Evaluating spectral diversity approaches for tree species diversity mapping in hemi-boreal forests using Sentinel-2 and biodivMapR

Arathi Biju [1,2], Oleksandr Borysenko [1], Holger Virro [2], Jean-Baptiste Féret [3], Jan Pisek [1], Evelyn Uuemaa [2]

[1] Tartu Observatory, University of Tartu, Tõravere, Estonia - (arathi.biju, oleksandr.borysenko, jan.pisek)@ut.ee
[2] Landscape Geoinformatics Lab, Department of Geography, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia - (holger.virro, evelyn.uuemaa)@ut.ee
[3] TETIS, INRAE, AgroParisTech, CIRAD, CNRS, Université Montpellier, Montpellier, France - jb.feret@teledetection.fr

Keywords: Spectral Variation Hypothesis, Spectral species, Spectral diversity, Sentinel-2, Biodiversity.

Abstract

Spectral diversity mapping is increasingly used to estimate biodiversity from satellite imagery, but the effect of input feature type remains poorly tested. We evaluated how different Sentinel-2 input feature types - surface reflectance, Principal Component Analysis (PCA) transformed surface reflectance, and spectral indices - affect biodivMapR-based tree diversity mapping across three hemi-boreal forest landscapes in Estonia. Outputs were validated against field-measured tree species Shannon diversity using Spearman correlation and bootstrap resampling. PCA transformed surface reflectance produced the strongest and most consistent correlations across sites, spectral indices performed slightly better than surface reflectance in two sites, and interquartile range-based filtering had little effect. The results identify input feature type selection as a key parameter in Sentinel-2 spectral diversity workflows.

1. Introduction

Remote sensing-based spectral diversity has emerged as a scalable proxy for field-measured biodiversity (Torresani et al., 2019; Wang and Gamon, 2019; Fassnacht et al., 2022), grounded in the Spectral Variation Hypothesis (SVH), which posits that greater spectral heterogeneity in remotely sensed imagery reflects greater ecological diversity on the ground (Palmer et al., 2002). While studies have examined methodological choices known to influence spectral diversity mapping performance (Schmidtlein and Fassnacht, 2017; Robertson et al., 2023), the effect of input feature type, specifically the choice between surface reflectance, Principal Component Analysis (PCA) transformed surface reflectance, and spectral indices, has not been empirically tested.

The biodivMapR package (Féret and de Boissieu, 2020) implements the spectral species concept for biodiversity mapping and acknowledges both PCA transformed reflectance and spectral indices as valid input options. However, little empirical evidence exists regarding how these alternative input representations influence spectral diversity estimates and their relationship with observed biodiversity. The package documentation recommends that users evaluate different approaches against field observations (Féret and de Boissieu, 2026), but systematic comparisons remain scarce. More broadly, reproducible biodiversity monitoring workflows based entirely on open satellite data and open-source software remain under-evaluated, particularly with respect to methodological parameterisation choices.

To address this gap, we aim to compare three input feature types for spectral diversity computation, across three ecologically distinct hemi-boreal forest sites in Estonia and validate them against field-measured tree species diversity. We address two research questions:

(1) Does PCA-based dimensionality reduction produce stronger correlations with field-measured tree species diversity than surface reflectance without PCA or spectral indices used as biodivMapR inputs?

(2) Are results consistent across ecologically distinct hemi-boreal forest sites?

2. Materials and methods

2.1 Study area

We selected three study sites in Estonia that represent contrasting hemi-boreal forest landscapes: Otepää, Soomaa, and Rapla (Figure 1a). Together, they span a gradient from topographically heterogeneous mixed forests to wetland-dominated forest mosaics and more homogeneous managed production forests.

Otepää (Figure 1b) represents a glacial moraine upland with pronounced topographic heterogeneity and mixed nemoral forests. The landscape contains a mosaic of uplands, depressions, wetlands, lakes, and small agricultural fields. Forest stands include spruce-, pine-, and birch-dominated forests as well as species-rich mixed stands (Arold, 2005; Remm, 2005).

Soomaa (Figure 1c) represents a peatland and floodplain forest complex shaped by strong hydrological gradients. Bogs, swamp forests, floodplain forests, and semi-natural wetland mosaics dominate the landscape. Pine, birch, and alder are the main tree taxa, with locally important broadleaved floodplain forests also present (Palisaar, 2006; Keskkonnaamet, 2016).

Rapla (Figure 1d) represents a lowland managed forest mosaic typical of intensively managed hemi-boreal production forests. Compared with Otepää and Soomaa, Rapla provides a more homogeneous and management-influenced forest setting.

