

Materials and Methods

2.1 Study areas

The study was conducted in three protected areas (PAs) in India (Figure 2.1): Shoolpaneshwar Wildlife Sanctuary (SWS) and Vansda National Park (VNP), both situated in Gujarat, and Mudumalai Tiger Reserve (MTR), located in Tamil Nadu. These areas have been declared as protected areas, facing variable rainfall patterns and temperatures as well as increasing anthropogenic pressure. The physiographic and climatic conditions of the three PAs are outlined in Table 2.1, showcasing a gradient in climatic conditions among them. Specifically, there is a discernible increase in rainfall moving from SWS to MTR. Based on the observed variability in the rainfall, SWS and VNP are referred to as drier PAs, and MTR as wetter PA.

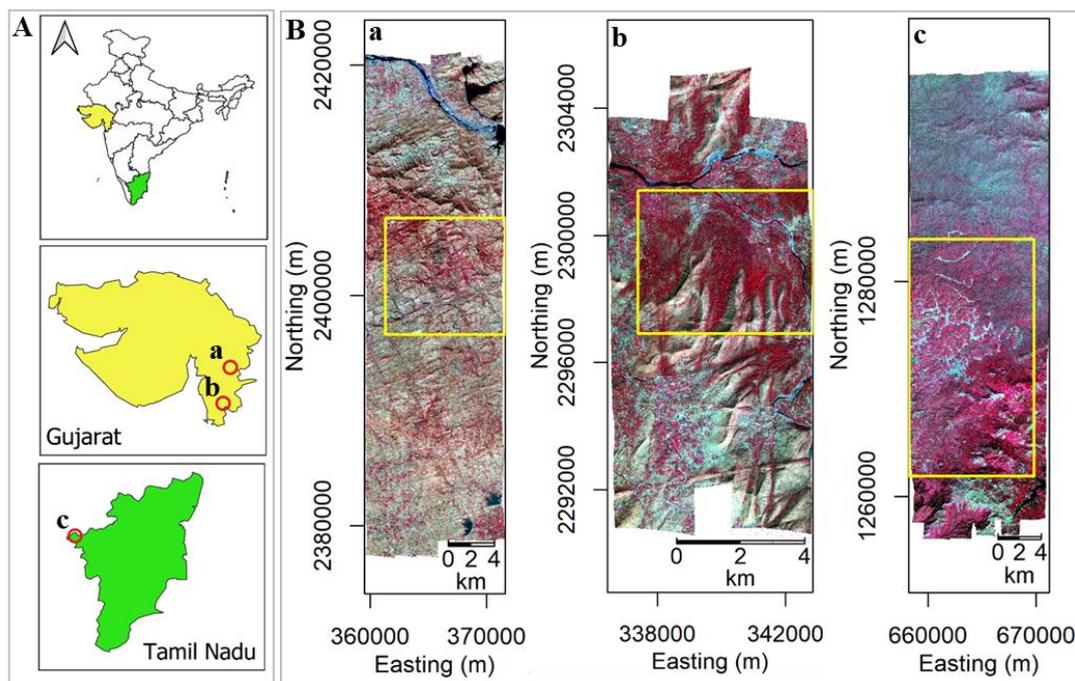


Figure 2.1| A. Location of three protected areas (PAs) in India. B. False color composite image of each PA (862.28 nm; red, 651.92 nm; green, 551.74 nm; blue) obtained from AVIRIS-NG datasets, with the yellow box on image indicating the region of interest generated as a subset based on field survey. (a) Shoolpaneshwar Wildlife Sanctuary (SWS), (b) Vansda National Park (VNP), and (c) Mudumalai Tiger Reserve (MTR).

Table 2.1| Location, topography, and decadal mean of rainfall and temperature of three PAs.

PAs	Location		Average Elevation (m)	Rainfall (mm)	Temperature (°C)	
	Latitude (°N)	Longitude (°E)			Min	Max
SWS	21.88	73.65	287	1140.13	26.88	27.21
VNP	20.82	73.44	169	1511.38	26.45	26.70
MTR	11.73	76.46	1233	1636.98	24.64	25.36

Source: Rainfall: (Funk et al., 2015), Temperature: (Dee et al., 2011) and Average elevation: (JARVIS, 2008).

2.1.1 Description of each Protected Area

Protected Area 1: Shoolpaneshwar Wildlife Sanctuary (SWS)

SWS is situated in the Narmada District of Gujarat, India, at the westernmost end of the Satpura mountain range along the southern bank of the Narmada River. The eastern boundary is shared by the Indian states of Madhya Pradesh and Maharashtra (Nirmal Kumar et al., 2005). The hills in the sanctuary are part of the Satpuras and Vindhayan, with the Rajpipla hills dominating the landscape. The highest peak, Dhaman Mal, reaches an elevation of approximately 882 m in the eastern part. The topography is undulated, featuring continuous and discontinuous mountain regions interspersed with valleys, streams, and agricultural clearings (Christian & Krishnayya, 2009). Geologically, the area consists of basaltic rocks, part of the Deccan trap, scattered with inter-montane valleys dating back from the upper Cretaceous to the lower Eocene (Nirmal Kumar et al., 2005).

The region experiences three distinct seasons: summer (March to June), winter (November to February), and the southwest monsoon (July to October). The temperature of the region reaches its highest point in the summer months (43 °C), and drops to 10 °C during peak winter months of December-January. The decadal mean of rainfall and temperature is presented in Table 2.1. During the monsoon, the sanctuary is covered with numerous streams and rivulets, some of which are ephemeral, forming small pools on rocky river beds used by wildlife. SWS is known for its natural habitat, tribal population, and role as a catchment basin for nearby water sources (Sabnis & Amin, 1992).

The major forest types in SWS are Tropical dry deciduous forests (Champion & Seth, 1968; Roy & Roy, 2015), including very dry teak forests, southern dry mixed deciduous forests, dry bamboo, and tropical riverine forests. Riverine forests are found along the Terav River, Narmada Rivulets, and the Dev River. In the interior of the forest, there are some agricultural fields and experimental silvicultural research plots. Agricultural fields surrounding the villages and human settlements were also observed.

SWS has rich floral and faunal diversity due to its geographical location, climate, topography, and altitudinal variation. An earlier study reported floral diversity comprising 575 species (Nirmal Kumar et al., 2005). According to Pradeepkumar, (1993), the dominant species include *Tectona grandis* and *Dendrocalamus strictus*, evenly spread across the area. All other tree species show a heterogeneous distribution. The sanctuary supports a wide range of fauna, including 32 species of mammal, 198 species of bird, and various insects. The area is home to several animal species including the sloth bear, leopard, rhesus macaque, mongoose, four-horned antelope, barking deer, rusty-spotted cat, wild dog, sambar, Indian porcupine, pangolin and flying squirrel (Nirmal Kumar et al., 2005).

