Received: 26 July 2023 | Revised: 19 January 2024 | Accepted: 21 February 2024

DOI: 10.1111/avsc.12778

Applied Vegetation Science IAVS

RESEARCH ARTICLE

Prediction of species richness and diversity in sub-alpine grasslands using satellite remote sensing and random forest machine-learning algorithm

Katlego Mashiane[1] | Abel Ramoelo[2,3] | Samuel Adelabu[2]

[1]Department of Geography, Natural and Agricultural Science, Phuthaditjhaba, South Africa

[2]Department of Geography, University of the Free State, Bloemfontein, South Africa

[3]Centre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Hatfield, South Africa

Correspondence
Katlego K. Mashiane, Department of Geography, Natural and Agricultural Science, QwaQwa Campus, Phuthaditjhaba 9866, South Africa.
Email: mashianekk@ufs.ac.za

Co-ordinating Editor: Duccio Rocchini.

Abstract

Aims: Remote-sensing approaches could be beneficial for monitoring and compiling essential biodiversity data because they are cost-effective and allow for coverage of large areas over a short period. This study investigated the relationship between multispectral remote-sensing data from Landsat 8 and Sentinel-2 and species richness and diversity in mountainous and protected grasslands.

Locations: Golden Gate Highlands National Park, Free State, South Africa.

Methods: *In-situ* data of plant species composition and cover from 142 plots with 16 relevés each were distributed across the study site and used to calculate species richness and Shannon–Wiener species diversity index (species diversity). We used a machine-learning random forest algorithm to optimize the prediction of species richness and diversity. The algorithm was used to identify the optimal spectral bands and vegetation indices for estimating species richness and diversity. Subsequently, the selected bands and vegetation indices were used to estimate species richness through random forest regression.

Results: This research found weak relationships between remote-sensing vegetation indices and the diversity metrics, but significant relationships were found between some spectral bands and diversity metrics. Moreover, using machine-learning random forest, the multispectral data sets exhibited strong predictive powers. In this investigation, near-infrared (NIR) seemed to be the most selected band for both sensors to explain species diversity in mountainous grasslands.

Main conclusions: This finding further ascertains the efficiency of optimizing high spatial resolution spectral information to estimate plant species richness and diversity. This research shows that NIR, Soil-Adjusted Vegetation Index (SAVI) and Enhanced Vegetation Index (EVI) are the most adequate for predicting species richness and diversity in mountainous grasslands with relatively good accuracies. Plant phenology and the choice of sensor affect the relationship between spectral information and species diversity variables.

KEYWORDS
biodiversity, conservation, grasslands, machine learning, remote sensing, species distribution modeling

1 | INTRODUCTION

Biodiversity is essential for maintaining ecosystem functioning and services and supporting human well-being (Oliver et al., 2015). Plant diversity in grassland ecosystems contributes to multiple ecosystem services at local and landscape scales, making conservation efforts within and among ecological communities imperative (Hautier et al., 2018). However, monitoring biodiversity at various scales is a common conservation challenge in protected areas (Ferreira et al., 2011). Different approaches to monitoring vegetation, that is, *in-situ* species and remote-sensing approaches, are complex with pros and cons; hence, a hybrid or coupling of procedures is necessary (Lausch et al., 2018). Remote-sensing approaches could be beneficial for monitoring essential biodiversity variables because they are cost-effective and allow for coverage of large areas over a short period in contrast to *in-situ* methods (Lausch et al., 2018). Consistent vegetation monitoring forms the basis of wildlife conservation, which is essential as environmental changes are initially observed in the vegetation (Brown et al., 2013).

Evidence of biodiversity loss and ecosystem functioning is compelling, but the issue remains contentious (Cardinale et al., 2012). As a significant conservation issue, biodiversity loss has received varied scientific reporting and views worldwide (Cardinale et al., 2018). For example, a recent synthesis of time-series data suggests species richness is decreasing in some locations and increasing in others but not changing on average (Cardinale et al., 2018). A wide range of ecological viewpoints regarding biodiversity loss exist, but few empirical tests exist (Naeem et al., 1995). Cardinale et al. (2018) argue that the lack of scientific evidence unequivocally supporting biodiversity loss is mainly attributable to (1) low-quality of data, and (2) lack of spatial representation and failure to account for the main drivers, especially biological invasion (Cardinale et al., 2018). Quantifying and predicting biodiversity's spatial and temporal distribution has become increasingly important, especially given global change (Oliver et al., 2015).

