

Backscatter and coherence analysis using space borne C-band data for forest characterization

Arunima Singh1*, S.K.P. Kushwaha2, and Shashi Kumar3

¹Forestry and Ecology Department, Indian Institute of Remote Sensing, Dehradun, Uttarakhand - 248001, India

²Geomatics Group, Department of Civil Engineering, Indian Institute of Technology, Roorkee, Uttarakhand - 247667, India

³Photogrammetry and Remote Sensing Department, Indian Institute of Remote Sensing, Dehradun, Uttarakhand - 248001, India

*Email: singharunima92@gmail.com

(Received: Aug. 30, 2019; in final form: May 19, 2020)

Abstract: Synthetic Aperture Radar (SAR) has shown immense potential in the area of forestry and proved to be one of the important tools of remote sensing in the characterization of a forest. Paper reports finding of a study carried out to characterize the forest area in the Dudhwa National Park, Uttar Pradesh, India using optical and SAR data. Backscatter (polarization VV and VH) and coherence images were generated with the help of a pair of datasets (Sentinel-1A of 29 May 2018 and 10 June 2018) for classification. Classification of SAR data was carried out using two classifiers i.e., Random Forest (RF) and K-Dimensional (KD) tree K-Nearest Neighbors (KNN) classifier. Results of these classifications of Sentinel-1A C-band datasets were analyzed and evaluated with respect to a reference map of forest type map prepared using Maximum Likelihood Classification (MLC) of Sentinel-2 dataset (22 March 2018). The accuracy achieved for MLC classification is 82.3%. The Random Forest accuracy is 75.2%, and the correlation is 0.7464, whereas the KD KNN accuracy is 82.5%, and having a correlation value of 0.829. So, the backscatter and coherence values can be used for the classification of forest area.

Keywords: Maximum Likelihood Classifier (MLC), KD tree KNN, Random Forest (RF), C-band, Sentinel-1A and Sentinal-2

1. Introduction

Forests are an important component of the earth system. Forest helps in the recycling of water, purification of air, enable oxygen and absorption of carbon dioxide from the air, etc. Forests are not a single entity as they also possess other resources and make an entire ecosystem. The maintenance of an ecosystem requires every single unit to function properly. The forest ecosystem is getting disturbed due to human interference and overexploitation of the resources. Monitoring, preservation, and maintenance of such a system is the utmost necessity. Remote sensing techniques provide unique solutions in the large-area assessment of forest resources as compared to time-consuming ground surveys.

Optical remote sensing data have been used for discriminating forest type and density (Hirschmugl et al., 2018), however, this technique is limited by sensing more of surface characteristics of the forest canopy (De Souza Mendes, 2019). Radar remote sensing is emerging technology in the field of forestry. Mapping and monitoring of dry tropical forests using SAR data are more useful in regions associated with frequent cloud cover. It has been observed that during mapping of a forest, a low incidence angle of the sensor is sensitive to biomass whereas high incidence angle of the sensor is sensitive to detect deforestation. The deforested area of the forest can be mapped using both optical and SAR datasets (Rahman and Sumantyo, 2010). The accuracy of the dominant forest type classification is fairly good but the mixed deciduous or coniferous species classification is comparatively erroneous (Saatchi and Rignot, 1997). The main focus of this study is to classify the forest area using Sentinel-1 C-band data at the species level using two machine learning algorithms and to generate a forest type map of the Dudhwa National Park. The classification

was carried out by generating backscatter and coherence values.

1.1 Classification Techniques

1.1.1 Maximum Likelihood Classifier

The Maximum Likelihood Classifier (MLC) works on the Gaussian theory of probability distribution (Paola and Schowengerdt, 1995). The discriminating function for each of the class is shown in the equation no 1 where X is data vector, n is number of bands, Σ_i is covariance matrix of class i, U_i is the mean vector of class i.

$$g_i(X) = p(X|w_i)p(w_i)$$

= $\frac{p(w_i)}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \cdot e^{-(1/2)(X-U_i)^T \Sigma_i^{-1}(X-U_i)}$ 1

1.1.2 Random Forest Classifier

The random forest deals with the tree predictors and further the resampling technique is used for the construction, the best split selected among all the attributes after the random sampling of all the attributes (Du et al., 2015).

