Vegetation Community Composition and Species–Environment Relationships Along an Elevational Gradient in South-Central Bhutan

Abstract

Questions: Understanding how plant communities vary along elevation is essential for predicting biodiversity responses to environmental change. We asked how community composition, diversity, and species–environment relationships differ among vegetation strata along an elevational gradient, and which environmental variables best explain regeneration patterns.

Location: Sarpang District, south-central Bhutan; unmanaged subtropical to cool broadleaved forests spanning approximately 260 to 1 964 m a.s.l.

Methods: We analysed vegetation data from four strata (trees, shrubs, herbs, regeneration) across 220 plots. We constructed community matrices, quantified environmental variables, applied non-metric multidimensional scaling and canonical correspondence analysis, and tested group differences with permutational multivariate analysis of variance. Indicator species analysis and machine-learning models (random forest and gradient boosting) assessed regeneration patterns.

Results: Community composition differed among forest types in all strata (R^2 0.017 to 0.050; $p = 0.001$). Environmental predictors explained 3.2 to 3.8 percent of variation in trees and shrubs and were not significant for herbs and regeneration. Trees had the highest Shannon diversity (1.391 ± 0.595), whereas herbs had the lowest (0.325 ± 0.451). For regeneration richness, random forest showed modest cross-validated performance (root mean square error 1.165 ± 0.182; coefficient of determination 0.142 ± 0.040).

Conclusions: Forest types impose a statistically detectable but limited macro-environmental signal across strata. Community assembly is dominated by fine-scale heterogeneity and vertical decoupling among layers, and regeneration is only weakly predictable from coarse environmental gradients. These results provide a baseline for monitoring unmanaged broadleaved forests within the sampled 260–1 964 m range.

Keywords: elevational gradient, plant community, tropical mountain forest, species–environment relationships, community assembly, diversity, regeneration, ordination, canonical correspondence analysis, indicator species

1 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 (Lomolino, 2001; Rahbek, 2005). 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 (Grytnes and Vetaas, 2002; McCain, 2009). In forest ecosystems, steep abiotic gradients combined with fine-scale variation in soils, canopy structure, and disturbance regimes generate pronounced compositional turnover and vertically stratified responses among trees, shrubs, and herb layers (Gilliam, 2007; Sundqvist et al., 2013). Mountain floras therefore often exhibit high beta diversity over short distances, reflecting both dispersal limitation and niche differentiation along climatic and edaphic gradients (Qian et al., 2005). Recent research has increasingly focused on disentangling the relative roles of environmental filtering, functional traits, and biotic interactions in shaping elevational community organization (Siefert et al., 2013; Kraft et al., 2015). Nevertheless, many studies continue to focus primarily on canopy trees, despite mounting evidence that understorey strata respond more sensitively to microclimatic and edaphic variation and may follow distinct compositional trajectories along the same gradient (Gilliam, 2007; Lenoir et al., 2008). Regeneration layers are especially informative in this context because they integrate recent recruitment processes and may foreshadow future changes in forest composition under ongoing climatic warming (Clark et al., 2014).

The eastern Himalaya represents a global centre of biodiversity characterized by extreme elevational relief, complex terrain, and monsoonal climates that generate sharp forest transitions and substantial species turnover (Myers et al., 2000; Grytnes and Vetaas, 2002). Within this region, numerous studies have documented steep elevational zonation and marked diversity patterns, while broad-scale syntheses emphasize the region's sharp biodiversity gradients and habitat heterogeneity (Myers et al., 2000; Grytnes and Vetaas, 2002). Despite this growing body of work, community-level investigations integrating multiple vegetation strata across Himalayan elevational gradients remain comparatively uncommon. Few studies explicitly compare species–environment relationships among canopy and understorey layers, and regeneration dynamics are seldom analysed in parallel with diversity and compositional patterns, despite their central importance for forest resilience and long-term reorganization (Sundqvist et al., 2013; Clark et al., 2014).

Here, we present a multi-stratum assessment of plant community composition, diversity, and species–environment relationships along an elevational gradient in south-central Bhutan. We test whether forest types differ in composition across four strata—trees, shrubs, herbs, and regeneration—identify macro-environmental correlates of community structure, and quantify elevational trends in alpha and beta diversity together with indicator species. We also use machine-learning models as exploratory tools to evaluate which measured variables are associated with regeneration richness. By integrating multiple vegetation strata within a unified framework, this study provides one of few multi-stratum assessments in the eastern Himalaya and explicitly evaluates scale mismatch between macro predictors and microsite processes.

2 Methods

2.1 Study area

The study was conducted in unmanaged forests of Sarpang District in south-central Bhutan under the jurisdiction of Divisional Forest Office, Sarpang. 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

(Grytnes and Vetaas, 2002; Körner, 2007) (Figure 1).

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 (Bookhagen and Burbank, 2010). 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 (Grytnes and Vetaas, 2002). The landscape also forms part of a strategic ecological linkage among Bhutan's protected areas, although the sampled forests themselves lie largely outside formal management regimes.

2.2 Sampling design

Vegetation sampling followed a stratified random design along the elevational gradient extending from Shershong (approximately 260 m a.s.l.) in the south to Singye (approximately 1 964 m a.s.l.) in the north. A total of 220 plots were established across forest types to ensure representation of major vegetation zones and environmental conditions.

Each plot measured 20 × 20 m (400 m^2), a size widely used in forest community studies to characterise tree assemblages and stand structure, with a nested 2 × 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. Plots were distributed approximately evenly among elevational bands and forest types to ensure balanced representation of vegetation zones. Within each band, plot locations were stratified by forest type and positioned using random bearings and distances from access transects, subject to terrain safety and forest accessibility constraints.

2.3 Vegetation data collection

All free-standing woody individuals with diameters at breast height (DBH) > 10 cm and height > 1.3 m were 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 $\leq$ 1.3 m in height, irrespective of stem diameter, whereas saplings were defined as individuals $>$ 1.3 m in height with DBH $\leq$ 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.

Species identifications were verified using regional floras and herbarium reference material, and taxonomic nomenclature was cross-checked against the World Flora Online database (accessed 6 Feb 2026). Structural attributes, including basal area, DBH-class distributions, and height-class distributions, were derived to characterize stand structure and regeneration dynamics across forest types and elevational gradients.