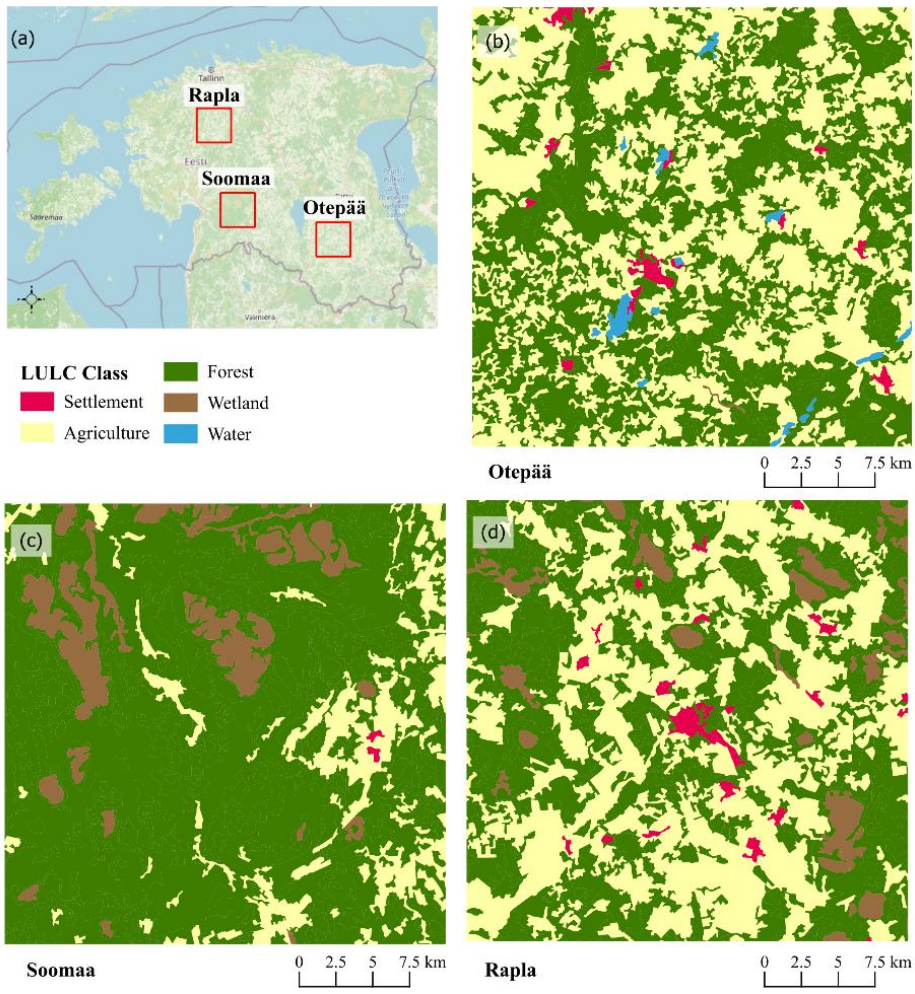

Figure 1. Location of the three 30 km × 30 km study sites in Estonia. Panel (a) shows site locations, panels (b)-(d) show broad land-use/land-cover groups from CORINE Land Cover 2018 (Copernicus Land Monitoring Service, 2020).

2.2 Satellite data

We used Copernicus Sentinel-2 Level-2A surface reflectance products from the European Space Agency for the 2024 summer growing season (June-August) (European Space Agency, 2021). Seasonal median composites were generated in Google Earth Engine (Gorelick et al., 2017) following Torresani et al. (2019), who found that SVH-based relationships between spectral variation and tree species diversity were strongest in summer. We masked clouds with CloudScore+, an operational cloud-masking product for Google Earth Engine that performs comparably to or better than several alternative approaches (Rifai and Harintaka, 2024; Lee et al., 2025). We applied the same preprocessing workflow to all three sites. Before calculating spectral diversity, we masked the imagery to forested areas using a forest mask derived from Estonian Forest Registry maintained by the Estonian Environment Agency (Estonian Environment Agency, 2025) to limit the analysis only to forested area where we had validation data (see next section).

2.3 Validation data

We obtained field-based validation data from the Estonian Forest Registry and retained only records updated from 2015 onward (Estonian Environment Agency, 2025) because each forest registry plot is field checked at least once every 10 years. To reduce temporal mismatch with the 2024 Sentinel-2 imagery, we removed plots affected by forest clear-cutting between 2015 and 2024 using Hansen Global Forest Change v1.12 dataset (Hansen et al., 2013). We applied a 30 m inward buffer as a filtering step to identify plots large enough to contain at least one full 3 × 3 moving window. Plots with a residual area below 900 m² were removed. This filtering step reduced the influence of class boundaries and mixed pixels, where spectral diversity values may partly reflect neighbouring land-cover types rather than the forest conditions inside the validation plot. For the retained plots, we extracted mean spectral diversity values using the original plot extents. After filtering, the final validation dataset contained 3,758 forest validation plots in Otepää, 10,259 in Soomaa, and 4,322 in Rapla.