Villages, mainly in the form of hamlets, are inhabited by the Tadvi and Vasava tribes, whose traditional lifestyle relies on hunting, gathering forest resources, and practicing shifting agriculture (Nirmal Kumar et al., 2005). Domestic animal herds include cows, buffaloes, goats, and sheep reared in the Sanctuary. A major source of income for the local villagers is from the harvest of minor forest products (MFP) such as gathering leaves of *Diospyros melanoxylon*, gum, honey, fruits of *Terminalia bellirica* and *Phyllanthus emblica*, and flowers of *Madhuca longifolia*. A visual representation of the PA is shown in Figure 2.2.



Figure 2.2| Field photographs depicting landscape characteristics of SWS.

Protected Area 2: Vansda National Park (VNP)

VNP is located in the Dangs District of Gujarat, India. The landscape of the park is characterized by hills with altitudes ranging from 110 to 360 m, extending from the Sahyadri mountain ranges. In 1986, the State Forest Department, Government of Gujarat, declared a forest area of 23.99 km² as a National Park under the Wildlife (Protection) Act, 1972. The northeastern boundary of the park is formed by the Waghai-Bilimora railway line and the concurrent Ambica-Khapri River, while the southern boundary is marked by the Navtad-Waghai State Highway. The western side is delineated by the Navtad-Kala Amba Road. VNP, along with Purna Wildlife Sanctuary, constitutes the northern zone of the Western Ghats in Gujarat, covering 4.2% of the total geographical area of the Western Ghats zone in Gujarat (V. Kumar et al., 2013).

The climate is tropical, with three distinct seasons: summer (March to mid-June), winter (October to February), and monsoon (from mid-June to October). July experiences the highest rainfall. Intermittent showers occur from November to January and March to May. The decadal mean of rainfall and temperature are given in Table 2.1. Temperatures begin to rise in the latter half of February, with May being the hottest month, featuring a mean daily maximum temperature of around 40 °C and a mean daily

minimum temperature of approximately 26 °C. December is the coldest, with a mean minimum of about 16 °C.

The major forest types in VNP include tropical dry deciduous forests and tropical moist deciduous forests (Champion & Seth, 1968; Roy & Roy, 2015). According to Champion & Seth, (1968), this forest has subgroups such as dry teak forest, moist teak forest, southern dry and moist mixed deciduous forest, and dry bamboo brakes. Agricultural practices within the park are confined to certain areas. Plantations of *Tectona grandis*, *Dendrocalamus strictus*, and *Mangifera indica* were observed. VNP supports rich flora and fauna, with a recorded total of 108 different tree species, 51 shrub species, 64 climbers, 202 herbs, and 25 grass species (Vyas, 2004). Additionally, the park serves as a habitat for numerous higher vertebrates (Singh et al., 2000). The landscape view of the PA is represented in Figure 2.3.



Figure 2.3| Field photographs depicting landscape characteristics of VNP.

Protected Area 3: Mudumalai Tiger Reserve (MTR)

The Mudumalai Tiger Reserve is situated on the border between Tamilnadu and Karnataka, India, and is a part of the Nilgiri Biosphere Reserve (NBR), designated as both an Elephant and a Tiger Reserve (Sukumar et al., 2005). It is surrounded by Bandipur National Park (874 km²), Wayanad Wildlife Sanctuary (344 km²), and Singara Reserve forests. The terrain elevation ranges from 440 m to 1260 m above mean sea level, featuring undulating and varied topography, including hills, valleys, ravines, watercourses, and swamps. The soils are of both red and black loam types, with the base rock being of the igneous type (George et al., 1988).

The region experiences three distinct seasons: summer (March to mid-May), winter (November to February), and monsoon (mid-May to October). The decadal mean of rainfall and temperature are given in Table 2.1. During the summer months, the mean daily temperature reaches a maximum of 35 °C, dropping to 12 °C during winter (December-January). A significant portion of the reserve receives rain from the southwest monsoon between June and September, while another part of the reserve experiences rain from the northeast monsoon between October and November (Suresh et al., 2010).

Tropical dry deciduous forests, tropical moist deciduous forests, and tropical semi-evergreen forests are the types of forests found in MTR (Champion & Seth, 1968; Roy & Roy, 2015). These include dry and moist teak forests, southern dry and moist mixed deciduous forests, and dry and moist bamboo brakes (Champion & Seth, 1968). The western part of MTR exhibits semi-evergreen vegetation, while the eastern part features a dry deciduous cover. The northern portion is characterized by dry deciduous forest, while the southern part displays moist deciduous vegetation. The history and characteristics of dry season fires (Kodandapani et al., 2008), as well as the flora and vegetation types of MTR (Suresh, 2006), have been thoroughly documented. The average fire-return period is 6 years in tropical dry deciduous forests and 10 years in dry thorn forests (Kodandapani et al., 2008). In the tropical dry deciduous forest of MTR, the 50-ha Mudumalai Forest Dynamics Plot (MDFP) is centrally located, with a more comprehensive description provided by Sukumar et al., (2004).

Due to both topographical and climatic variations, the study area is considered unique in its species composition and biodiversity. Plantation of *Camellia sinensis*, *Coffea Arabica*, *Grevillea robusta*, and *Eucalyptus globulus* were observed. The dominant tree species observed here are *Tectona grandis*, *Dalbergia latifolia*, *Lagerstroemia lanceolata*, *Anogeissus latifolia*, and *Terminalia crenulata* (Verma & Jayakumar, 2015). The fauna comprises a diverse wildlife population, including elephants, gaur, tigers, leopards, wild dogs, deer, etc. with varied avifauna and reptiles. Human settlements are scattered throughout the northeastern and central parts of the study area. The landscape view of the PA is depicted in Figure 2.4.



Figure 2.4| Field photographs depicting landscape characteristics of MTR.

2.1.2 Workflow

The workflow illustrating the methods employed in this study is given in Figure 2.5.