Plant species richness and diversity are key ecosystem indicators because these diversity metrics can explicitly observe ecosystem health (Symstad & Jonas, 2011). They also describe ecosystem health, stability, and resilience and can be used for monitoring plant species (Lausch et al., 2018), especially in conservation areas. Hence, the Group of Earth Biodiversity Observation Network identifies taxonomic diversity as one of the vital biodiversity variables (Pereira et al., 2013). Remote-sensing techniques could measure vegetation characteristics that are suitable for discriminating species turnover and floristic composition (Lausch et al., 2018). Determining species using remote sensing depends on multiple biological and physical factors, appropriate data, and modeling algorithms (Richter et al., 2016). However, remote-sensing approaches could be beneficial for monitoring essential biodiversity variables because they are cost-effective and allow for coverage of large areas with a quick turnaround time. Studies demonstrated an increase in the strength of the relationship between species alpha diversity and remotely sensed spectral heterogeneity when accounting for species' relative abundances. This improved the capability of local species diversity estimations, especially while using spectral information in addition to the commonly used spectral indices (Rocchini et al., 2007).

Multispectral sensors have limitations on properties such as species identification; for example, they are inferior to hyperspectral data in spectral information (Lyon & Huete, 2016). As such, alpha diversity is commonly predicted and mapped based on the spectral variation hypothesis (SVH), which starts with a heterogeneity map from a satellite sensor image correlated with field sampling data (Rocchini et al., 2007, 2016). In grassland studies, the spectral properties of grass species may be difficult to detect because of the similarity in the taxa, mainly due to broadband remote sensing's inability to identify slight differences in green vegetation (Lyon & Huete, 2016). Furthermore, weak and moderate relationships between species richness, diversity, and spectral vegetation indices (VIs) are derived from remote sensors (Rocchini et al., 2007). This seemingly universal observation may be because VIs are poorly related to structural properties. For example, a study by Ingram et al. (2005) excluded the Normalized Difference Vegetation Index (NDVI) from the predictive analysis because of its weak correlation with species' structural features. However, the relationship between remotely sensed data and species diversity is scale-dependent (Rocchini et al., 2007). Sentinel-2 Multispectral Instrument (MSI) has shown enormous potential for vegetation mapping because it is one of the multispectral sensors that can acquire images with 13 spectral bands (Torresani et al., 2019), including red-edge bands (Xulu et al., 2021). In addition, its free and open data policy may benefit nature conservation in developing and under-resourced countries (Torresani et al., 2019).

Remote sensing has had some success as a tool for predicting and detecting species richness and diversity; this success is mainly due to the application of NDVI. However, predicting species richness and diversity remains a significant application for remote sensing in managing biodiversity and ecosystems. Landsat carries two sensors, that is, the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). Landsat OLI measures the visible, near-infrared (NIR) and short-wave infrared spectral information related to plant properties, especially chlorophyll (Nagendra et al., 2010). In contrast, the Sentinel-2 MSI sensor may improve the accuracy of explaining plant biodiversity variables by including red-edge bands that allow for chlorophyll characterization and computation of novel indices for biodiversity mapping (Puletti et al., 2018).

Despite the availability of remote-sensing data and applications, their use has yet to be widely incorporated by managers and researchers working in biodiversity monitoring (Reddy, 2021). Many studies have been conducted to map species diversity using remote-sensing data sets and study designs, resulting in varying outcomes and accuracies (Schmidtlein and Fassnacht, 2017). Data sets of coarser spatial resolution perform better when estimating species diversity than data sets of finer spatial resolution (Gessner et al., 2015). Nevertheless, increased spectral resolution is deemed beneficial for improving the estimation of species diversity (Xulu et al., 2021); however, hyperspectral remote-sensing data are cumbersome, and most of their bands are redundant (Vihervaara et al., 2017). Methodological advances such as machine-learning algorithms may present significant opportunities for utilizing multispectral data sets with high accuracies. The accuracy difference between hyperspectral and multispectral bands could be much higher for grass species (Gessner et al., 2015).

The loss of biodiversity is now more prevalent for many biomes across the globe due to overexploitation and land use transformations; thus, monitoring and modeling biodiversity on a local scale via remote sensing can aid in abating this crisis (Cardinale et al., 2018). Monitoring patterns of species diversity over time is essential for decision-making in conservation, especially in the context of unprecedented global change. Remote sensing is one of the most cost-effective approaches to identifying biodiversity hotspots and predicting changes in good time. This study aims to analyze different satellite sensors to predict species richness and diversity.

