Deciphering tropical tree communities using earth observation data and machine learning

Rahul Bodh[1], Hitendra Padalia[1,*], Divesh Pangtey[1], Ishwari Datt Rai[1], Subrata Nandy[1] and C. Sudhakar Reddy[2]

[1]Indian Institute of Remote Sensing (ISRO), Dehradun 248 001, India
[2]National Remote Sensing Centre (ISRO), Hyderabad 500 037, India

Publicly available EO datasets offer new possibilities to generate biodiversity information at the community composition level, an essential biodiversity variable, beyond forest type. We demonstrated the potential of Sentinel-2, GEDI LiDAR canopy height and ALOS-DEM in discriminating and classifying tropical tree communities in the Western Himalayas, India. For this, tree communities were first identified based on the ordination of field data and subsequently classified using satellite data applying machine learning, i.e. random forest (RF). From the three forest types in the study area, eight distinct tree communities were identified for which classification accuracy increased from single date (75.17%) to multi-date images (85.33%) and further by applying feature selection (88.17%). Whereas the best classification accuracy of 94.66% was achieved when canopy height and topographic variables were also considered. The findings suggest that RF is suitable for mapping tree communities by combining Sentinel-2 with GEDI and DEM parameters.

Keywords: Biodiversity, canopy height, machine learning, remote sensing, tropical forest.

TROPICAL forest ecosystems that are home to at least two-thirds of the Earth's terrestrial biodiversity are facing high rates of loss due to a variety of anthropogenic disturbances[1,2]. Tree community composition is the most robust and sensitive indicator of degradation in natural forests[3], making it invaluable for conservation planning and assessing changes[4]. Satellite remote sensing (RS) has become an integral tool for mapping and monitoring biodiversity at different scales. There is an emphasis on developing Earth Observation (EO) methods to address essential biodiversity variables (EBVs) in order to monitor biodiversity changes from regional to global scales[5]. Community composition has been identified as one of the EBVs for which satellite RS-based approaches must be used for better mapping and monitoring[6]. The high diversity, topographic variability and lack of cloud-free data are particular challenges in the community-level mapping of tropical forests[7].

Recent studies highlighted the potential utility of Sentinel-2 (S2) data in classifying vegetation composition[8] and tree species[9]. Better spatial, spectral (13 bands, including three red-edge bands) and temporal (five-days) characteristics, consistent availability and open access of S2 make it one of the best options for routine monitoring of tropical forests. In this context, there is a need to explore the potential of S2 for classifying and mapping tree communities in tropical ecosystems.

Optical RS imagery explicitly describes vegetation horizontal structure, while they also differ in vertical structure due to resource competitiveness. With regard to tree communities in mountainous regions, a spatial correlation exists between diversity patterns and topographic heterogeneity[10]. Several studies have demonstrated the benefits of integrating airborne LiDAR[11,12] and topographic parameters[13,14] for improving vegetation classification in different ecosystems. Improvements in the accuracy and resolution of freely available space-borne LiDAR data (Global Ecosystem Dynamics Investigation (GEDI)) and DEM (Advanced Land Observing Satellite (ALOS-2)) open up possibilities for improving vegetation classification from regional to global scales.

Machine learning (ML) methods have been explored for their ability to enhance vegetation classification by incorporating multi-temporal datasets[8], vegetation indices[15], vegetation structure[12] and topographic variables[13]. Several studies have observed that ML performs better than conventional statistical classifiers, especially for complex and high-dimensional RS datasets[16]. Random Forest (RF) is the most widely used non-parametric ML method due to its high reliability and accuracy in classification applications[8,10].

The objectives of the present study are to develop a model to (i) identify the possible range of tree communities occurring in the study area using the ordination technique; (ii) assess and compare the classification accuracy of RF using different combinations of spectral variables from multi-temporal S2 data, and (iii) analyse the performance of RF classification when canopy height model (CHM) and topographic variables are integrated with multi-temporal S2 data.

*For correspondence. (e-mail: hitendra@iirs.gov.in)