We used the dominant tree species composition reported by the Estonian Forest Registry to calculate one plot-level field-measured tree species Shannon diversity value for each forest validation plot across the three study sites. We treated the recorded species shares as proportional abundances, p_i, and calculated Shannon diversity as:

$$H' = -\sum_{i=1}^{S} p_i \ln(p_i) \quad (1)$$

where S = number of recorded tree species
p_i = proportional contribution of species i

This index accounts for both richness and evenness and is widely used as a field α-diversity metric in forest remote-sensing studies (Shannon and Weaver, 1949; Végh and Tsuyuzaki, 2021).

2.4 Spectral diversity mapping

We calculated spectral diversity in R with the biodivMapR package (Féret and De Boissieu, 2020), which implements the spectral species concept (Féret and Asner, 2014). This approach follows the SVH, which links spectral heterogeneity to ecological heterogeneity and biodiversity (Palmer et al., 2002; Fassnacht et al., 2022; Torresani et al., 2024).

We compared three input feature types. First, we applied PCA to ten Sentinel-2 bands: blue (B02), green (B03), red (B04), red-edge bands (B05, B06, B07), near-infrared bands (B08 and B8A), and shortwave-infrared bands (B11 and B12), and retained the first three components as they explained approximately 96% of variance in the spectral feature space. We computed PCA separately for each site. Second, we used the same ten spectral bands directly without dimensionality reduction. Third, we calculated three spectral indices using the spinR package within the biodivMapR workflow: the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI), and LAI_SAVI, a Soil-Adjusted Vegetation Index (SAVI)-based estimate for leaf area index (LAI) (Huete, 1988; Gao, 1996; Féret and de Boissieu, 2026; Végh and Tsuyuzaki, 2021; Kacic and Kuenzer, 2022). These three feature types allowed us to compare dimensionality reduction, raw spectral input, and ecologically interpretable indices within the same workflow. Hereafter, we refer to the Sentinel-2 band input without PCA as surface reflectance, the PCA applied band input as PCA transformed surface reflectance, and the spectral indices-based input as spectral indices.

All analyses used imagery that had already been masked to forest. In biodivMapR, we derived spectral species by k-means clustering with 40 clusters, meaning that the algorithm divided the forest pixels into 40 spectral classes based on their reflectance or index values (Féret and Asner, 2014; Féret and de Boissieu, 2020). For each study site and input feature type, we sampled 100,000 valid forest pixels to train the k-means model and then assigned all remaining valid pixels to the closest spectral species. We then calculated spectral Shannon diversity from the relative abundance of spectral species within a 3 × 3 moving window. We processed each site in tiles and mosaicked the outputs into final site-level spectral Shannon diversity maps.

Additionally, we recorded processing time for each input feature type across the three 30 km × 30 km study sites. Runtime was measured from script start to completion on a Fedora 40 Linux workstation with 128 GB RAM and an Intel i9 processor. This provides a general comparison of computational differences among input feature types, although runtime may vary depending on the number of valid forest pixels, raster tiling, parallel processing settings, and computing environment.

We also tested the optional interquartile range (IQR) filter in biodivMapR. This filter masks pixels with unusually high or low spectral values before clustering, using thresholds of Q1 - 4 × IQR and Q3 + 4 × IQR computed over the full input dataset. We ran all three input feature types with and without this filter to test whether it affected spectral diversity estimates in forest-masked imagery.

2.5 Statistical analysis

We used Spearman rank correlation as the primary measure of association because our main aim was to test whether validation plots with higher spectral Shannon diversity also tended to have higher field-measured tree species Shannon diversity. In other words, we were mainly interested in whether the input feature types correctly ranked validation plots from lower to higher diversity, even when the exact values did not align in a linear, one-to-one manner. This use of Spearman's correlation is consistent with recent biodiversity studies that evaluate monotonic relationships between proxy-based and field-based diversity measures (Schmidtlein and Fassnacht, 2017; Hauser et al., 2021; Botero-Cañola et al., 2024).

We used bootstrap resampling to estimate uncertainty in the Spearman correlation between field-measured Shannon diversity and each spectral diversity output (Efron, 1979). For each study site and input feature type, we resampled the validation plots with replacement, recalculated Spearman's rank correlation, and reported the mean bootstrapped correlation with its 95% confidence interval.

We implemented the workflow using open-source software: biodivMapR v2.4 in R for spectral diversity mapping (Féret and De Boissieu, 2020), QGIS v3.40.11 for visual interpretation (QGIS Development Team, 2024), and Python packages for field Shannon diversity calculation, validation, and statistical analysis, mainly NumPy, SciPy, and GeoPandas (Harris et al., 2020; Virtanen et al., 2020; GeoPandas Development Team, 2025). This supports methodological transparency and reproducibility.

3. Results

3.1 Relationship between spectral and field-measured diversity

All three input feature types showed positive associations between spectral Shannon diversity and field-measured tree species Shannon diversity across the three study sites (Figure 2). The strength of this relationship varied by the feature type and study site, but PCA transformed surface reflectance consistently produced the strongest overall associations.