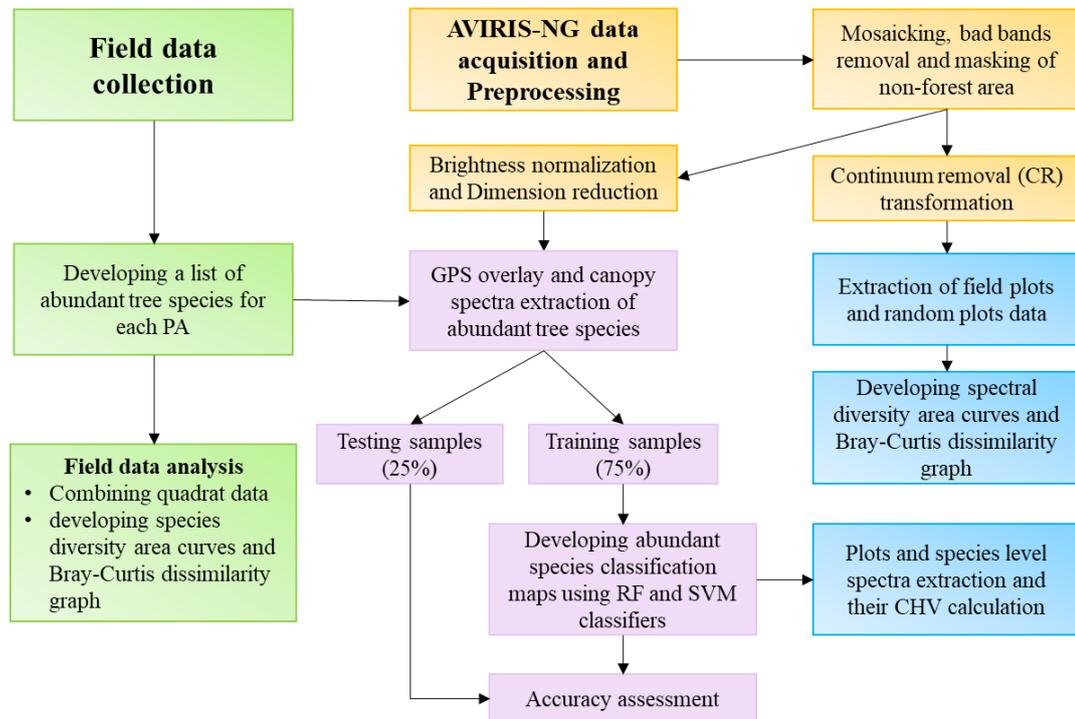


Figure 2.5| Flowchart of the carried-out methodology.

2.2 Data collection

2.2.1 Field survey and data collection

The initial field surveys at each PA were conducted approximately ± 5 days from the date of AVIRIS-NG image acquisition. These field surveys involved the recording of both quantitative and qualitative data on vegetation covers. During the preliminary field visits, it was observed that individual trees of some of the frequently occurring species showed canopy spread of > 4 meters. Before the AVIRIS-NG data collection, the provided technical information indicated a spatial resolution of 4-8 meters. Considering this, quadrats with dimensions of $8\text{ m} \times 8\text{ m}$ were randomly placed along the forest trails, with some deliberately placed at a distance from human trails. Additionally, three to five polygons of varying sizes ($500\text{--}750\text{ m}^2$) were also established. These field quadrats and additional polygons were laid along the length and width of the flight path

(Figure 2.6). The determination of polygon sizes and the number of quadrats at each PA were primarily based on observed vegetation features and expert comments from local forest personnel.

In each quadrat, data on tree species diversity, density, height, and canopy spread were recorded (Figure 2.7). Height measurements were obtained using a Vertex Hypsometer (Haglof, Vertex IV), while canopy spread and the diameter of tree trunks at 1.37 meters above ground were measured using a meter tape. Tree species identification relied on information from published field manuals, consultations with local residents, forest department personnel, and published records. Trees with a diameter > 4 cm at breast height (DBH) in each quadrat and polygon were counted. A portable global positioning system (GPS) device (Garmin Ltd., Olathe KS, USA) was used to geolocate trees and field plots, with an instrument accuracy of less than 3 meters.

In the subsequent months, PAs were revisited to compile comprehensive tree species inventories. Qualitative observations indicated variations in the progression and duration of the senescent phase among species and across the three PAs. Some species exhibited greener crowns and full foliage during the data acquisition period. The recorded tree species were categorized as evergreen or deciduous, common (found in more than one PA), and PA-specific (seen only in one PA). Drier PAs were characterized mostly by deciduous species with a few evergreen ones, while wetter PA supported deciduous species with a higher proportion of evergreen species. The fieldwork was carried out by a team of research fellows from the Ecology laboratory. It can be seen in Chaurasia et al., (2020) and Chaurasia et al., (2021).

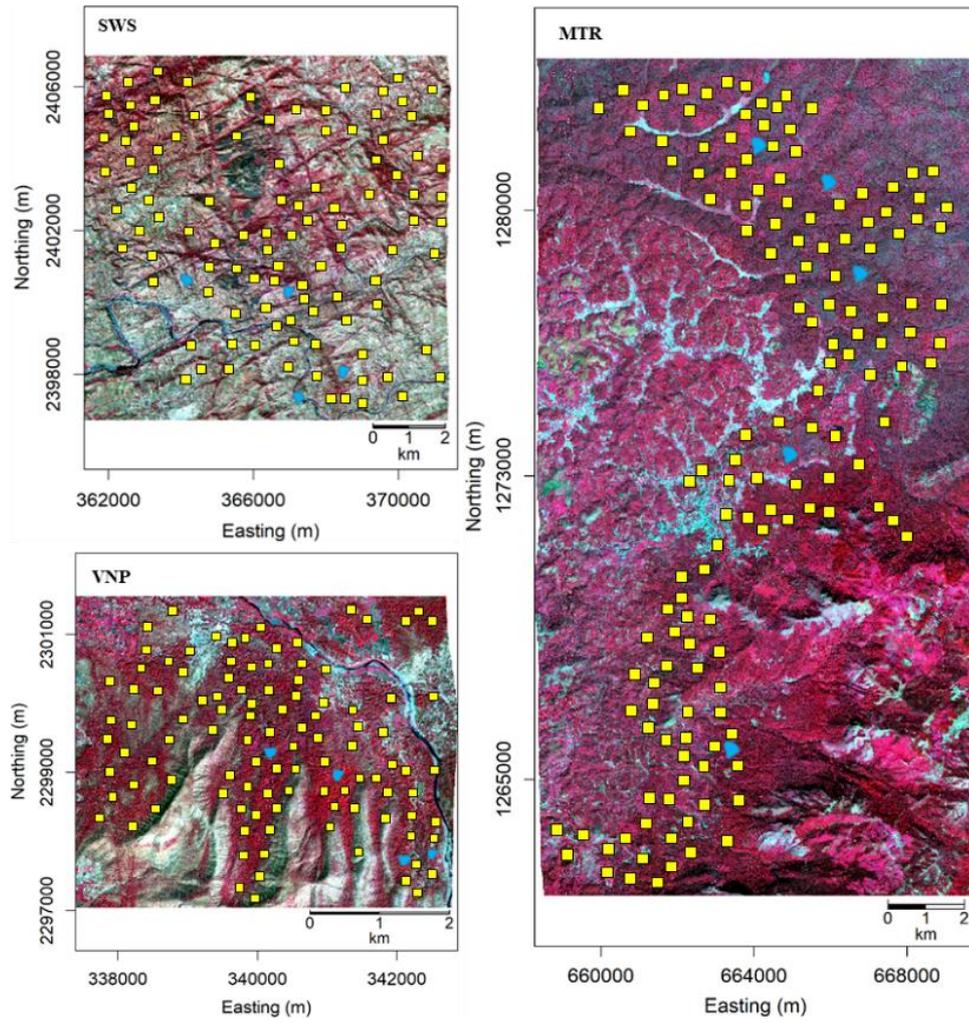


Figure 2.6| Sampled quadrats (yellow) and polygons (blue) generated during the field study.