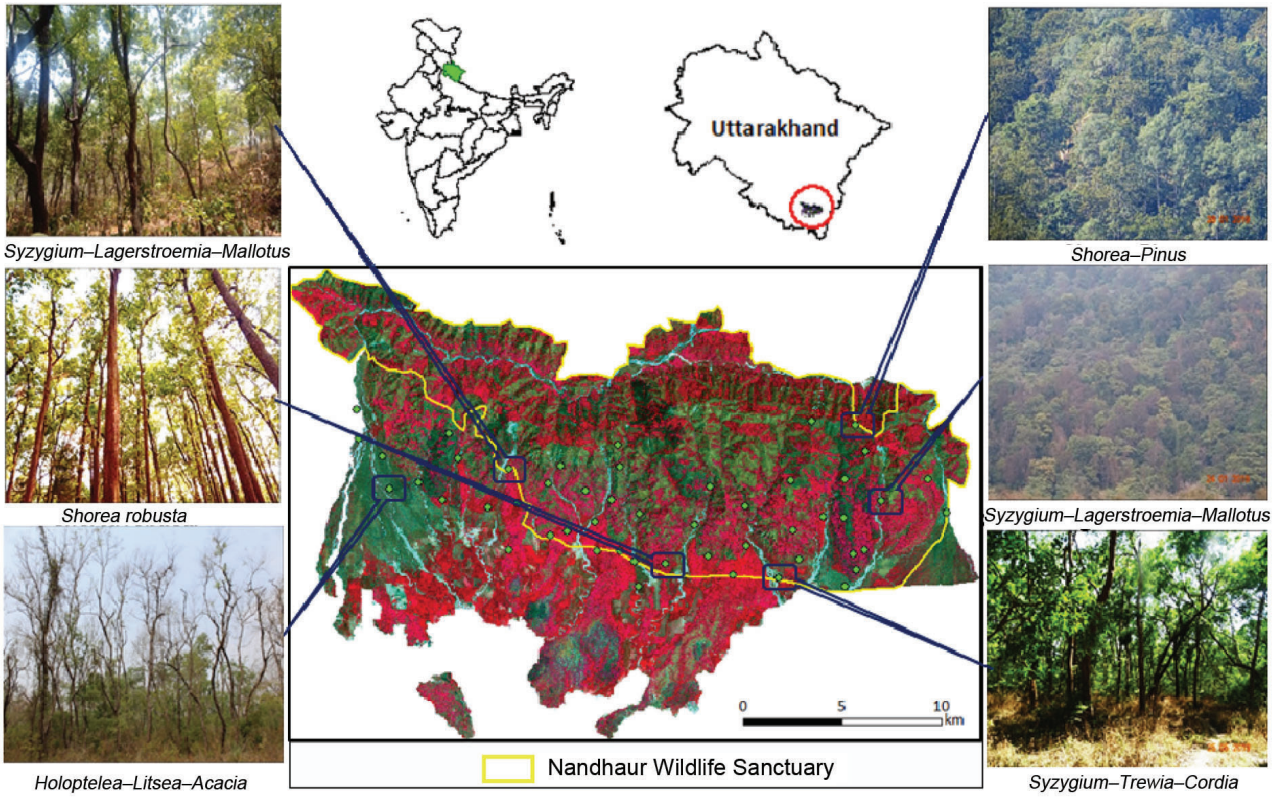

Figure 1. Location map of the study area in India, and distribution of sample plots and tree communities.

Materials and methods

Study area

The study was conducted in the Nandhaur landscape, which includes the Nandhaur Wildlife Sanctuary and adjoining areas, falling in the Nainital and Champawat districts of Uttarakhand in the foothills of Western Himalaya, India (Figure 1). The physiography is mountainous with elevation ranging from 150 to 1400 m amsl. The sites fall under Cwa (temperate: dry winter and hot summer) of the Köppen–Geiger climate classification[17]. The mean annual temperature is 23°C, while the mean annual precipitation is 940 mm. According to Champion and Seth[18], tropical moist and dry deciduous forests dominate the study area. It also constitutes important habitats and a corridor for tigers, elephants, and other large mammals. Nandhaur Wildlife Sanctuary (270 sq. km) is a critical conservation unit and part of the Shivalik Elephant Reserve.

Sampling design and field inventory

The sampling design included the distribution of 0.1 ha (31.62 m × 31.62 m) sample plots across different combinations of elevation, aspect and moisture categories. A pilot study was conducted to determine the total number of sampling units to be laid based on variance in the species richness using the formula of Chacko[19]

$$n = \frac{t^2 \times \mathrm{CV}^2}{(\mathrm{SE}\%)^2}, \quad (1)$$

where n is the number of sample plots, CV the coefficient of variation of tree species richness, SE% the standard error percentage (10) and t is the statistical value at 95% significance level. Plot inventory followed the internationally standardized protocols and methodologies established by the Amazon Forest Inventory Network, RAINFOR[20].

Classification of tree communities using ordination technique

Two-way indicator species analysis (TWINSPAN)[21] was used to identify the patterns of compositional variation of tree species in the region. It is based on the partitioning of the ordination axis resulting from reciprocal averaging. This is followed by a discriminant function for assigning the sites to either side of the dichotomy based on indicator species. This process is repeated for each cluster obtained from this portioning until a predefined stopping criterion is met, i.e. minimum group size for division and maximum level of divisions. TWINSPAN classification was carried

Table 1. Spectral, topographic and structural variables extracted from Sentinel-2, ALOS-PALSAR DEM and GEDI LiDAR, used for developing tree community classification models

Feature/dataset	Independent variables	Reference
Sentinel-2 band reflectance	B2 (blue) 490 nm	Sentinel-2 User
	B3 (green) 560 nm	Handbook, 2015
	B4 (red) 665 nm	
	B5 (vegetation red edge 1) 705 nm	
	B6 (vegetation red edge 2) 749 nm	
	B7 (vegetation red edge 3) 783 nm	
	B8 (near infrared) 842 nm	
	B8a (near infrared) 865 nm	
	B11 (short wave infra-red) 1610 nm	
	B12 (short wave Infrared) 2190 nm	
Spectral indices	Normalized difference vegetation (NDVI)	30
	Enhanced vegetation index (EVI1 and EVI 2)	31
	Red edge normalized difference vegetation (NDRE)	32
	Green chlorophyll index (CI green)	33
	Red-edge Chlorophyll Index (CI red-edge)	33
	Difference vegetation index (DVI)	34
Topographic variables (ALOS-PALSAR DEM)	Elevation	ALOS Data Users
	Slope	Handbook, 2008
	Aspect	
Structural variables	Canopy height model developed by integrating GEDI LiDAR, multi-date Landsat 8 OLI and SRTM DEM using Random Forest (RF)	

out based on plot-level abundance data of tree species using PC-ORD software (version 4.20).