In Otepää, Spearman's ρ was 0.24 for surface reflectance, 0.26 for spectral indices, and 0.29 for PCA transformed surface reflectance. In Soomaa, correlations were ρ = 0.25 for surface reflectance, ρ = 0.28 for spectral indices, and ρ = 0.30 for PCA transformed surface reflectance. Rapla showed the strongest relationships overall, with ρ = 0.38 for surface reflectance, ρ = 0.38 for spectral indices, and ρ = 0.41 for PCA transformed surface reflectance. Across the three study sites, PCA transformed surface reflectance produced the highest correlations, while Rapla showed stronger spectral diversity–field diversity relationships than Otepää and Soomaa.

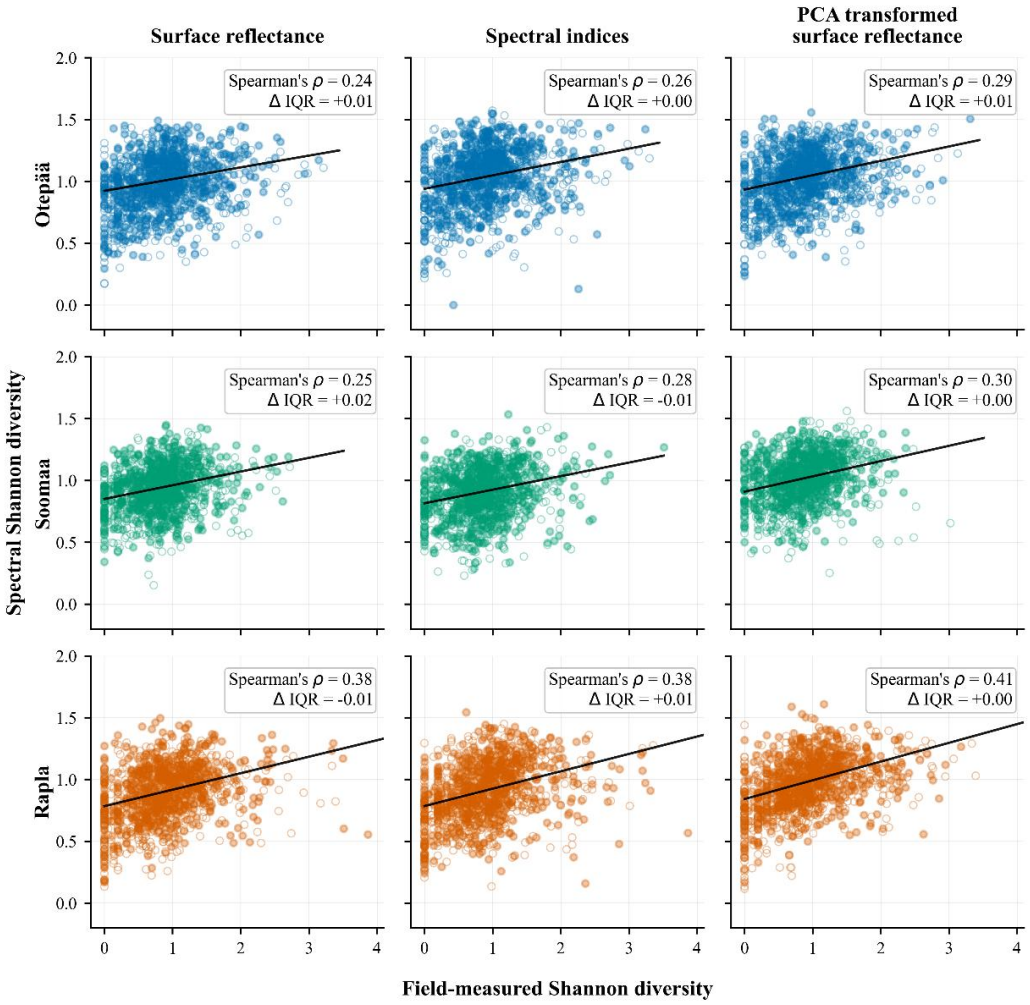

Figure 2. Relationship between field-measured Shannon diversity and spectral Shannon diversity across the three study sites and three input feature types. Δ IQR represents the change in Spearman's ρ after IQR filtering, calculated as $\rho_{IQR} - \rho_{non\text{-}IQR}$

Applying the optional IQR filtering step produced only minor changes in correlation strength across sites and input feature types (Figure 2). In most cases, the change in Spearman correlation after IQR filtering was close to zero, indicating that this pre-processing step had little influence on the relationship between spectral and field-measured diversity in the forest-masked imagery used here. This was consistent with the low proportion of pixels removed by the IQR filter across all study sites and input feature types.

