Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Vegetation Community Composition and Species–Environment Relationships Along an Elevational Gradient in Sarpang District, Bhutan

Wangdi Wangdi

wangdiues@gmail.com

Ministry of Energy and Natural Resources, Royal Government of Bhutan

Rupesh Subedi

Royal University of Bhutan

Kezang Choden

Royal University of Bhutan, Paro College of Education

Research Article

Keywords: elevational gradient, plant community, tropical mountain forest, forest type, scale mismatch, diversity, regeneration, ordination, PERMANOVA, indicator species

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10185208/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Vegetation Community Composition and Species–Environment Relationships Along an Elevational Gradient in Sarpang District, Bhutan.

Wangdi[1,*], Rupesh Subedi[2], Kezang Choden[3]

[1]Forest Resources Planning and Management Division,

Department of Forests and Park Services, Ministry of Energy and Natural Resources,

Royal Government of Bhutan, Thimphu, Bhutan.

[2]Royal University of Bhutan, Thimphu District, Bhutan.

Associate Lecturer (Forest Science).

[3]Royal University of Bhutan,

Paro College of Education, Bhutan. Research Officer.

*Correspondence: Wangdi Wangdi, Forest Resources Planning and Management Division, Department of Forests and Park Services, Ministry of Energy and Natural Resources, Royal Government of Bhutan, Thimphu, Bhutan. Email: wangdiues@gmail.com

ORCID: Wangdi Wangdi, 0009-0007-7726-1742; Rupesh Subedi, 0000-0003-0417-0296; Kezang Choden, 0009-0004-1245-5931

Abstract

Understanding variation in mountain forest plant communities requires distinguishing broad elevational structure from local environmental variation. Vegetation data from four strata—trees, shrubs, herbs, and regeneration—were analysed across unmanaged broadleaved forests in Sarpang Dzongkhag. Community matrices were combined with \ terrain variables, bioclimatic variables and vegetation indices. Non-metric multidimensional scaling, PERMANOVA with dispersion diagnostics, environmental vector fitting, indicator species analysis and exploratory machine-learning models were used to evaluate community structure, environmental associations and regeneration predictability. Community composition differed among forest types in all strata, but effect

sizes were small (PERMANOVA R2 = 0.017-0.050), and shrub and herb results were partly confounded by heterogeneous dispersion. Environmental vector fitting identified weak associations for trees, herbs and regeneration, while shrubs showed stronger but still exploratory associations with elevation, evapotranspiration, temperature and macroclimatic gradients. Indicator species represented a small proportion of taxa in all strata. Regeneration richness was weakly predictable: random forest explained only 0.081 +/- 0.057 of cross-validated variance, whereas gradient boosting performed worse than a mean predictor. The results show that broad forest-type classes contain a detectable but limited compositional signal and do not support strong claims about sharply separated forest communities. The study provides a baseline for long-term monitoring and highlights the need for direct plot-level soil, canopy, disturbance, and microclimate measurements in Himalayan Forest studies.

Keywords: elevational gradient, plant community, tropical mountain forest, forest type, scale mismatch, diversity, regeneration, ordination, PERMANOVA, indicator species

Introduction

Elevational gradients constitute powerful natural laboratories for understanding how climate, topography, and habitat heterogeneity interact to structure plant communities. Because temperature, moisture availability, and growing-season length vary predictably with altitude, mountain systems have long served as test beds for ecological theory addressing species richness, environmental filtering, and community turnover [1, 2]. Across regions, richness–elevation relationships are frequently unimodal but vary with spatial extent, taxonomic scope, and climatic context, indicating that no single mechanism universally governs elevational diversity patterns [3, 4]. In forest ecosystems, steep abiotic gradients combined with variation in soils, canopy structure, and disturbance regimes generate pronounced compositional turnover and vertically stratified responses among trees, shrubs, and herb layers [5, 6]. Mountain floras therefore often exhibit high beta diversity over short distances, reflecting both dispersal limitation and niche differentiation along climatic and edaphic gradients [7]. Recent research has increasingly focused on disentangling the relative roles of environmental filtering, functional traits, and biotic interactions in shaping elevational community organization [8, 9]. Nevertheless, many studies continue to focus primarily on canopy trees, despite mounting evidence that

58 understorey strata respond more sensitively to microclimatic and edaphic variation and
59 may follow distinct compositional trajectories along the same gradient [5, 10].
60 Regeneration layers are especially informative in this context because they integrate
61 recent recruitment processes and may foreshadow future changes in forest composition
62 under ongoing climatic warming [11].

63 The eastern Himalaya represents a global centre of biodiversity characterized by extreme
64 elevational relief, complex terrain, and monsoonal climates that generate sharp forest
65 transitions and substantial species turnover [3, 12]. Within this region, numerous studies
66 have documented steep elevational zonation and marked diversity patterns, while broad-
67 scale syntheses emphasize the region's sharp biodiversity gradients and habitat
68 heterogeneity [3, 12]. Despite this growing body of work, community-level investigations
69 integrating multiple vegetation strata across Himalayan elevational gradients remain
70 comparatively uncommon. Few studies explicitly compare species–environment
71 relationships among canopy and understorey layers, and regeneration dynamics are
72 seldom analysed in parallel with diversity and compositional patterns, despite their central
73 importance for forest resilience and long-term reorganization [6, 11].

74 This study presents a multi-stratum assessment of plant community composition, diversity,
75 and species–environment relationships in unmanaged broadleaved forests of Sarpang.
76 Rather than treating elevation alone as a complete mechanism, the sampled gradient is
77 used to evaluate three explicit questions and associated working expectations. First, do
78 broad forest-type and elevational classes show detectable compositional structure across
79 trees, shrubs, herbs, and regeneration? Second, how much community variation is
80 explained by measured topography and climate variables in the complex mountain terrain?
81 Third, is regeneration richness predictably related to the available stand, topographic, and
82 macroclimatic predictors? Machine-learning models were used only as exploratory
83 predictive tools. This framework positions the study as a baseline assessment of broad
84 community structure and species compositional changes with changes in climatic and
85 topographic features.

Methods

Study area

The study was conducted in unmanaged forests of Sarpang Dzongkhag. The region spans a pronounced elevational gradient from approximately 153 to 3 500 m a.s.l., encompassing subtropical broadleaved forests at lower elevations (100–500 m), warm broadleaved forests at mid-elevations (500–1 500 m), and cool broadleaved forests between 1 500 and 3 000 m. Such steep altitudinal transitions typify Himalayan landscapes and generate strong climatic and edaphic gradients over short distances [3, 13] (Figure 1).

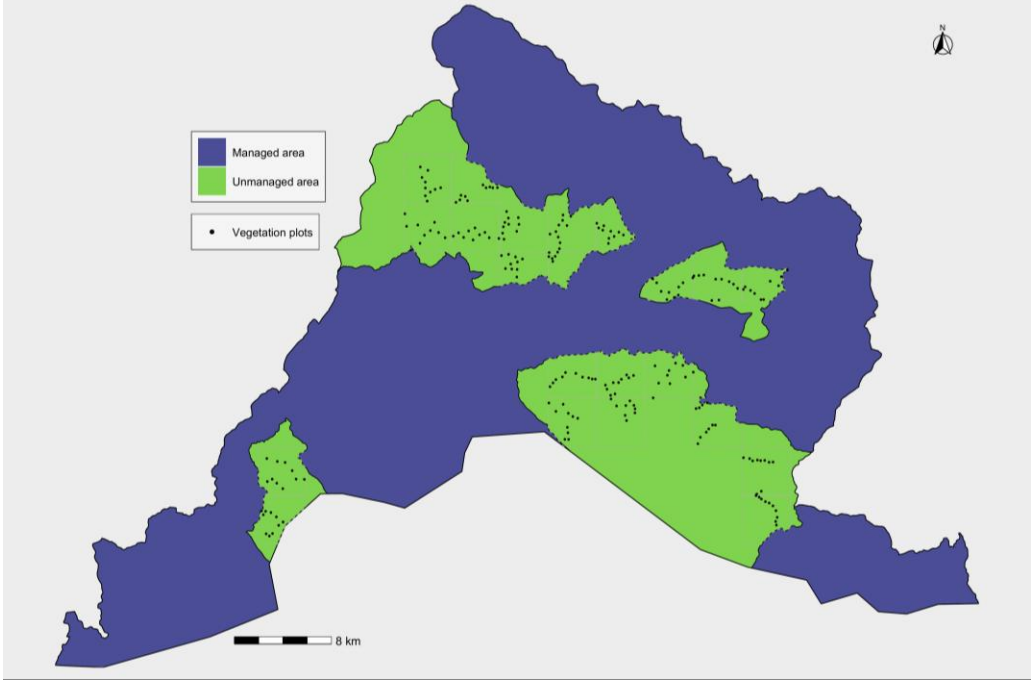


Fig. 1 Study area and vegetation sampling plots in Sarpang District. Green polygons indicate unmanaged forest areas and black points represent sampled vegetation plots along the elevational gradient

This broader regional range provides ecological context; all statistical inference in this study is based on sampled plots spanning approximately 260 to 1 964 m a.s.l.

