)	Longitude (E)	Plot no.	Latitude (N)	Longitude (E)
1	24° 47' 10.70"	87° 48' 0.36"	9	24° 50' 4.56"	87° 47' 30.12"
2	24° 48' 7.46"	87° 43' 10.40"	10	24° 48' 41.04"	87° 47' 39.12"
3	24° 48' 36.78"	87° 44' 57.34"	11	24° 50' 54.24"	87° 48' 21.6"
4	24° 48' 50.85"	87° 46' 9.34"	12	25° 0' 10.8"	87° 49' 17.04"
5	24° 49' 27.48"	87° 47' 44.88"	13	24° 53' 46.32"	87° 36' 11.16"
6	24° 49' 26.76"	87° 47' 45.24"	14	24° 50' 57.48"	87° 42' 0"
7	24° 49' 36.12"	87° 47' 18.96"	15	24° 53' 2.04"	87° 46' 46.2"
8	24° 50' 4.92"	87° 47' 28.32"			

Several previous studies have employed the DT algorithm for LU/LC classification and mapping crop types [15, 20, 35]. The illustration of DT classifier has been provided in Fig. 4. In this study, NDVI images of January, February, and March months obtained from Landsat-8 (OLI) have been used in DT classifier for wheat acreage mapping. As the OLI sensors data have higher spatial resolution than the MODIS, we used the OLI data for acreage estimation. The NDVI thresholds were established for each month for known GPS locations (at least four plots) of wheat plots collected during field work using GPS. The rules were generated for those GPS locations, and these rules were applied to extract wheat crop across the Sahibganj District.

3.5.3 Statistical Wheat Yield Prediction Model

The wheat area (hectares) and production (tonnes) data over the period 2002–2016 of Sahibganj District have been used to develop a statistical regression model. Wheat area was plotted against the wheat production, and by adding a linear trend line, a production-based statistical regression of wheat acreage model has been developed (Fig. 5). Equation 4 was developed for the Sahibganj District based on the simple regression model between area and production of wheat. The slope and coefficient were applied to predict wheat production for the year 2017. The correlation coefficient (r) was 0.85 and (p value < 0.001) obtained between wheat production and area.

$$y = 1.019x + 579.5 \quad (1)$$

where y is the estimated production (tonnes) and x is the area (hectares).

4 Results

4.1 MODIS-Based NDVI Profile and Phenological Parameters

The NDVI during December 2016 to April 2017 has been plotted in Fig. 6 to capture various stages of wheat such as emergence, tillering (wintering), flowering (milking), maturing (ripening), and harvesting. Based on the wheat crop calendar of the Sahibganj District, it is usually sown during last week of November, while harvested during end of March. Our results from NDVI profile indicate that the emergence of wheat was observed in early December and tillering (wintering) was observed in mid-January. Peak growth stage of wheat was observed with flowering and milking during the period mid-January to mid-February. The maturing and ripening was observed in mid-February to mid-March followed by harvesting during end of March or first week of April. Based on NDVI profile, we can infer that the various growth stages of wheat were well captured, which were in line with the crop calendar information of the Sahibganj District.

Additionally, we have plotted the NDVI profile for the four GPS points of wheat plots taken during field investigations to generate key phenological parameters such as maximum growing rate, maximum peak value, and maximum fading rate (Fig. 7). The emergence stage of wheat was observed when the NDVI was ~ 0.3 to 0.35. The tillering/wintering stage noticed when the NDVI was ~ 0.4 to 0.5. The peak growth stage of the wheat was noticed when the NDVI value was close to ~ 0.5 to 0.65. It was observed that during maturing/ripening and harvesting stage, the NDVI value reduces to 0.35 to 0.45 and 0.25 to 0.3, respectively. The derived NDVI profile showed three key phenological features (maximum growing rate, maximum peak value, and maximum fading rate) and their corresponding dates such as maximum growing date, maximum peak date, and maximum fading date. The detailed characteristics of phenological features were presented in Table 4.