2 | METHODS

2.1 | Study area

The study was conducted in the Golden Gate Highlands National Park (GGHNP) in the northeastern Free State province, South Africa (Figure 1). The park comprises 32,758.35 ha and lies in the range 28°27′–28°37′ S and 28°33′–28°42′ E. The park is in mountainous grasslands at the foothills of the Drakensberg and forms part of the mesic highveld grassland with marked variation in geology, topography, and rainfall. The soil types in the park include shallow rocky soils (Glenrosa and Mispah), deep drainage lines (Oakleaf), well-developed sand soils (Hutton and Clovelly), and clayey structured soils (Milkwood and Tambakulu)

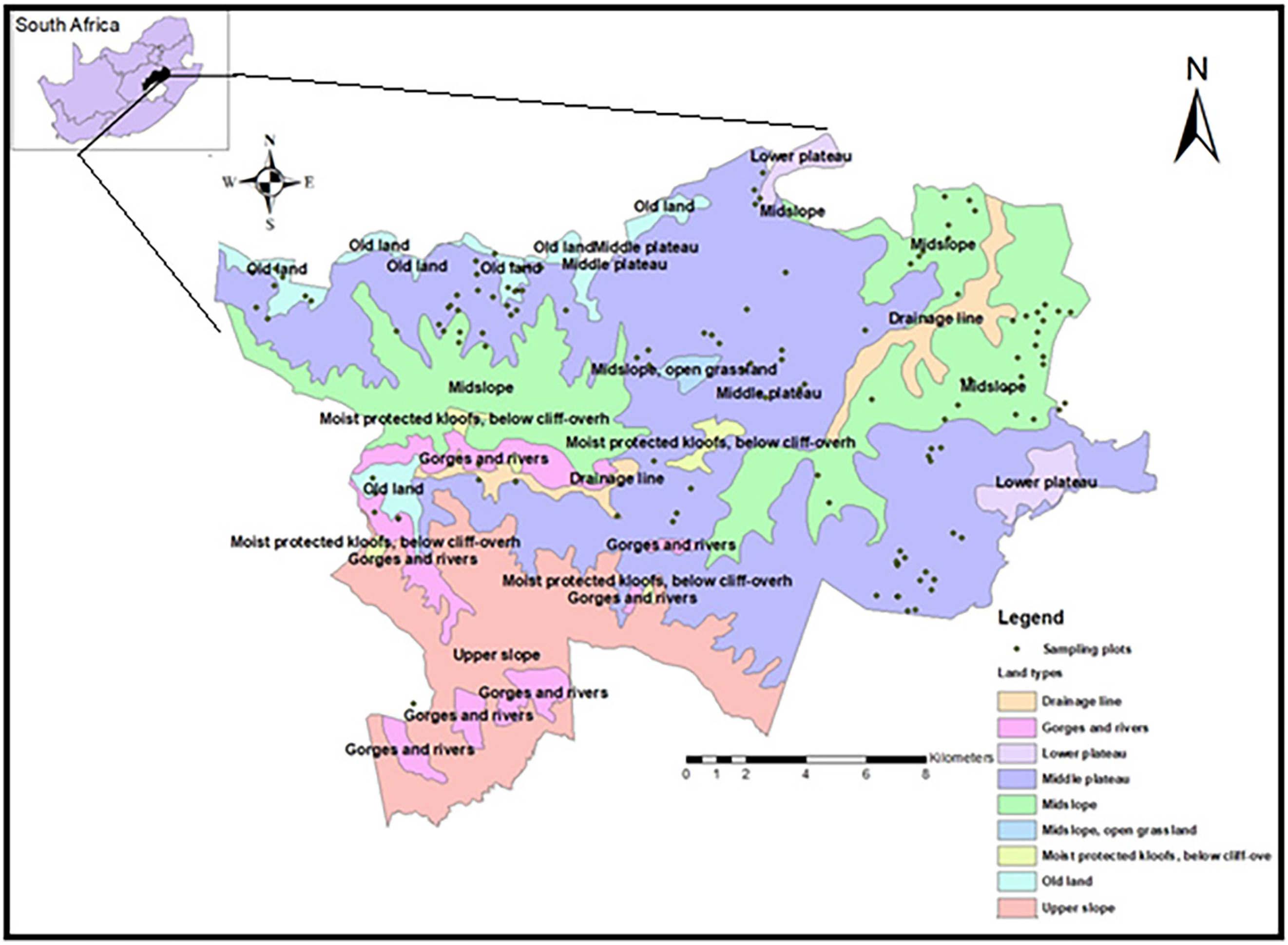

FIGURE 1 Map of the study area.

(SANParks, 2020). The park is characterized by summer rainfall, temperate summers, and cold winters. The rainfall season stretches from September to April, with a mean annual rainfall ranging from 800 mm to 2000 mm. The park lies between 1892 m and 2829 m a.s.l. and comprises the following subalpine mesic grasslands units: Eastern Free State sandy grasslands (Gm 4), Basotho montane shrubland (Gm 5), Lesotho highveld basalt grassland (Gd 8), and Northern Drakensberg highveld (Gd5) (Mucina & Rutherford, 2006). The park is home to multiple antelope species and is used for grazing by domesticated cattle and wild animals.

