SOYBEAN CROPLAND MAPPING USING MULTI-TEMPORAL SENTINEL-1 DATA

Mamta Kumari 1,*, C.S. Murthy 1, Varun Pandey 1, G.D. Bairagi 2

1 Agricultural Sciences & Applications Group, RS & GIS – Applications Area, National Remote Sensing Centre, Indian Space Research Organization, Hyderabad, India
2 Madhya Pradesh Council of Science and Technology, Bhopal, Madhya Pradesh, India
* mamta9507@gmail.com

Commission III, WG III/10

KEY WORDS: Soybean, Sentinel-1, SAR, Classification, Phenology, SVM

ABSTRACT:

Soybean, a high value oilseed crop, is predominantly grown in the rainfed agro-ecosystem of central and peninsular India. Accurate and up-to-date assessment of the spatial distribution of soybean cultivated area is a key information requirement of all stakeholders including policy makers, soybean farmers and consumers. A methodology for timely assessment with high precision of soybean crop using satellite data is yet not operational in India. In this scenario, synthetic aperture radar (SAR) has been shown to be a reliable form of gathering crop information, especially during monsoon season. In this work, repeat coverage from the C-band Sentinel-1 satellite over Ujjain district, Madhya Pradesh is used for in-season soybean crop mapping along with other agricultural land-cover types. The data were processed through four steps: (a) data preprocessing, (b) constructing smooth time series backscatter data, (c) soybean crop classification using knowledge-based decision rule classifier and support vector machines (SVM) and (d) accuracy assessment. The results indicated that the smooth VH backscatter profiles reflected the temporal characteristics of soybean crop growing in the study region. Phenological characteristics were also derived from the smoothed S-1 VH backscatter time series to segregate early and late sown soybean. This information was used as an input to a decision-rule classifier and SVM in order to classify the input data into soybean and other crops. An overall accuracy of more than 80% using SVM and 75% using rule based classifier, in Ujjain district was achieved. These results demonstrate the scope for using time-series S-1 VH data to develop an operational soybean crop-monitoring framework.

1. INTRODUCTION

Soybean (Glycine max), being a high value oilseed, is also known as the “Golden Bean” of the twentieth century. In the recent past, soybean cultivation has increased manifold as compared to any other oilseed crop in India and stands next only to groundnut. Soybean is primarily grown in the rainfed agro-ecosystem of central and peninsular India. About 53 per cent of the cropped area falls under this crop in Madhya Pradesh. In recent years, recurrent aberrant rainfall pattern and intermittent dry spell, both in time and space, were observed in these parts of country. These discrepancies inhibited the successful cultivation of soybean crop and leads to significant annual variation in cultivated area of soybean. Access to timely and reliable data on crop distribution and its condition assumes importance for enabling decision-makers to make informed decisions with respect to imports, exports, and other policy matters. Since, soybean is a short duration kharif season crops, mapping of soybean crops and their growth assessment using optical data is difficult due to consistent clouds. Mapping and monitoring of soybean crop is important due to various reasons such as, this crop is not covered in FASAL project; there is requirement from the Ministry of Agriculture to increase the number of crops for crop inventory; economic values of soybean crop are very high; and soybean crop is highly relevant from the crop insurance perspective.

Optical remote sensing is one of the most commonly used data source for deriving crop information such as crop type map, biophysical parameters, crop growth stages, yield etc. Crop leaves are very sensitive to visible and infrared regions. However, persistent cloud cover during the monsoon season makes the availability of optical data difficult. This is almost impossible to obtain a cloud-free optical image during this period in a subtropical country like India. In this scenario, Synthetic Aperture Radar (SAR) is the only option to map the kharif season crops. Contrary to optical sensors, SAR sensors are capable to acquire images under all weather conditions which make it appropriate for crop mapping and monitoring during rainy season (Liao et al., 2018). However, the use of SAR data has not been established for agricultural applications, as compared to optical data. There are several reasons, such as, the complexity, cost and availability of SAR data, in addition to the difficulty of data interpretation. The recently launched Sentinel-1 with the advantage of free distribution provides more opportunities to derive crop type map. Being the high spatial and temporal resolution data, it allows not only to discriminate different crop types but also to assess their growth status.