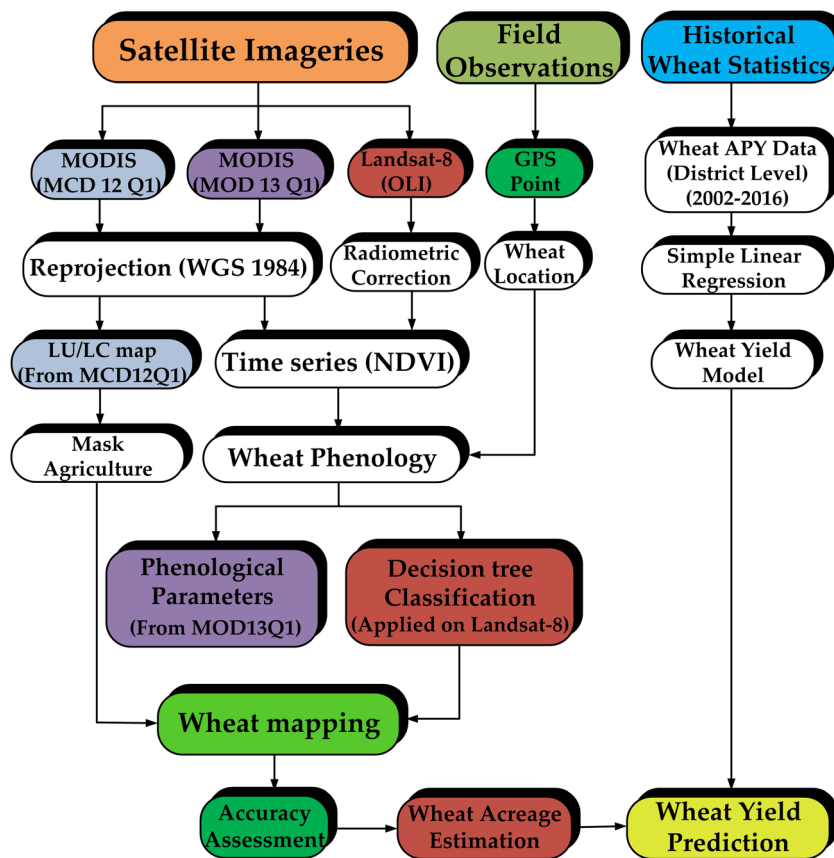
The maximum growing rate that belongs to the local maximum value of the first derivative on the left side was detected on first week of January 2017 and when the average NDVI was 0.52. The maximum NDVI value during the growing season (maximum peak value) was detected on first week of February 2017, wherein the average NDVI value was 0.66 that belongs to flowering and milking stage. The maximum fading rate indicating the local minimum value of first derivative curve on the right side was noticeable on second week of March 2017 with mean NDVI of 0.49 and it belongs to maturing and ripening stage. These features describe the typical wheat crop specific phenological patterns, and enable to identify the wheat growing pattern using the RS satellite data.

4.2 Wheat Acreage Estimation Based on DT Classifier

In the study area, there were limited Landsat-8 data available, and thus, there was a limitation in generating NDVI profile at 16-day interval. However, due its better spatial resolution (30 m), we generated NDVI profile and computed NDVI thresholds. Despite poor temporal resolution of Landsat-8 (OLI), it can be useful to compare the NDVI profile with that of MODIS-based NDVI profile. The Landsat-8 (OLI)-based NDVI profile showed that the maximum growing date, maximum peak date, and maximum fading date of wheat were well comparable with the MODIS-based NDVI profile. This understanding helps to generate NDVI threshold values based on Landsat-8 (OLI) data and further to derived rules for DT classifier.

To estimate wheat acreage in the Sahibganj District, we employed DT classification method, wherein Landsat-8 (OLI) data has been used because of its higher spatial resolution. The NDVI threshold values used for generating rules for wheat crop mapping were shown in Table 5. We found that the NDVI threshold values were having different ranges depending on the various growth stages of wheat (i.e., emergence stage to harvesting stage). Notably, these threshold values must be developed accurately so that the accuracy of acreage estimation will be higher.

Fig. 3 The detailed methodology adopted for this study which employs both satellite data and ground-based data to estimate acreage under wheat and to predict wheat yield



The wheat acreage map has been produced using the DT classifier and presented in Fig. 8. Notably, the wheat crop was observed in the upper northern part and south-east region of the district. The croplands along the Ganga River and other tributaries (lower part of southern region) have dominantly covered with wheat crop because of easy access to water via irrigation facility and fertile soils. However, the wheat plots were sparse over the middle and western portion of the district as these regions were predominantly covered with forest patches along the Rajmahal Hills. In this hilly portion, farmers mainly used the water from minor reservoirs to irrigate the crops, which were also verified during the field visit.

According to the satellite-based wheat acreage map of Sahibganj District, the total area of wheat was estimated as

~3870 ha for the year 2016–2017 (Rabi season) (Table 6). As per the DES [6], the mean wheat acreage of the Sahibganj District was reported as ~2846 ha (mean over 2002–2016). The minimum and maximum of wheat acreage over the period 2002–2016 were between 1245 and 5431 ha. As per Eq. 4, the predicted wheat production was calculated as ~4523 t for the year 2016–2017 during Rabi season (winter). As per the DES [6] statistics, it reported the mean production of ~3482 t (mean over 2002–2016), while production can vary from 1617 to 5616 t. The large inter-annual variability can be explained by the variation of environmental factors, soil factors, crop