1.1.3 KD Tree KNN Classifier

A new method is K- Nearest Neighbor (KNN) based on the maximal margin principle. A function is defined for the given points (x) in the defined dimensional input parameters. The entire set of training samples are ordered with respect to the given points (Blanzieri and Melgani, 2008). The algorithm used for the classification of the binary classification problems is shown in equation no 2 where y_{r_x} (i) is the class label of the ith nearest training sample.

$$kNN(x) = sign\left(\sum_{i=1}^{k} y_{r_x(i)}\right) \quad \dots \dots \quad 2$$

© Indian Society of Geomatics

1.1.4 SAR remote sensing for forest vegetation

Remote sensing sensors are of two types i.e., active and passive sensors. An active sensor can transmit electromagnetic (em) pulses through a transmitter to illuminate and it requires a receiver to receive the backscatter signals from the illuminated area of the Earth surface. The passive sensors are those which can not transmit electromagnetic pulses and they are completely dependent on any other source of energy to do the imaging. The sensors used in optical, thermal, and passive microwave remote sensing techniques are passive by nature that can measure the radiation illuminating from an object. Synthetic Aperture Radar (SAR) is an active radar imaging system that has been widely used for biophysical characterization of forest vegetation (Kumar et al., 2019, 2017; Tomar et al., 2019). The basic principles are the transmission of high-frequency pulses and recording of returning echoes to generate high spatial resolution images for various applications (Shimada et al., 2010). SAR Interferometry (InSAR) is a technique that utilizes the acquisition of two SAR data with the same SAR geometry over an area on the Earth's surface. InSAR-based approaches are using phase information for precise measurement of height and displacement. The InSAR phase component is proportional to the vertical structure of the scene under observation, in particular, relief topography can be obtained by means of two interferometric SAR surveys (López-Martínez et al., 2012). The accuracy of the coherence magnitude is obtained as a function of the number of pixels averaged and the number of independent samples per pixel (Touzi et al., 1999). A recent study shows, classification of the forest can be done using Sentinel-2 successfully and gives a fine classification result for forest types on large scale (Puletti et al., 2017).

Forest degradation is an important issue in global environmental studies. Changes in land cover classes such as a degraded forest are analyzed by supervised analysis and backscatter spatial statistics (De Grandi et al, 2015). A better approximation of relative canopy density than optical data derived canopy density is obtained using SAR C-band data which can penetrate through forest vegetation and make it possible to extract information of crown components (Varghese et al., 2016). Optical remote sensing data have been used in the past in discriminating forest type, density, and extent. Optical sensors have major limitations in providing information related to forest structure as they can only detect forest canopy (Kumar, 2009).

Main et al., (2016) have investigated the relationships between hyper- temporal C-band ASAR data and woody structural parameters. Sentinel-1 radar backscatter provides useful observations during the cloudy monsoon season. The spatio-temporal variations have been analyzed using sentinel-1 and sentinel-2 data for agriculture practices (Ferrant et al., 2017). Though remote sensing and GIS techniques are efficient in the assessment of forest health, the major advantage of SAR remote sensing is the assessment of biophysical parameters for landscapes that primarily depends upon traditional field methods (Chandola, 2014).

2. Study area

The Dudhwa National park is a part of the Dudhwa Tiger reserve and Kishanpur Wildlife sanctuary. The location of the park is on the Indo-Nepal border in the Nighasan Tehsil of District Lakhimpur Kheri in Uttar Pradesh, India and lies between 28'18' to 28'42'N latitudes and 80'28' to 80'57' E longitudes (Figure 1). The total area of the park is around 768.62 km². The park is divided into two buffers, the southern and northern buffer which comprises 124.01 km² and 660.23 km² area. These areas were once part of the North Kheri Forest Division. It was declared as National Park by the State Government in 1975. The total grassland present is 19% of the forest area, the third-largest habitat in India and wetland comprising of rivers, streams, lakes, and marshes.

Dudhwa National Park is one of the most diverse forests in India, it comprises more of the endangered species, obligate species of tall wet grasslands, and species of restricted distribution. The total area is further subdivided into two parts, Core Zone and Buffer Zone, the former is 490.2 km² whereas later is having 190.03 km² and the total area taken for the study is 768.62 km². This is one of the most threatened ecosystems of India. The forest of the study area has been classified into two subcategories, tropical deciduous forest, and tropical evergreen forest.