Previous studies performed the classification on Sentinel-1 data for land cover mapping by utilizing backscatter values of single dual polarimetric data (Abdikan et al. 2016, Whelen et al., 2017), especially to map rice crop (Son et al., 2018, Nguyen et al., 2017) and yielded maps with descent accuracy. However, no studies have been carried out to map soybean crop using Sentinel-1 data and very few studies are done using other SAR data (McNairn et al., 2014). In this study, potential of dual polarized Sentinel-1 backscatter data for in-season mapping of soybean crop is investigated. We analyzed multi-temporal C-band VH and VH polarized Sentinel-1 SAR data to evaluate the backscatter behavior of soybean crop and to map its spatial distribution in Ujjain district, Madhya Pradesh.
2. METHODOLOGY

2.1 Study area

Ujjain district is situated in the heart of Malwa Plateau, Madhya Pradesh at a general elevation of 527 meter above mean sea level. The district area extends between the parallels of latitude 22° 50' and 23° 46' North and between the meridians of longitude 75° 08' and 76° 16' East. The normal annual rainfall of Ujjain district is 914.5 mm. Ujjain district receive maximum rainfall (about 92.1%) during southwest monsoon period i.e. June to November. The total geographical area of the district is 6,130.23 Sq. Km and divided in seven tehsil and six blocks. The irrigation facilities in Ujjain district are moderate and groundwater is the main source of irrigation in the district. Ujjain district is mainly agriculture-based district and its cropping pattern is diversified. Black cotton soils with heavy to light texture are found in the whole area. Soybean is the most dominant kharif season crop. Other crops grown in kharif season are urad, maize, vegetables etc.

Fig 1. Study area: Ujjain district

2.2 Data collection

The input datasets for the analysis are listed hereunder:

2.2.1 Sentinel 1A data: This study used dual-polarized (VV/VH) C-band imagery from the European Space Agency’s (ESA) Sentinel-1A, covering 2017 soybean crop growing season. The Sentinel-1A operates at a radar wavelength of about 5.6 cm (corresponding to frequency 5.405 GHz; C-band), with a repeat coverage of 12 days. It has four imaging modes: interferometric wide swath (IW), extra wide swath (EW), strip map (SM), and wave (WV). This study used the S1A_IW_GRDH product, which contains VH and VV polarizations with a swath width of approximately 250 km and a spatial resolution of 10 m; which provides backscatter normalized with respect to area. Eleven dates (22 nos. images) with a descending orbit during June to October 2017 were used in the analysis. The incidence angle of the VH and VV images acquired over the study region ranges from approximately 37–45°. The details of the data collection dates are: 01st June, 13th June, 25th June, 19th July, 31st Jul, 12th Aug, 24th Aug, 05th Sep, 17th Sep, 29th Sep, 23rd Oct.

2.2.2 Optical data and Ancillary data: For the study area, multi-temporal Resourcesat-2 AWIFS15-days NDVI composite and Landsat-8 OLI data for the year 2017 (June to October) were used to produce reference classification maps for validating the Sentinel-1A soybean crop map. The Resourcesat 2 AWiFs15-days NDVI composite data of 56 m resolution was obtained from NRSC. The Landsat 8 OLI data of 30 m resolution was downloaded from the USGS website.

2.2.3 Ground reference data: Field information on the distribution of different land cover classes in general and soybean crop in particular, has been collected. A mobile application - FASAL app – developed by NRSC was used during field visit during September 2017. Various information such as crop type, crop sowing and harvesting dates, field moisture status, plant density etc., were collected during field visit, using mobile app. This aided in faster and more efficient collection of ground information about the crop along with field photographs, as well as in building up a geodatabase with which could be directly uploaded on to Bhuvan geoportal.

2.3 Satellite data processing

SAR derived time-series backscatter data are the key datasets in this project as per the methodology framework depicted in Figure 2.

2.3.1 Data pre-processing: The Sentinel-1A data were pre-processed using ESA’s Sentinel’s Application Platform (SNAP). Firstly, radiometric calibration was performed to convert digital pixel values of VV and VH amplitude into sigma naught ($\sigma^0$) values that can be directly related to the radar backscatter of the scene. Radiometric calibration is required to remove the errors due to fluctuations in the transmitted power, receiver gains, system noise and the illumination pattern of the antenna. Speckles from the $\sigma^0$ data were filtered using adaptive Lee filter with 5 × 5 moving window size to make the interpretation of features easy. The terrain correction was applied to compensate the distortions due to the side-looking geometry of images. In this study, the Range Doppler Terrain correction algorithm that uses the elevation data from the 1 arc-second DEM product from the Shuttle Radar Topography Mission (SRTM) was used for geocoding and terrain correction. In this process, data are resampled and reprojected to the Universal Transverse Mercator (UTM) coordinate system (zone 43 N) with a grid of 10 m spacing maintaining the 20 m spatial resolution. The data were then mosaicked and subset over the study region. Because Sentinel-1 data are not fully polarimetric, polarimetric decomposition methods were not available to be included in the analysis. Both polarizations were examined separately as well as in combination.