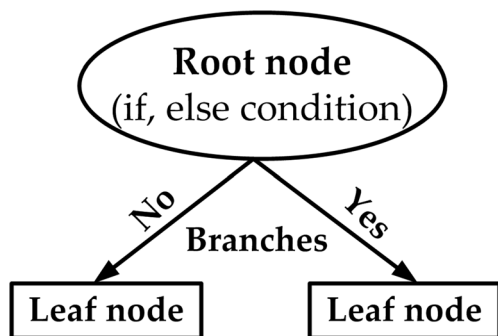


Fig. 4 An illustration of decision tree classifier used in this study

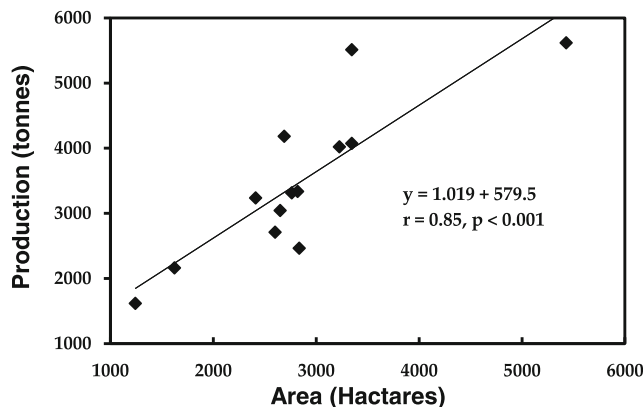
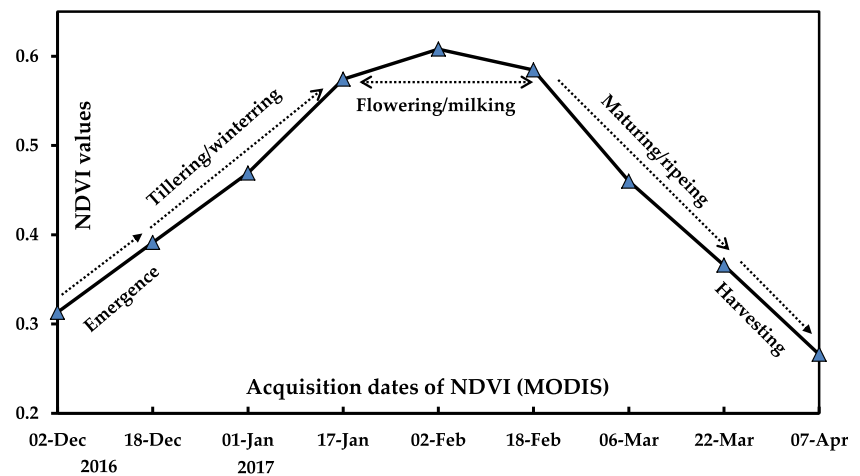


Fig. 5 Wheat crop production plotted against wheat area of Sahibganj District. The data were pooled from the years 2002 to 2016

Fig. 6 Growth stages of wheat based on NDVI values of MOD13Q1. The various growth stages of wheat have been shown using the NDVI profile (average NDVI values of wheat plot nos. 1, 2, 3, and 4)



management practices, and natural hazards (land degradation, large-scale mining, and floods). We also observed that the production of wheat is higher along the rivers, while lower along the mountainous areas of Rajmahal Hills. We find that the predicted production was relatively overestimated than the mean production of the district. However, the deviation was within the limit of maximum range of wheat production. The predicted yield of 1.17 t/ha was well comparable with the DES yield data (1.24 t/ha). Overall, the wheat yield was underestimated by 0.07 t/ha.

Accuracy assessment is vital to endorse the digital image classification results. In this study, accuracy assessment has been performed using the ten GPS points of wheat plots, which was collected during field visit. The prepared wheat acreage map based on DT classifier showed a satisfactory result with a fair overall accuracy of 80%, wherein the producer and user accuracy were calculated as of 78% and 70%, respectively (Table 7). The kappa coefficient was 0.7, indicating that the classified map was quite reliable, and thus, the accuracy of acreage estimation can be acceptable.

5 Discussions

In India, wheat is a vital crop that paved way for the national food security. Its area is about 70% of that of rice but

its production accounts for ~92% of that of rice [6]. Wheat also contributed largely during the Green Revolution in India. As cultivation of wheat employs about ~80% of total workforce of the Jharkhand State, it becomes imperative to understand the growth stages of wheat including its various phenological features. So the RS data offer great potential in estimating crop phenology, extent, and yield at various scales owing to their synoptic and repetitive coverage. Over the last few decades, phenology and RS-based vegetation indices (e.g., NDVI, EVI, LAI) have been used to derive phenological metrics [12, 18, 24]. Crop phenological transitions (e.g., emergence, heading, anthesis) help not only to track crop growth but also to estimate crop production. RS-based spectral profile can also aid in the discrimination of crop types because each crop has an obvious distinct temporal and spectral signatures. Therefore, the value of crop growth dynamics and phenology is more evident when treating the spectral profile separately for each crops and may provide substantial information on yields [7, 42, 49, 51]. For deriving phenological metrics, a known limitation of this study was the coarse spatial resolution of satellite data (MODIS NDVI products: 250 m), but it has better temporal resolution [25, 48, 49]. This presents significant challenges where average plot sizes can be far smaller than pixel sizes of the RS data. The aforementioned limitation