2.4 Environmental variables

Topographic attributes (latitude and longitude), including elevation, slope, and aspect, were recorded in the field using handheld Global Positioning System receivers, clinometers, and compasses. Spatial climatic surfaces describing air temperature, precipitation, evapotranspiration, and water balance were extracted from the national gridded climate dataset produced by Dorji et al. (2025), developed by CSIRO and distributed with a permanent digital object identifier (`https://doi.org/10.25919/pec2-hs50`).

Environmental values were extracted at plot coordinates using the dataset's native grid and historical baseline layers (1986–2015), consistent with the metadata provided for the archived climate product.

These climate layers were used consistently across all spatial and statistical analyses.

2.5 Data preparation

Field records were curated in tabular format to compute species abundance, frequency, basal area and importance value index 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 (Shannon, 1948; Simpson, 1949; Pielou, 1966).

2.6 Multivariate community analyses

Floristic variation among plots was examined using non-metric multidimensional scaling (NMDS) based on Bray–Curtis dissimilarities (Bray and Curtis, 1957; Kruskal, 1964). Species–environment relationships were assessed using canonical correspondence analysis (CCA), which constrains ordination axes by linear combinations of climatic and topographic predictors (ter Braak, 1986). Ordination axes were interpreted through correlations between site scores and environmental variables.

Community differences among forest types were tested using permutational multivariate analysis of variance (Anderson, 2001). 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 (Warton et al., 2012).

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 (Warton et al., 2012). NMDS was used for pattern visualization rather than as a stand-alone test of group separation.

Prior to CCA, collinearity among predictors was assessed using variance inflation factors, and variables exceeding a threshold of 10 were sequentially removed following deterministic selection rules. Statistical significance of constrained axes and individual predictors was evaluated using 999 permutations.

Indicator species analysis was applied to identify taxa significantly associated with forest types using the indicator value method, with significance assessed at $\alpha = 0.05$ and 999 permutations (Dufrêne and Legendre, 1997).

2.7 Regeneration dynamics and predictive modelling

Seedling and sapling data were analysed to quantify regeneration patterns and spatial structure across forest types and elevations. Environmental drivers 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 (Breiman, 2001; Chen and Guestrin, 2016). 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.

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 modest, variable-importance results were interpreted as exploratory indicators of potential drivers rather than definitive causal rankings.

2.8 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 (ter Braak, 1986; Anderson, 2001; Dufrêne and Legendre, 1997). 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.

2.9 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.

Upon submission, data and code supporting the results will be deposited in a public repository with a digital object identifier in accordance with data-sharing policy, with anonymized access provided during double-blind peer review.

3 Results

3.1 Community composition and ordination

Non-metric multidimensional scaling (NMDS) showed modest but detectable compositional differentiation among forest types across strata (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). We explored multiple dimensional solutions, and stress values remained consistently low, confirming ordination stability.

The near-zero stress values for trees, herbs, and regeneration likely arise from matrix properties that ease low-dimensional rank fitting (for example, high tied dissimilarities and sparse compositional structure), so ordinations were interpreted conservatively as broad summaries rather than precise geometric distances (Kruskal, 1964; Clarke, 1993). We therefore cross-checked NMDS patterns against PERMANOVA, dispersion tests, and constrained ordination instead of relying on visual separation alone.

Shrub and herb assemblages displayed the greatest dispersion in ordination space and the strongest alignment with fitted climatic vectors, particularly temperature, evapotranspiration and precipitation (Figure 2B–C). Tree assemblages were comparatively compressed along NMDS axes (Figure 2A), while regeneration plots showed broad scatter (Figure 2D), together indicating vertical decoupling and strong neighbourhood-scale heterogeneity.

These stratified responses mirror documented altitudinal forest-type transitions along

Bhutanese dry valley slopes (Wangda and Ohsawa, 2006a).

3.2 Community differences among forest types

Community composition differed significantly among forest types across all vegetation strata (Table 1). Tree assemblages exhibited modest but statistically significant separation (PERMANOVA $R^2 = 0.031$, $F = 3.38$, $p = 0.001$), with homogeneous dispersions (betadisper $p = 0.291$), indicating that group differences reflected centroid shifts rather than unequal within-group variability (Appendix S2).

Shrub communities showed the strongest differentiation among forest types ($R^2 = 0.050$, $F = 5.05$, $p = 0.001$), 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 1).

The greater sensitivity of shrub and herb layers relative to trees is consistent with reported altitudinal zonation and vegetation transitions in Himalayan forests (Wangda and Ohsawa, 2006a; Grierson and Long, 2001). 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.

3.3 Species–environment relationships

Canonical correspondence analysis revealed that selected environmental variables—aspect, evapotranspiration, latitude, longitude and slope—explained a small but statistically significant fraction of constrained inertia in trees and shrubs. The percentage of total inertia explained was 3.2% for trees ($p = 0.008$) and 3.8% for shrubs ($p = 0.001$; Table 2; Figure 5). For herbs and regeneration, constrained models explained 3.4% of total inertia but were not significant at $\alpha = 0.05$ ($p = 0.125$ and 0.076, respectively) (Appendix S9).

CCA biplots illustrated contrasting environmental associations among strata (Figure 5). Tree communities aligned primarily with slope and aspect gradients, shrub assemblages were most strongly oriented toward evapotranspiration and latitudinal position, while herbaceous vegetation responded to combined slope–climate axes. Regeneration patterns were weakly associated with evapotranspiration and aspect.

The low proportion of constrained inertia across all strata is expected in steep, heterogeneous Himalayan terrain and is consistent with substantial unmeasured microsite drivers operating within broader altitudinal transitions (Grytnes and Vetaas, 2002; Wangda and Ohsawa, 2006a).

3.4 Diversity patterns

Alpha diversity differed markedly among vegetation strata (Table 3; Figure 3). 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).

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 4). Such patterns are consistent with documented altitudinal zonation of Bhutanese dry-valley forests and reported regeneration dynamics of dominant trees (Wangda and Ohsawa, 2006a,b). 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.

Species richness varied along the elevational gradient for all strata (Figure 7). 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 (Acharya et al., 2011; Grytnes and Vetaas, 2002; Wangda and Ohsawa, 2006a).

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.

3.5 Indicator species

Indicator-species analysis detected significant associations with elevational classes across all strata (Table 4). Shrubs contained the largest proportion of indicator taxa (15 of 101 species; 14.9%), followed by herbs (10 of 134; 7.5%), regeneration (8 of 109; 7.3%) and trees (10 of 221; 4.5%) (Appendix S4).