2.2 | Field data collection

The land type map of GGHNP was used as the first stratification. Sampling sites of homogenous grass patches were then located in a randomly stratified manner. Thirty-sic vegetation sampling plots (30 m × 30 m) ranging between three and five per site (six sites with five plots and two sites with three plots; Figure 2a) were placed randomly within the homogenous grass patches. Sixteen (1 m × 1 m) quadrats were set systematically within each sampling plot at every 10 m along four parallel rows (Figure 2). The above data set was merged with another data set from a different sampling program comprising 12 sites with 106 plots. The latter was collected using a 100-step points method from four transects located within a plot of 30 m × 30 m, and all species were recorded at every step point (Figure 2b). The standard data set comprised 142 plots with 13 sites, seven of which were the same and six different. The distribution of sampling plots across the land types in GGHNP is given in Figure 3.

The taxonomic composition and cover of the vegetation were used to derive species richness and diversity per plot. The values from each plot were averaged to attain mean species richness per site (Table 1). Species richness and diversity were computed using the statistical packages *vegan* and *plyr* in R studio, which employed the diversity and an applied function for species diversity and richness, respectively (Oksanen, 2017). Species diversity was calculated using the Shannon–Wiener Index (Equation 1), where p_i is the proportion of the species within the sampling units. Species richness was determined by adding all species from each quadrat and averaging by the number of quadrats in each plot to obtain the average plot value (Oksanen, 2017):

$$H' = -\sum p_i{}^* \ln p_i. \tag{1}$$

2.3 | Remote-sensing data collection

Satellite images from Sentinel-2 and Landsat 8 data sets were extracted and processed from the JavaScript code editor Google Earth Engine (GEE). All images in this research were nearly cloud-free. The mean spectral land surface reflectance value of images was filtered using monthly dates from January to March and the average of January to March to assess phenology's effects on species diversity's predictiveness. From the extracted spectral images of both sensors, vegetation indices [NDVI, Soil-Adjusted Vegetation Index (SAVI), Simple Ratio (ST), and Enhanced Vegetation Index (EVI) in Equations 1–4] were calculated within GEE.

$$\text{NDVI} = (\text{NIR} - R)/(\text{NIR} + R), \tag{2}$$

$$\text{SAVI} = \left[(\text{NIR} - R)/(\text{NIR} + R + L)\right]^* (1 + \text{L}), \tag{3}$$

$$\text{SR} = \text{NIR}/\text{Red}, \tag{4}$$

$$\text{EVI} = G^* \left[(\text{NIR} - R)/(\text{NIR} + \text{RED}^* R - \text{C2}^* B + L)\right]. \tag{5}$$

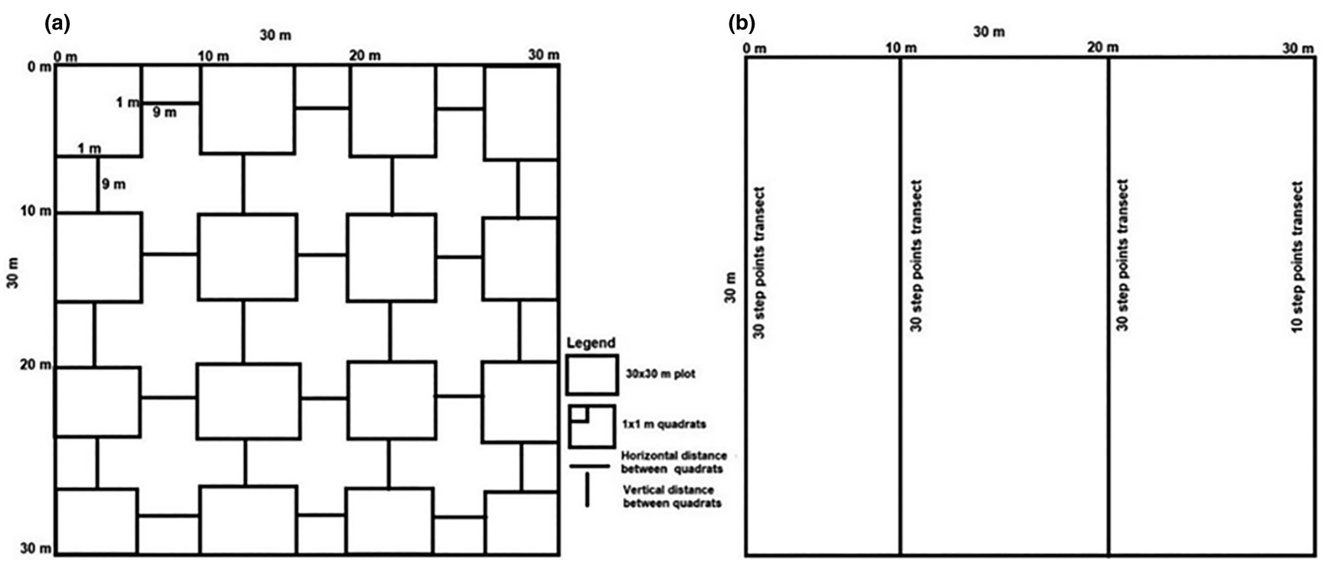

FIGURE 2 (a) Quadrat and (b) step-point-based vegetation plot design.