Site	Surface reflectance	Spectral indices	PCA transformed surface reflectance
Otepää	37	30	30
Soomaa	54	44	45
Rapla	33	18	28
Mean	**41**	**30**	**34**

Table 1. Processing time for the 30 km × 30 km site-level spectral diversity workflow using three Sentinel-2 input feature types. Runtime is reported in minutes, rounded to the nearest minute.

The processing time comparison showed moderate differences among the three Sentinel-2 input feature types (Table 1). Spectral indices had the shortest mean processing time (30 min), showing a slight time advantage over PCA transformed surface reflectance (34 min), while surface reflectance was the slowest (41 min). This suggests a small computational advantage for spectral-index-based inputs in this workflow.

3.2 Bootstrap uncertainty in spectral diversity-field diversity relationships

Bootstrap resampling showed positive Spearman correlations between field-measured Shannon diversity and spectral diversity for all sites and input feature types (Figure 3). Rapla produced the highest correlations across all three input types, while Soomaa and Otepää showed lower correlations.

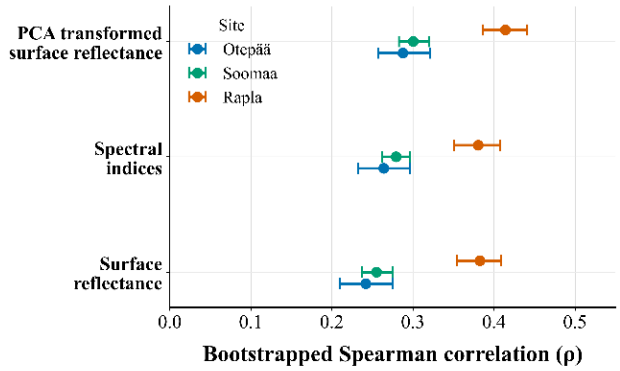

Figure 3. Bootstrapped Spearman correlations between field-measured Shannon diversity and spectral Shannon diversity by input feature type and study site.

The width of the confidence intervals differed among sites. Otepää showed the widest confidence intervals, indicating higher uncertainty in the estimated correlations. Soomaa showed the narrowest intervals, while Rapla showed intermediate uncertainty. This suggests that the strength of the spectral diversity–field diversity relationship was most variable in Otepää and most stable in Soomaa.

Across input feature types, PCA transformed surface reflectance consistently produced the highest mean bootstrapped correlations in all three sites. Spectral indices produced higher correlations than surface reflectance in Otepää and Soomaa, while both input types performed similarly in Rapla. Overall, PCA transformed surface reflectance showed the strongest performance, followed by spectral indices and then surface reflectance.

4. Discussion

This study shows that the choice of Sentinel-2 input feature type strongly influences the performance of biodiversity-oriented spectral diversity mapping in biodivMapR (Féret and Asner, 2014; Féret and de Boissieu, 2020). Across all three sites, PCA transformed surface reflectance produced the strongest association with field-measured tree species Shannon diversity, showing that dimensionality reduction improved the performance of the workflow (Féret and Asner, 2014; Féret and de Boissieu, 2020; Kacic and Kuenzer, 2022). Spectral indices showed a smaller and more site-dependent advantage over surface reflectance, while IQR filtering had little effect on the results.

The bootstrap confidence intervals also suggest that the stability of the relationship between spectral Shannon diversity and field-measured Shannon diversity varied among sites. Otepää showed the highest uncertainty, which may reflect its more heterogeneous and fragmented landscape structure. Even though the imagery was forest-masked, medium-resolution Sentinel-2 pixels near forest edges may still include mixed canopy and non-canopy signals, especially in smaller forest patches. Soomaa showed narrower confidence intervals, suggesting more stable estimates, but its wetland-dominated forests may have weakened the direct link between spectral diversity and field-measured diversity. Water and canopy moisture strongly influence reflectance in the near- and shortwave-infrared regions (Gao, 1996), and spectral diversity metrics can also be affected by vegetation cover, canopy structure, litter, soil, and other background signals rather than plant diversity alone (Hauser et al., 2021; Fassnacht et al., 2022). Rapla, with more managed and relatively drier forest stands, may have provided a clearer canopy signal, which could explain why it produced the strongest correlations across all input feature types.

Although PCA transformed surface reflectance performed best, this advantage should be interpreted with some caution. Since PCA was computed separately for each site, the retained components do not necessarily represent the same ecological gradients across Otepää, Soomaa, and Rapla. This is a known disadvantage of PCA-based spectral diversity workflows, where selected components may vary between images and require user interpretation or component selection before diversity mapping (Féret and Asner, 2014; Féret and de Boissieu, 2020). This limits strict between-site comparability of PCA-based spectral diversity maps and reduces their interpretability in multi-site applications. In contrast, spectral indices retain an ecological meaning wherever they are applied, because they represent the same vegetation- or water-related properties across sites. They can also be generated without site-specific component selection, making them easier to integrate into large-scale automated workflows. Although the processing time comparison showed only a slight advantage for spectral indices over PCA transformed surface reflectance, this difference may become more relevant over larger spatial extents, where repeated PCA computation and component selection could increase processing time. This remains to be tested through a formal benchmark across larger areas.