Acquisition of satellite datasets and processing

S2 level-1C cloud-free images for winter (January 2018), spring (March 2018), and summer (May 2017) seasons were downloaded from the Sentinel's Scientific Data Hub (https://scihub.copernicus.eu/) of the European Space Agency[22]. The level-1C images were subsequently converted to level-2A top-of-canopy (TOC) surface reflectance using the atmospheric correction module in Sen2Cor (version 2.4) processor. The digital image processing and calculation of vegetation indices (VIs) were carried out using ArcGIS (10.3).

ALOS phased array type L-band synthetic aperture radar (PALSAR) RTC (radiometrically terrain corrected) DEM was downloaded from Alaska Satellite Facility (https://asf.alaska.edu/) at 12.5 m resolution and topographic variables, viz. elevation, slope and aspect were derived. Table 1 provides information about the satellite datasets used in this study.

Mapping tree canopy height with GEDI and Landsat 8

To incorporate the canopy height characteristics of tree communities in the classification, we developed a CHM by integrating GEDI LiDAR observations, multi-date Landsat 8 OLI images/vegetation indices and Shuttle Radar Topography Mission (SRTM) DEM at 30 m using RF regression method. These variables were used for modelling due to their spatial and temporal relevance with field inventory data. The highest quality GEDI L2A (RH95) samples available from March 2019 till July 2021 from the power beam with beam sensitivity ≥0.9, quality flag = 1, collected during nighttime, were used for the model. Of the eight bands available in Landsat 8 OLI, the spectral information from bands 2 to 7 corresponding to blue (0.45–0.51 μm), green (0.53–0.59 μm), red (0.64–0.67 μm), near-infrared (NIR; 0.85–0.88 μm) and two shortwave infrared bands (SWIR1, 1.57–1.65 μm and SWIR2, 2.11–2.29 μm) at 30 m spatial resolution were used for analysis. Spectral bands and NDVI of the cloud-free Landsat 8 images for 5 April 2019, 15 November 2019, 22 March 2020, 17 November 2020, and 21 February 2021 were used as independent variables to extrapolate the estimates of GEDI LiDAR.

Selection of optimal independent spectral variables for tree community type mapping

Removing the irrelevant or redundant variables from the large datasets results in higher classification accuracy. Recursive feature elimination (RFE), integrated with RF was implemented for the selection of optimum predictors constituting the S2 bands and vegetation indices based on the importance ranking of the RF model, i.e. mean decrease accuracy (MDA) and mean decrease gini (MDG). MDA was computed by permuting the prediction error on the out-of-bag (OOB) portion of the data. MDG is the total decrease in node impurities measured by the Gini index from splitting on the variable, averaged over all trees. The RFE method improves the prediction performance of the

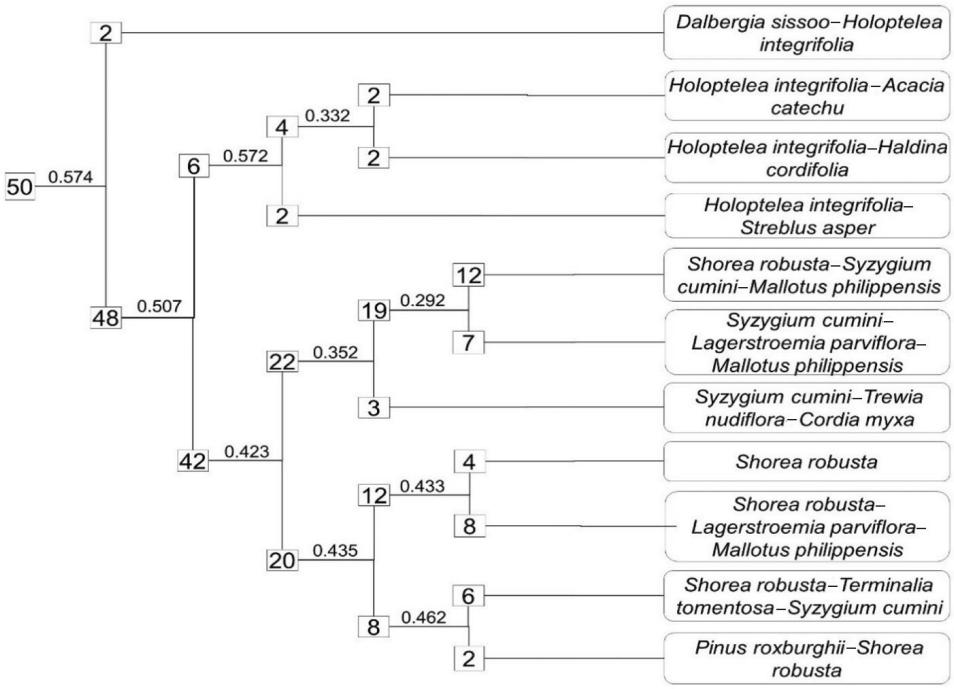

Figure 2. TWINSPAN dendrogram with interpreted clusters up to the sixth level of division. Numerals in the box represent the number of sample plots and the decimal numbers represent data heterogeneity in each level of division.

predictor variables by progressively eliminating the least promising variables. This method is performed iteratively on the reduced dataset using a cross-validation function until 3% of the number of variables remains. The set of predictors showing the least RMSE after cross-validation was selected as the optimum for classification. The 'randomForest' package in RStudio was used to run the RF model.