Fig 2. Framework of methodology showing steps for soybean crop classification in study area

2.3.2 Time series filtering to construct smooth time series backscatter data: The backscatter time-series data
possess the short-term influence of environmental conditions and inherent noises in the Sentinel-1A data (such as noise due to speckle). Filtering is required to reduce these noises that can potentially affect the classification results. In current study, wavelet transform, an adaptive noise reduction method, is used to cope up the problem. It has certain advantages such as to preserve the full non-linearity and non-stationarity in physical space, which is more conducive to data denoising and information extraction. It has been extensively used for the denoising the time series of satellite data (Sakamoto et al. 2005; Galford et al. 2008; Martínez & Gilbert 2009; Chen. Huang et al. 2011; Chen. Son et al. 2011). The wavelet transform $W(s, r)$ of a signal $x(t)$ can be expressed as follows:

$$W(s, r) = \int x(t)\psi^*(\frac{t-r}{s})d(t),$$

where $s > 0$ and $r \in R$; $x(t)$ is the analyzed input signal; $\psi(t)$ is the mother wavelet; and $\tau$ rare scaling and translation parameters. A mother wavelet, Coiflet function (with order 4) was preferred over Daubechies and Symlet functions for signal decomposition, to get time profile of backscatter data for deriving the crop phenological information (Sakamoto et al. 2005). The processed output is a smoothed backscatter signal, which will be used for the extraction of different phenological stages of soybean (e.g., a start of the season, heading time, and length of season); in order to discriminate the soybean pixels from the other land cover classes.

### 2.3.3 Extraction of Vegetation Phenology Parameters

The three phenological indicators, viz., date of beginning of season (DoS), date of maximum backscatter (DoM), and Length of the season (LoS) are derived in this step. This allows delineating the areas as “early planted soybean and late planted soybean”, and all other areas were placed in a generalized “non-soybean” class. These parameters are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition and Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoS</td>
<td>During the crop growing season, the date of the beginning of season is defined as the first localminima in time-series of smooth VH backscatter data</td>
</tr>
<tr>
<td>DoM</td>
<td>During the growing season, the date when backscatter reaches a maximum value is defined as the local maxima in time-series of smooth VH backscatter data. This date must come after the date of the beginning of season, where it reaches its local minimum backscatter value.</td>
</tr>
<tr>
<td>LoS</td>
<td>The length of the season is defined as the number of days difference between DoM and DoS.</td>
</tr>
</tbody>
</table>

Table 1. Phenological parameters for the rule-based classification

### 2.3.4 Training data selection

This study focused on delineating soybean crops (early and late planted) in Ujjain district. The non-cropped areas, including orchards, other crops, forests, built-up areas and water bodies, were masked out using the smooth VH backscatter data to limit the analysis to the cropped areas. The Support vector machine (SVM) used training samples extracted from the filtered time series VH backscatter data. The fuzzy C-mean clustering algorithm (Dunn 1973; Bezdek 1981) was applied to classify the extracted pixels into several clusters. The clusters reflecting the pattern of the soybean crop being analyzed were selected to train SVM for classification.

### 2.3.5 Support vector machines

Support Vector Machines (SVMs), a non-parametric statistical learning method, has recently been used in image classification for various applications. Being a supervised learning method, the SVM need user-defined parameters and each parameter has different impact on kernels hence the classification accuracy of SVMs is based upon the choice of the parameters and kernels. The algorithm projects training data of the input space into a high-dimensional space using a kernel function in which the classes can be separable.

The basic idea behind SVMs is to separate the two different classes through a hyperplane which is specified by its normal vector $‘w’$ determining the hyperplane orientation and the bias ‘b’. The optimal hyperplane can be given as $\langle w, x \rangle + b = 0$ and the corresponding decision function is $f(x) = sgn(\langle w, x \rangle + b)$, where $x$ is a point lying on the hyperplane. The sign of $f(x)$ depends on the side of the hyperplane where the sample lies.

For a non-linear case, a kernel function $K(\mathbf{X}, \mathbf{w})$ is introduced to enable an efficient computation. The ‘kernel trick’ can be applied since all feature vectors only occurred in dot products. In current study, radial basis function (RBF) was used for SVM classification of the VH backscatter data as it has demonstrated to produce more accurate results than other functions such as linear and polynomial kernels (Camps-Valls and Bruzzone 2005) for classification of remotely sensed data (Huang et al. 2002; Keerthi and Lin 2003; Pal and Mather 2005). The parameters such as cost parameter (C), gamma ($\gamma$), bias term ($\tau$) were used in SVM classification. The optimal values of these parameters were identified to increase the classification accuracy, by various permutations and combinations of these parameters in different models for sensitivity analysis of SVM architecture.