The main flora consists of Sal (Shorea robusta), Asna (Terminalia tomentosa), Shisam (Dalbergia sisso), Jamun (Syzygium cumini), and Gular (Ficus glomerata). The grasslands are prominent features of the park. The wetlands constitute the third major habitat type.

3. Methodology

The processing of the datasets has been done in the Sentinel Application Platform (SNAP) version 5.0 software and interpretation, analysis is done in the R-studio. The pre-processing steps, generation of backscatter image, and coherence image are done in the SNAP software (Figure 2).

3.1 Backscatter image generation

The single look complex (SLC) data of Sentinel-1A was preprocessed to split the scene and to get a deburst product. The SLC data suffers from the slant range ambiguity due to the side looking nature of the SAR system. To minimize the slant range ambiguity the slant to ground range conversion is done. Slant to groundrange conversion includes the re-projection of Single Look Complex (SLC) data from slant- range onto a flat ellipsoid surface. The process redistributes the SLC data in range with equal pixel spacing. Due to two spatial resolutions within a single SAR resolution cell, the shape of the resolution cell becomes rectangular. To generate squaLopezre pixel shape the Multilooking operator of the SNAP v5.0 was used.

To obtain an amplitude image these two channels (real and imaginary) are combined to obtain the composite signal intensity and the pixels of the resultant image show amplitude values. The processing to obtain amplitude images from complex SAR data includes the square of the sum of the square of the real and imaginary values of the complex pixel values of Sentinel-1 data. The radiometric calibration was performed to generate a normalized backscatter cross-section image. The backscatter image represents an actual radar return from a resolution cell that is a coherent sum of backscatter contribution due to different objects within the SAR resolution cell. The backscatter image was orthorectified with the help of the Range Doppler Terrain Correction algorithm of the SNAP v 5.0 tool to minimize the ambiguity in the SAR data due to the topographic variations of the scene. The digital elevation model (DEM) with a resolution of 1 arc-second (~ 30 meters) of the Shuttle Radar Topography Mission (SRTM) was used in the Orthorectification procedure. The backscatter cross-section image was normalized with a local angle of incidence image to get actual radar returns from the resolution cell.

3.2 Classification of the SAR data using Random Forest Classifier using backscatter values

The classification has been done using backscatter values as the threshold in the random forest classifier and the KD KNN Classifier and later on the accuracy of both the classifiers is compared. The coherence is defined as the complex correlation between two SAR images. The coherence image is generated using two datasets i.e., master and slave image. InSAR coherence is a measure of the decorrelation of two co-registered SAR data acquired in interferometric mode. The interferometric SAR datasets are co-registered, stacked and multi looked. Coregistration was performed at the sub-pixel level that is an essential requirement of interferometric processing of the SAR data. To minimize the speckles from the coregistered product a total of four filters were tried and the refined Lee filter was used for removing speckles and noise from the data.

The coherence image shows the correlation between the pixels of master and slave images. The pixel values of a coherence image lie between 0 and 1. The pixel value 1 of a coherence image shows the total correlation and if the value lies near to zero then it shows total decorrelation. The objects that are stable and not going to change in the interferometric data acquisition show maximum coherence and unstable structure shows low coherence. The stable scatterers like rocks and urban structures behave like a stable structure and these objects appear brighter in the coherence image. Forest vegetation shows moderate to low coherence because the top canopy surface of the forest vegetation is not a completely stable structure.



Figure 1: Location map of the study area



Correlation values help in the identification of the density of the forest. Due to wind velocity variation and natural structure, forests shows high decorrelation. The settlements or urban areas also show a high correlation.

3.4 Satellite data used

The datasets used to carry out this research are shown in table-1.