Fig. 7 Wheat phenological parameters obtained from 16-day NDVI composite data (MOD13Q1). The four lines represent wheat plot locations (plot nos. 1, 2, 3, and 4) collected during field visit

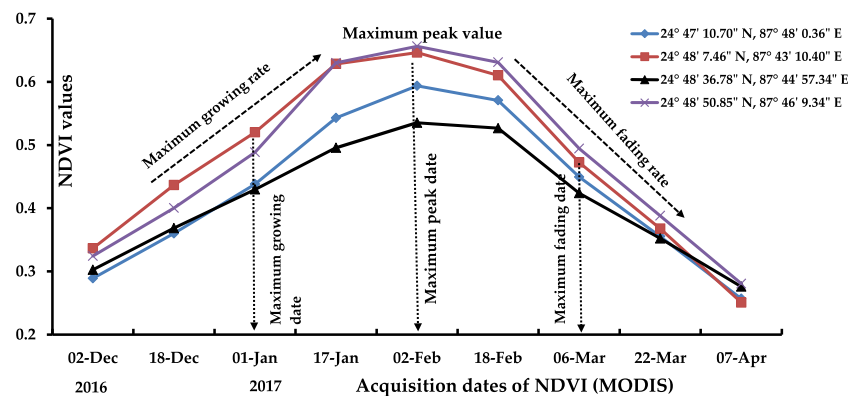


Table 4 NDVI-based profile representing wheat phenology metrics and their characteristics

Features	Phenological interpretation	Average NDVI and date	Growth stages
Maximum growing rate	The local maximum value of the first derivative curve on the left side	0.52 (first week, January 2017)	Wintering
Maximum peak value	Maximum NDVI value during a growing season	0.66 (first week, February 2017)	Milking
Maximum fading rate	The local minimum value of the first derivative curve on the right side	0.49 (second week, March 2017)	Maturing

can be observed in most of the states of India (Bihar, Jharkhand, Odisha, Chhattisgarh, etc.), where most of the farm holdings are marginal.

Hence, the present study was focused on estimating wheat acreage and yield by using RS-based data, wherein phenological metrics (Landsat-based) and DT algorithms were employed, in an intensive agricultural region, Sahibganj District (Jharkhand state) in India. Our findings suggest that MODIS-derived phenological metrics from NDVI data can be utilized to derive phenological transitions of wheat despite limited by lower spatial resolution. These derive phenological transitions can be compared with lower temporal resolution data, such as Landsat-8 OLI data. Our results suggest that the phenological transitions and growth phases, such as maximum growing date, maximum peak date, and maximum fading date of wheat, were well comparable with MODIS-derived phenological transitions. In this context, several studies have applied NDVI time-series data to capture phenological transitions, and this approach is being used to monitor the phenology of various agricultural crops in Asia, Europe, and America [8, 13, 27, 50]. The knowledge gained through phenological transitions can be used for generating rules for DT classifier. Our key findings suggest that by combining phenological information and DT classifier, it is possible to obtain accurate crop types map based on multi-temporal satellite data. We found that the developed wheat acreage map of Sahibganj District has overall accuracy of 80%. In this respect, Friedl and Brodley [10] have suggested that DT classifier can consistently outperform the supervised algorithms, such as maximum likelihood and linear discriminant classifiers in regard to classification accuracy. DT algorithms are non-parametric and, so, more robust to provide crop type maps. Pal and Mather [29] obtained an accuracy of ~86% in

identifying crop using DT classifier, while Punia et al. [39], Sharma et al. [43], and Palchowdhuri et al. [30] reported an accuracy of ~90% for crop type mapping. Notably, phenological information and DT approaches were used to estimate acreage of crop in Sahibganj District of Jharkhand. The approaches of this case study can be applied in other provinces of India to map acreages of crops and further, upscaling the crop yields. The embedded field-based GPS locations of wheat plots further strengthened the outcomes of acreage of wheat and thereby, the accuracy of predicting yields.