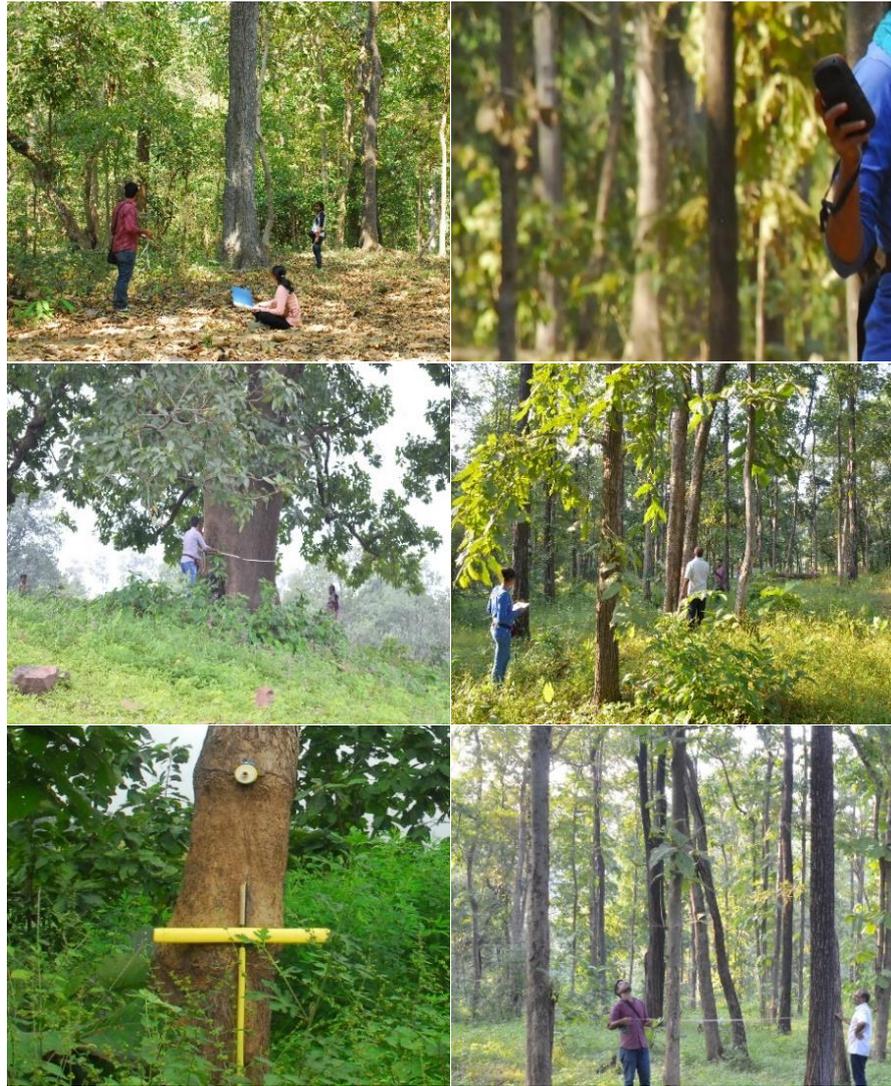


Figure 2.7| Photographs showing activities carried-out during field work.

2.2.2 AVIRIS-NG data and preprocessing

The Airborne Visible/InfraRed Imaging Spectrometer Next Generation (AVIRIS-NG) is a new generation airborne sensor of the National Aeronautics and Space Administration (NASA) developed by Jet Propulsion Laboratory (<https://aviris-ng.jpl.nasa.gov/>). As a part of the Indian Space Research Organization (ISRO)-NASA joint initiative for the HYperSpectral Imaging (HYSI) program, the first phase of the airborne hyperspectral campaign was organized with AVIRIS-NG payload to cover an extensive 22,840 sq. km area across 57 sites in India over 84 days from December 16, 2015, to March 6, 2016 (Bhattacharya et al., 2019). These sites represent diverse fields such as forestry, agriculture, horticulture, geology, coastal areas, oceans, rivers, snow,

etc. Additionally, coordinated field campaigns were conducted, involving researchers from universities and various ISRO centers located in Ahmedabad, Dehradun, Hyderabad, and the Indian Institute of Technology, who collected *in situ* data.

AVIRIS-NG captured data across wavelengths ranging from 380 to 2510 nm with 5 nm spectral sampling. Mounted on a King Air B200 at elevations between 4000–8000 m, the sensor generated pixels ranging in size from 4 to 8 m based on the respective flight altitudes. The AVIRIS-NG data used in this study had a spatial resolution of 4 meters for the three PAs. Comprehensive details regarding AVIRIS-NG image data acquisition and sensor specifications are given in Table 2.2 and Table 2.3 respectively (Hati et al., 2021). The total number of flight lines and geographical area covered at each of the PAs were as follows: six flight lines covering 501.88 sq. km for SWS, three flight lines covering 82.36 sq. km for VNP, and ten flight lines covering 540.70 sq. km for MTR.

The preprocessing procedure involved the generation of Level-0 (L0) and Level-1(L1) radiance data, which were subsequently utilized to generate Level-2 (L2) surface reflectance data comprising 425 bands at 5 nm intervals. Additionally, the dataset underwent atmospheric correction (Thompson et al., 2015). Due to the influence of topography and anisotropic reflectance by vegetation, airborne imaging spectroscopy data were susceptible to strong sun–sensor surface illumination effects. These effects were addressed using the procedures outlined in Soenen et al., (2005) and Wanner et al., (1995), resulting in imagery that were topographically and bidirectional reflectance distribution function (BRDF) corrected. Thus, the AVIRIS-NG dataset used in this study was atmospherically, topographically and BRDF corrected. All of these data preprocessing steps were conducted by the Phil Townsend lab at the University of Wisconsin, Madison, USA, with the associated code available at <https://github.com/EnSpec/HyTools-sandbox>.

Table 2.2| Flight details of AVIRIS-NG data.