Annual precipitation is high and strongly seasonal, reflecting monsoon circulation intensified by orographic uplift, a defining feature of the eastern Himalayan foothills [14]. Rugged terrain and variable slope exposure create pronounced microclimatic

heterogeneity, which is expected to contribute to high beta diversity and fine-scale community turnover along the gradient [3].

Sampling design

Vegetation sampling followed a stratified random design along the sampled elevational gradient extending from Shershong (approximately 260 m a.s.l.) in the south to Jigmechoeling (approximately 1 964 m a.s.l.) in the north. Plots were allocated to broad forest-type classes i.e. subtropical forest, warm broadleaved forest and cool broadleaved forest. These classes were assigned from existing vegetation-zone descriptions, elevation, physiognomy and field recognition of dominant broadleaved forest conditions. They were therefore treated as operational vegetation-zone strata rather than sharply bounded floristic communities.

Each plot measured 20 $\times$ 20 m (400 m^2), a size widely used in forest community studies to characterise tree assemblages and stand structure, with a nested 2 $\times$ 2 m subplot used to survey herbaceous vegetation. Fieldwork was undertaken between March and November to encompass seasonal variability in understorey communities while avoiding periods of peak monsoon inaccessibility. Within each broad vegetation zone, plot locations were positioned using random bearings and distances from access transects, subject to terrain safety, access and intact-forest constraints. The design therefore supports inference about unmanaged forest stands sampled along this gradient, but not a fully spatially balanced test of all possible microhabitats.

To assess potential sampling bias, plot representativeness was summarised by elevation band, forest type, slope class and aspect class; the dependence between forest type and elevation band was quantified with a chi-square test and Cramer's V; and the spatial arrangement of plots was evaluated using nearest-neighbour distances and the Clark-Evans aggregation index (Appendix S12).

Vegetation data collection

All free-standing woody stems with diameters at breast height (DBH) > 10 cm and height > 1.3 m was recorded as trees. DBH was measured using diameter tapes, and total height was estimated by using hypsometers. Shrubs were assessed for presence, height,

and lateral spread, while herbaceous species were quantified within subplots using percentage cover, frequency, and maximum height.

Regeneration comprised all woody individuals below the canopy layer that had not yet reached the tree stratum. Seedlings were defined as individuals ≤ 1.3 m in height, irrespective of stem diameter, whereas saplings were defined as individuals > 1.3 m in height with DBH ≤ 10 cm. Species exceeding 10 cm DBH were assigned to the tree layer. Regeneration richness therefore integrated both seedling and sapling cohorts, representing early demographic stages that are particularly sensitive to microsite conditions and neighborhood effects. All plant material included in this study was collected from natural, unmanaged forest stands in Sarpang Dzongkhag, Bhutan, along an elevational gradient ranging from approximately 260 m to 1,964 m a.s.l. No cultivated, planted, or externally sourced plant material was included. Exact geographic coordinates of all sampling plots were recorded using handheld GPS units and are provided in the supplementary dataset.

Voucher specimens representing the wild plant species collected during fieldwork were prepared and deposited in the herbarium collection maintained by the Divisional Forest Office, Sarpang, Department of Forests and Park Services, Royal Government of Bhutan. Each specimen was assigned a unique field collection code. Formal herbarium accession or deposition numbers are not yet available because specimen cataloguing is ongoing. Species identification was carried out by Mr. Ngawang Gyeltshen using the *Flora of Bhutan* and authenticated herbarium reference specimens. Taxonomic nomenclature was verified using World Flora Online. Where necessary, uncertain identifications were confirmed through comparison with authenticated herbarium specimens.

Structural attributes, including basal area, diameter-at-breast-height class distributions, and height-class distributions, were derived to characterize stand structure and regeneration dynamics across forest types and elevational gradients.

Environmental variables

Plot coordinates were recorded in the field using handheld Global Positioning System receivers. Terrain predictors used in the final analyses were extracted from an ASTER

digital elevation model in DrukRef 03/Bhutan National Grid at 29 m resolution, from which elevation, slope and aspect were derived. Aspect was transformed to eastness and northness for modelling. Historical bioclimatic variables were extracted from 250 m surfaces for 1986–2015, including annual mean temperature and annual precipitation. To establish broad water-availability indices, we extracted reference evapotranspiration (ETo) and water balance metrics using Google Earth Engine

To add finer-scale remotely sensed habitat information, Sentinel-2 surface reflectance was used to calculate the Normalized Difference Vegetation Index (NDVI) at 10 m and the Normalized Difference Moisture Index (NDMI) and Normalized Burn Ratio (NBR) from 20 m shortwave-infrared bands. These indices were treated as vegetation-condition and moisture/disturbance proxies, not as direct measurements of canopy openness or disturbance. Climatic layers were interpreted as macroclimatic context, not as measurements of plot-scale microclimate. Canopy openness, soils, disturbance history and microsite moisture were not measured and could not be included directly in the explanatory models (Table 2).

Because several terrain and climatic predictors are strongly intercorrelated (Appendix S19), fitted environmental variables are interpreted throughout as correlates of community structure rather than as independent causal drivers.

Table 2. Environmental predictors used in analyses and their interpretation. Climate variables are macroclimatic or broad water-availability predictors, whereas Sentinel-2 variables are remotely sensed vegetation-condition proxies.

Predictor group	Resolution	Interpretation
ASTER elevation, slope, aspect	29 m	Terrain context; aspect transformed to eastness and northness for modelling
Historical bioclim variables	250 m	Macroclimatic context for 1986–2015; summarized with principal components where needed
Sentinel-2 NDVI	10 m	Vegetation greenness proxy
Sentinel-2 NDMI and NBR	20 m	Vegetation moisture/disturbance-condition proxies derived from SWIR bands

Data preparation

Field records were organized into structured datasets to compute species abundance, frequency, basal area, and importance value index (IVI) separately for each vegetation stratum. Alpha diversity was quantified using the Shannon–Wiener index (H′), Simpson's diversity index (1 − D), and Pielou's evenness (J) for each forest type and vegetation layer [15, 16, 17].

Because the tree (400 $m^{2)}$ and herb (4 m^2) sampling areas differ, among-stratum diversity was additionally compared using incidence-based richness estimators (Chao2 and jackknife) and coverage-standardised richness, and differences among strata were tested with Kruskal-Wallis tests (Appendix S16). The unimodal versus monotonic form of each richness-elevation relationship was tested with generalised additive models.

Multivariate community analyses

Floristic variation among plots was examined using non-metric multidimensional scaling (NMDS) based on Bray–Curtis dissimilarities [18, 19]. NMDS was used to visualise broad compositional structure and overlap among groups, not as a formal test of separation. Environmental associations with NMDS structure were evaluated using vector fitting with 999 permutations. This analysis was interpreted as an exploratory summary of association strength, not as a causal model of community assembly.

Because near-zero NMDS stress can arise from sparse, heavily tied dissimilarity matrices rather than from strong low-dimensional structure, ordination adequacy was assessed using Shepard diagrams, the distribution of Bray-Curtis dissimilarities, counts of tied distances and species-poor plots, and multiple random starts. Configurations were cross-checked against Jaccard (presence-absence) and Hellinger-transformed ordinations (Appendix S13).

Community differences among forest types were tested using permutational multivariate analysis of variance (PERMANOVA) [20]. NMDS was computed with two dimensions ($k = 2$) and a fixed random seed (42) to ensure reproducibility. Environmental vectors were fitted using permutation tests with 999 randomizations. PERMANOVA was conducted with 999 permutations, and homogeneity of multivariate dispersions was evaluated using permutation tests on distance-to-centroid values [21].

Where the overall test was significant, pairwise PERMANOVA among forest types was computed with Benjamini-Hochberg-adjusted p-values, together with pairwise dispersion tests, between-centroid distances and within-group distance-to-centroid values (Appendix S14).