The practical advantage of spectral indices is also supported by Perrone et al. (2023), who found that the vegetation-index-based metric sdNDVI performed comparably to PCA-based Rao's Q in a landscape-scale analysis, suggesting that vegetation-index information can sometimes capture biodiversity-relevant gradients as effectively as more complex PCA-based spectral diversity metrics. Although Perrone et al. (2023) compared continuous spectral diversity metrics rather than spectral species directly, their finding supports the broader interpretation that index-based inputs can be useful when reproducibility, interpretability, and cross-site comparability are important.

The performance of spectral indices relative to surface reflectance was more site dependent. Spectral indices produced slightly stronger correlations in Soomaa and Otepää but performed similarly to surface reflectance in Rapla. This suggests that their advantage depends less on the use of indices alone and more on whether the selected index combination captures the dominant ecological gradients of each landscape. Previous studies have shown that different spectral indices and optical traits do not contribute equally to biodiversity estimation, because they emphasize different aspects of vegetation condition, canopy structure, pigment content, productivity, or moisture status (Torresani et al., 2021; Hauser et al., 2021; Gomarasca et al., 2024). In Soomaa, where hydrological and wetland-related gradients strongly structure forest conditions, indices linked to vegetation and moisture may provide a more meaningful summary of the spectral signal than in a more homogeneous managed forest landscape such as Rapla. This suggests that the value of spectral indices is partly landscape dependent. For example, Gomarasca et al. (2024) found that NIRv-based Rao's Q was less affected by saturation and bare soil contribution than NDVI-based Rao's Q, suggesting that alternative productivity-related indices such as NIRv may improve spectral diversity estimates in some ecosystems. Future work could test alternative combinations of spectral indices, including NIRv and moisture-sensitive indices, to determine whether other index sets provide a stronger correlation with field biodiversity. It would also be useful to examine whether combining spectral indices with canopy-level functional trait proxies improves the relationship with field-measured species diversity.

The strength of the observed relationships should also be interpreted in the context of the broader spectral-diversity literature. The moderate relationships observed here are consistent with earlier work (Torresani et al., 2024). Previous reviews of the SVH emphasized both the potential of spectral diversity as a biodiversity proxy and the strong dependence of reported relationships on ecosystem type, spatial scale, spectral input, and methodological choices (Fassnacht et al., 2022; Torresani et al., 2024; Wang and Gamon, 2019). For example, Schmidtlein and Fassnacht (2017) showed that the relationship between spectral variability and plant species richness varied strongly across space and season, while Perrone et al. (2023) found significant but relatively weak landscape-scale relationships between spectral diversity metrics and both species richness and functional diversity. Similarly, Fassnacht et al. (2022) emphasized that spectral variation can be influenced by several factors other than species richness, including phenology, physiological status, vegetation identity, and habitat composition. The moderate but consistent correlations observed in this study are in line with previous SVH research and should not necessarily be interpreted as poor performance. Rather, they indicate that spectral diversity products are most appropriate for mapping relative biodiversity patterns and supporting field-based monitoring, rather than replacing field observations directly (Wang and Gamon, 2019).

The limited effect of IQR filtering has practical implications for workflow design. After restricting the imagery to forested areas, this additional filtering step changed correlation values only minimally. In forest-only imagery, pixels flagged as outliers may still represent meaningful ecological variation rather than noise, for example differences in canopy structure, moisture conditions, or species composition. Excluding such pixels could also remove useful ecological signal and weaken the relationship between spectral and field-measured diversity. For this reason, and because the observed effect was negligible, IQR filtering may not be necessary in forest-masked biodiversity workflows unless the imagery contains stronger non-ecological spectral extremes.

Several methodological limitations should also be acknowledged. First, we used a single cluster number and a single moving window size for all three study sites. Different landscapes may benefit from different parameter settings, and the optimal cluster number or moving window size may vary with spectral complexity, forest structure, and average plot size (Féret and Asner, 2014; Kacic and Kuenzer, 2022; Robertson et al., 2023). Second, forest masking does not fully remove mixed-pixel effects. At Sentinel-2 resolution, pixels along forest boundaries may still contain spectral contributions from adjacent land-cover types or structurally different forest edges. This may influence spectral diversity estimates, particularly in fragmented landscapes or when field plots are close to patch boundaries. Although we used a 30 m inward buffer to exclude plots that were too small or irregular to contain at least one full moving window, zonal statistics were extracted from the retained plots using their original extents. A site-specific sensitivity analysis of cluster number and moving window size would therefore be a valuable next step.