Among canopy trees, *Schima wallichii* and *Castanopsis indica* were characteristic of low- to mid-elevation forests, whereas *Alnus nepalensis* was associated with higher elevations. These patterns accord with floristic zonation documented in Bhutanese mountain forests, where evergreen broadleaved species dominate lower belts and pioneer or cool-adapted taxa become more frequent upslope (Wangda and Ohsawa, 2006a; Grierson and Long, 2001).

3.6 Regeneration modelling

Random forest produced modest predictive performance for regeneration richness (Table 5; Figure 6), with cross-validated $R^2 = 0.142 \pm 0.040$ and RMSE 1.165 ± 0.182. Gradient boosting showed near-zero to negative explanatory power ($R^2 = -0.009 \pm 0.096$; RMSE = 1.261 ± 0.192) (Appendix S5).

Fold-level diagnostics (Appendix S7) showed moderate variability in RF performance (R^2 range: 0.096–0.193; RMSE range: 0.999–1.476) and unstable XGBoost performance (R^2 range: -0.154 to 0.106; RMSE range: 1.102–1.570), reinforcing that coarse predictors provide only weak predictability of recruitment.

Variable-importance rankings indicated that shrub-layer attributes, together with temperature, elevation and evapotranspiration, were the strongest correlates of regeneration richness (Figure 6; Appendix S6). Ecologically, this pattern is consistent with neighbourhood-scale filtering and local biotic interactions outweighing coarse macro-environmental gradients for early life stages (Wangda and Ohsawa, 2006b).

Given the low explanatory power, these models are interpreted as exploratory ecological inference rather than deterministic prediction. Important drivers—such as canopy openness, disturbance history, soil properties or fine-scale moisture regimes—were not fully captured by the available predictors, a common limitation in forest-regeneration modelling in complex mountain terrain (Elith et al., 2008).

4 Discussion

4.1 Stratified community turnover along elevational gradients

Our analyses indicate clear vertical stratification in community turnover, with understorey assemblages (shrubs and herbs) more variable than canopy trees. This pattern mirrors observations from other Himalayan forests where understorey communities respond strongly to microclimatic and edaphic heterogeneity, whereas long-lived canopy trees integrate conditions over broader temporal and spatial scales (Wangda and Ohsawa, 2006a; Grytnes and Vetaas, 2002). Forest-type effects were statistically detectable but 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 are compatible with high beta diversity and fine-grained heterogeneity in montane systems (Acharya et al., 2011; Carpenter, 2005). Importantly, interpretation 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 centroid shifts among forest types. This cross-stratum contrast supports vertical decoupling, where canopy differentiation is detectable but modest and understorey turnover is dominated by neighbourhood-scale heterogeneity.

4.2 Environmental drivers and scale of control

Direct gradient analyses (CCA) showed that aspect, slope, evapotranspiration, and spatial position explained only a small fraction of floristic variation ($\approx$3–4% in trees and shrubs). Rather than indicating weak study design, this low constrained inertia is expected in steep Himalayan terrain, where microtopography, canopy gaps, soils, and disturbance history vary at much finer scales than gridded macro predictors (Grytnes and Vetaas, 2002; Wangda and

Ohsawa, 2006a). The contrasting orientations among strata reinforce this scale mismatch. Tree communities aligned mainly with slope and aspect, while shrubs and regeneration were more associated with evapotranspiration and latitude; herbs tracked mixed slope–climate axes. Together, these patterns indicate that macro-environmental gradients are detectable but limited, and that community assembly is dominated by microsite processes and vertical decoupling among strata.

4.3 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 our 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 (Grierson and Long, 2001; Wangda and Ohsawa, 2006a). 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 (Grytnes and Vetaas, 2002). 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 (Grytnes and Vetaas, 2002). The fact that our 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.

4.4 Indicator species and floristic transitions

Indicator species analysis revealed fewer significant canopy indicators than shrub and herb indicators, supporting vertical decoupling in habitat filtering strength. Low tree indicator specificity is consistent with broader niche amplitudes in canopy taxa across the sampled gradient. In contrast, stronger shrub and herb differentiation implies finer-scale filtering in understorey microsites. This contrast aligns with Himalayan zonation as overlapping distributions rather than discrete boundaries (Ohsawa, 1995; Wangda and Ohsawa, 2006a).

4.5 Regeneration dynamics and model interpretation

Machine-learning results were used as exploratory ecological evidence on regeneration controls. The best model (random forest) explained only about 14% of cross-validated variance ($R^2 \approx 0.14$), indicating weak predictability from the available coarse predictors. Shrub-layer structure emerged as the strongest correlate, supporting neighbourhood-scale filtering during early recruitment. This interpretation is consistent with Bhutanese forest observations that local stand structure influences regeneration (Wangda and Ohsawa, 2006a,b). The low predictive power also indicates missing microsite drivers, such as canopy openness, soil properties, disturbance history, and fine-scale moisture (Elith et al., 2008). Accordingly, these models should not be read as deterministic forecasts, but as evidence that regeneration is only weakly predictable from macro-environmental gradients in this landscape.

4.6 Limitations and novelty of the study

Several caveats merit consideration. First, climate variables were derived from interpolated surfaces (Dorji et al., 2025) and may not capture plot-level microclimate in steep terrain. Second, heterogeneous dispersion in shrub and herb strata means that part of PERMANOVA significance reflects spread differences, not only centroid shifts. Third, regeneration measurements represent a temporal snapshot and may miss episodic establishment dynamics of slow-growing montane trees. Spatial autocorrelation was not explicitly modeled, which may influence fine-scale inference, and future work could apply spatially explicit approaches. Even with these constraints, this study provides one of few multi-stratum assessments in the eastern Himalaya. The main contribution is to show that broad-scale environmental effects are statistically detectable but limited, while fine-scale heterogeneity, neighbourhood filtering, and vertical decoupling dominate assembly patterns.

5 Conclusions

Across 260–1 964 m a.s.l. in south-central Bhutan, forest types imposed a statistically detectable but limited macro-environmental signal on community composition ($R^2 = 0.017$–0.050). Understorey strata showed greater turnover and higher dispersion than canopy trees, indicating strong vertical stratification and decoupling among layers.