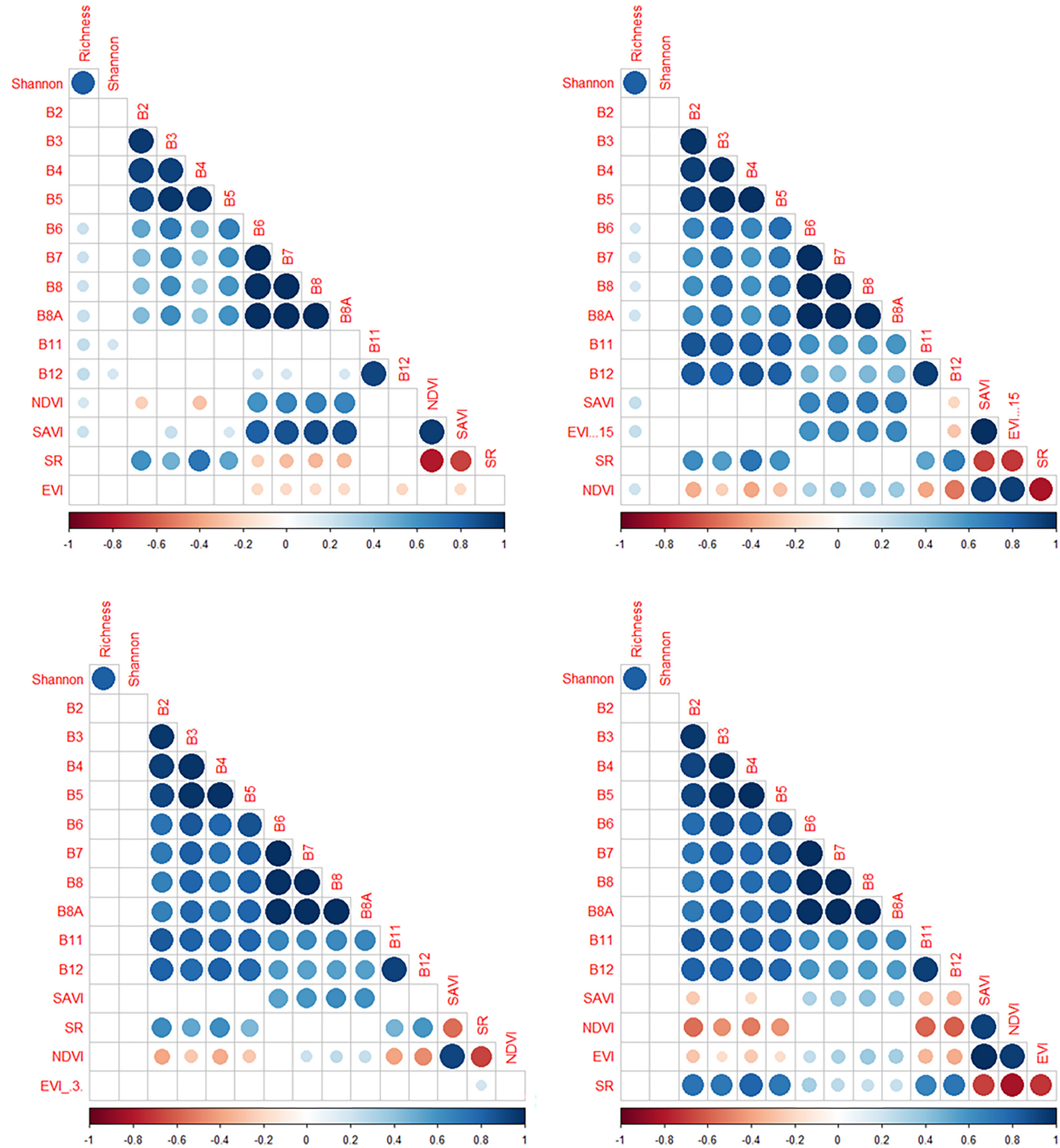

FIGURE 3 Correlation coefficients between species richness (SR) and diversity (SW) and vegetation were calculated from all possible combinations of Sentinel-2 images in January (top left), February (top right), March (bottom left), and January–March (bottom right) (crossed-out correlation mean the relationship is not significant)

TABLE 1 Descriptive statistics of average species richness and diversity (H') in the study area per plot.