Overall, these results show that input feature type selection is not only a technical preprocessing choice, but an important source of variation in spectral diversity mapping. PCA transformed surface reflectance provided the strongest within-site relationships with field-measured diversity, while spectral indices offered a more interpretable and transferable input option with performance advantage over surface reflectance in some landscapes. This distinction is relevant for open-source biodiversity monitoring workflows, where users often need to balance maximum local performance with reproducibility and cross-site comparability.

5. Conclusions

This study demonstrates that spectral input selection is a key parameter in Sentinel-2-based spectral diversity mapping of tree diversity. By comparing surface reflectance, PCA transformed surface reflectance, and spectral indices across three hemi-boreal forest landscapes, we show that PCA transformed surface reflectance produced the strongest and most consistent correlations with field-measured Shannon diversity. This highlights the value of PCA for extracting dominant biodiversity-relevant spectral variation while reducing redundancy in multispectral data. At the same time, spectral-index-based spectral diversity retained meaningful relationships with field diversity and offered practical advantages for reproducible, interpretable, and scalable workflows. Their performance varied among sites, indicating that the usefulness of index-based inputs depends on how well the selected indices capture local ecological gradients. IQR filtering had little effect, suggesting that additional outlier filtering may not improve results once cloud masking and forest masking are applied. Overall, we recommend PCA-based spectral species when the goal is to maximise site-level agreement with field observations, and spectral-index-based spectral species when transferability, automation, and ecological interpretability are priorities. Future work should test additional index combinations and canopy functional trait proxies to improve the predictability and robustness of spectral diversity mapping across forest landscapes.

Acknowledgements

This work was funded by Estonian Ministry of Education and Research, Centre of Excellence for Sustainable Land Use (TK232). This work contains modified Copernicus Sentinel data 2024.

References

Arold, I., 2005: Eesti maastikud. Tartu Ülikooli Kirjastus, Tartu.

Botero-Cañola, S., Wilson, K., Garcia, E., Heisey, M., Reeves, L.E., Burkett-Cadena, N.D., Romagosa, C., Sieving, K.E., Wisely, S.M., 2024: Acoustic indices track local vertebrate biodiversity in a subtropical landscape. Ecological Indicators, 166, 112292. doi.org/10.1016/j.ecolind.2024.112292.

Copernicus Land Monitoring Service, 2020: CORINE Land Cover 2018 vector, Europe, 6-yearly, version 2020_20u1. European Environment Agency. doi.org/10.2909/71c95a07-e296-44fc-b22b-415f42acfdf0.

Efron, B., 1979: Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1-26. doi.org/10.1214/aos/1176344552.

Estonian Environment Agency, 2025: Metsaportaal: Estonian Forest Register portal. register.metsad.ee/ (1 November 2025).

European Space Agency, 2021: Sentinel-2 MSI Level-2A Product Definition. Issue 4.9, S2-PDGS-MPC-L2A-PDD. European Space Agency. step.esa.int/thirdparties/sen2cor/2.10.0/docs/S2-PDGS-MPC-L2A-PDD-V14.9-v4.9.pdf (1 February 2026).

Fassnacht, F.E., Müllerová, J., Conti, L., Malavasi, M., Schmidtlein, S., 2022: About the link between biodiversity and spectral variation. Applied Vegetation Science, 25, e12643. doi.org/10.1111/avsc.12643.

Féret, J.-B., Asner, G.P., 2014: Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecological Applications, 24(6), 1289-1296. doi.org/10.1890/13-1824.1.

Féret, J.-B., de Boissieu, F., 2020: biodivMapR: an R package for α- and β-diversity mapping using remotely sensed images. Methods in Ecology and Evolution, 11(1), 64-70. doi.org/10.1111/2041-210X.13310.

Féret, J.-B., de Boissieu, F., 2026: Produce features from a raster. biodivMapR documentation, jbferet.github.io/biodivMapR/articles/biodivMapR_02.html (15 March 2026).

Gao, B.-C., 1996: NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257-266. doi.org/10.1016/S0034-4257(96)00067-3.

GeoPandas Development Team, 2025: GeoPandas v1.1.1. Zenodo. doi.org/10.5281/zenodo.15750510.

Gomarasca, U., Duveiller, G., Pacheco-Labrador, J., Ceccherini, G., Cescatti, A., Girardello, M., Nelson, J.A., Reichstein, M., Wirth, C., Migliavacca, M., 2024: Satellite remote sensing reveals the footprint of biodiversity on multiple ecosystem functions across the NEON eddy covariance network. Environmental Research: Ecology, 3(4), 045003. doi.org/10.1088/2752-664X/ad87f9.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017: Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27. doi.org/10.1016/j.rse.2017.06.031.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013: High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850-853. doi.org/10.1126/science.1244693. Data available online at: glad.earthengine.app/view/global-forest-change (17 February 2026).

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020: Array programming with NumPy. Nature, 585, 357-362. doi.org/10.1038/s41586-020-2649-2.