CCA explained only about 3–4% of variation, which is consistent with steep Himalayan terrain where microsite heterogeneity is high and many local drivers are unmeasured. PERMANOVA interpretation was therefore conditioned on dispersion structure, with mixed centroid–spread effects in heterogeneous strata.

Regeneration was only weakly predictable from coarse predictors (random forest CV $R^2 \approx 0.14$), and shrub-layer attributes emerged as the strongest correlates. This supports neighbourhood-scale filtering during early recruitment and reinforces that macro gradients alone do not capture regeneration dynamics.

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.

References

Acharya, B. K., B. Chettri, and L. Vijayan (2011). Distribution pattern of trees along an elevational gradient of Eastern Himalaya, India. *Acta Oecologica 37*, 329–336.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology 26*, 32–46.

Bookhagen, B. 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.

Bray, J. R. and J. T. Curtis (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs 27*, 325–349.

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5–32.

Carpenter, C. (2005). The environmental control of plant species density on a Himalayan elevation gradient. *Journal of Biogeography 32*, 999–1018.

Chen, T. and C. Guestrin (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Clark, J. S., D. Bell, C. Chu, B. Courbaud, M. Dietze, M. Hersh, et al. (2014). High-dimensional coexistence based on individual variation: A synthesis of evidence. *Ecological Monographs 84*, 3–42.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology 18*, 117–143.

Dorji, T. et al. (2025). Bhutan high-resolution climate surfaces (historical 1986–2015 and future 2015–2100). CSIRO Data Collection.

Dufrêne, M. and P. Legendre (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs 67*, 345–366.

Elith, J., J. R. Leathwick, and T. Hastie (2008). A working guide to boosted regression trees. *Journal of Animal Ecology 77*, 802–813.

Gilliam, F. S. (2007). The ecological significance of the herbaceous layer in temperate forest ecosystems. *BioScience 57*, 845–858.

Grierson, A. J. C. and D. G. Long (1983–2001). *Flora of Bhutan*. Edinburgh, UK: Royal Botanical Garden Edinburgh and Royal Government of Bhutan.

Grytnes, J.-A. 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.

Körner, C. (2007). The use of 'altitude' in ecological research. *Trends in Ecology and Evolution 22*, 569–574.

Kraft, N. J. B., 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.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika 29*, 115–129.

Lenoir, J., 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.

Lomolino, M. V. (2001). Elevation gradients of species density: Historical and prospective views. *Global Ecology and Biogeography 10*, 3–13.

McCain, C. M. (2009). Global analysis of bird elevational diversity. *Global Ecology and Biogeography 18*, 346–360.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent (2000). Biodiversity hotspots for conservation priorities. *Nature 403*, 853–858.

Ohsawa, M. (1995). Latitudinal pattern of mountain vegetation zonation in eastern Asia and its relation to climate. *Vegetatio 121*, 3–10.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology 13*, 131–144.

Qian, H., 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.

Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters 8*, 224–239.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal 27*, 379–423.

Siefert, A., 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.

Simpson, E. H. (1949). Measurement of diversity. *Nature 163*, 688.

Sundqvist, M. K., 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.

ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology 67*, 1167–1179.

Wangda, P. and M. Ohsawa (2006a). Gradational forest change along the climatically dry valley slopes of Bhutan in the midst of humid eastern Himalaya. *Plant Ecology 186*, 109–128.

Wangda, P. and M. Ohsawa (2006b). 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.

Warton, D. I., T. W. Wright, and Y. Wang (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution 3*, 89–101.

Tables

Table 1: 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

Table 2: Summary of canonical correspondence analyses by vegetation stratum. Environmental variables were selected following iterative variance inflation factor screening (threshold = 10). Significance of constrained inertia was assessed using 999 permutations.

Stratum	Sites	Species	Env. vars	Total inertia	Constrained	% Explained	p
Trees	216	221	5	35.117	1.109	3.2	0.008
Shrubs	194	101	5	27.832	1.045	3.8	0.001
Herbs	206	134	5	68.475	2.309	3.4	0.125
Regeneration	205	109	5	58.906	1.987	3.4	0.076

Table 3: 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

Table 4: 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	10	4.5
Shrubs	194	101	3	15	14.9
Herbs	206	134	3	10	7.5
Regeneration	205	109	3	8	7.3

Table 5: 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	13	192	0.142 ± 0.040	1.165 ± 0.182	0.907
XGBoost	13	192	-0.009 ± 0.096	1.261 ± 0.192	0.942