Tree community type mapping using ML classification

Tree communities obtained from the ordination analysis were classified using the RF classifier. RF is an ensemble of decision trees built on a bootstrap sample of the training data and chooses the best split at each node using a randomly selected subset of predictor variables[23]. The RF was considered for classifying the multi-source and non-linear RS data in the present study as: (i) it is flexible and can handle many variables and a large amount of missing data; (ii) the curse of dimensionality in the data is naturally reduced by selecting the most optimum variable subset with minimum OOB error and (iii) the bootstrapping procedure improves the efficiency of model performance by decreasing the variance of the model without increasing the bias by selecting many classification trees.

The classification was performed on different combinations of S2 images of different seasons, followed by optimum predictors of S2 bands and spectral indices obtained from RFE. Finally, the optimum predictors were integrated with topographical variables and CHM independently. Classification accuracy for all the datasets was evaluated using confusion matrix. The kappa statistics was also used to compare the true agreement between classes that occur on the ground versus those classified by the classifier rather than those that occur by chance.

Results and discussion

Tree communities obtained from ordination analysis

The TWINSPAN classification at the sixth hierarchical level of division based on tree species composition and abundance recognized 11 tree communities. The final number of divisions was chosen based on expert-based validation[24], as the number of clusters cannot be set manually. Among the identified tree communities, four were dominated by *Shorea robusta* (Sal), two by *Syzygium cumini* (Jamun), one by *Pinus roxburghii* (Chir pine), and the rest belonged to the riverine community. The *Dalbergia sissoo* and *Acacia catechu* communities were not prominent and formed a sparse open forest mixed with grasses along the river beds. Further, the classification of riverine communities separately did not provide a meaningful result. So, we combined the riverine communities into *Holoptelea integrifolia–Litsea* sp.*–Acacia catechu*, making only eight woody communities for further analysis. Figure 2 shows a dendrogram of the result obtained from TWINSPAN.

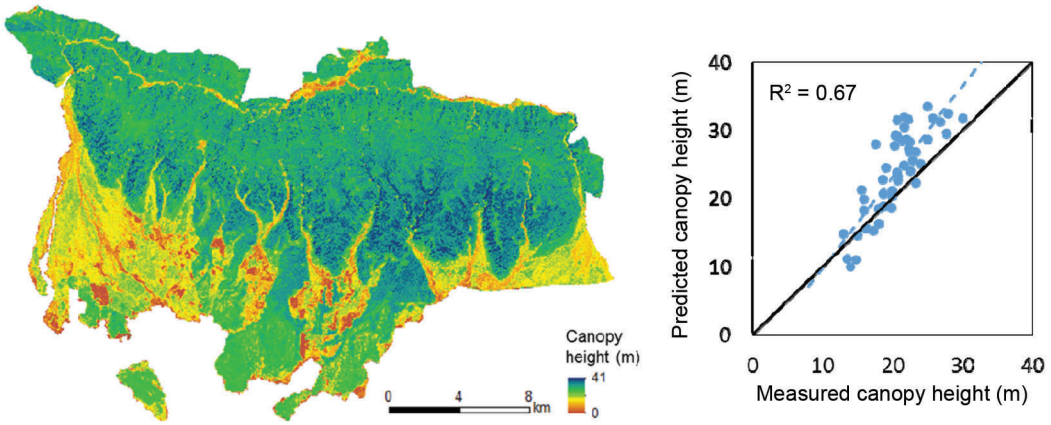

Figure 3. Canopy height map derived using GEDI LiDAR based on Random Forest (RF) regression model.

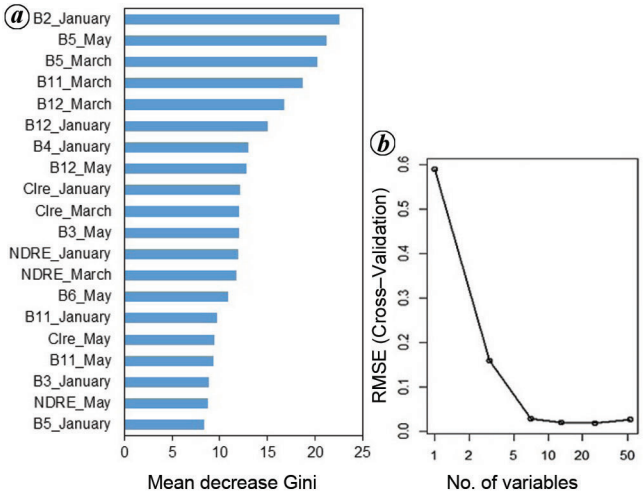

Figure 4. RF for the optimization of independent variables: (***a***) variable importance index and (***b***) selection of optimum number of variables based on the least RMSE for ten-fold cross-validation.

Canopy height modelling

After filtering out the less reliable GEDI samples, a total of 9405 sample footprints were used for the generation of CHM. Independent validation of predicted canopy height was done using field-measured maximum stand height of 50 plots of 0.1 ha collected during 2019–20 (Figure 3). A promising relationship was observed between the measured and predicted canopy height ($r^2 = 0.67$) with an RMSE of 5.12 m. The predicted canopy height tends to be overestimated as we move from low canopy to higher canopy. A mean difference of 3.8 m was observed between the predicted and measured canopy height.