Hauser, L.T., Timmermans, J., van der Windt, N., Sil, Â.F., de Sá, N.C., Soudzilovskaia, N.A., van Bodegom, P.M., 2021: Explaining discrepancies between spectral and in-situ plant diversity in multispectral satellite earth observation. Remote Sensing of Environment, 265, 112684. doi.org/10.1016/j.rse.2021.112684.

Huete, A.R., 1988: A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309. doi.org/10.1016/0034-4257(88)90106-X.

Kacic, P., Kuenzer, C., 2022: Forest biodiversity monitoring based on remotely sensed spectral diversity - A review. Remote Sensing, 14(21), 5363. doi.org/10.3390/rs14215363.

Keskkonnaamet, 2016: Soomaa National Park. Keskkonnaamet. ISBN 978-9949-582-04-6.

Lee, C.B., Peralta Brichtova, A.C., Roca, M., Murray, T., Bolívar Rodríguez, O.D., Cerra, D., Muller-Karger, F.E., 2025: Comparing cloud mask products for seagrass mapping over Sentinel-2 imagery: toward a first national seagrass map for Venezuela. Journal of Geophysical Research: Machine Learning and Computation, 2(2), e2024JH000559. doi.org/10.1029/2024JH000559.

Palisaar, J., 2006: The floodplain meadows of Soomaa National Park, Estonia. Vegetation - Dispersal - Regeneration. Dissertation, Universität Regensburg. doi.org/10.5283/epub.10478.

Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., Wohlgemuth, T., 2002: Quantitative tools for perfecting species lists. Environmetrics, 13(2), 121-137. doi.org/10.1002/env.516.

Perrone, M., Di Febbraro, M., Conti, L., Divísek, J., Chytrý, M., Keil, P., Carranza, M.L., Rocchini, D., Torresani, M., Moudrý, V., Símová, P., Prajzlerová, D., Müllerová, J., Wild, J., Malavasi, M., 2023: The relationship between spectral and plant diversity: disentangling the influence of metrics and habitat types at the landscape scale. Remote Sensing of Environment, 293, 113591. doi.org/10.1016/j.rse.2023.113591.

QGIS Development Team, 2024: QGIS Geographic Information System. Open-Source Geospatial Foundation. qgis.org (1 June 2026).

Rifai, M., Harintaka, 2024: Integration of Cloud Score+ with Sentinel-2 Harmonized for land use and land cover classification using machine learning algorithms. IOP Conference Series: Earth and Environmental Science, 1418, 012039. doi.org/10.1088/1755-1315/1418/1/012039.

Robertson, K.M., Simonson, E., Ramirez-Bullon, N., Poulter, B., Carter, R., 2023: Effects of spatial resolution, mapping window size, and spectral species clustering on remote sensing of plant beta diversity using biodivMapR and hyperspectral imagery. Journal of Geophysical Research: Biogeosciences, 128(7), e2022JG007350. doi.org/10.1029/2022JG007350.

Shannon, C.E., Weaver, W., 1949: The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Schmidtlein, S., Fassnacht, F.E., 2017: The spectral variability hypothesis does not hold across landscapes. Remote Sensing of Environment, 192, 114-125. doi.org/10.1016/j.rse.2017.01.036.

Torresani, M., Rocchini, D., Sonnenschein, R., Zebisch, M., Hauffe, H.C., Heym, M., Pretto, F., Tonon, G., 2019: Estimating tree species diversity from space in an alpine conifer forest: the Rao's Q diversity index meets the spectral variation hypothesis. Ecological Informatics, 52, 26-34. doi.org/10.1016/j.ecoinf.2019.04.001.

Torresani, M., Feilhauer, H., Rocchini, D., Fava, F., Marcantonio, M., Ricotta, C., Tonon, G., 2021: Which optical traits enable an estimation of tree species diversity based on the Spectral Variation Hypothesis? Applied Vegetation Science, 24(2), e12586. doi.org/10.1111/avsc.12586.

Torresani, M., Rossi, C., Perrone, M., Hauser, L.T., Féret, J.-B., Moudrý, V., Simova, P., Ricotta, C., Foody, G.M., Kacic, P., Feilhauer, H., Malavasi, M., Tognetti, R., Rocchini, D., 2024: Reviewing the Spectral Variation Hypothesis: Twenty years in the tumultuous sea of biodiversity estimation by remote sensing. Ecological Informatics, 82, 102702. doi.org/10.1016/j.ecoinf.2024.102702.

Végh, L., Tsuyuzaki, S., 2021: Remote sensing of forest diversities: the effect of image resolution and spectral plot extent. International Journal of Remote Sensing, 42(15), 5985-6002. doi.org/10.1080/01431161.2021.1934596.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261-272. doi.org/10.1038/s41592-019-0686-2.

Wang, R., Gamon, J.A., 2019: Remote sensing of terrestrial plant biodiversity. Remote Sensing of Environment, 231, 111218. doi.org/10.1016/j.rse.2019.111218.