PAs	Image acquisition date	Image Acquisition Time (GMT +5:30)	Flight Elevation (km)	Flight footprint length (km)	Geographical area (sq. km)	Cloud cover
SWS	08-Feb-16	12:58:43 – 14:01:58	4.15	44.85	501.88	Clear
VNP	09-Feb-16	11:02:17 – 11:57:46	4.16	13.32	82.36	Clear
MTR	05-Jan-16	11:02:47 – 13:21:39	4.83	42.22	540.70	Clear to hazy

Table 2.3| Sensor specifications of AVIRIS-NG data.

	Specifications	AVIRIS-NG
Spectral information	Range	380 to 2510 nm
	Position	5 nm
	Response	1 to 1.5 X sampling
	Calibration	± 0.1 nm
Radiometric information	Range	0 to max Lambertian
	Precision (SNR)	> 2000 @ 600 nm
	—	> 1000 @ 2200 nm
	Accuracy	95% (<5% uncertainty)
	Linearity	≥ 99% characterization
Spatial information	Range	34 field of view ³³
	Sampling	1 milliradian
	Response	1 to 1.5 X sampling
	Sample Distance	0.3 m to 20 m
	Geom Model	Full 3 Axes cosines
	Uniformity	Spectral Cross-Track
	Spectral-IFOV-Variation	> 95% Spectral Direction

The topographically and BRDF-corrected images of each PA were mosaicked into a single image using Environment for Visualizing Images (ENVI v 5.3) software. Subsequently, a subset was extracted from this mosaic for each PA, serving as the region of interest for tree species classification. The forest area of each subset in the respective PAs is 17.23 sq. km (SWS), 10.92 sq. km (VNP), and 141.67 sq. km (MTR).

Out of the available 425 bands, 366 usable bands were retained for analysis after eliminating noisy (< 411 nm) and water vapour absorption bands (1348–1428, 1778–1949 nm) due to their low spectral information content for vegetation. Based on visual and field observations, non-forest areas were masked by applying NDVI threshold

values of 0.4 (SWS and VNP) and 0.6 (MTR) to each image. The forest pixels having > threshold values were retained for further processing. The NDVI formula used is given in Equation 1.

$$NDVI = \frac{\rho_{860} - \rho_{650}}{\rho_{860} + \rho_{650}} \quad (1)$$

These thresholds were visually determined following Dahlin, (2016) and selected to exclude pixels likely to be mixtures of soil and vegetation from the analysis. The variation in NDVI threshold values across PAs was due to the observed differences in canopy greenness and senescent phases during the flight pass over the three PAs. Following Dahlin, (2016), a continuum removal (CR) transformation was also applied to minimize overall brightness variation. CR, widely used in studies linking plant functional traits to imaging spectroscopy (Dahlin et al., 2013; Féret & Asner, 2014), involves fitting a convex hull over each spectrum, setting the convex hull to 1.0, and subtracting the original spectrum. This normalization of reflectance spectra ensures that CR spectra maintain the same number of dimensions as the original, with end values set to one. Spectra exhibiting zero variance were excluded, while the remaining CR-transformed spectra were considered for further spectral diversity analysis.

2.3 Data analysis

2.3.1 Field data analysis

The field-collected data from 8 m × 8 m quadrats were aggregated by combining 10 quadrats to form each field plot. This aggregation resulted in 20 field plots for SWS, 19 for VNP, and 21 for MTR. Examination of the field data revealed that 21–23 species were seen frequently at each PA, collectively accounting for more than 85% of the forest cover and thus considered as abundant species. The field plot data for analysis were systematically organized into matrices, representing the abundance values of each species for every field plot (Kindt & Coe, 2005). These matrices were generated for all recorded species and abundant species of each PA. All analyses of field plots were conducted in R version 4.0.5 (R Core Team, 2021), using the “BiodiversityR” package (Kindt, 2023).

Estimation of the total number of species in the survey area

In instances where the objective is to estimate the number of species in a vast area, it is acknowledged that the entire area cannot be surveyed. Due to the inherent challenge of many species remaining unseen or undetected in sample-based surveys (Kindt & Coe, 2005; Oksanen, 2018), the determination of total species was estimated using various estimators to provide a more comprehensive understanding of species richness patterns. To extrapolate the total number of species in three PAs, four non-parametric estimators (Chao, Jackknife 1, Jackknife 2, and Bootstrap) were evaluated.

The expectation is that using a range, rather than a single value, will provide a better characterization of tree species richness estimates for a PA. These estimators estimate the number of unseen species and incorporate them into the observed species richness. Equations for these estimators are given below (Oksanen, 2018). In the following, S_P is the extrapolated richness in a pool, S_0 is the observed number of species in the collection, a_1 and a_2 are the number of species occurring only in one or only in two sites in the collection, p_i is the frequency of species i , and N is the number of sites in the collection. The variants of extrapolated richness are:

$$\begin{array}{ll}
 \text{Chao bias-corrected} & S_P = S_0 + \frac{a_1(a_1-1)}{2(a_2+1)} \frac{N-1}{N} \\
 \text{First order jackknife} & S_P = S_0 + a_1 \frac{N-1}{N} \\
 \text{Second order jackknife} & S_P = S_0 + a_1 \frac{2N-3}{N} - a_2 \frac{(N-2)^2}{N(N-1)} \\
 \text{Bootstrap} & S_P = S_0 + \sum_{i=1}^{S_0} (1 - p_i)^N
 \end{array}$$

These estimators generally perform well when there are a moderate number of rare species, but their accuracy can decrease if the number of rare species is very high. Chao estimator is specifically designed to handle datasets with many rare species and will often give higher estimates in such cases. Bootstrap estimator can smooth out the effect of rare species by averaging over many resampled datasets, but it may still be influenced by the presence of rare species.

Rank-abundance curve

Rank-abundance curves provide a straightforward method for analyzing diversity patterns (Kindt & Coe, 2005). To assess potential changes in the tree community across the three PAs, rank-abundance curves were constructed using field plot data from each

PA. Also known as Whittaker plots (Whittaker, 1965), these curves arrange recorded species from most to least abundant along the x-axis, with their proportional or percentage abundance displayed on the y-axis. To accommodate all recorded species on a single graph, the rank abundance curve is typically presented in a log transformation. These curves distinctly illustrate variations in species richness (width) and highlight differences in evenness among assemblages (shape).