Optimal satellite variables for tree community classification

In this study, 890 sample points (70% training and 30% testing) were used for the RF model. For optimization of independent spectral variables (51 variables), the RF algorithm was tested for different Mtry and Ntree values iteratively. Since it is a classification, the node size was kept at 1. The Ntree value of 480 and Mtry of 7 were found to contain the lowest OOB prediction error. MDA and MDG were computed as measures of variable importance ranking. Using RFE with ten-fold cross-validation, the lowest RMSE was obtained for a set of 20 variables from MDG (Figure 4).

The mean spectral response of the tree communities was analysed for the S2 images of three seasons (Figure 5). The SWIR region showed the highest variability in the spectral response among classes, followed by the red edge and NIR region, while the blue, green and red regions showed minimum separability. With regard to the season, the data obtained for May showed the highest spectral variability among classes, followed by March.

Performance of ML classifier

Table 2 shows an increase in the performance of RF on using single to multi-date imagery (overall accuracy, 75.17–85.33%). The high spectral variation in the summer and spring seasons and the contrast of winter and spring seasons helped in discriminating communities and increasing the classification accuracy. Similar observations were reported by Macintyre *et al.*[8] in mapping floristic communities of Western Australia, where a multi-temporal feature set comprising autumn and spring images returned the highest accuracy.

Among the two variable importance measures obtained from RF, optimum predictors from MDG obtained higher classification accuracy (88.17%). High dominance of red-edge bands/indices and SWIR was observed in the important variable measures. This likely reflects the importance of differences in moisture and chlorophyll content of the species in different phenophases and soil conditions. Similar observations have been reported in vegetation mapping

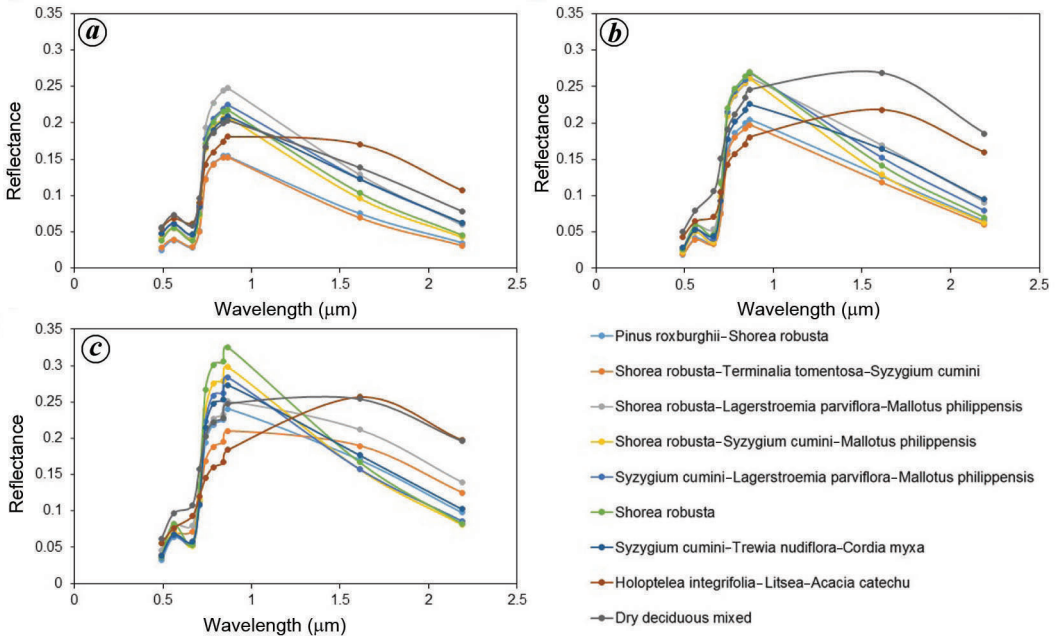

Figure 5. Mean spectral signature of all target classes for S2 images of (***a***) January, (***b***) March and (***c***) May. Mean spectral reflectance is shown on the *y*-axis and wavelength of the band on the *x*-axis.

Table 2. Classification accuracy of RF for all the datasets

Dataset description	No. of variables	RF			
		UA	PA	OA	KC
January	10	70.51	71.86	68.50	0.59
March	10	73.28	74.38	71.33	0.62
May	10	76.88	79.70	75.17	0.65
January + May	20	79.77	80.09	78.33	0.68
March + May	20	78.70	75.19	77.37	0.64
January + March	20	81.10	82.43	80.00	0.70
January + March + May	30	85.92	87.39	85.33	0.75
Mean decrease accuracy (MDA)	20	87.43	87.54	86.67	0.76
Mean decrease Gini (MDG)	20	88.40	89.02	88.17	0.77
MDG + DEM	23	91.92	91.69	91.17	0.80
MDG + CHM	21	93.32	93.74	93.17	0.82
MDG + DEM + CHM	24	95.20	94.8	94.66	0.84

PA, Producers' accuracy; UA, Users' accuracy; OA, Overall accuracy; KC, Kappa coefficient.

over a range of ecosystems[8,25,26]. The red-edge bands/indices are sensitive and show relation to the leaf structure and chlorophyll content of plants[27,28] and are found suitable for mapping heterogeneous landscapes, especially by including red-edge-associated vegetation indices.