Figures

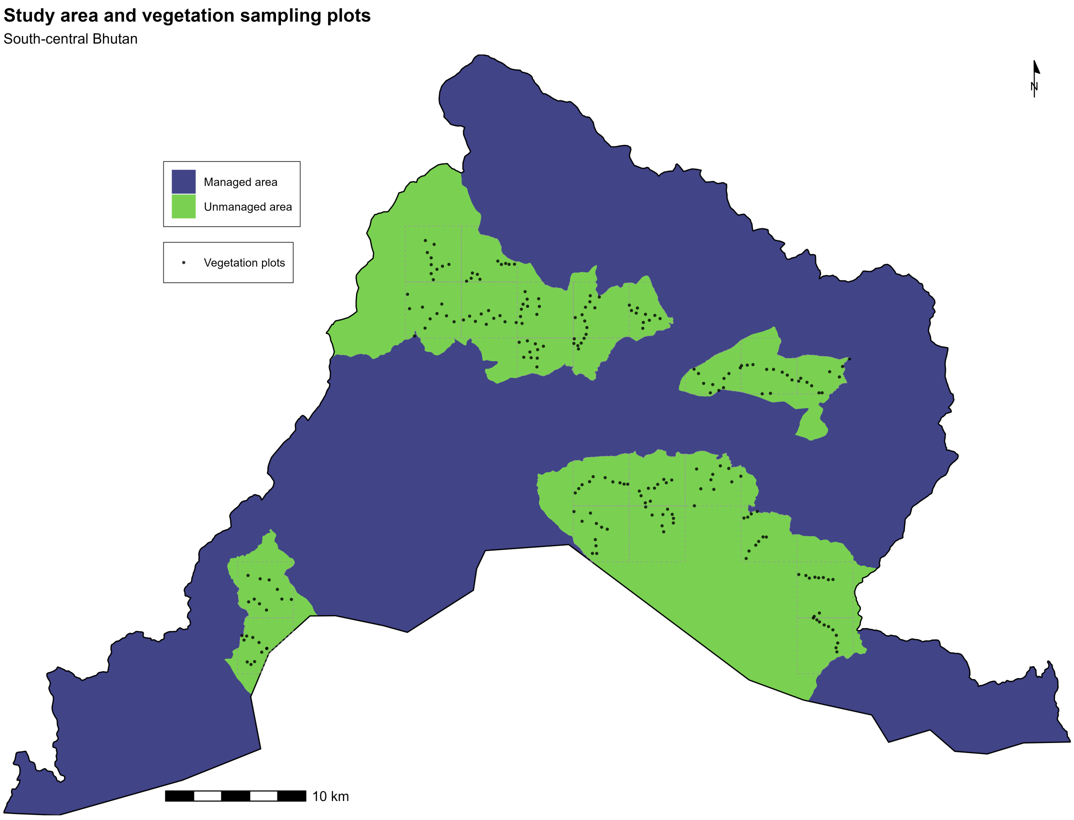

Figure 1: Study area and vegetation sampling plots in south-central Bhutan (Sarpang District). Green polygons indicate unmanaged forest areas and black points represent sampled vegetation plots along the elevational gradient.

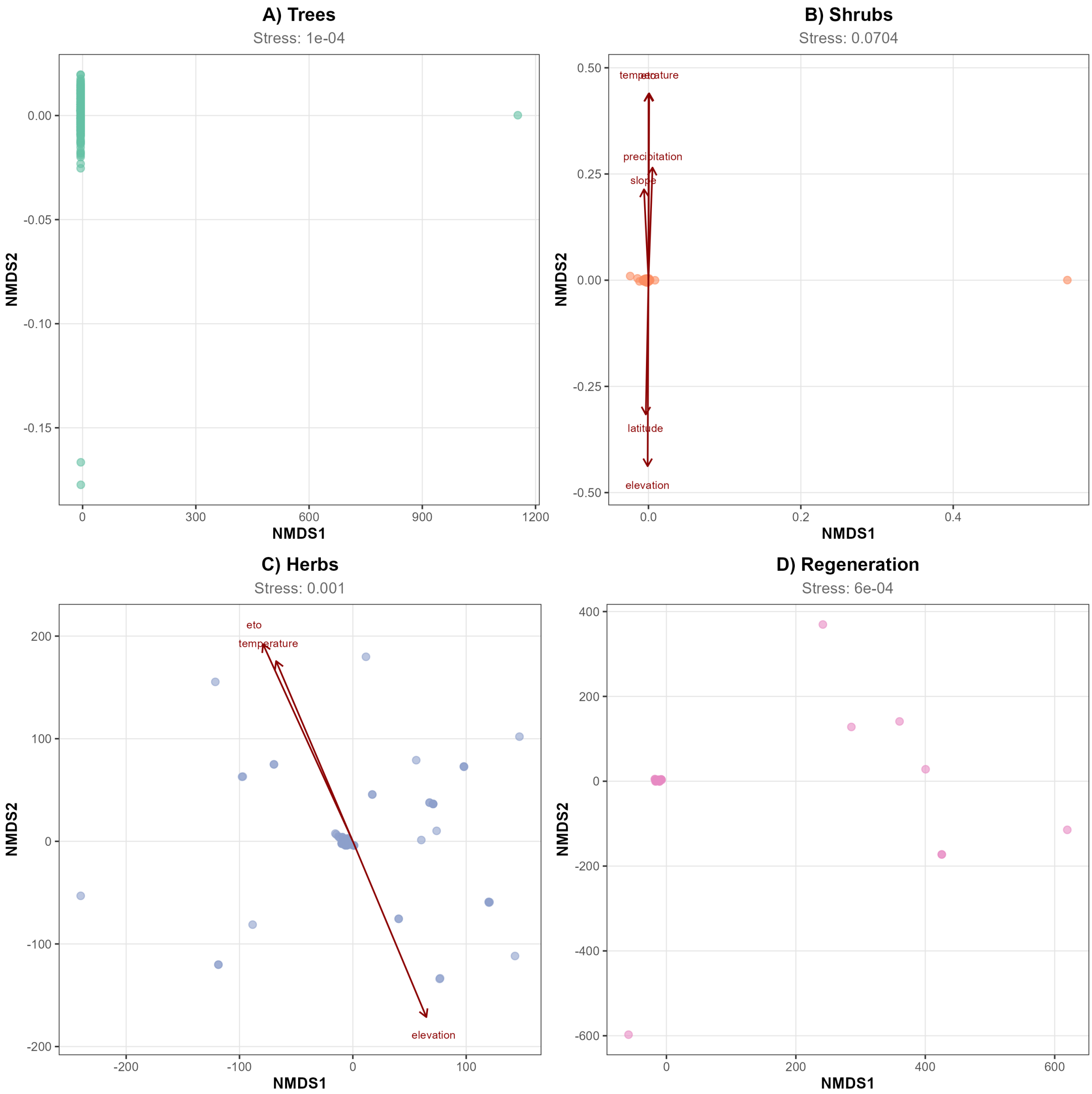

Figure 2: Non-metric multidimensional scaling (NMDS) ordinations of vegetation plots for four strata: (A) trees, (B) shrubs, (C) herbs and (D) regeneration. Points represent sample plots coloured by forest type; polygons delineate convex hulls for each forest type. Red arrows indicate fitted environmental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