We further employed simple linear regression technique to predict yields of wheat, wherein we utilized time-series available area and production statistics from DES database (2017). In this respect, we found that predicted wheat yield of 1.17 t/ha for 2016–2017 Rabi (winter) season. These results were well comparable with the DES database (2017) that showed yield of 1.24 (mean of 2002–2016) tonnes/ha. To our knowledge there were no literatures for the Sahibganj Province with respect to wheat acreage map and yield prediction using RS data and yield model. Therefore, this unique study could be beneficial for acreage estimation and crop monitoring for the Sahibganj Province. Furthermore, studies have suggested that statistical regression models using observed yield statistics and vegetation indices at certain growth stage of the growing season can give an indication of crop yield [3, 45]. This study has used simple linear yield model on the basis of long-term area and production data, wherein satellite-derived acreage was used as an input to the yield model. Albeit, this is a simple approach but it facilitate to estimate yield accurately. Nevertheless, other statistical models, such as linear and non-linear models, can be utilized to predict crop yields. Forecasting methods, such as logistic and Gompertz, were also available to predict wheat yields [37].

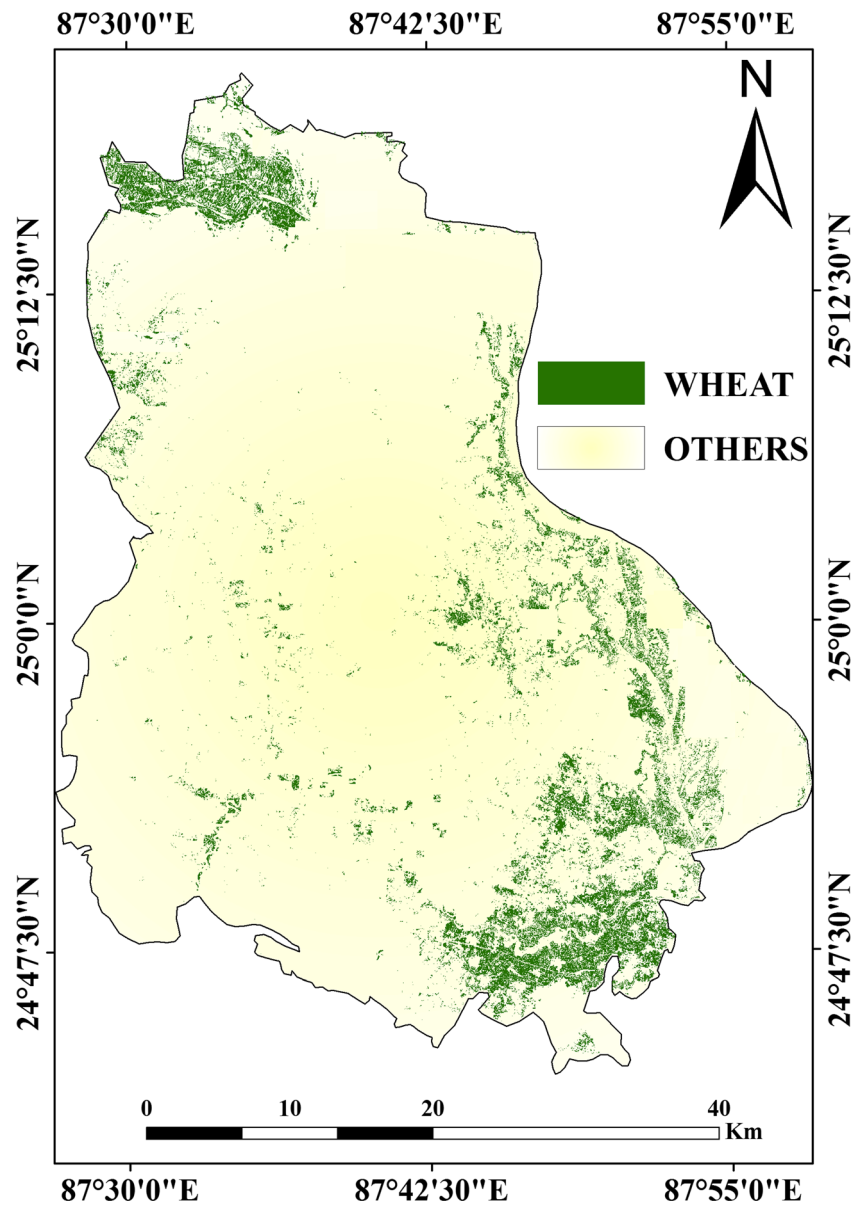
Keeping in view the importance of wheat, quantitative assessment of APY was critical for a wide range of applications and for policy makers. Further, quantitative assessment at regional level was extremely vital as it form the basis for economic and policy planning by state and central governments. Accurate maps of wheat crop extent are vital for its yield estimation from RS-based data as the extent maps can be used to diminish the errors introduced by non-agricultural land covers. So, generating accurate wheat maps can form the first stage of RS-based yield analysis.