S2 was not sufficient to capture all the variability present in the tree communities with similar species composition. To segregate such classes, other parameters of vegetation, such as structural and topographical variables, must be used to explain variability. The addition of CHM resulted in further explanation of the variability and increased the classification accuracy (5%). The highest improvement in accuracy was observed for *Syzygium cumini–Trewia nudiflora–Cordia myxa* (8%) and *Syzygium cumini–Lagerstroemia parviflora–Mallotus philippensis* (7%) community types.

The addition of topographic variables (elevation, slope and aspect) provided an improvement in overall accuracy (>2%). The highest increase in accuracy was observed for *P. roxburghii–S. robusta* (7%), *S. robusta–L. parviflora–M. philippensis* (5%) and *S. robusta–S. cumini–M. philippensis* (4%) communities which occupy the upper and lower regions of the study area respectively. Similar observations were also reported in other studies, especially in the mountainous ecosystem[13,29].

Besides the similarity in community composition, spectral mixing/confusion between classes is also possible due to the edge effect. However, the transition between tree

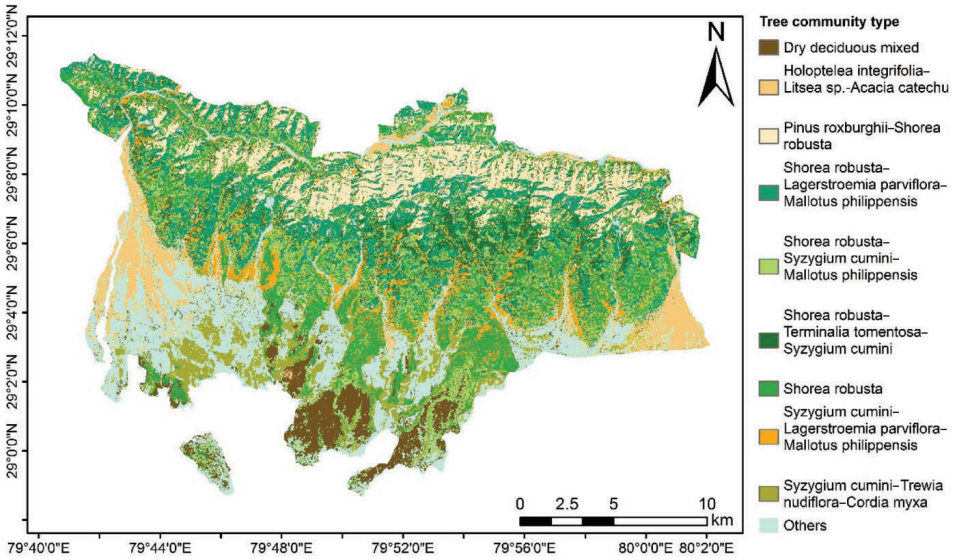

Figure 6. Tree community type map prepared using the best combination of predictor variables.

Table 3. Area statistics of target classes inside and outside the Nandhaur Wildlife Sanctuary, Uttarakhand, India

Vegetation type	Class	Protected area (sq. km)	Unprotected area (sq. km)	Study area (sq. km)
Moist deciduous forest	*S. robusta–Terminalia tomentosa–Syzygium cumini*	34.9	4.7	39.6
Shorea robusta-dominated	*S. robusta–Lagerstroemia parviflora–Mallotus philippensis*	62.6	10.0	72.5
	S. robusta–S. cumini–M. philippensis	36.1	12.5	48.6
	S. robusta	35.2	20.5	55.8
Syzygium cumini-dominated	*S. cumini–L. parviflora–M. philippensis*	19.0	7.2	26.2
	S. cumini–Trewia nudiflora–Cordia myxa	2.1	21.5	23.6
Sub-tropical forest	*Pinus roxburghii–S. robusta*	51.7	3.3	55.0
Dry deciduous forest	Dry deciduous mixed	0.5	24.4	24.8
	Holoptelea integrifolia–Litsea sp.*–Acacia catechu*	12.9	25.6	38.5
Others		11.3	55.2	66.5
	Total area	266.3	184.6	450.9

communities is continuous rather than discrete, so it would be difficult to separate them effectively using the present dataset. Our approach effectively classified the compositional variation in tree communities; however, slightly lower accuracy was obtained in the transition zones, especially for the *S. robusta* and *S. cumini*-dominated communities.

The best classification result (94.67%) was observed when CHM and topographic variables were integrated with the optimum predictor variables of S2. The map generated using the present approach can be utilized to identify the composition and distribution patterns of tree communities in the region.

Spatial patterns of tree communities of the Nandhaur landscape

The final classified map showed eight tree communities obtained from the field data and a dry deciduous mixed community (Figure 6). For the dry deciduous mixed community, sample plots could not be laid out; so training points were taken with the help of the characteristic spectral signature of vegetation on RS imagery and the additional GPS locations taken during the field inventory. The area statistics showed the dominance of moist deciduous tree communities inside the Nandhaur wildlife sanctuary. Majority of the area outside the Sanctuary is dominated by dry deciduous tree communities, where plantation activities are also carried out. Table 3 shows the area statistics of tree communities both inside and outside the protected area.