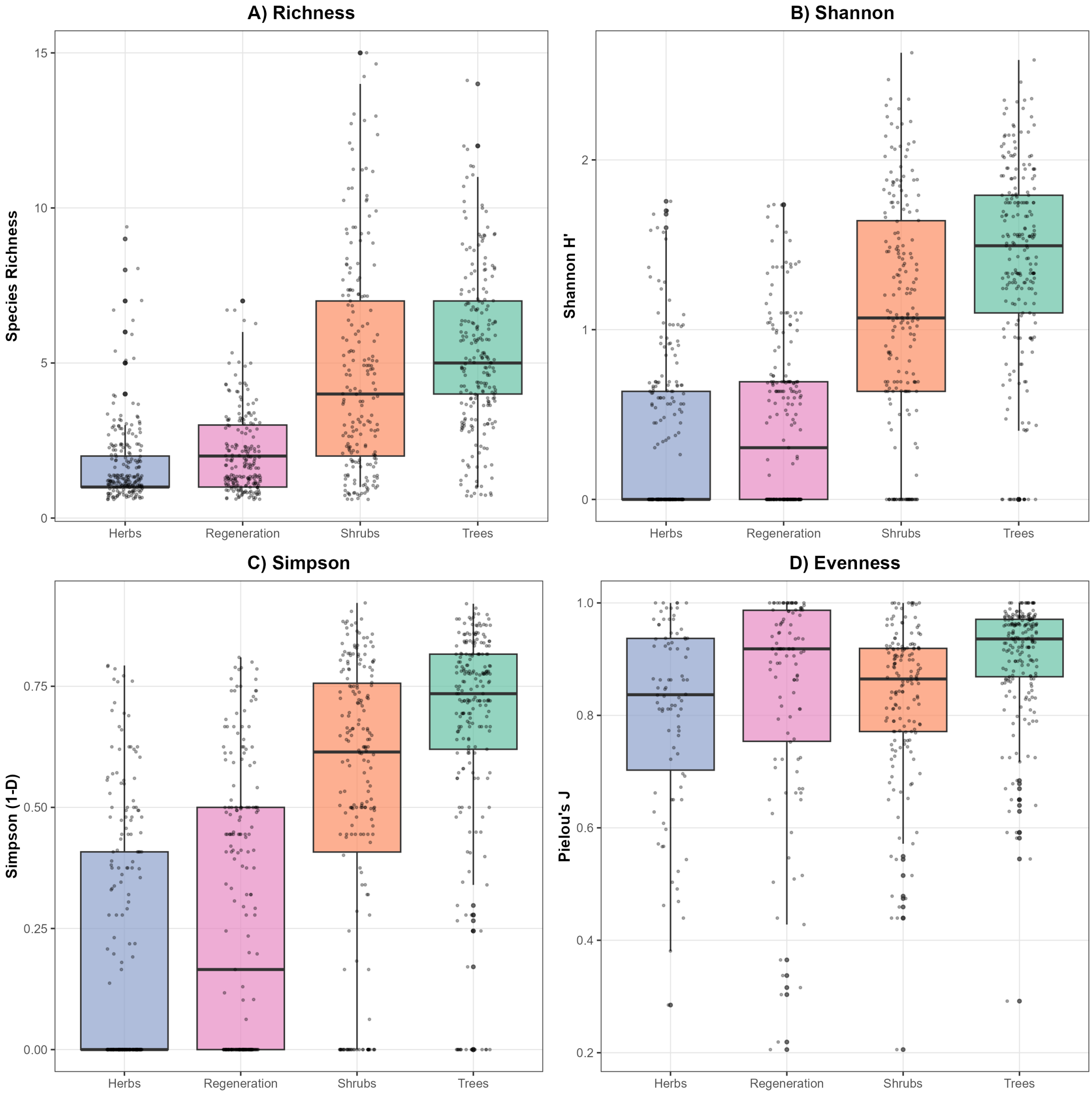

Figure 3: Alpha-diversity metrics across vegetation strata. Panels show (A) species richness, (B) Shannon diversity (H'), (C) Simpson diversity ($1-D$) and (D) Pielou's evenness (J). Boxplots depict medians and interquartile ranges; points represent individual plots.

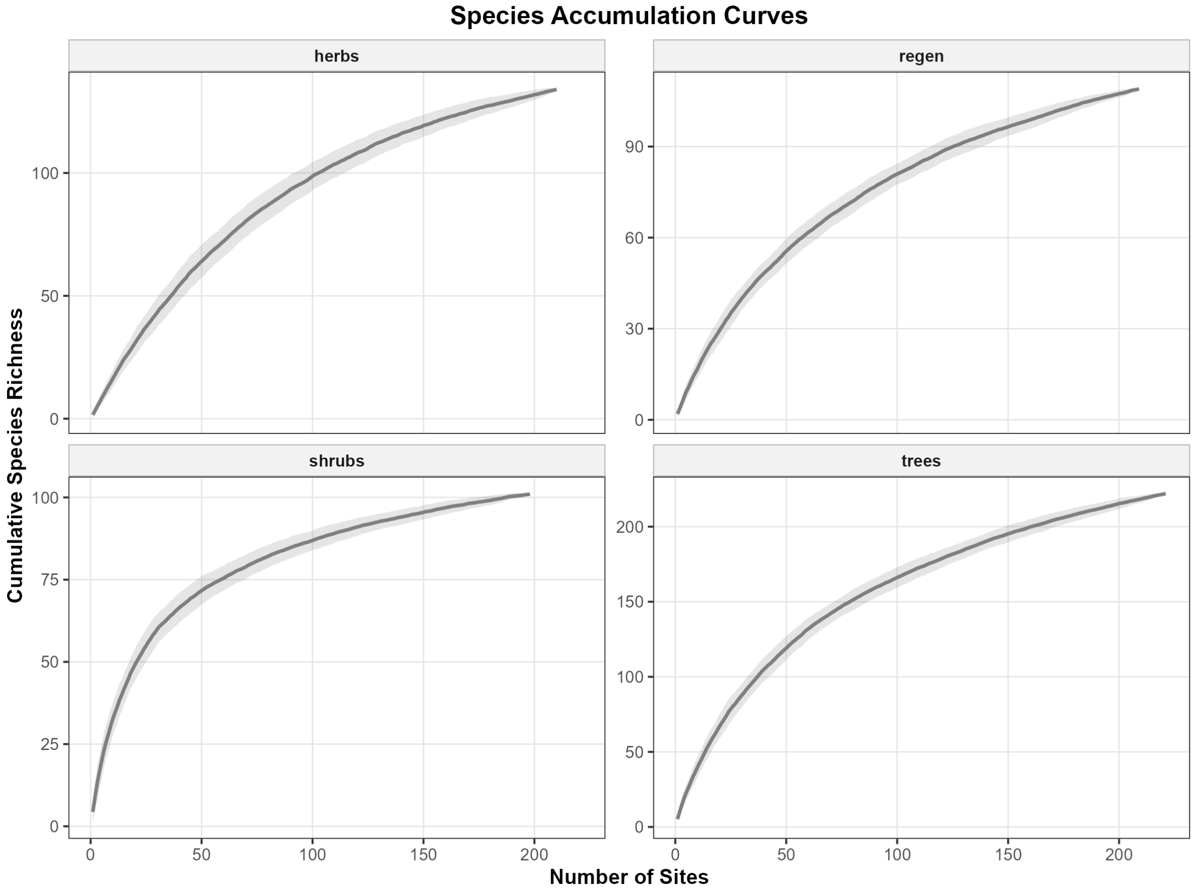

Figure 4: Sample-based species accumulation curves for herbs, regeneration, shrubs and trees. Solid lines indicate mean richness across random permutations; shaded envelopes show 95% confidence intervals.