Species diversity-area curves

Species diversity-area curves (Dahlin, 2016) also referred to as species accumulation curves depict the species richness for different plot combinations (Kindt & Coe, 2005). To estimate and compare species richness in each PA, species diversity-area curves were generated. These curves illustrate the average pooled species richness for all possible combinations of plots. The curve continues to rise as long as new species are discovered with increasing sampling efforts (Ugland et al., 2003). When no new species are discovered, even with an increase in the number of sample plots, the curve will reach an asymptote. This signifies that the species inventory was adequate to comprehensively capture the total species richness within the corresponding PA (Ugland et al., 2003). In this study, species diversity-area curves were developed for all the field plots in each PA using the accumulation exact method (Kindt & Coe, 2005) with 1000 permutations. These curves were created for all the recorded species as well as the abundant species in each PA.

Bray-Curtis dissimilarity

Exploring the intricacies of species composition goes beyond the insights provided by species diversity area curves. While these curves lack information on species overlap, Kindt & Coe, (2005) suggest that measures of similarity or dissimilarity become crucial in understanding the diversity dynamics. It is also emphasized that the quantification of biodiversity is incomplete without addressing the spatial variation in species composition (Rocchini et al., 2016). To measure the differences in species composition among plots (beta diversity), various dissimilarity methods exist, with Bray-Curtis dissimilarity (Bray & Curtis, 1957) as one of the most effective dissimilarity measures (Southwood & Henderson, 2009). This method takes into account species abundance

(Chao et al., 2005). Ecological distance summarizes the differences into a single distance statistic (Kindt & Coe, 2005).

In this study, Bray-Curtis dissimilarity was calculated for field plots of each PA, and graphs were created to examine the compositional dissimilarity between any pair of field plots within a PA. The values of the Bray-Curtis dissimilarity matrix range from 0 to 1, depending on the degree of similarity between existing species in a pair of plots (0 for 100% similarity, 1 for complete dissimilarity). Bray-Curtis graphs, akin to species diversity area curves, were created for all recorded species and abundant species in each PA. These graphs provide a visual representation of the compositional dissimilarity, offering insights into the spatial variation and diversity dynamics within the studied PAs.

2.3.2 Data preparation for classification

Additional class to the identified species classes

In addition to the 21–23 abundant species, an extra four to six species were identified in each PA, collectively covering 8%–10% of the respective forest area. These species exhibited a relatively low number of individuals recorded during field surveys, and showed less variation in physiognomy and canopy spectra. Consequently, they were grouped into a single class and given the name ‘others’. The combined contribution of the abundant species and the ‘others’ accounted for 88%–93% of the total forest cover in each PA. The class ‘others’ was referred to as an abundant species in the classification process. As a result, the total number of species considered for classification in each PA was 24 for SWS, 23 for VNP, and 22 for MTR, combining both the abundant species and the ‘others’ class. This grouping strategy aimed to capture the overall biodiversity and spectral characteristics within each PA, acknowledging the ecological significance of both abundant and less numerous species in the classification efforts.

Crown level spectra extraction

A crown-level spectral library was developed manually for all the abundant tree species. The spatial position of different tree species was determined by overlaying GPS

coordinates onto an AVIRIS-NG false-color composite image in ENVI 5.3 software. Using the Region of Interest (ROI) tool in ENVI 5.3 software, spectra of each species were extracted from the image. The false-color composite image of the area was used to ensure that no pixels were collected outside the canopy of each species. Specific criteria were established for extracting crown-level spectra for each species, taking into account tree crown dimensions and prevailing growing conditions. The diameter of the tree crown had to be sufficiently large, encompassing at least one pixel (4 m). Abundant species were often encountered with a canopy spread exceeding 5 m or as pure patches with a spread exceeding 15 m (Table 2.4). To mitigate potential spectral mixing from other species, selected trees were predominantly of a single tree species, allowing for the extraction of pure pixels for classification. For the classification process, shapefiles containing crown-level spectra of abundant tree species were extracted from the image for each PA.

Table 2.4| Measured biophysical parameters of the tree species at each PA.

PAs	Height (m)			Canopy area (m ²)			DBH (m)		
	Mean (±SD)	Min	Max	Mean (±SD)	Min	Max	Mean (±SD)	Min	Max
SWS	12.23 ± 4.90	2.60	25.80	57.76 ± 67.32	2.41	613.80	0.42 ± 0.29	0.04	1.80
VNP	18.31 ± 6.29	3.50	35.00	71.12 ± 68.71	2.55	446.93	0.38 ± 0.22	0.05	1.58
MTR	22.83 ± 7.57	6.00	39.00	74.61 ± 74.44	2.41	638.20	0.68± 0.35	0.04	1.98

Spectral variation and regions selection

For the spectral analysis of tree species, only those species with a minimum abundance of five individuals in each PA were chosen. Using ENVI 5.3 software, the spectral values (reflectance) for each abundant species across the full spectrum were extracted. Subsequently, the mean reflectance of each abundant species was plotted against the spectral bands, creating a visual representation of the species' spectral characteristics within each PA. This step not only facilitated a detailed exploration of individual species' spectral behavior but also provided insights into the overall spectral variations

across the study areas. To enhance the clarity and efficacy of the subsequent classification process, spectral regions where the spectra of different species overlapped were eliminated. This step aimed to mitigate potential confusion arising from spectral similarities, thereby ensuring more accurate discrimination of species based on their unique spectral signatures (Fassnacht et al., 2016; Thenkabail et al., 2004).

The determination of optimal hyperspectral narrow bands for vegetation studies involves a thorough literature review (L. Kumar et al., 2001; Schmidt & Skidmore, 2003; Thenkabail et al., 2004; Vaiphasa et al., 2005). Additionally, visual observation was used to identify spectral regions where species exhibited separability. The selected regions, chosen for their significance for vegetation, were incorporated into numerous studies due to their absorption and reflectance characteristics across various spectrum regions (Ahmad et al., 2021).

Brightness normalization

Brightness normalization (BNORM) serves as a reflectance normalization method designed to correct brightness gradients in spectral bands (Jänicke et al., 2020). By implementing BNORM, variations in brightness are eliminated, while emphasizing the inherent structural characteristics of the spectrum. This is achieved by dividing each image spectrum by its sum over all bands (Berman et al., 2004; Collings et al., 2010). In this study, BNORM was specifically applied to the bands within the selected spectral regions. This approach ensured that any brightness discrepancies across these bands were rectified, allowing for a more accurate and consistent analysis of the spectral data.

Moreover, in addition to BNORM, the use of continuum removal (CR) normalization was also incorporated (Jänicke et al., 2020). This technique involves dividing each image spectrum by its respective convex hull, aiming to eliminate brightness differences and enhances the absorption features akin to BNORM. However, upon comprehensive evaluation, it was observed that BNORM consistently outperformed CR in terms of classifier performance. Therefore, BNORM was chosen as the preferred normalization method for this study.