The tree communities showed diverse relationships with the input variables used in the final classification (Figure 7). There was significant variation in the distribution pattern of tree communities with the type and season of spectral bands/indices. Tree communities showed the highest variability in the distribution pattern with elevation. The canopy height variable had a distinct distribution pattern for *S. cumini*-associated communities, which were spectrally less separable from Sal-associated communities.

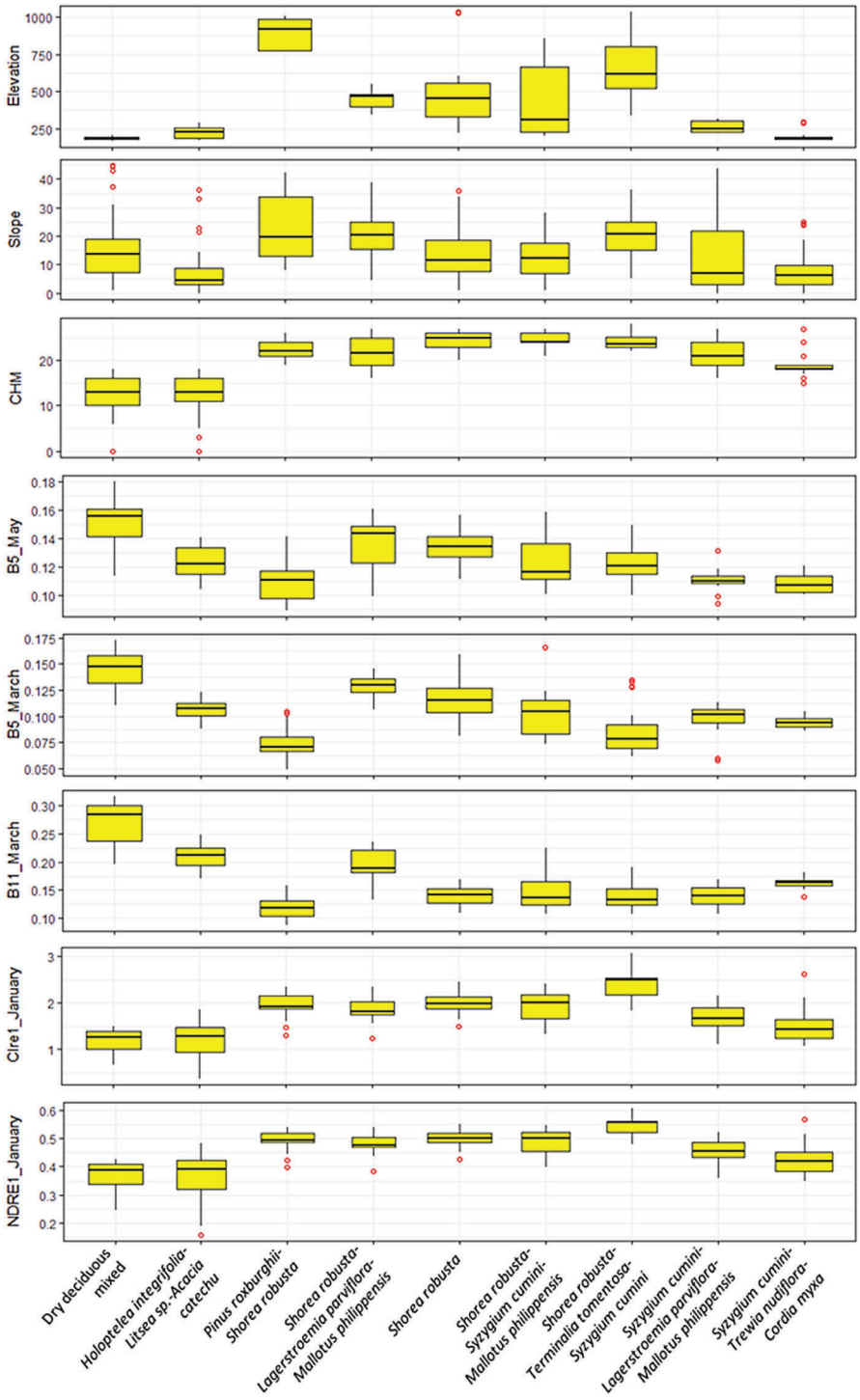

Figure 7. Spectral separability of tree community types for important variables.

Conclusion

In this study, tropical tree communities of the study area were first systematically identified based on field data, and thereafter classification was done using multi-temporal satellite datasets of the dry season. Reasonably good classification accuracy was obtained by choosing the most optimal predictors among multi-season spectral bands and ratios. We found that by incorporating CHM and topographic variables, tree community classification can be further enhanced. Tree communities associated with topographic features in the mountainous terrain and differences in canopy height owing to different site conditions play an important role in determining the spatial distribution of tropical tree communities with strong seasonality. The findings of this study suggest that RF is highly efficient in discriminating

tree communities using multi-sensor RS datasets. The present approach can be replicated in other ecosystems with publicly available satellite datasets utilized in this study.

Conflict of interest: The authors declare that there is no conflict of interest relationships that could have appeared to influence the work reported in this paper.