To support valid interpretation, PERMANOVA results were interpreted jointly with effect sizes (R^2) and dispersion tests. In strata with heterogeneous dispersion, PERMANOVA was interpreted as reflecting combined centroid and spread effects rather than pure location shifts [21]. NMDS was used for pattern visualization rather than as a stand-alone test of group separation.

To reduce redundancy among climate variables, the 19 bioclimatic variables were summarized using principal components before ordination-based environmental fitting and predictive modelling. Because preliminary analyses showed low constrained inertia, constrained ordination with variance-inflation factor (VIF) screening was retained only as a supplementary diagnostic. Indicator species analysis was applied to identify taxa significantly associated with low-, mid- and high-elevation groups using the indicator value method, with significance assessed at $\alpha = 0.05$ and 999 permutations [22]. These groups were used to evaluate whether broad elevational zonation produced clear floristic indicators, complementing the forest-type comparisons used in PERMANOVA.

Spatial dependence was examined explicitly using Moran's I on plot-level diversity, a Mantel test and Mantel correlogram relating community dissimilarity to geographic distance, and distance-based Moran's eigenvector maps (dbMEM) within a variation-partitioning framework that separated forest-type from purely spatial components of community variation (Appendix S15). Regeneration models were additionally evaluated with spatially blocked cross-validation (Appendix S18).

Regeneration dynamics and predictive modelling

Seedling and sapling data were analysed to quantify regeneration patterns and spatial structure across forest types and elevations. Environmental correlates of regeneration richness were modelled using random forest and extreme gradient boosting algorithms, both of which are widely applied for non-linear ecological prediction problems [23, 24].

Predictors included terrain variables, macroclimatic variables, Sentinel-2 vegetation-condition indices and tree/shrub diversity metrics. Random forest models were fitted with 500 trees, and predictor importance was quantified as the percentage increase in mean squared error following permutation. Gradient boosting models were parameterized with 100 boosting rounds, a maximum tree depth of four and a learning rate of 0.1, and predictor influence was assessed using relative gain values.

In addition to richness, regeneration density and the seedling-to-sapling ratio were summarised by forest type; adult-juvenile correspondence was assessed for the dominant canopy taxa *Schima wallichii, Castanopsis indica and Alnus nepalensis*; and adult versus regeneration composition was compared using Mantel and Procrustes tests (Appendix S18). An XGBoost hyperparameter search was used to confirm that the weak predictive performance was not an artefact of default settings.

Model performance was evaluated using five-fold cross-validation with a fixed random seed (42) to ensure reproducibility. Predictive accuracy was summarized using cross-validated error statistics and explained variance, and response curves were inspected to verify ecologically plausible relationships between regeneration patterns and key climatic or topographic predictors.

Because predictive performance was weak, variable-importance results were interpreted only as exploratory correlates. They were not used to infer causal drivers of regeneration.

Software environment

All analyses were conducted in R. Community-ecology workflows, including ordination, dissimilarity-based testing and indicator-species analysis, were implemented primarily using functions described in the ecological literature underlying widely adopted multivariate methods [20, 22, 25]. Results from selected multivariate classifications and indicator-species procedures were independently verified in PC-ORD version 5 to confirm analytical consistency. Microsoft Excel was used for preliminary data curation and quality-control screening prior to statistical analysis.

Reproducibility and data stewardship

Reproducibility was ensured through systematic archiving of raw field observations, spatial environmental layers, curated analysis tables and derived outputs. All distance measures, ordination settings, permutation schemes and machine-learning hyperparameters were documented in metadata files accompanying the archived datasets. Intermediate products, including ordination scores, regeneration prediction surfaces, cross-validation partitions and variable-importance tables were retained to permit independent replication of all analytical steps. Data and code supporting the results of this study are available in the project repository: https://github.com/wangdiues/Vegetation_Ecology.

Results

Community composition and ordination

Non-metric multidimensional scaling (NMDS) showed broad overlap among forest types across strata, with only weak visual differentiation (Figure 2). Stress values were uniformly low (trees = 0.0001; shrubs = 0.070; herbs = 0.001; regeneration = 0.0006), indicating a low-stress two-dimensional rank-order fit of Bray–Curtis dissimilarities rather than a perfect ordination (Appendix S3). Multiple dimensional solutions were explored, and stress values remained consistently low, confirming ordination stability.

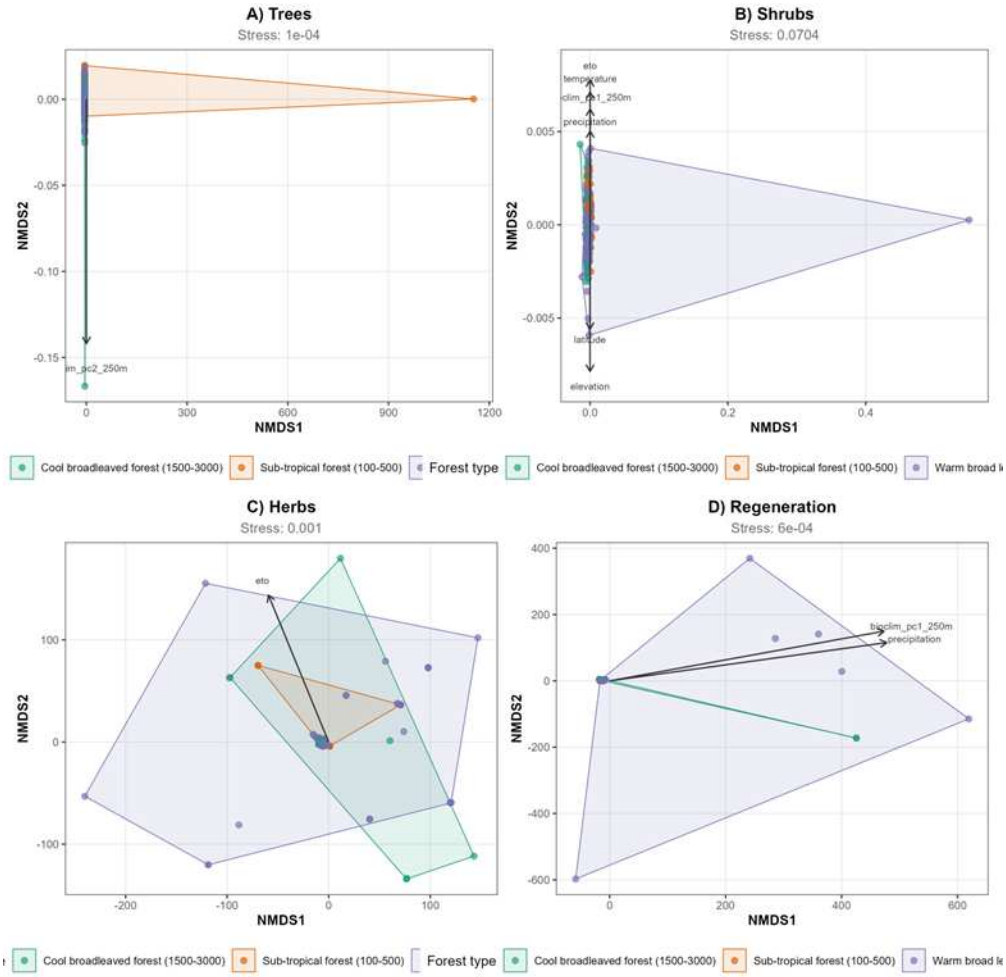


Fig. 2: Non-metric multidimensional scaling (NMDS) ordinations of vegetation plots for four strata: (A) trees, (B) shrubs, (C) herbs and (D) regeneration. Points represent sample plots coloured by forest type; polygons delineate convex hulls for each forest type and show broad overlap rather than discrete group separation. Red arrows indicate fitted environmental vectors significant at $p < 0.05$ based on 999 permutations. Stress values are shown for each ordination

The near-zero stress values for trees, herbs, and regeneration may arise partly from matrix properties that ease low-dimensional rank fitting (for example, high tied dissimilarities and sparse compositional structure), although such values should be treated cautiously in ecological community data. Ordinations were therefore interpreted conservatively as broad summaries rather than precise geometric distances [19, 26].

NMDS patterns were cross-checked against PERMANOVA, dispersion tests, and constrained ordination instead of relying on visual separation alone.