### 2.3.6 Accuracy Assessment

For validation and evaluation of classification results, standard accuracy assessment measures i.e., kappa coefficient, overall accuracy, omission error, and commission error were used.

### 3. RESULTS & DISCUSSIONS

The current study attempts to delineate kharif soybean crop area using SAR data. The 12-day time series VH backscatter from Sentinel-1A SAR has been effectively used to delineate the soybean crop area from other land use land cover types, in view of soybean being short duration crop and grown in the during monsoon season. The present study was carried out in Ujjain district, Madhya Pradesh. The results obtained in this study are described in this section.

#### 3.1 Temporal soybean backscatter signature from Sentinel-1 SAR Data

Figure 3 shows the colour composites, which are created by using the multi-temporal SAR acquisitions in order to highlight the temporal characteristics within the soybean fields. The red, blue, and green coolers in the figure correspond to the images acquired during the sowing (June/July), flowering/pod formation (August/September) and post harvested (October) period, respectively.
The profiles produced from the wavelet transform of VH backscatter data revealed temporal discriminations between two types of soybean crop (early sown and late sown) and those of other land use cover types (e.g., orchards, water body and built-up area). The VH backscatter intensities of soybean crops showed a gradual increase of backscatter values after sowing on reaching the maximum value during the pod formation of crops. The backscatter intensities of soybean crops during the sowing period were approximately −21 dB and those during the pod formation periods (i.e., approximately 50 days after sowing) were −15 dB, respectively. After sowing period, soybean crop height increase with the unfolding of trifoliate leaves till the start of flowering stage. Thus, the backscatter intensity showed a significant increase during this vegetative phase due to the greater the contribution of VH from volume scattering in soybean plant’s internal canopies. After pod formation period, the soybean plant was characterized by a decrease in the number of leaves (i.e., ripening stages), the backscatter intensities slightly decreased until the harvesting period. The VH backscatter intensities from orchards and built-up areas were relatively high. The backscatter values of water were also distinguishable from those of soybean crop and other land use cover types, showing very low VH backscatter values throughout the year. Only VH backscatter temporal profile shown consistent temporal pattern and showed an increase in backscatter with crop growth, however, VV backscatter and derivatives of backscatter reflected the inconsistent and abrupt changes with crop growth in temporal backscatter profile.

3.2 Characteristics of smooth VH backscatter LUC profiles

Fig 4 illustrated the some noise in raw VH, VV, VH/VV ratio and VH+VV sum backscatter profiles in the signature of soybean crop growth profile. Among all these backscatter profiles, VH backscatter profile showed maximum potential by resembling the growth profile of soybean. Therefore, only VH backscatter profile was chosen for further analysis; to reduce the computation time and complexity. The smooth curves (for VH backscatter) obtained from the wavelet transform indicated that the noise was significantly mitigated, as shown in Fig 5. The cross-polarized VH backscatter was more sensitive to the growth of soybean crop than the like-polarized VV backscatter, characterizing the temporal variations of the soybean crop during the season. The more sensitivity of the soybean in the cross-polarized VH backscatter was mainly attributed to the depolarization of signal that occurred within the soybean canopy during the multiple reflection of the incoming signal inside the vegetation volume.

3.3 Thresholds Selection for rule based classification

The temporal backscatter signatures is shown in Figure 6 served to define specific thresholds which were applied to the Sentinel-1 time series in order to generate soybean crop area maps for the growing season of 2017.

To map the different type of soybean crop areas based on sowing time, the first two key parameters, DoS- backscatter at the beginning of season and DoM- the peak (maximum) of VH backscatter are very important. The amplitude backscatter difference (Δσ; the difference between DoS and DoM) also plays a significant role to decide the threshold. The corresponding positions (date) and backscatter values of DoS showed in the Figure 8.0. The peak and valley of VH backscatter (i.e., DoS, DoM) within the one soybean-growing cycle can be identified by local maxima. To eliminate unrealistic peaks, a threshold for VH backscatter is required. These two parameter thresholds (i.e., DoM ≥ 80 and Δσ ≥ 30) were selected based on a conservative deduction from training data. The third parameter, length of the season (LoS) which is defined as the temporal distance (number of days difference) between the date of beginning of season and the date when maximum backscatter value is recorded. This temporal distance has to be greater than the shortest possible soybean growing cycle and smaller than the longest possible soybean growing
cycle. From crop calendar and expert knowledge about soybean growing season in the study areas, the threshold of temporal distance which is about 70 and 110 days, respectively. If all three conditions are met then the pixel is classified as any type of soybean, otherwise as non-soybean area. To discriminate within soybean crop types, DoS and harvesting time window was identified to decide whether it is early sown or late sown soybean crop.