Table 5 NDVI threshold values and rules used in DT classifier to extract wheat crop acreage

Months	NDVI threshold	Rules
January	0.39–0.45	$(b_1 > 0.39 \text{ and } < 0.45), (b_2 > 0.58 \text{ and } < 0.63),$ and $(b_3 > 0.29 \text{ and } < 0.34)$
February	0.58–0.63	
March	0.29–0.34	

The variables b_1 , b_2 , and b_3 represent January, February, and March, months, respectively

Fig. 8 Wheat acreage map of Sahibganj District derived from Landsat-8 (OLI) data using DT classifier



6 Conclusions

This study indicates that phenological parameters derived from RS-based NDVI profile were capable of identifying the wheat crop. The distinct phenological transition profile of

Table 6 Comparisons of wheat APY between predicted and reported by DES (2017) for the Sahibganj District

Wheat (APY)	Predicted	DES (mean)	DES (min–max)
Area (hectares)	3870	2846	1245–5431
Production (tonnes)	4523	3482	1617–5616
Yield (tonnes/ha)	1.17	1.24	0.87–1.55

As per the DES, the mean, minimum, and maximum values have been shown

wheat makes feasibility for wheat mapping and acreage estimation accurately as the NDVI profile separate the wheat from non-agriculture land covers. By employing DT classifier, acreage of wheat map was generated, and further, the wheat map was evaluated against GPS points of wheat plots. The accuracy assessment results showed that the developed wheat acreage map has overall accuracy of 80%. As other natural vegetation classes overlap each other on the condition (rules)

Table 7 Accuracy assessment parameters of wheat map obtained from DT classifier

Overall accuracy	80%
Producer accuracy	78%
User accuracy	70%
Kappa coefficient	0.7

created for DT algorithms, it was hard to discriminate the wheat map with > 90% of overall accuracy. Notably, the classified map was quite reliable, and thus, the accuracy of acreage estimation can be acceptable within the limit of errors.

The DT approach used to produce the acreage map based on the usages of the multi-temporal data (Landsat-8) is strictly non-parametric and based on the expert knowledge. Thus, the DT classifier is very subjective in nature as the accuracy highly depends on threshold based knowledge. Based on the Landsat-8 (OLI) sensor data, the acreage of wheat was estimated as ~3870 ha for the year 2016–2017 (Rabi season) (Table 6). Further, wheat yield over the Sahibganj District for the year 2016–2017 has been predicted as ~1.17 t/ha, wherein total wheat production has been predicted as ~4523 t. As per the DES [6] report, the production ranged from 1617 to 5616 t in the district. Our predicted values were more than the mean values, but it was within the range of maximum production (5616 t) reported by DES [6]. We concluded that satellite-derived acreage estimation can be successfully preferred instead of the traditional-based CCE survey. It is a reliable, cost-effective, and time-saving technique, which could help to the decision and policy maker by giving an early estimation of yield by using statistical regression method.

Acknowledgements Author's thanks to United State Geological Survey (USGS) for providing the MODIS and Landsat-8 OLI sensor data. We also thank to Anniket Rajak and Abhishek Rajak for the support during field data collection. Authors also thank to anonymous reviewers for their constructive comments.

Author's Contributions Conceived, designed research, analyzed data, and wrote the manuscript: BRP and AKR.