1. Giam, X., Global biodiversity loss from tropical deforestation. *Proc. Natl. Acad. Sci. USA*, 2017, **114**, 5775–5777.
2. Barlow, J. *et al.*, Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature*, 2016, **535**, 144–147.
3. Imai, N. *et al.*, Tree community composition as an indicator in biodiversity monitoring of REDD+. *For. Ecol. Manage.*, 2014, **313**, 169–179.
4. Ferrier, S., Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst. Biol.*, 2002, **51**, 331–363.
5. Pettorelli, N. *et al.*, Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing Ecol. Conserv.*, 2016, **2**, 122–131.
6. Vihervaara, P. *et al.*, How essential biodiversity variables and remote sensing can help national biodiversity monitoring. *Global Ecol. Conserv.*, 2017, **10**, 43–59.
7. Erinjery, J. J., Singh, M. and Kent, R., Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. *Remote Sensing Environ.*, 2018, **216**, 345–354.
8. Macintyre, P., van Niekerk, A. and Mucina, L., Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *Int. J. Appl. Earth Obs. Geoinf.*, 2020, **85**, 101980.
9. Breidenbach, J. *et al.*, National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Can. J. For. Res.*, 2021, **51**, 365–379.
10. Young, B., Yarie, J., Verbyla, D., Huettmann, F. and Herrick, K. C. F., Modeling and mapping forest diversity in the boreal forest of interior Alaska. *Landsc. Ecol.*, 2017, **32**, 397–413.
11. Nordkvist, K., Granholm, A. H., Holmgren, J., Olsson, H. and Nilsson, M., Combining optical satellite data and airborne laser scanner data for vegetation classification. *Remote Sensing Lett.*, 2012, **3**, 393–401.
12. Rees, H., Nyström, M., Nordkvist, K. and Olsson, H., Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation. *Int. J. Appl. Earth Obs. Geoinf.*, 2014, **27**, 81–90.
13. Abdollahnejad, A., Panagiotidis, D., Joybari, S. S. and Surovỳ, P., Prediction of dominant forest tree species using QuickBird and environmental data. *Forests*, 2017, **8**, 42.
14. Hościło, A. and Lewandowska, A., Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sensing*, 2019, **11**, 929.
15. Dorigo, W., Lucieer, A., Podobnikar, T. and Čarni, A., Mapping invasive *Fallopia japonica* by combined spectral, spatial, and temporal analysis of digital orthophotos. *Int. J. Appl. Earth Obs. Geoinf.*, 2012, **19**, 185–195.
16. Khatami, R., Mountrakis, G. and Stehman, S. V., A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. *Remote Sensing Environ.*, 2016, **177**, 89–100.
17. Peel, M. C., Finlayson, B. L. and Mcmahon, T. A., Hydrology and earth system sciences updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 2007, **11**, 1633–1644.
18. Champion, H. G. and Seth, S. K., *A Revised Survey of the Forest Types of India*, Manager of Publication, 1968.
19. Chacko, V. J., *A Manual on Sampling Techniques for Forest Surveys*, Manager of Publications, Government of India, New Delhi, 1965.
20. http://www.rainfor.org/ (accessed on 2 April 2019).
21. Hill, M. O., A FORTRAN Program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Twinspan, 1979.
22. https://scihub.copernicus.eu/ (accessed on 16 December 2019).
23. Breiman, L., Statistical modeling: the two cultures. *Stat. Sci.*, 2001, **16**, 199–215.
24. Rapinel, S., Rossignol, N., Hubert-Moy, L., Bouzillé, J. B. and Bonis, A., Mapping grassland plant communities using a fuzzy approach to address floristic and spectral uncertainty. *Appl. Veg. Sci.*, 2018, **21**, 678–693.
25. Ramoelo, A., Cho, M., Mathieu, R. and Skidmore, A., Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl. Remote Sensing*, 2015, **9**, 094096.
26. Immitzer, M., Vuolo, F. and Atzberger, C., First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 2016, **8**, 166.
27. Clerici, N., Valbuena Calderón, C. A. and Posada, J. M., Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *J. Maps*, 2017, **13**, 718–726.
28. Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H. and Sampson, P. H., Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Trans. Geosci. Remote Sensing*, 2001, **39**, 1491–1507.
29. Dorren, L. K. A., Maier, B. and Seijmonsbergen, A. C., Improved landsat-based forest mapping in steep mountainous terrain using object-based classification. *For. Ecol. Manage.*, 2003, **183**, 31–46.
30. Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W., *Monitoring vegetation systems in the Great Plains with ERTS*, NASA Special Publication, USA, 1974, vol. 351(1), p. 309.
31. Jiang, Z., Huete, A. R., Didan, K. and Miura, T., Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing Environ*., 2008, **112**, 3833–3845.
32. Gitelson, A. and Merzlyak, M. N., Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol*., 1994, **143**, 286–292.
33. Gitelson, A. A., Gritz, Y. and Merzlyak, M. N., Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol*., 2003, **160**, 271–282.
34. Tucker, C. J., Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing Environ.*, 1979, **8**, 127–150.

ACKNOWLEDGEMENTS. This study was carried out as a part of the 'Biodiversity characterization at community level using EO data' project supported by the Department of Biotechnology (DBT), New Delhi and Department of Space (DOS), Bengaluru. We thank the European Space Agency for providing the S2 data, Alaska Satellite Facility for the ALOS-PALSAR DEM data and National Aeronautical and Space Administration, USA for providing the GEDI data. We also thank the Uttarakhand Forest Department for help in field data collection. This work was funded by DBT, New Delhi and DOS, Bengaluru.

Received 22 April 2022; revised accepted 21 November 2022

doi: 10.18520/cs/v124/i6/704-712