Table 1: Datasets used in the study

Parame te rs	Data set-1	Data set-2	Data set-3
Sensor	SENTINEL-1	SENTINEL-1	SENTINEL-2
Date	29-05-2018	10-06-2018	22-04-2018
Orbit	22112	22287	
Track	165	165	
Product	S1A_IW_SLC_1SDV	S1A_IW_SLC_1SDV	
Polarization	VV+VH	VV+VH	
Swath	IW2	IW2	
Node	Descending	Descending	
Wavelength	5.6 cm	5.6 cm	

4. Results and discussions

4.1 Supervised classification of Dudhwa National Park using Sentinel 2 data

The vegetation type map (Figure 3) is prepared using supervised classification and maximum likelihood classifier using the Sentinel-2 data set. This map is used as a reference for assessing the accuracy of classification achieved using Sentinel 1A datasets. The training datasets are used and the accuracy obtained is 82.3 %. The classes used for the classification are Eucalyptus, Sal, Sal mixed, fallow land, settlements, grassland, scrubland, agriculture, waterbody, dry riverbed, open land. The maximum likelihood classifier works on the principle of the probability density function and it classifies the pixels into different classes based on their probability of finding that pixel into that particular class.

4.2 Backscatter values analysis of the VH and VV intensity for species-level mapping

Backscatter images of VH and VV (Figure 4 and Figure 5) are generated using the workflow shown in figure 2 and further analysis of the backscatter values is done for the classification and identification of the species classified in the optical data. The backscatter values are required to compare the classification accuracy because this gives the geometry and the structure of the tree. Also, the darker areas show the forest area comprising of different species with different values as well as the fallow land with the presence of moisture, whereas the brighter portion of the backscatter image shows settlement and open dry land. In this study, the information retrieved from the canopy of the tree is used for the classification of trees because C-band gives better canopy structure information than optical sensors.

4.2.1 Backscatter values of VH band

The cross-polarization is more sensitive to the forest zone because of the volume scattering phenomenon in a

forested area. The VH band is vertical transmit and horizontally received. Due to multiple scattering in the forest system, the cross polarised bands are more robust as compared to the co-polarised. The scattering from the forest occurs in the three components, that is, surface, double bounce, and direct scattering. The surface scattering occurs due to sparse vegetation and scattering from the ground, whereas the double bounce results due to the interaction of band from ground and trunk, and the direct scattering occurs from the crown of the tree. Since the C-band wavelength is comparatively shorter so, it cannot penetrate deep inside the forest so most of the scattering is obtained from the crown or little deeper till branches.

4.2.2 Backscatter values of VV band

The backscatter image generated using VV band is useful to discriminate built-up and dry riverbed and open land from other classes as these areas look much brighter than the other features. This is vertical transmit and vertical receive since the built-up features show double-bounce scattering.

The VV band is the vertical transmit and vertically received co-polarised wave which is best for the discrimination between the settlement and the open land and also can discriminate the vegetation within the builtup or the fallow land.

4.3 Backscatter maps of VV and VH bands to characterize forest area of Dudhwa National park using Random Forest Classifier and KNN Classifier

The Random Forest Classification of the area is done and the classification accuracy is satisfactory, it can be interpreted that the maximum error is 0.0085 in the classification of vegetation classes, especially in the plantation group (Figure 6), this is because of the mixing of the pixels of two interrelated classes. The KD KNN Classification has been done on the basis of different classes as in Random forest classification (Figure 7). The classes mainly taken are built-up, Eucalyptus, Sal, Salmixed, River bed, Open grassland, etc. But due to the mixing of pixel information of different vegetation types the class is further merged as vegetation class. The accuracy achieved with classifier is 82.5%, this accuracy came for three different classes as Built-up, Vegetation, and Open grassland. The accuracy of these classes is same for random forest classifier. The backscatter values of the segregating classes is mentioned in figure 4 and figure 5 for VH and VV bands respectively. The accuracy, correlation, and error rate are depicted in table 2 and statistical values of the KNN classifiers are given in table 3. This is for the Random forest classifier. It is clearly shown that the error rate is maximum for the Eucalyptus and Teak plantation. From both the classifier, few of the pixels get mixed for different classes, the backscatter values are nearly equal for these classes and hence, the accuracy is varied between the classifier and also for the mentioned classes. The comparative analysis of the two classifiers can be done. The Eucalyptus plantation, Teak plantation, Settlement, and Agriculture

show pronounce accuracy in RF whereas it is poor in KNN. The classes such as scrubland, open land, waterbody are well segregated in the KNN classifier as a comparison to RF. The class-wise segregation performance of different classifiers is different and overall classification accuracy is different.