Funding Information This research was supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST) project grant no. YSS/2015/000801. The first author (BRP) was funded by a SERB, while the second author (AKR) was involved as a member.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Atzberger C (2013) Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sens* 5:949–981. <https://doi.org/10.3390/rs5020949>
- Baby A, Shekh AM (2005) Field calibration and evaluation of crop simulation model InfoCrop to estimate wheat yields. *J Agrometeorol* 7:199–207
- Bolton DK, Friedl MA (2013) Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric For Meteorol* 173:74–84. <https://doi.org/10.1016/j.agrformet.2013.01.007>
- Brown ME, de Beurs KM, Marshall M (2012) Global phenological response to climate change in crop areas using satellite remote sensing of vegetation, humidity and temperature over 26 years. *Remote Sens Environ* 126:174–183. <https://doi.org/10.1016/j.rse.2012.08.009>
- Chen Z, Li S, Ren J et al (2008) Monitoring and management of agriculture with remote sensing. In: *Advances in Land Remote Sensing*. Springer, Dordrecht, pp 397–421
- Directorate of Economics & Statistics, DAC&FW (DES) (2018) Agricultural statistics at a glance. Department of agriculture, cooperation and farmers welfare, ministry of agriculture and farmers welfare. Government of India, Government of India <https://eands.dacnet.nic.in/Default.htm>. Accessed 20 April 2018
- Duncan JMA, Dash J, Atkinson PM (2015) The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Front Environ Sci* 3:1–16. <https://doi.org/10.3389/fenvs.2015.00056>
- Durgun Y, Gobin A, Van De Kerchove R, Tychon B (2016) Crop area mapping using 100-m Proba-V time series. *Remote Sens* 8: 585. <https://doi.org/10.3390/rs8070585>
- Duveiller G, Defoumy P (2010) A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sens Environ* 114:2637–2650. <https://doi.org/10.1016/j.rse.2010.06.001>
- Friedl MA, Brodley CE (1997) Decision tree classification of land cover from remotely sensed data. *Remote Sens Environ* 61:399–409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7)
- Galford GL, Mustard JF, Melillo J et al (2008) Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens Environ* 112:576–587. <https://doi.org/10.1016/j.rse.2007.05.017>
- Ganguly S, Friedl MA, Tan B et al (2010) Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. *Remote Sens Environ* 114:1805–1816. <https://doi.org/10.1016/j.rse.2010.04.005>
- González-Gómez L, Campos I, Calera A (2018) Use of different temporal scales to monitor phenology and its relationship with temporal evolution of normalized difference vegetation index in wheat. *J Appl Remote Sens* 12:1. <https://doi.org/10.1117/1.JRS.12.026010>
- Haldar AK, Srivastava R, Thampi CJ, Sarkar D, Singh DS, Sehgal J, Velayutham M (1996) Soils of Bihar for optimizing land use (soils of India series). National Bureau of soil Survey and land use planning, Nagpur, p 70
- Hao P, Wang L, Niu Z (2015) Comparison of hybrid classifiers for crop classification using normalized difference vegetation index time series: a case study for major crops in North Xinjiang, China. *PLoS One* 10:e0137748. <https://doi.org/10.1371/journal.pone.0137748>
- Huete A, Didan K, Miura T et al (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* 83:195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Jharkhand Economic Survey (JES) (2018) Jharkhand economic survey 2018-19, planning cum finance department, centre for fiscal studies, government of Jharkhand. https://financejharkhand.gov.in/pdf/JES_2018_19_ebook.pdf. Accessed 12 March 2019
- Jiang Z, Huete AR, Chen J et al (2006) Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens Environ* 101:366–378. <https://doi.org/10.1016/j.rse.2006.01.003>
- Kalra N, Aggarwal PK, Singh AK, et al (2006) Methodology for national wheat yield forecast using wheat growth model, WTGROWS, and remote sensing inputs. In: Kuligowski RJ, Parihar JS, Saito G (eds) In Proc. SPIE 6411, agriculture and hydrology applications of remote sensing. 641106. <https://doi.org/10.1117/12.697698>