	Number of samples	Minimum	Maximum	Median	Mean	Variance	Standard deviation	Coefficient of variance
Species richness	142.00	3.00	17.00	9.00	8.97	12.11	3.48	0.39
Species diversity (H')	142.00	0.64	2.45	1.63	1.63	0.16	0.40	0.25

All the spectral images and indices were imported into RStudio, and the GPS coordinates of each sampling plot were used to obtain values. This was achieved using the extract function from the raster library in RStudio.

3 | DATA ANALYSIS

A correlogram depicting Pearson correlations between species richness and diversity, S2 MSI bands, and VIs, including the averages of January–March, was computed to show relationships between explanatory and response variables. Random forest (RF) is a machine-learning algorithm that improves the regression and classification trees by combining a large set of decision trees. In this study, this technique was used to select the optimal variables that can be used to estimate species richness and diversity. Subsequently, a set of selected optimal variables was input into the RF model to predict species richness and diversity.

Three parameters were used in the RF model; the number of regression trees (n_{tree}) is based on the value of the observations called bootstrap sample (500 in this study). The m_{try}, which refers to the number of predictors to be tested at each node, was set at the square root of the input variables used in the model, which is 4.

To validate the performance of the RF regression model, the data set was split into 70% for training and 30% for testing, respectively. The training data set was used to develop an RF model that could estimate the response variables, while the test data set was used to validate the final model. A one-to-one relationship between the observed and predicted species diversity and the richness of test and training data was fitted for cross-validation. The coefficient of determination (R^2) and root-mean-squared error (RMSE) were used to assess the predictive performance.

4 | RESULTS

There was a positive, albeit weak, relationship between species richness, Sentinel-2 bands (6–12) and indices NDVI and SAVI, while for January species diversity was related to bands 11 and 12 only. In February, none of the Sentinel-2 bands and indices had a relationship with species diversity; however, this relationship was positive for species richness. There was no relationship between species richness, diversity, Sentinel-2 bands, and vegetation indices for March and the average of January to March (Figure 3).

The variables of interest, that is species richness and diversity, exhibited a positive, albeit moderate (ranging $R=0.2$–0.3) relationship with all Landsat 8 OLI bands for January. This relationship further weakened in February ($R=0.1$–0.2) and March, where the red band (Band 4) was the only band relating to the species diversity for the former month. The relationship between species diversity, richness, Landsat 8 bands, and indices strengthened again when monthly band averages were used. Notably, the Landsat 8-derived indices were separate from the variables of interest (Figure 4).

4.1 | Predicting species richness and diversity

The selected variables needed for optimal species richness and diversity prediction had notably low R^2 and high RMSE. The significant Landsat 8 variables (Table 2) optimally explaining species richness were EVI for January (RMSE = 3.621, $R^2=0.044$) and February (RMSE = 3.493, $R^2=0.080$), and NIR for March (RMSE = 4.126, $R^2=0.008$) and January–March (RMSE = 3.935, $R^2=0.006$). Moreover, the selected Sentinel-2 variables explaining species richness were SR for January (RMSE = 3.928, $R^2=0.009$), NIR for February (RMSE = 3.872, $R^2=0.001$), red-edge for March (RMSE = 3.711, $R^2=0.001$), and NIR for January–March (RMSE = 3.657, $R^2=0.014$). The Landsat 8 variables explaining species diversity were EVI for January and February, SAVI for March, and NIR for January–March. The Sentinel-2 variables for explaining species diversity were red-edge 1 for January, February, and January–March, and SR for March. In January, the most significant band selected was EVI.

The RF model of Landsat-8 variables explained 87% of the species diversity in January, 89% in February, 90% in March and 79% in January–March. For species richness, it explained 90% in January, 88% in February and 87% in March and 81% in January–March. On the other hand, the species richness and diversity variation explained by Sentinel-2 variables ranged between 82% and 91% (Table 3). The test exhibited similar predictive performance to the training data set across the two satellites for most months.