Coherence map of VV and VH bands between 4.4 29/5/2018 (Master image) and 10/6/2018 (Slave Image) The coherence map shows different values for the different classes and the value ranges between 0 to 1, the higher coherence value shows the built-up areas whereas the lower values shown by the vegetation classes. This is the reason; the built-up area looks brighter in the image whereas vegetation parts look dull. Since C-band cannot penetrate deep through the canopy till the ground, so all we can get is the canopy information, the geometrical properties of the leaves, and the scattering behavior determines the backscatter values among the various vegetation types. The coherence image of both VV and VH bands gives information of correlation and decorrelation between the various features on the ground in the two different imagery which help in the identification of features. The coherence of VV and VH bands are generated and the values obtained are further analyzed. The VV band is able to separate the urban and grassland area from the forest (Figure 8 and Figure 10). The VH band is not effectively showing the clear cut demarcation between these classes (Figure 9 and Figure 11).

Table 2: The statistical values of the Random Forest classification of Sentinel-1 product of Dudhwa National Park

Class Type	Accuracy	Precision	Correlation	Error
Eucalyptus	0.559	0.4	0.483	0.44
Teak	0.58	0.147	0.473	0.41
Settlement	0.852	0.32	0.319	0.14
Agriculture	0.943	0.151	0.12	0.05
Scrubland	0.9161	0.076	0.085	0.083
Open land	0.9814	0.728	0.714	0.018
Water	0.998	0.842	0.841	0.0017

 Table 3: The statistical values of the KNN Classifier of Sentinel-1 product of Dudhwa National Park

Class	Accuracy	Precision	Correlation	Error
Eucalyptus	0.44	0.32	0.321	0.56
Teak	0.35	0.132	0.355	0.52
Settlement	0.753	0.22	0.303	0.2
Agriculture	0.921	0.12	0.09	0.15
Scrubland	0.901	0.062	0.65	0.093
Open land	0.885	0.625	0.55	0.025
Water	0.865	0.785	0.75	0.0035



Figure 3: Supervised Classification Map of Dudhwa National park of different Forest types



Figure 4: Backscatter map of VH band



Figure 5: Backscatter map of VV band



Figure 6: Classification map with the Random Forest Classifier using backscatter value



Figure 7: Classification map with the KNN Classifier using backscatter value



Figure 8: Coherence Map of VV band of Dudhwa National park



Figure 9: Coherence Map of VH band of Dudhwa National park



Figure 10: Backscatter values range for different classes of VV band



Figure 11: Backscatter values range for each class of VH band

5. Conclusions and recommendations

The classifications were done using two classification algorithms i.e., RF and KNN. The backscatter and coherence image information is used for the classification. Coherence proved to be efficient in the segregation of vegetative and non-vegetative classes. The built-up and vegetation areas show higher values of backscatter in the VH band. The accuracy assessment is done and based on that, it has shown that RF gave pronounced results as compared to KNN. It can be concluded that the mixed species in the forest area are not able to segregate due to mixing in the information of backscatter and coherence, whereas, the dominant species can be easily separated from the other classes. Multitemporal observations of different seasons and advanced techniques like deep learning may further improve in the classification of forest Species.

References

Blanzieri, E. and F. Melgani (2008). Nearest neighbor classification of remote sensing images with the maximal margin principle. IEEE Transactions on Geoscience and Remote Sensing, 46(6), 1804–1811.

Chandola, S. (2014). Polarimetric SAR Interferometry for Forest Aboveground Biomass Estimation, M.Sc. (Geoinformation Science and Earth Observation) Thesis, ITC Netherlands and IIRS, Dehradun, India, 62p.

De Grandi, E.C., E. Mitchard, I.H. Woodhouse and G.D. De Grandi (2015). Spatial wavelet statistics of SAR

backscatter for characterizing degraded forest: A case study from Cameroon. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(7), 3572–3584,

https://doi.org/10.1109/JSTARS.2015.2420596

De Souza Mendes, F., D. Baron, G. Gerold, V. Liesenberg and S. Erasmi (2019). Optical and SAR remote sensing synergism for mapping vegetation types in the endangered Cerrado / Amazon ecotone of Nova. Remote Sensing, 10(11), 1161.