20. Kantakumar LN, Neelamsetti P (2015) Multi-temporal land use classification using hybrid approach. *Egypt J Remote Sens Space Sci* 18:289–295. <https://doi.org/10.1016/j.ejrs.2015.09.003>
21. Kastens J, Kastens T, Kastens D et al (2005) Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sens Environ* 99:341–356. <https://doi.org/10.1016/j.rse.2005.09.010>
22. Konecny G (2003) Geoinformation: remote sensing, photogrammetry and geographic information systems. Taylor & Francis
23. Krishnan P, Sharma RK, Dass A et al (2016) Web-based crop model: web InfoCrop—wheat to simulate the growth and yield of wheat. *Comput Electron Agric* 127:324–335. <https://doi.org/10.1016/j.compag.2016.06.008>
24. Lobell DB, Asner GP (2004) Cropland distributions from temporal unmixing of MODIS data. *Remote Sens Environ* 93:412–422. <https://doi.org/10.1016/j.rse.2004.08.002>
25. Lobell DB (2013) The use of satellite data for crop yield gap analysis. *Field Crop Res* 143:56–64. <https://doi.org/10.1016/j.fcr.2012.08.008>
26. Mohabey DP, Sharma NK, Kumar A (2015) Change detection in LU/LC of Sahibganj district, Jharkhand, India using remote sensing & GIS. *IJARSGG* 3(2):59–70
27. Moulin S, Kergoat L, Viovy N, Dedieu G, (1997) Global-Scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. *J Climate* 10:1154–1170
28. Nagy A, Feher J, Tamas J (2018) Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. *Comput Electron Agric* 151:41–49. <https://doi.org/10.1016/j.compag.2018.05.035>
29. Pal M, Mather PM (2001) Decision tree based classification of remotely sensed data. In *Proceedings of the 22nd Asian Conference on Remote Sensing (ACRS)*, Singapore, 5–9 November 2001. Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore, Singapore Institute of Surveyors and Valuers (SISV), Asian Association on Remote Sensing (AARS), Singapore, pp 245–248
30. Palchowdhuri Y, Valcarce-Diñeiro R, King P, Sanabria-Soto M (2018) Classification of multi-temporal spectral indices for crop type mapping: a case study in Coalville, UK. *J Agric Sci* 156:24–36. <https://doi.org/10.1017/S0021859617000879>
31. Pan Y, Li L, Zhang J et al (2012) Winter wheat area estimation from MODIS-EVI time series data using the crop proportion phenology index. *Remote Sens Environ* 119:232–242. <https://doi.org/10.1016/j.rse.2011.10.011>
32. Panda SS, Hoogenboom G, Paz JO (2010) Remote sensing and geospatial technological applications for site-specific management of fruit and nut Crops: a review. *Remote Sens* 2:1973–1997. <https://doi.org/10.3390/rs2081973>
33. Parida BR (2006) Analysing the effect of severity and duration of agricultural drought on crop performance using Terra/MODIS satellite data and meteorological data. Dissertation, The International Institute for Geo-Information Science and Earth Observation
34. Parida BR, Collado WB, Borah R et al (2008a) Detecting drought-prone areas of rice agriculture using a MODIS-derived soil moisture index. *GI Sci Remote Sens* 45:109–129. <https://doi.org/10.2747/1548-1603.45.1.109>
35. Parida BR, Oinam B, Patel NR et al (2008b) Land surface temperature variation in relation to vegetation type using MODIS satellite data in Gujarat state of India. *Int J Remote Sens* 29:4219–4235. <https://doi.org/10.1080/01431160701871096>
36. Parihar JS (2016) FASAL concept in meeting the requirements of assessment and forecasting crop production affected by extreme weather events. *Mausam* 67(1):93–104
37. Parmar RS, Rajarathinam A, Patel HK, Patel KV (2016) Statistical modeling on area, production and productivity of cotton (*Gossypium* spp.) crop for Ahmedabad region of Gujarat state. *J Pure Appl Microbiol* 10:751–759
38. Patel NR, Bhattacharjee B, Mohammed AJ et al (2006) Remote sensing of regional yield assessment of wheat in Haryana, India. *Int J Remote Sens* 27:4071–4090. <https://doi.org/10.1080/01431160500377188>
39. Punia M, Joshi PK, Porwal MC (2011) Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. *Expert Syst Appl* 38:5577–5583
40. Reed BC, Schwartz MD, Xiao X (2003) Remote sensing phenology. In: *Phenology of ecosystem processes: applications in global change research*. Springer, New York, pp 365–381
41. Saxena R, Bhardwaj V, Kalra N (2006) Simulation of wheat yield using WTGROWS in northern India. *J Agrometeorol* 8(1):87–90
42. Sehgal VK, Rajak DR, Chaudhary KN, and Dadhwal VK, (2002) Improved regional yield prediction by Crop growth monitoring system using Remote Sensing derived crop phenology, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRS & SIS)*, Vol. 34, Part 7, Resource and Environment Monitoring, Hyderabad, India
43. Sharma R, Ghosh A, Joshi PK (2013) Decision tree approach for classification of remotely sensed satellite data using open source support. *J Earth System Sci* 122:1237–1247. <https://doi.org/10.1007/s12040-013-0339-2>
44. Singh PK, Singh KK, Baxla AK, Rathore LS (2015) Impact of climatic variability on wheat yield prediction using DSSAT v 4.5 (CERES-wheat) model for the different agroclimatic zones in India. In: Singh AK, Dagar JC, Arunachalam A et al (eds) *Climate change modelling, planning and policy for agriculture*. Springer, New Delhi, pp 45–55
45. Son NT, Chen CF, Chen CR et al (2014) A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agric For Meteorol* 197:52–64. <https://doi.org/10.1016/j.agrformet.2014.06.007>
46. U.S. Geological Survey (2018) Landsat 8 (L8) data users handbook (version 3, LSDS 1574). USGS Earth Resources Observation and Science, Sioux Falls
47. Vieira MA, Formaggio AR, Rennó CD et al (2012) Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sens Environ* 123: 553–562. <https://doi.org/10.1016/j.rse.2012.04.011>
48. Vrieling A, de Beurs KM, Brown ME (2011) Variability of African farming systems from phenological analysis of NDVI time series. *Clim Chang* 109:455–477. <https://doi.org/10.1007/s10584-011-0049-1>
49. Vrieling A, de Leeuw J, Said M (2013) Length of growing period over Africa: variability and trends from 30 years of NDVI time series. *Remote Sens* 5:982–1000. <https://doi.org/10.3390/rs5020982>
50. Yu L, Liu T, Bu K, Yan F, Yang J, Chang L, Zhang S (2017) Monitoring the long term vegetation phenology change in Northeast China from 1982 to 2015. *Sci Rep* 7(1):14770. <https://doi.org/10.1038/s41598-017-14918-4>
51. Zheng Y, Wu B, Zhang M, Zeng H (2016) Crop phenology detection using high spatio-temporal resolution data fused from SPOT5 and MODIS products. *Sensors* 16:2099. <https://doi.org/10.3390/s16122099>