Dimensionality reduction

The hyperspectral imagery has huge data dimensionality which leads to extensive data processing and more complex computations. To address this challenge, the brightness-normalized spectral bands of each PA were subjected to forward Minimum Noise Fraction (MNF) transformation (Green et al., 1988). This transformation was implemented to mitigate the high data dimensionality issue in hyperspectral images and enhance the identification of bands with maximum variance. The MNF transformation effectively reduces the high-dimensional data into a more manageable, low-dimensional form without loss of information.

This process involves two consecutive Principal Component (PC) transformations. Firstly, the noise covariance matrix is computed to decorrelate and rescale the noise from the data, referred to as "noise whitening". Subsequently, eigen decomposition is applied to the modified matrix to organize the bands based on their signal-to-noise ratio. The first PC transformation primarily focuses on whitening the noise, while the second standard PC transform arranges bands in descending order of significance, from the highest eigenvalue (>1) containing the most significant bands to the lowest eigenvalue (close to 1) with noisy bands (Ballanti et al., 2016).

The resulting set of MNF bands was ranked solely based on variance within demarcated forested areas. While ENVI's data dimensionality wizard allows for automated estimations of MNF band coherence, using default calculations carries a risk of both over- and underestimating dimensionality. To address these concerns, a visual analysis of MNF band images was conducted. The analysis revealed that the first 15 MNF bands for VNP and the first 20 bands for SWS and MTR demonstrated coherence and provided more informative content. These selected bands were utilized as inputs for the classification of abundant species.

Data sampling

The crown-level spectra extracted for each abundant species of each PA were divided into training (75%) and testing (25%) samples, using a simple random sampling technique with replacement. To address the issue of imbalanced classification, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training

samples (Chawla et al., 2002). SMOTE tackles the imbalance by duplicating classes with fewer samples, thereby augmenting the minority data population. The application of this approach resulted in enhanced classification outcomes.

2.3.3 Classification of hyperspectral data

Abundant species mapping using Random Forest (RF) and Support Vector Machine (SVM)

A comparative classification approach was utilized to evaluate the performance of two machine learning algorithms (RF and SVM). The MNF-transformed spectra of training samples for all abundant species of each PA were subjected to supervised classification using the RF classifier (Breiman, 2001) and SVM classifier (Cortes & Vapnik, 1995). Both of these non-parametric classification algorithms are widely recognized as valid and effective methods for hyperspectral remote sensing data classification (Melgani & Bruzzone, 2004; Mountrakis et al., 2011; Waske et al., 2012; Wu et al., 2023). The classification was carried out using the “caret” package (Kuhn et al., 2022), which offers a standard syntax for executing various machine-learning algorithms, simplifying the process of systematically comparing different classifiers (Ghosh et al., 2014; Maxwell et al., 2018). Additionally, the “raster” (Hijmans et al., 2023), and the “themis” (Hvitfeldt, 2022) packages were also utilized.

The popularity of the RF classifier arises from its user-friendly interface and capability to produce accurate and robust results, even with small sample sizes and high-dimensional feature spaces (Scornet et al., 2015). RF is an ensemble classifier that uses bootstrap aggregation (bagging) to construct multiple decision trees, utilizing the best subset of input variables at each node (Genuer et al., 2010). During the bagging process, a portion of the original data is selected with replacement to develop each tree, while the remaining samples, representing the out-of-bag (OOB) sample, are reserved for model validation (Adelabu et al., 2015; Maxwell et al., 2018). The aggregation of OOB predictions provides the mean squared error of the model (Friedman, 2001).

The RF classifier in the “caret” package utilizes the “randomForest” package (Liaw & Wiener, 2002). It is straightforward to implement, requiring only two tuning

parameters: the number of trees (ntree) and the number of features (mtry), which are randomly selected at each split in the tree-building process (Ghosh et al., 2014; Probst et al., 2018). Accordingly, the RF classifier was tuned by varying values of tuning parameters at each PA, with mtry parameter set to two, and ntree values ranging from 300 to 700 for the three PAs.

SVM is a kernel-based classifier designed to identify the optimal hyperplane in n -dimensional feature space with the highest margin between classes by introducing the concept of the kernel function (Cortes & Vapnik, 1995). The SVM classifier in the “caret” package utilizes the “kernlab” package (Karatzoglou et al., 2004). In this study, considering the complexity of the dataset, the radial kernel function (Ghosh et al., 2014; Wu et al., 2023) was employed. The training of the “svmRadial” method involves tuning two parameters: cost of constraints violation (C) and sigma (σ). The parameters were tuned with a range of values to identify the optimal model, setting the value of C to five, and sigma values ranging from 0.05 to 0.7 for the three PAs. The resulting classification maps of abundant species underwent an accuracy assessment procedure, concluding with a comparative analysis of the outputs from the two classifiers.

2.3.4 Accuracy assessment

Accuracy assessment is an important aspect of remote sensing data analysis. Standard methods were chosen by using a confusion matrix as the basis for comparison. This matrix delineates accurately identified pixels and facilitates the evaluation of class accuracy through metrics such as overall accuracy, producer's accuracy, user's accuracy, and kappa coefficient (Congalton & Green, 2019). The overall accuracy, representing the total classification accuracy, is computed by summing all correctly classified values and dividing them by the total number of values. Producer's accuracy measures the probability of correctly classifying a feature on the ground, determined by dividing pixels correctly classified in each class by the sample pixels for that class. Conversely, user's accuracy assesses the likelihood that a pixel labeled as a specific class on the map genuinely belongs to that class. It is calculated by dividing accurately classified pixels by the total pixels classified in that class. The kappa coefficient, a discrete multivariate method in accuracy assessment, ranges from -1 to 1. A value of 0 suggests no improvement over random classification, while a negative value indicates

a classification significantly worse than random. A value near 1 signifies a classification significantly better than random.

The accuracy of the classifications was also assessed using the receiver operating characteristics (ROC) technique (DeLeo, 1993) and its derived area under the curve (AUC) (Bradley, 1997; Hanley & McNeil, 1982). The ROC graph, a two-dimensional representation of a classifier's performance (Fawcett, 2006), is generated by calculating sensitivity (equation 2) and specificity (equation 3) for each possible classification threshold, where

$$\text{Sensitivity} = \frac{a}{(a+c)} \quad (2)$$

$$\text{Specificity} = \frac{b}{(b+d)} \quad (3)$$

Here, a and d signify true positives and true negatives for a given classification respectively, and b and c represent corresponding false positives and false negatives. Sensitivity reflects the probability of correctly classifying a pixel belonging to a specific class, while specificity measures the probability of accurately classifying a pixel from a different class. In this way, the best-performing classification would be the one with the highest possible value of both sensitivity and specificity. The ROC graph was developed for each class, and from it, the corresponding AUC was calculated. AUC serves as a quantitative performance score, measuring the probability that a randomly selected positive sample is correctly classified with higher suspicion than a randomly selected negative sample, ranging between 0.5 (random assignment) and 1 (perfect classification).