Diagnostic analysis confirmed that the low stress values reflected matrix properties rather than strong compositional structure: tied dissimilarities accounted for 98.1-99.3% of plot pairs in the trees, herbs and regeneration matrices and 275 plots contained at most one species, while Jaccard and Hellinger-transformed ordinations produced comparably low stress (maximum 0.07; Appendix S13). The ordinations are therefore reported as broad rank-order summaries and not as precise geometric representations.

Shrub and herb assemblages displayed the greatest dispersion in ordination space and were the only strata with several significant fitted environmental vectors (Figure 2B–C). Tree assemblages were comparatively compressed along NMDS axes (Figure 2A), while regeneration plots showed broad scatter (Figure 2D). These patterns indicate high within-stratum heterogeneity and limited visual support for discrete forest-type clusters.

These stratified responses mirror documented altitudinal forest-type transitions along dry valley slopes [27].

Community differences among forest types

Community composition differed significantly among forest types across all vegetation strata, but effect sizes were small (Table 3). Tree assemblages exhibited a modest difference (PERMANOVA $R^2 = 0.031$, $F = 3.38$, $p = 0.001$), with homogeneous dispersions (betadisper $p = 0.291$), indicating that this result is consistent with a small centroid shift among broad forest-type classes (Appendix S2).

Table 3. Results of permutational multivariate analysis of variance (PERMANOVA) and tests of multivariate dispersion for vegetation strata across forest types. p-values are based on 999 permutations.

Stratum	Sites	Groups	R^2	F	p	Disp. F	Disp. p	Homogeneous
Trees	216	3	0.031	3.38	0.001	1.24	0.291	Yes
Shrubs	194	3	0.050	5.05	0.001	12.93	0.001	No
Herbs	206	3	0.017	1.70	0.001	15.05	0.001	No
Regeneration	205	3	0.021	2.14	0.001	1.17	0.319	Yes

Shrub communities had the largest PERMANOVA effect size among strata ($R^2 = 0.050$, $F = 5.05$, $p = 0.001$), although dispersions were heterogeneous ($p = 0.001$), suggesting that both location and spread contributed to observed differences. Herb and regeneration assemblages also differed significantly ($R^2 = 0.017$ and 0.021, respectively; both $p = 0.001$), but only the herb stratum showed heterogeneous dispersion (Table 3).

In strata with heterogeneous dispersion, PERMANOVA reflects combined centroid and spread effects. Accordingly, shrub and herb results were interpreted as mixed location–dispersion signals, whereas homogeneous dispersions for trees and regeneration indicate that their compositional differences more closely represent centroid shifts. The results therefore support weak broad-scale compositional structure, not sharply separated forest-type communities.

Pairwise comparisons among forest types (Appendix S14) remained significant after Benjamini-Hochberg correction in 12 of 12 tested pairs, but effect sizes were uniformly small (R2 <= 0.089) and between-centroid distances were small relative to within-group dispersion, reinforcing the interpretation of weak rather than discrete forest-type structure.

Forest type was associated with elevation band (*Cramer's V = 0.26*), confirming the expected overlap and partial confounding of broad zones along the gradient. Plots showed a clustered spatial arrangement (*Clark-Evans R = 0.37*; median nearest-neighbour distance 0.44 km). Community-geographic distance-decay was weak (Mantel r = 0.004-0.088), and variation partitioning attributed only a small, largely shared fraction of community variation to pure space (<= 0.009) versus pure forest type (<= 0.032; Appendix S15). Spatial structure therefore does not overturn the weak-gradient interpretation.

Species–environment relationships

Vector fitting against NMDS ordinations showed weak and stratum-specific environmental associations (Table 4; Figure 3). Tree composition was associated only weakly with the second bioclimatic principal component ($R^2 = 0.027$, $p = 0.036$). Herb composition was weakly associated with evapotranspiration ($R^2 = 0.039$, $p = 0.018$), and regeneration

showed weak associations with the first bioclimatic principal component and precipitation ($R^2 \leq 0.028$, $p \leq 0.047$).

Table 4. Exploratory environmental vector-fitting results for selected variables fitted to NMDS ordinations. Values are permutation-based vector-fit R^2 statistics and should be interpreted as association strength, not causal explanatory power.

Stratum	Variable	Interpretation	R^2	p
Trees	Bioclim PC2	macroclimate gradient	0.027	0.036
Shrubs	Elevation	terrain gradient	0.130	0.001
Shrubs	ETo	water-availability index	0.127	0.001
Shrubs	Temperature	macroclimate	0.106	0.001
Shrubs	Bioclim PC1	macroclimate gradient	0.081	0.001
Shrubs	Precipitation	macroclimate	0.053	0.001
Herbs	ETo	water-availability index	0.039	0.018
Regeneration	Bioclim PC1	macroclimate gradient	0.028	0.044
Regeneration	Precipitation	macroclimate	0.027	0.047

Shrub assemblages showed the clearest environmental signal, with significant fitted vectors for elevation ($R^2 = 0.130$), evapotranspiration ($R^2 = 0.127$), temperature ($R^2 = 0.106$), the first bioclimatic principal component ($R^2 = 0.081$), latitude ($R^2 = 0.066$), and precipitation ($R^2 = 0.053$). Even in shrubs, these associations explained only a modest fraction of ordination structure and should be interpreted as exploratory correlations rather than direct evidence of environmental control.

The weak fitted-vector results indicate that the measured terrain, 250 m macroclimatic variables and Sentinel-2 vegetation-condition indices captured only a small component of community variation. This limits inference about species–environment relationships and points to substantial unmeasured microsite drivers operating within broader altitudinal transitions [3, 27].

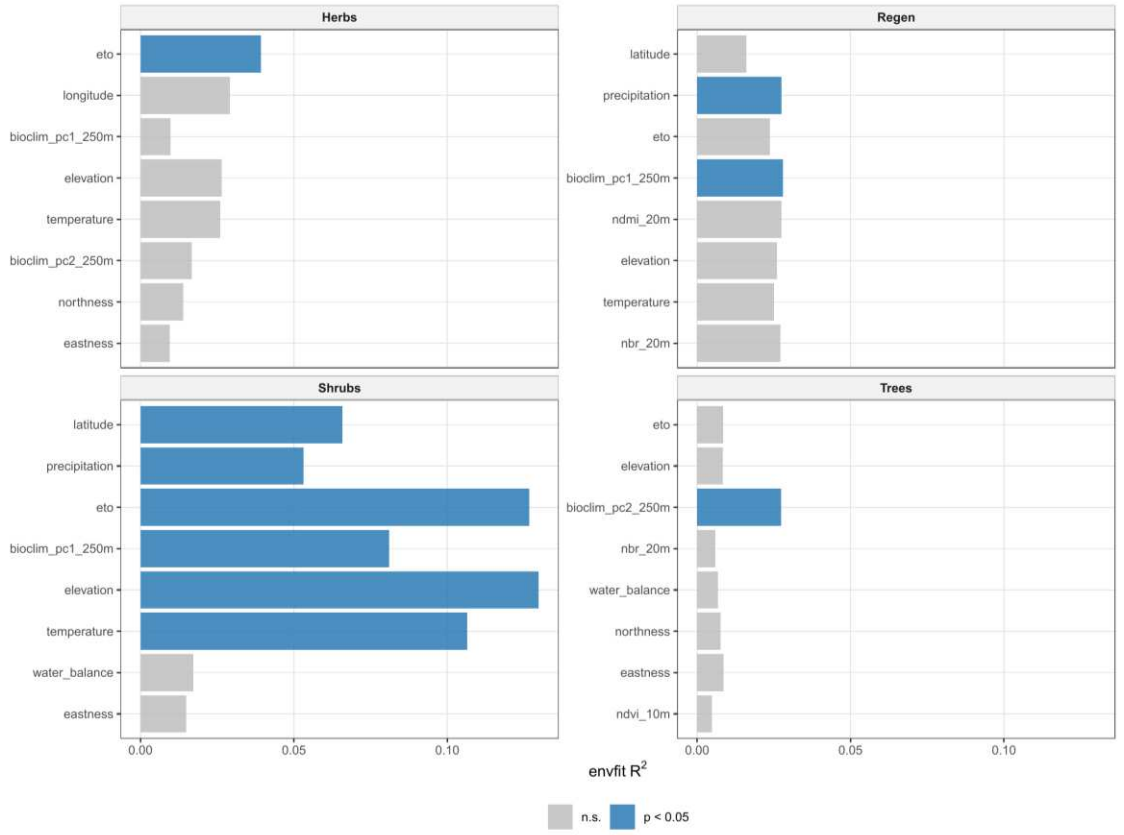


Fig. 3 Exploratory environmental vector-fitting results for NMDS ordinations by vegetation stratum. Bars show vector-fit R^2 values for the strongest fitted variables in each stratum; blue bars indicate $p < 0.05$ based on 999 permutations and grey bars indicate non-significant associations. These values summarize association strength and are not interpreted as causal drivers.