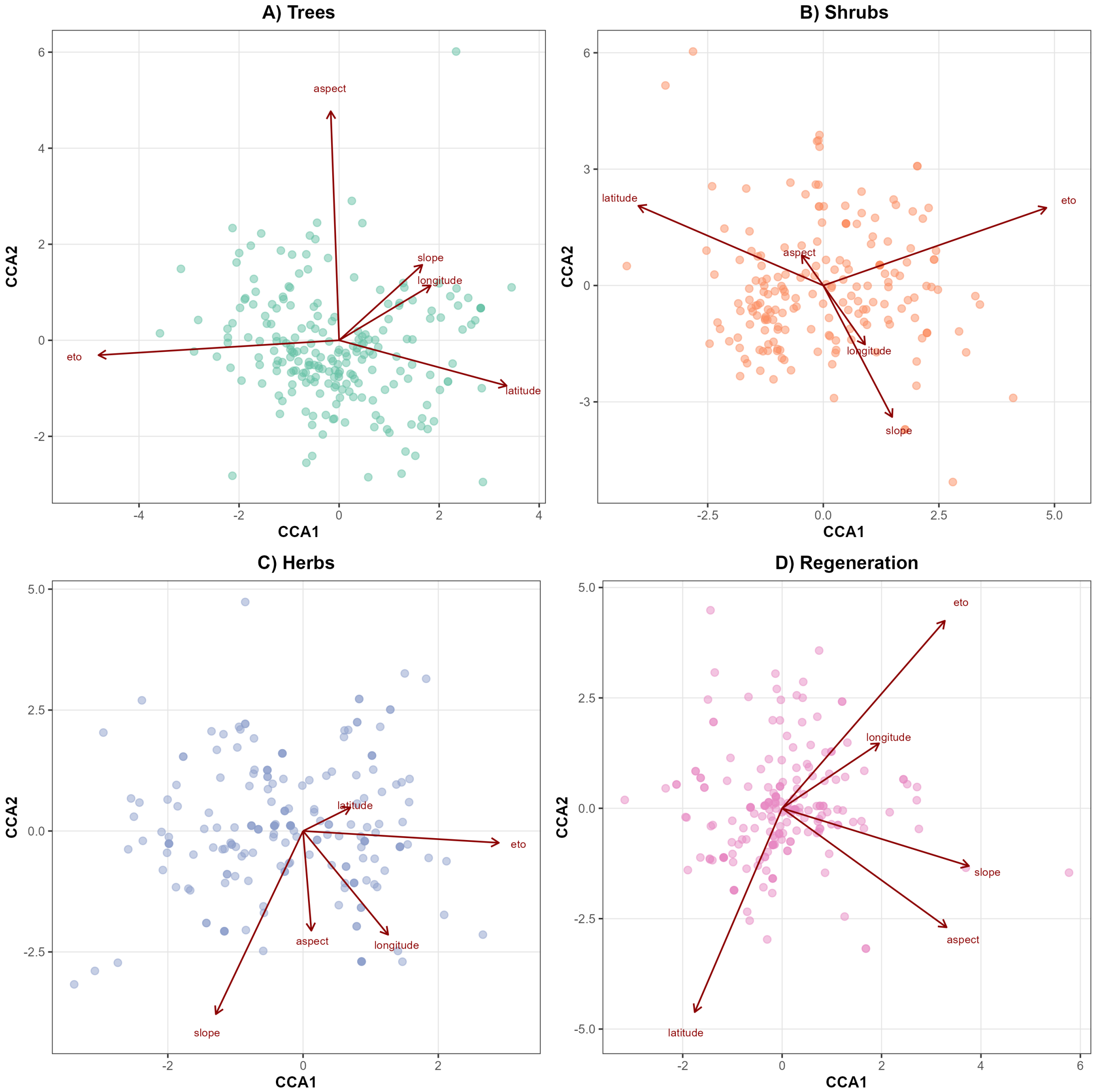

Figure 5: Canonical correspondence analysis (CCA) biplots for (A) trees, (B) shrubs, (C) herbs and (D) regeneration. Points represent plots and arrows denote environmental predictors retained after variance inflation factor screening (threshold = 10). Arrow length indicates the strength of correlations with ordination axes.