5 | DISCUSSION

The relationship between species counts and spectral information varies based on the spatio-temporal dynamics of the area under study (Schmidtlein & Fassnacht, 2017). Our study corroborates that this relationship weakens with the end of summer. Spectral bands are seldom tested individually for their relationship with species counts; mainly spectral variability based on band derivatives is used (Rocchini et al., 2007; Schmidtlein & Fassnacht, 2017). Our study shows that the spatial resolution of a sensor does not lead to a positive relationship between spectral information and species count variables, nor does it affect the predictive performance; instead, Landsat 8 bands are related better to species richness and diversity than Sentinel-2's. This may be due to saturation issues associated with the relatively low radiometric resolution of sensors. Notably, the vegetation indices derived from Landsat 8 bands had no relationship with species richness and diversity contrary to those of Sentinel-2. Normally, the dispersion measures of conventional vegetation indices are used as spectral information to relate to species count variables (Rocchini et al., 2010).

The response of species richness and diversity to ecological drivers often reflects diverse outcomes, depending on external factors in an ecosystem and the measuring methods (Symstad & Jonas, 2011). However, for these diversity metrics to be ideal, grassland indicator estimation methods ought to be improved and augmented (Symstad & Jonas, 2011). Machine-learning algorithms allow for exploring

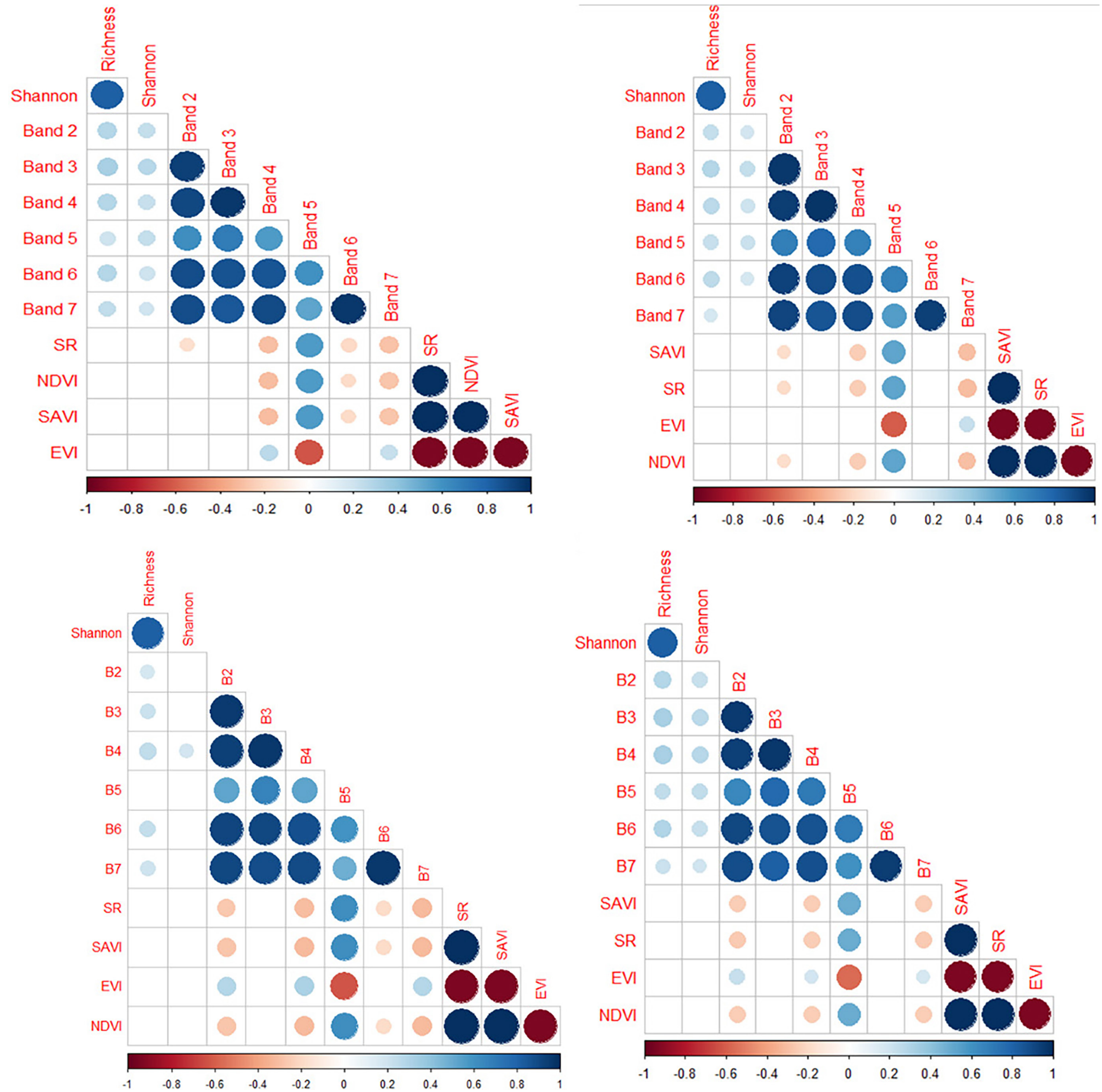

FIGURE 4 Correlation coefficients between species richness (SR) and diversity (SW) and vegetation were calculated from all possible combinations of Landsat 8 images in January (a), February (b), March (c), and January–March (d) (a crossed-out correlation mean the relationship is not significant).