Diversity patterns

Alpha diversity differed markedly among vegetation strata (Table 5; Figure 4). Trees exhibited the highest mean species richness (5.30 ± 2.57 species per plot) and Shannon diversity (1.391 ± 0.595), followed by shrubs (richness = 4.77 ± 3.43; Shannon = 1.063 ± 0.706), regeneration (richness = 2.00 ± 1.26; Shannon = 0.443 ± 0.499) and herbs (richness = 1.79 ± 1.32; Shannon = 0.325 ± 0.451). Pielou's evenness was highest for trees (0.903) and lowest for herbs (0.804).

Table 5. Mean alpha-diversity metrics by vegetation stratum. Values are mean ± standard deviation.

Stratum	Plots	Richness	Shannon	Simpson	Evenness
Trees	221	5.30 ± 2.57	1.391 ± 0.595	0.665	0.903
Shrubs	198	4.77 ± 3.43	1.063 ± 0.706	0.524	0.832
Regeneration	209	2.00 ± 1.26	0.443 ± 0.499	0.264	0.839
Herbs	210	1.79 ± 1.32	0.325 ± 0.451	0.193	0.804

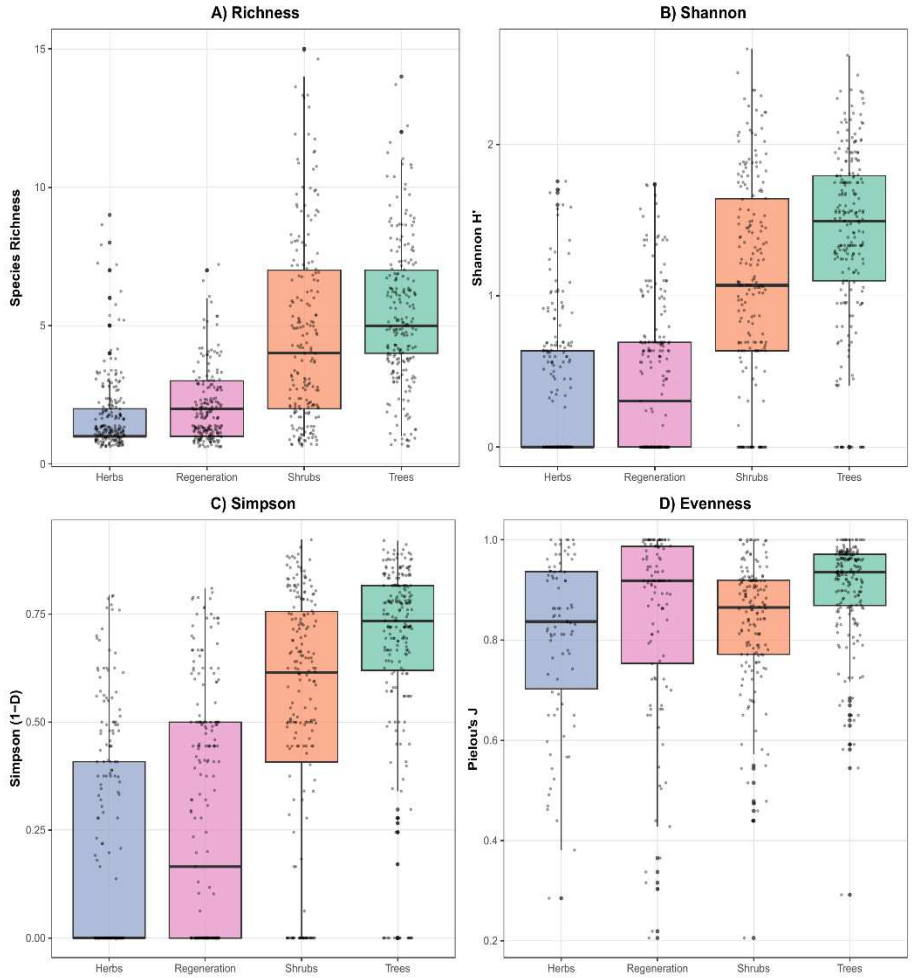


Fig. 4 Alpha-diversity metrics across vegetation strata. Panels show (A) species richness, (B) Shannon diversity (H'), (C) Simpson diversity ($1 - D$) and (D) Pielou's evenness (J). Boxplots depict medians and interquartile ranges; points represent individual plots

Species accumulation curves approached asymptotes most rapidly for shrubs and regeneration, whereas trees and herbs continued to increase steadily with sampling effort,

indicating incomplete saturation of the regional species pool in these strata (Figure 5). Such patterns are consistent with documented altitudinal zonation of Bhutanese dry-valley forests and reported regeneration dynamics of dominant trees [27, 28]. Continued species accumulation in tree and herb strata suggests that additional sampling would likely yield further taxa, particularly rare canopy species and ephemeral understorey herbs, indicating substantial compositional turnover.

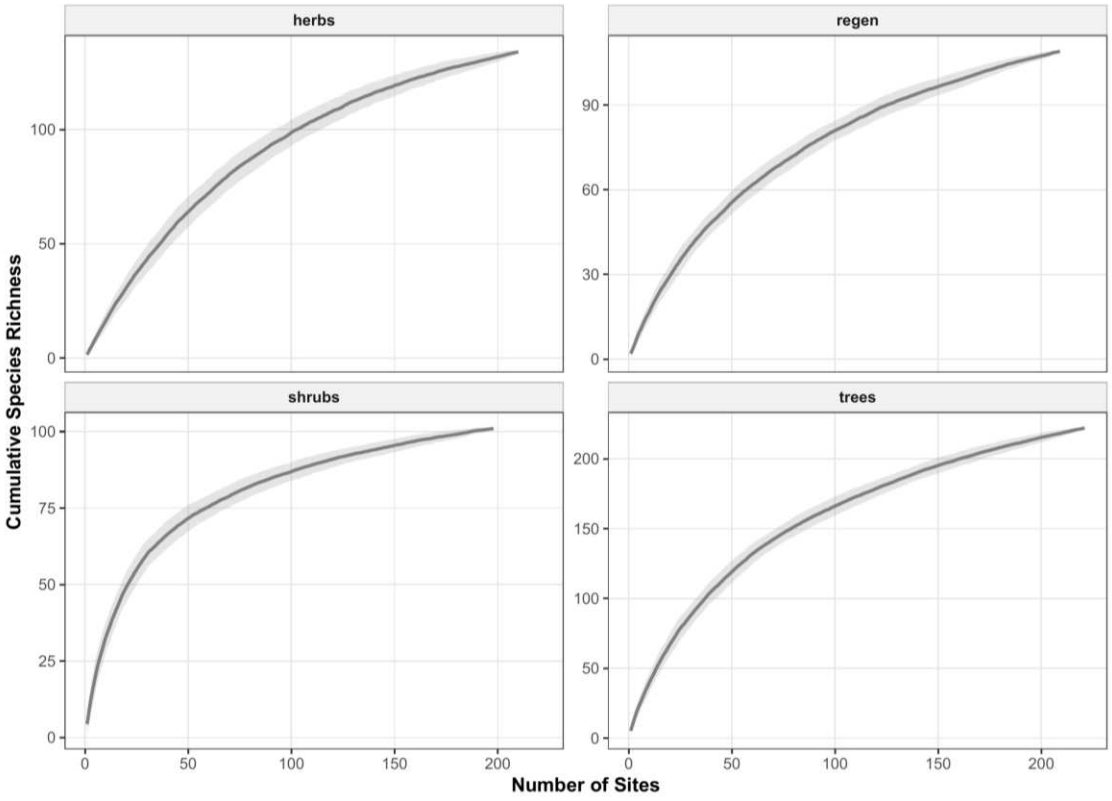


Fig. 5 Sample-based species accumulation curves for herbs, regeneration, shrubs and trees. Solid lines indicate mean richness across random permutations; shaded envelopes show 95% confidence intervals

Species richness varied along the elevational gradient for all strata (Figure 6). Shrubs displayed a unimodal mid-elevation peak, whereas trees showed comparatively stable richness across the sampled range, and herbs and regeneration exhibited weaker, irregular trends. These responses align with Himalayan studies reporting mid-elevation richness peaks and altitudinal zonation in woody layers [3, 27, 29].