Du, P., A. Samat, B. Waske, S. Liu and Z. Li (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features, ISPRS Journal of Photogrammetry and Remote Sensing, 105, 38–53. https://doi.org/10.1016/j.isprsjprs.2015.03.002

Ferrant, S., A. Selles, M. Le Page, P.A. Herrault, C. Pelletier, A. Al-Bitar, S. Mermoz, S. Gascoin, A. Bouvet, M. Saqalli, B. Dewandel, Y. Caballero, S. Ahmed, J.C. Maréchal and Y. Kerr (2017). Detection of irrigated crops from Sentinel-1 and Sentinel-2 data to estimate seasonal groundwater use in South India. Remote Sens. 9. https://doi.org/10.3390/rs9111119

Hirschmugl, M., C. Sobe, J. Deutscher and M. Schardt (2018). Combined use of optical and synthetic aperture radar Data for REDD + Applications in Malawi. Land, 4, 1–17. https://doi.org/10.3390/land7040116

Kumar, S., 2009. Retrieval of forest parameters from Envisat ASAR data for biomass inventory in Dudhwa National Park,U.P.,India. M.Sc.(Geoinformation Science and Earth Observation) Thesis, ITC Netherlands and IIRS, Dehradun, India, 80p.

Kumar, S., R.D. Garg, H. Govil, S.P.S. Kushwaha, 2019. PolSAR-decomposition-based extended water cloud modeling for forest aboveground biomass estimation. Remote Sens. 11, 1–27. https://doi.org/10.3390/rs11192287

Kumar, S., U.G. Khati, S. Chandola, S. Agrawal, S.P.S. Kushwaha, 2017. Polarimetric SAR Interferometry based modeling for tree height and aboveground biomass retrieval in a tropical deciduous forest. Adv. Sp. Res. 60, 571–586. https://doi.org/10.1016/j.asr.2017.04.018

López-Martínez, C., X. Fabregas and A. Alonso-González (2012). Analysis and validity of the PolInSAR line model on forested areas. In EUSAR 2012; 9th

Main, R., R. Mathieu, W. Kleynhans, K. Wessels, L. Naidoo and G. Asner (2016). Hyper-temporal C-band SAR for baseline woody structural assessments in deciduous savannas. Remote Sensing, 8(8), 661.

Paola, J.D. and R.A. Schowengerdt (1995). A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Transactions on Geoscience and Remote Sensing, 33(4), 981–996. https://doi.org/10.1109/36.406684

Puletti, N., F. Chianucci and C. Castaldi (2017). Use of Sentinel-2 for forest classification in Mediterranean environments. Annals of Silvicultural Research, $\theta(0)$, 1–7. https://doi.org/10.12899/asr-1463

Rahman, M.M. and J.T.S. Sumantyo (2010). Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. Applied Geomatics, 3(2), 113–121. https://doi.org/10.1007/s12518-010-0026-9

Saatchi, S.S. and Rignot, E. (1997). Classification of Boreal forest cover types using SAR images. Remote Sensing of Environment, 3(60), 270-281.

Shimada, M., M. Ohki and H. Noguchi (2010). Incidence angle dependence of PALSAR repeat pass interferometry. In 8th European Conference on Synthetic Aperture Radar (pp. 1–4).

Tomar, K.S., S. Kumar and V.A. Tolpekin (2019). Evaluation of hybrid polarimetric decomposition techniques for forest biomass estimation, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3712–3718. https://doi.org/10.1109/JSTARS.2019.2947088

Touzi, R., A. Lopes, J. Bruniquel and P.W. Vachon (1999). Coherence estimation for SAR imagery, *I*EEE Transactions on Geoscience and Remote Sensing, 37(1 PART 1), 135–149. https://doi.org/10.1109/36.739146

Varghese, A.O., A. Suryavanshi, A. and A.K. Joshi (2016). Analysis of different polarimetric target decomposition methods in forest density classification using C band SAR data. International Journal of Remote Sensing, 37(3), 694–709.

https://doi.org/10.1080/01431161.2015.1136448