![Image of smoothed temporal VH backscatter profile from late and early sown soybean](a)

**3.4 Comparison of Sentinel-1A derived spatial distribution of soybean crop and classification accuracies**

The mapping results obtained from the classification of the smooth time series Sentinel-1A VH backscatter data using rule based and SVM indicated comparable spatial distributions of soybean crop (both early and late planted) with those from the ground reference data. In study area, soybean crop is extensively grown in *kharif* season and occupied most of the study region, with considerable variability in sowing time. The classification maps produced from rule based and SVM (Fig 7) were evaluated with the ground reference data (200 points) to assess the accuracy of the mapping methods. The optimal values of SVM parameters were identified (Kernel= RBF, $\gamma =0.2$, $C =100$, Pyramid level=0.2, Classification probability = 0.5) for soybean crop classification in study area. The results indicated that the mapping results obtained from SVM was slightly better than those from rule based. The overall accuracy and Kappa coefficient achieved from SVM were respectively 83.1% and 0.74, while the values from rule based were 76.3% and 0.68, respectively (Table 2). There were some error sources exaggerating the mapping results, possibly attributed to the small variation of incidence angle, ground reference error and mixed-pixel issues. We could see that soybean fields, relatively small and scattered, were mixed with other land use cover types, such as other crops (urad) having same seasonal pattern, orchards, canals and rivers in settlement areas. The mixed-pixel and boundary effects could complicate the radar backscattering coefficients, causing temporal uncertainty in discriminating between signatures of soybean crops and other land-use types, consequently leading to misclassification of soybean crop and exaggerating mapping errors in these areas. Moreover, the 12-day temporal resolution of Sentinel-1A is relatively long to capture phenological information of soybean crop (e.g. sowing and maximum growth periods), which might lead to an increased uncertainty in soybean crop type classification, especially areas of high variability in sowing. These results were reaffirmed by the MP government’s soybean crop area statistics of Ujjain district with the relative error in area values of 16% and 8.5% for rule based and SVM, respectively.

![Image of crop map showing spatial distribution of soybean along with other land use cover type derived by (a) SVM (b) Rule based classifier](b)

**Fig 7. Crop map showing spatial distribution of soybean along with other land use cover type derived by (a) SVM (b) Rule based classifier**

<table>
<thead>
<tr>
<th>Year</th>
<th>Soybean area (MP state agriculture portal) (lakh ha)</th>
<th>Sentinel-1 derived soybean area (lakh ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>by SVM classifier</td>
<td>by Rule based classifier</td>
</tr>
<tr>
<td>2014-15</td>
<td>4.33</td>
<td>-</td>
</tr>
<tr>
<td>2015-16</td>
<td>4.55</td>
<td>-</td>
</tr>
<tr>
<td>2017-18</td>
<td>4.82</td>
<td>5.15</td>
</tr>
<tr>
<td>Overall accuracy (%) (Kappa statistic)</td>
<td>83.1 (0.74)</td>
<td>76.3 (0.68)</td>
</tr>
</tbody>
</table>

**Table 2. Accuracy assessment**

**4. CONCLUSIONS**

The current study investigated the feasibility of multi-temporal Sentinel-1A VH backscatter data for soybean crop classification in Ujjain district, Madhya Pradesh. The results indicated that the smooth VH backscatter profiles (wavelet transform-based) could represent the temporal variations in growth profile of soybean crop in the study region. Moreover, the smooth VH backscatter time series could also segregate the early and late sown soybean. As the mixed-pixel, sowing time variability and crop with similar phenology could lower the level of classification accuracy, this study demonstrated that rule based and SVM classifier confirmed the validity for mapping soybean in the region. The overall accuracy and Kappa coefficient achieved from rule were, respectively, 76.3% and 0.68, while the values from SVM were 83.1% and 0.74, respectively. These results demonstrate the scope for using time-series Sentinel-1 VH data to develop an operational large-scale soybean crop mapping framework.
There is further scope of improving the classification accuracy through complementary use of cloud free optical data (such as Sentinel-2) in combination with SAR data.

REFERENCES


Duy Ba Nguyen 1, 2, and Wolfgang Wagner 2 European Rice Cropland Mapping with Sentinel-1 Data: The Mediterranean Region Case Study, Water 2017, 9, 392; doi: 10.3390/w9060392.