Beta diversity (Whittaker's β) was highest for herbs (75.04), followed by regeneration (54.63), trees (41.86) and shrubs (23.89), indicating pronounced species turnover in the herb layer relative to the regional species pool.

Alpha diversity differed significantly among strata (Kruskal-Wallis chi-square = 311.2, p = 3.79e-67). Incidence-based estimators indicated incomplete inventories, particularly for species-rich strata (trees: 222 observed vs Chao2 ~ 292, 76.1% detected; herbs: 134 vs 149, 89.8% detected; Appendix S16). Beta-diversity partitioning showed that turnover, rather than nestedness, dominated compositional differences (herb-layer turnover 100% of total beta; Appendix S17), confirming genuine species replacement along the gradient rather than a sampling artefact.

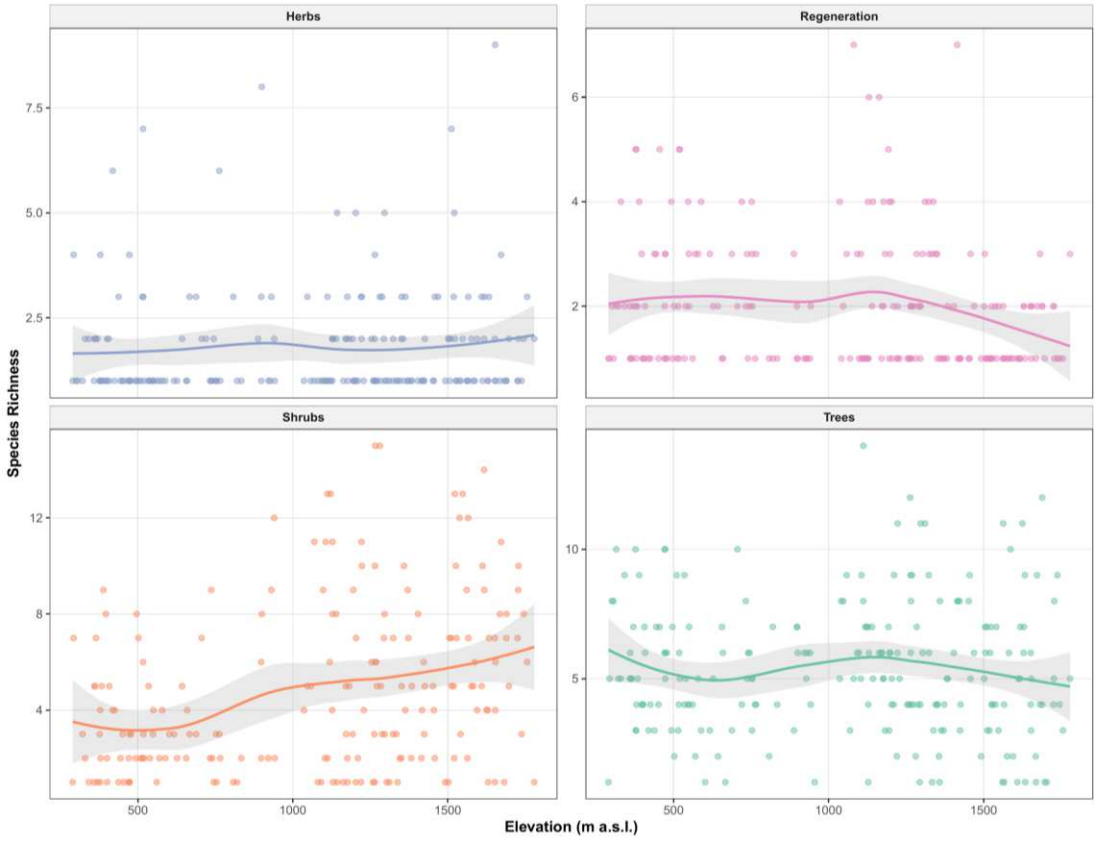


Fig. 6. Variation in plot-level species richness along the elevational gradient for herbs, regeneration, shrubs and trees. Curves represent locally weighted regression fits with 95% confidence band

Indicator species

Indicator-species analysis detected significant associations with elevational classes across all strata, but the proportion of significant indicators was low (Table 6). Shrubs contained the largest proportion of indicator taxa (14 of 101 species; 13.9%), followed by regeneration (9 of 109; 8.3%), herbs (10 of 134; 7.5%) and trees (12 of 221; 5.4%) (Appendix S4).

After Benjamini-Hochberg correction, the number of significant indicators fell from 45 to 3 across the four strata (Appendix S4), underscoring that broad elevational and forest-zone classes carry only a limited, FDR-robust floristic indicator signal.

Table 6. Summary of indicator-species analysis by stratum. Groups were defined by elevational class (low, mid and high). Significance was assessed using 999 permutations at $\alpha = 0.05$.

Stratum	Sites	Species	Groups	Significant indicators	% Significant
Trees	216	221	3	12	5.4
Shrubs	194	101	3	14	13.9
Herbs	206	134	3	10	7.5
Regeneration	205	109	3	9	8.3

Among canopy trees, *Schima wallichii* and *Castanopsis indica* were characteristic of low- to mid-elevation forests, whereas *Alnus nepalensis* was associated with higher elevations. However, only 5.4% of tree species were significant indicators, showing that the elevational and forest-zone classes should not be treated as sharply defined canopy communities. The stronger shrub signal suggests somewhat greater understorey differentiation, but still within broadly overlapping forest zones.

Regeneration modelling

Random forest produced weak predictive performance for regeneration richness (Table 7; Figure 7), with cross-validated $R^2 = 0.081 \pm 0.057$ and RMSE 1.202 ± 0.156. Gradient boosting performed worse than a mean predictor ($R^2 = -0.035 \pm 0.096$; RMSE = 1.274 ± 0.166) (Appendix S5).

Table 7. Cross-validated performance of machine-learning models predicting regeneration richness across five folds. R^2 denotes coefficient of determination; RMSE, root mean square error; MAE, mean absolute error. Full fold-level output is provided in Appendix S7.

Model	Predictors	Obs.	CV R^2	CV RMSE	CV MAE
Random Forest	17	192	0.081 ± 0.057	1.202 ± 0.156	0.935
XGBoost	17	192	-0.035 ± 0.096	1.274 ± 0.166	0.951

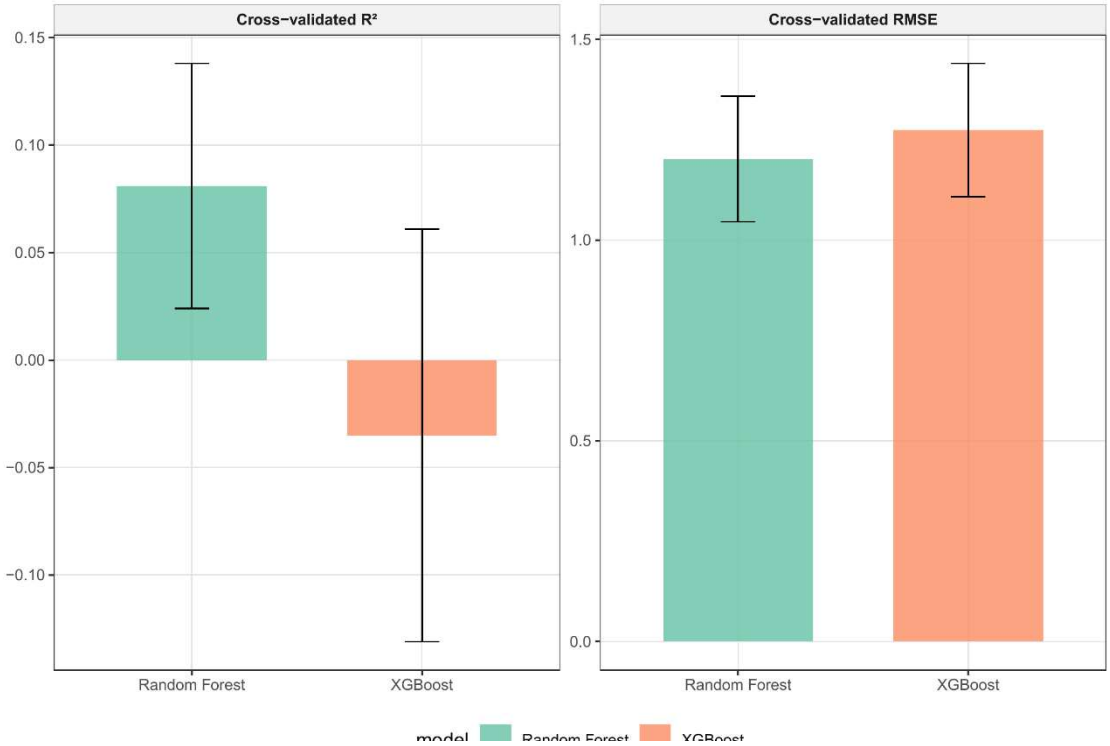


Fig. 7 Cross-validated performance of random forest and XGBoost models predicting regeneration richness. Bars show mean five-fold cross-validated R^2 and RMSE with standard deviations. Low or negative R^2 values indicate weak predictive performance from the available terrain, macroclimatic, remotely sensed and stand-level predictors

Fold-level diagnostics (Appendix S7) showed moderate variability in RF performance and unstable XGBoost performance, reinforcing that the available terrain, macroclimatic, remotely sensed and stand-level predictors provide only weak predictability of recruitment.