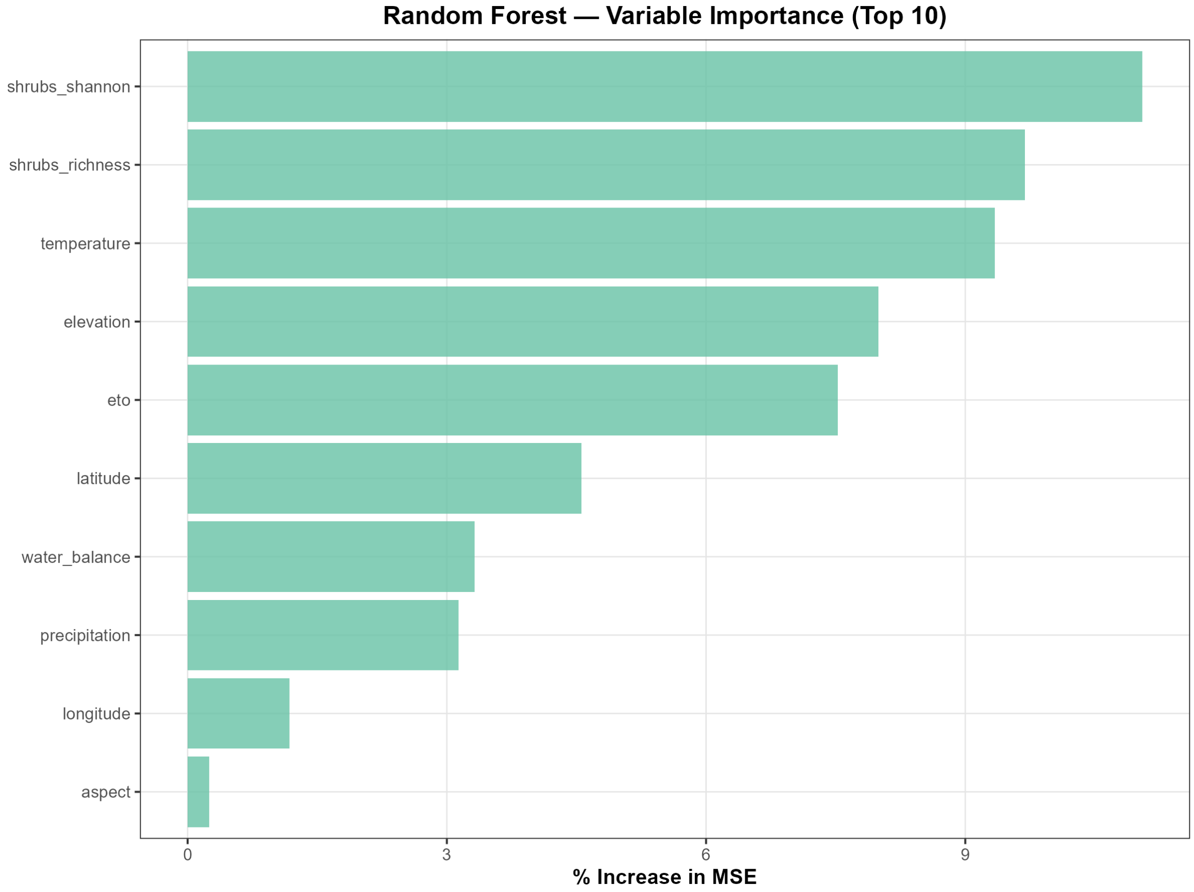

Figure 6: Random-forest variable-importance ranking for predictors of regeneration richness. Bars show percentage increase in mean squared error (%IncMSE) following permutation of each variable during five-fold cross-validation. Higher values indicate stronger influence on predictive accuracy.

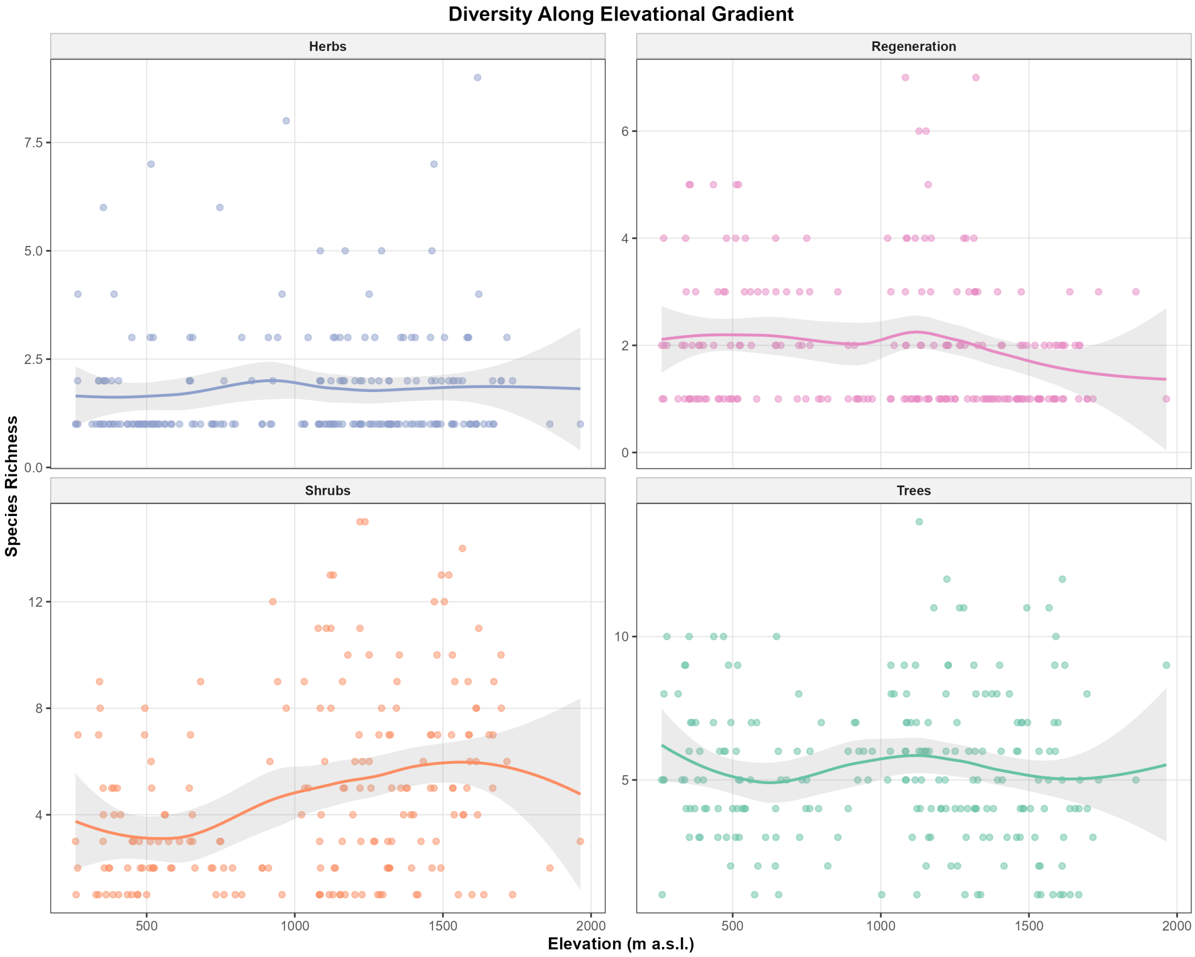

Figure 7: Variation in plot-level species richness along the elevational gradient for herbs, regeneration, shrubs and trees. Curves represent locally weighted regression fits with 95% confidence bands.

Supporting Information

Additional supporting information may be found online in the Supporting Information section:

Appendix S1. Variance inflation factor screening by stratum.

Appendix S2. Full PERMANOVA and dispersion outputs.

Appendix S3. Envfit results for all strata.

Appendix S4. Indicator species list.

Appendix S5. Machine-learning model parameters.

Appendix S6. Variable importance outputs.

Appendix S7. Cross-validation fold results.

Appendix S8. Diversity correlation matrix.

Appendix S9. Beta diversity metrics.

Appendix S10. Community matrix diagnostics.