2.3.5 Comparison of classifiers

A thorough visual examination was performed to detect potential areas of error and highlight differences between classification maps, which can be observed in the zoomed-in subset. The results obtained from the two classifiers were assessed using a confusion matrix, which considered overall accuracy, user's accuracy, and producer's accuracy (Congalton & Green, 2019).

The utility of the developed classification map was evaluated to determine its alignment with the observed distribution of both common and PA-specific species, as well as evergreen and deciduous species in the field. Additionally, the distribution of common abundant species across the three PAs was validated using the developed classification map, and proportional distribution maps were generated accordingly. This analysis aimed to ensure the accurate representation of the prevalence and distribution of commonly occurring tree species on the map. The verification process added depth to the assessment, contributing to a more nuanced understanding of the vegetation patterns within the studied PA.

2.4 Spectral data analysis

2.4.1 Correlation between species and spectral diversity

To explore the correlation between species and spectral diversity, various methodologies have been used (Chaurasia et al., 2020; Dahlin, 2016; Gholizadeh et al., 2018). In this study, both the CR-transformed images and the abundant species maps of each PA were used to assess the efficacy of CHV as a spectral diversity measure and its association with species diversity.

The CHV calculates the volume of pixels forming a convex hull, using the first three principal components (PC1, PC2, and PC3) obtained from spectral reflectance data (Dahlin, 2016). The CHV represents the volume of the convex hull enclosing these data points. Larger CHV values indicate greater spectral diversity.

Mathematically: Given a set P of n points in the plane (Rosén et al., 2014). To define the convex hull of P , denoted $\text{conv}(P)$, the convex hull is the largest convex polygon whose vertices are all points in P (Figure 2.8).

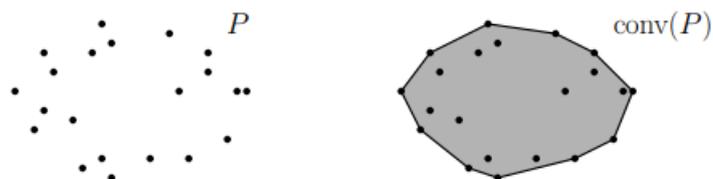


Figure 2.8| A point set and its convex hull.

CHV calculation for plots and species derived from classification maps

In this analysis, RF classification maps were utilized to select plots of uniform size (0.5 ha, $n = 6$), each containing a diverse range of species characterized by both low and high alpha diversity in the three PAs. After the selection of these plots, CR-transformed spectral data were extracted, and the CHV values were calculated at the plot level using the code provided by Dahlin, (2016). This provided valuable insights into whether and how the CHV values changed as the number of species in each plot increased. This method allowed for a comprehensive examination of how CHV values evolved in response to varying species composition within each plot.

As an alternative approach, an equal number of canopy spectra ($n = 10$) for each abundant species were extracted, aligning with the RF classification maps of each PA. Subsequently, these canopy spectra for each abundant species were added, and CHVs were calculated to explore the intricate one-to-one relationship between the number of abundant species and cumulative spectral diversity. Linear Regression Line was then used to test this relationship, providing a statistical assessment of the correlation between the number of species and their corresponding CHVs. This dual-method approach, combining information from RF classification maps, provided a robust framework for investigating the dynamic interplay between species diversity and spectral diversity.

Spectral diversity area-curves for random plots

In this method, 175 plots (10×10 pixels) were randomly laid down on an NDVI-masked image in ENVI v 5.3, and their CR-transformed spectra were extracted from each PA to investigate the relevance of hyperspectral data in evaluating diversity without field study inputs. The CR-transformed spectral data (366 bands) from the sampled plots of each PA were then exported from ENVI to R for further analysis, following the code provided by Dahlin, (2016). The initial step involved reducing the dimensionality of the spectral data using Principal Component Analysis (PCA), where the first three principal components (PCs) accounted for the majority of overall variation (94%). As a measure of spectral diversity, CHV values and the sum of the

variance of the first three PCs were calculated (Dahlin, 2016). Spectral diversity-area curves were constructed for each PA using the sum of the variance of sampled plots.

The CHV, akin to the multivariate equivalent of range (Cornwell et al., 2006), estimates the volume of the trait space occupied by each species within a community, irrespective of distribution shape. The CHV values and the sum of the variance of the first three PCs calculated from the spectral data of this study were utilized as empirical proxies for demonstrating the functional trait diversity of tree species within each PA, as reported by Dahlin, (2016).

Spectral diversity-area curves for field plots

To establish the relationship between the species diversity measured in the field plots and the spectral diversity obtained from remote sensing measurements over the same areas, this approach involved extracting spectral data from areas falling within the GPS coordinates of the field plots (20 in SWS, 19 in VNP, and 21 in MTR) within each PA. CHV values were calculated from the CR-transformed spectral data of each field plot, and these values were utilized to construct a moving average-based spectral diversity-area curve, representing the spectral diversity of species recorded during the field study. In the subsequent step, spectral data for only abundant species were extracted, aligning with the classification map of each PA, and their corresponding CHV values were calculated. Given that the spectral data were derived from the classification map indicating abundant species in the regions of interest, the patterns identified through this approach were considered as abundant species spectral diversity-area curves. Additionally, Bray-Curtis dissimilarity graphs were also generated for both all recorded species and abundant species, using the CHVs calculated for the field plots in each PA.

Comparative analysis of species and spectral diversity area curves

A comparison was conducted between the patterns observed in the species diversity-area curve and the spectral diversity-area curves of both all recorded species and abundant species. Furthermore, these patterns were contrasted with the spectral diversity area-curve obtained from the summed variance of 175 randomly selected plots. Additionally, Bray-Curtis dissimilarity graphs generated from both field data and remote sensing data were compared to further enhance the understanding of the

ecological dynamics. The purpose of this comprehensive analysis was to identify noteworthy similarities or distinctions among the various representations of species and spectral diversity. It aimed to provide valuable insights into the relationship between these ecological parameters within the studied areas.

2.4.2 Intra- and Inter-species spectral diversity

To explore variations in both intra- and inter-species characteristics, abundant species maps derived from remote sensing data were utilized. For each of the common abundant tree species 500 spectra were extracted across three PAs, to assess the effectiveness of spectral data as indicators of functional trait diversity. The CHV of the spectral data for each common abundant species in each PA was then calculated and visualized in a graph. The hypothesis posited that the CHV values for each of the common tree species would exhibit minimal variation, suggesting resilience to environmental factors. This analysis aimed to elucidate the relationships between species diversity, spectral diversity, and the functional trait diversity of common tree species within the studied areas.