Variable-importance rankings suggested that shrub-layer diversity attributes, evapotranspiration, elevation and temperature contributed most to random-forest prediction (Appendix S6). Because the model explained little variance, these rankings are reported only to document exploratory correlations and should not be interpreted as evidence that these variables are dominant drivers of regeneration richness.

Regeneration density and stage structure varied among forest types, with an overall seedling-to-sapling ratio of 15.6 (Appendix S18), indicating active but bottlenecked recruitment. Adult-juvenile correspondence for the dominant canopy taxa was incomplete: regeneration co-occurred with conspecific adults in a fraction of occupied plots (Schima wallichii 0.25, Castanopsis indica 0.35, Alnus nepalensis 0). Adult and regeneration composition were only moderately concordant (Mantel r = 0.08, Procrustes correlation = 0.25, p = 0.001), and spatially blocked cross-validation did not improve regeneration-richness prediction (random-forest spatial CV R2 = -0.036), consistent with strong unmeasured microsite control of recruitment.

Given the low explanatory power, these models are not suitable for deterministic prediction or causal ecological inference. Important drivers—such as canopy openness, disturbance history, soil properties or fine-scale moisture regimes—were not captured by the available predictors, which is a design constraint for interpreting regeneration patterns in this study [30].

Discussion

Stratified community turnover along elevational gradients

The analyses indicate weak but detectable broad-scale community structure, with understorey assemblages (shrubs and herbs) more variable than canopy trees. This pattern is consistent with observations from other Himalayan forests where understorey communities may respond to local canopy, soil and disturbance conditions, whereas long-

lived canopy trees integrate conditions over broader temporal and spatial scales [3, 27]. Forest-type effects were limited in magnitude: PERMANOVA explained only 1.7–5.0% of compositional variation ($R^2 = 0.017 - 0.050$, $p = 0.001$). These low effect sizes, overlapping NMDS hulls and low indicator-species percentages show that the sampled forest types are broad vegetation-zone classes rather than discrete communities. Interpretation also depends on dispersion structure: in shrub and herb strata, heterogeneous dispersions indicate mixed location–spread effects, whereas tree and regeneration strata with homogeneous dispersions indicate clearer but still small centroid shifts among forest types.

Environmental correlates and scale of control

Environmental vector fitting showed that measured terrain, macroclimatic and remotely sensed vegetation proxies were weakly associated with community structure, except for a somewhat clearer shrub-layer signal. The bioclimatic layers represent macroclimatic context and cannot resolve canopy gaps, soils, disturbance history, cold-air drainage, or plot-scale moisture. Sentinel-2 indices added finer-scale vegetation-condition information, but they are proxies rather than direct measurements of canopy openness or disturbance. Consequently, the analyses support only the conclusion that broad environmental gradients and remotely sensed habitat conditions are weakly detectable. They do not allow strong inference about the fine-scale mechanisms that structure communities in steep Himalayan terrain.

Diversity patterns and beta-diversity

Marked differences in alpha diversity among strata further highlight the vertically structured nature of Himalayan broadleaved forests. Tree layers in study maintained high species richness and evenness across the gradient, reflecting the coexistence of multiple canopy dominants in these unmanaged stands – a feature repeatedly documented in Bhutan's warm and cool broadleaved forests [27, 31]. Shrub richness peaked at mid-elevations, consistent with many Himalayan studies that find unimodal diversity curves in the woody understorey around intermediate altitudes where moisture, temperature, and structural complexity are jointly optimized [3]. By contrast, herbs and regeneration

showed weak or irregular elevational trends in richness, coupled with elevated beta diversity relative to alpha diversity in the herb layer. This pattern implies that small-scale environmental mosaics, rather than smooth elevational forcing, structure ground-layer communities. High herb-layer beta diversity is widely reported in Himalayan forests and alpine systems, where aspect, litter depth, and gap dynamics generate sharp species turnover over short distances [3]. The fact that species accumulation curves did not fully asymptote for trees and herbs even after 221 plots suggests that a considerable portion of the regional species pool remained unsampled in these strata, consistent with substantial compositional turnover.

Indicator species and floristic transitions

Indicator species analysis revealed fewer significant canopy indicators than shrub indicators. This weak tree specificity is important because forest types were used as sampling strata: the result indicates that these strata are not sharply defined floristic communities at canopy level. Instead, the patterns align with Himalayan zonation as overlapping distributions rather than discrete boundaries [27, 32]. The shrub signal may reflect stronger understorey turnover, but this remains exploratory because unmeasured microsite variables were unavailable.

Regeneration dynamics and model interpretation

Machine-learning results were retained as exploratory prediction diagnostics, not as tests of regeneration mechanisms. The best model (random forest) explained only about 8% of cross-validated variance ($R^2 \approx 0.08$), and gradient boosting performed worse than a mean predictor. Variable-importance rankings therefore cannot establish that shrub-layer structure controls regeneration; they only identify variables that contributed to a weak predictive model. The low predictive power is consistent with missing microsite drivers, such as canopy openness, soil properties, disturbance history and fine-scale moisture [30]. Accordingly, the main inference is that regeneration richness was poorly predicted by the available macro-environmental, remotely sensed and stand-level variables.

Limitations and novelty of the study

Several design constraints limit inference. First, climate variables were derived from 250 m resolution bioclim rasters; these layers cannot represent plot-level microclimate in steep terrain. Second, Sentinel-2 indices improved the representation of fine-scale vegetation condition but did not replace direct measurements of canopy openness, soil properties, disturbance history or microsite moisture. Third, heterogeneous dispersion in shrub and herb strata means that part of PERMANOVA significance reflects spread differences, not only centroid shifts. Fourth, regeneration measurements represent a temporal snapshot and may miss episodic establishment dynamics of slow-growing montane trees. Spatial autocorrelation was examined explicitly using Moran's I, Mantel correlograms and dbMEM variation partitioning (Appendix S15) and had limited effect on fine-scale inference. Even with these constraints, this study provides one of few multi-stratum baseline assessments in the eastern Himalaya. The main contribution is to quantify the limited explanatory power of broad forest-type, macro-environmental and remotely sensed habitat predictors, while identifying where future plot-level measurements are needed.

Building on this baseline, the priority next measurements are: (i) canopy openness from hemispherical photography or densiometer readings; (ii) soil pH, organic carbon, total nitrogen, available phosphorus and potassium, texture and moisture; (iii) in-situ microclimate logging of temperature and humidity; (iv) standardised disturbance scoring for grazing, trails, lopping, canopy gaps, landslides and fire; and (v) permanent seedling and sapling plots to estimate survival and recruitment directly. These variables target the unmeasured microsite drivers identified here as the main source of unexplained compositional and regeneration variation.

For management, the gradational rather than discrete structure of these forests implies that conservation planning should treat the subtropical-to-cool broadleaved transition as a continuum rather than as sharply bounded types. Because lowland dominants such as *Schima wallichii* and *Castanopsis indica* and the higher-elevation *Alnus nepalensis* showed incomplete adult-to-regeneration correspondence, maintaining connectivity across the elevational gradient is likely to be important for allowing species to track

suitable conditions under continued warming, supporting the role of these unmanaged forests as an ecological linkage.

Conclusions

Across the study area, broad forest-type classes showed statistically detectable but limited compositional structure ($R^2 = 0.017–0.050$). Understorey strata showed greater turnover and higher dispersion than canopy trees, but shrub and herb PERMANOVA results were partly confounded by heterogeneous dispersion.

Environmental vector fitting showed mostly weak terrain, macroclimatic and remotely sensed habitat associations, with the clearest but still modest signal in shrubs. Regeneration was only weakly predictable from the available predictors (random forest CV $R^2 \approx 0.08$; gradient boosting $R^2 < 0$). Variable-importance rankings should be read as exploratory diagnostics from a weak model, not as causal evidence.

This study provides a baseline for long-term monitoring of unmanaged broadleaved forests in the sampled elevation range. Inference should remain restricted to this 260–1 964 m gradient and to broad forest-zone patterns unless future work adds plot-level soil, canopy, disturbance and microclimate measurements.

Declarations

Funding

This research received no specific external funding. Institutional support was provided through routine government research and conservation monitoring programs of the Department of Forests and Park Services, Royal Government of Bhutan.

Competing interests

The authors declare no competing financial or non-financial interests.

Author contributions

W: Conceptualization; methodology; software; formal analysis; data curation; visualization; writing—original draft; writing—review and editing. RS: Methodology;

software; formal analysis; data curation; visualization; supervision; conceptual refinement; writing—review and editing. KC: Field coordination; data validation; writing—review and editing.

Data availability

The data supporting this study are provided as Supplementary Information and in the project repository at https://github.com/wangdiues/Vegetation_Ecology. Processed community matrices, environmental predictors, statistical outputs and workflow files are archived to support transparency and reproducibility.

Ethics approval and consent to participate

Not applicable. This study did not involve human participants or animal experimentation. Voucher specimens of all wild plant species collected during fieldwork were prepared and deposited in the herbarium at the Divisional Forest Office, Sarpang, Bhutan. Each specimen was assigned a unique field collection code and will be linked to an official accession number upon completion of the herbarium cataloguing process.

Plant identification was carried out by Mr. Ngawang Gyeltshen using the published volumes of the Flora of Bhutan (Grierson & Long, 1983–2001), and taxonomic nomenclature was verified against the World Flora Online database, accessed on 6 February 2026. Where necessary, identifications were confirmed through comparison with authenticated herbarium specimens. All plant collection activities were conducted in accordance with the Forest and Nature Conservation Act of Bhutan and its applicable rules and regulations, under research permission granted by the Department of Forests and Park Services, Royal Government of Bhutan.

Consent to publish

Not applicable. This manuscript does not contain any human participant data, images, or personal information requiring consent for publication.

Acknowledgements

The authors acknowledge the Department of Forests and Park Services, Royal Government of Bhutan, for facilitating access to spatial datasets and supporting national-

scale vegetation assessments. The authors further thank colleagues involved in protected-area management and ecological corridor planning whose long-term conservation efforts underpin the spatial and ecological datasets used in this study.

References

[1] M. V. Lomolino. 2001. Elevation gradients of species density: Historical and prospective views. *Global Ecology and Biogeography* 10, 3–13. https://doi.org/10.1046/j.1466-822X.2001.00229.x.

[2] C. Rahbek. 2005. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters* 8, 224–239. https://doi.org/10.1111/j.1461-0248.2004.00701.x.

[3] J.-A. Grytnes and O. R. Vetaas. 2002. Species richness and altitude: A comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. *The American Naturalist* 159, 294–304. https://doi.org/10.1086/338542.

[4] C. M. McCain. 2009. Global analysis of bird elevational diversity. *Global Ecology and Biogeography* 18, 346–360. https://doi.org/10.1111/j.1466-8238.2008.00443.x.

[5] F. S. Gilliam. 2007. The ecological significance of the herbaceous layer in temperate forest ecosystems. *BioScience* 57, 845–858. https://doi.org/10.1641/B571007.

[6] M. K. Sundqvist, N. J. Sanders, and D. A. Wardle. 2013. Community and ecosystem responses to elevational gradients. *Annual Review of Ecology, Evolution, and Systematics* 44, 261–280. https://doi.org/10.1146/annurev-ecolsys-110512-135750.

[7] H. Qian, R. E. Ricklefs, and P. S. White. 2005. Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecology Letters* 8, 15–22. https://doi.org/10.1111/j.1461-0248.2004.00682.x.

[8] A. Siefert, C. Ravenscroft, M. D. Weiser, and E. E. Cleland. 2013. Functional beta-diversity patterns reveal deterministic community assembly processes in eastern North American trees. *Global Ecology and Biogeography* 22, 682–691. https://doi.org/10.1111/geb.12030.

[9] N. J. B. Kraft, O. Godoy, and J. M. Levine. 2015. Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences of the United States of America* 112, 797–802. https://doi.org/10.1073/pnas.1413650112.

[10] J. Lenoir, J.-C. Gégout, P. A. Marquet, P. de Ruffray, and H. Brisse. 2008. A significant upward shift in plant species optimum elevation during the twentieth century. *Science* 320, 1768–1771. https://doi.org/10.1126/science.1156831.

[11] J. S. Clark, D. Bell, C. Chu, B. Courbaud, M. Dietze, M. Hersh, and others. 2014. High-dimensional coexistence based on individual variation: A synthesis of evidence. *Ecological Monographs* 84, 3–42.

[12] N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. https://doi.org/10.1038/35002501.

[13] C. Körner. 2007. The use of “altitude” in ecological research. *Trends in Ecology and Evolution* 22, 569–574. https://doi.org/10.1016/j.tree.2007.09.006.

[14] B. Bookhagen and D. W. Burbank. 2010. Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface* 115, F03019. https://doi.org/10.1029/2009JF001426.

[15] C. E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[16] E. H. Simpson. 1949. Measurement of diversity. *Nature* 163, 688. https://doi.org/10.1038/163688a0.

[17] E. C. Pielou. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13, 131–144. https://doi.org/10.1016/0022-5193(66)90013-0.

[18] J. R. Bray and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27, 325–349. https://doi.org/10.2307/1942268.

[19] J. B. Kruskal. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29, 115–129. https://doi.org/10.1007/BF02289694.

[20] M. J. Anderson. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

[21] D. I. Warton, T. W. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* 3, 89–101. https://doi.org/10.1111/j.2041-210X.2011.00127.x.

[22] M. Dufrêne and P. Legendre. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs* 67, 345–366. https://doi.org/10.2307/2963459.

[23] L. Breiman. 2001. Random forests. *Machine Learning* 45, 5–32. https://doi.org/10.1023/A:1010933404324.

[24] T. Chen and C. Guestrin. 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785–794. https://doi.org/10.1145/2939672.2939785.

[25] C. J. F. ter Braak. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167–1179. https://doi.org/10.2307/1938672.

[26] K. R. Clarke. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18, 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x.

[27] P. Wangda and M. Ohsawa. 2006. Gradational forest change along the climatically dry valley slopes of Bhutan in the midst of humid eastern Himalaya. *Plant Ecology* 186, 109–128. https://doi.org/10.1007/s11258-006-9116-5.

[28] P. Wangda and M. Ohsawa. 2006. Structure and regeneration dynamics of dominant tree species along an altitudinal gradient in dry valley slopes of the Bhutan Himalaya. *Forest Ecology and Management* 230, 136–150. https://doi.org/10.1016/j.foreco.2006.04.027.

[29] B. K. Acharya, B. Chettri, and L. Vijayan. 2011. Distribution pattern of trees along an elevational gradient of Eastern Himalaya, India. *Acta Oecologica* 37, 329–336. https://doi.org/10.1016/j.actao.2011.03.005.

[30] J. Elith, J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

[31] A. J. C. Grierson and D. G. Long. 1983--2001. Flora of Bhutan.

[32] M. Ohsawa. 1995. Latitudinal pattern of mountain vegetation zonation in eastern Asia and its relation to climate. *Vegetatio* 121, 3–10. https://doi.org/10.1007/BF00044667.

Supplementary Information

Additional supplementary information may be found online in the Supplementary Information section:

.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix12.xlsx
- AppendixS01VIFallstrata.csv
- AppendixS06variableimportance.csv
- AppendixS03envfitall.csv
- Appendix13.xlsx
- AppendixS08diversitycorrelations.csv
- AppendixS05MLparameters.csv
- AppendixS02PERMANOVAfull.csv
- Appendix15.xlsx
- AppendixS10matrixdiagnostics.csv
- AppendixS07CVfoldresults.csv
- AppendixS11samplingclimategrid.csv
- AppendixS19predictorcorrelations.csv
- AppendixS17betapartition.csv
- Appendix16.xlsx
- Appendix18.xlsx
- AppendixS04indicatorspeciesfull.csv