automated environmental monitoring using these diversity metrics and remote-sensing data. Our research used the RF approach using species richness and diversity plot measurements corresponding to remote-sensing bands of Landsat 8, Sentinel-2, and VIs derived therefrom. The RF features selection of selected EVI, NIR, and SAVI as optimal remote-sensing variables, explaining species richness and diversity. This is not surprising given the moderate predictive performance of RF machine-learning algorithms on data sets with plant species richness (Adjorlolo & Botha, 2015). Furthermore, the spectral band selected in our study was consistent with the findings of other research where NIR explained 41% of the variation in species richness (Rocchini et al., 2007).

The RF model in this study managed to predict species richness and diversity with relatively high accuracies. This shows that using different model techniques can improve the predictive power of satellite remote-sensing (SRS) variables (Rocchini et al., 2007). Machine-learning SRS models are essential for species monitoring and aiding conversation in protected areas because of their cost effectiveness and present novel approaches to identifying biodiversity hotspots and predicting changes (Rocchini et al., 2016). This

TABLE 2 Optimal Landsat 8 and Sentinel-2 variables explaining species richness and diversity.

	Species richness			Species diversity			
Selected bands	RMSE	R^2	MAE	Month	RMSE	R^2	MAE
Landsat 8							
January				*Jan*			
NIR	3.386	0.04008	3.063	NIR	0.4485	0.005913	0.3631
ST	3.631	0.04269	3.015	SR	0.4367	0.001187	0.3547
EVI	3.621	0.04362	3.019*	EVI	0.4281	0.008509	0.349*
February				*Feb*			
NIR	3.462	0.001955	3.228	NIR	0.4735	0.07716	0.3788
SAVI	3.511	0.053272	2.949	SR	0.4666	0.06849	0.3768
EVI	3.493	0.080437	2.859*		0.4632	0.06849	0.3753*
Mar				*Mar*			
NIR	4.126	0.008531	3.456*	NIR	0.465	0.03137	0.3872
SAVI	4.109	0.027952	3.445	SR	0.4521	0.0136	0.3735*
EVI	4.087	0.028821	3.424	SAVI	0.458	0.03173	0.3782
Jan-Mar				*Jan-Mar*			
NIR	3.935	0.006241	3.287*	NIR	0.4244	0.0192	0.3519*
SR	4.002	0.006201	3.445	SR	0.4343	0.01147	0.3563
NDVI	4.01	0.011824	3.457	SAVI	0.4274	0.01579	0.3543
Sentinel-2							
January				*January*			
Red-edge 1	3.928	0.009438	3.34	Red-edge 1	0.4366	0.000819	0.3656*
Red-edge 4	3.936	0.026493	3.369	Red-edge 4	0.447	0.00168	0.3679
SR	3.873	0.010987	3.319*	SR	0.4404	0.0000055	0.3645
February				*February*			
Red-edge 1	3.872	0.001696	3.251*	Red-edge 1	0.4411	0.0000271	0.3464*
Red-edge 4	3.968	0.02039	3.374	Red-edge 4	0.4504	0.0010034	0.3587
SR	3.983	0.0293	3.43	SR	0.4428	0.0002043	0.3515
March				*March*			
Red-edge 1	3.733	0.001088	3.137	Red-edge 1	0.4303	0.007374	0.3534
Red-edge 4	3.711	0.001685	3.074*	Red-edge 4	0.4286	0.006517	0.3502
SWIR	3.77	0.000864	3.135	SR	0.4231	0.011342	0.3438*
Jan-Mar				*Jan-Mar*			
Red-edge 1	3.657	0.014519	3.076*	Red-edge 1	0.4325	0.0028	0.3593*
Red-edge 4	3.677	0.003288	3.079	Red-edge 4	0.4392	0.00000309	0.3671
SR	3.76	0.001014	3.146	SR	0.4366	0.000513	0.3631

Abbreviations: EVI, Enhanced Vegetation Index; MAE, Mean Absolute Error; NDVI, Normalized Difference Vegetation Index; NIR, near-infrared; RMSE, root-mean-squared error; SAVI, Soil-Adjusted Vegetation Index; ST, Simple Ratio; SWIR, Shortwave Infrared.

*, $p < 0.05$.

research shows that machine-learning algorithms can improve our predictions of plant alpha diversity compared to the commonly used SVH, providing an alternative for species mapping (Appendices S1–S4). Moreover, the study shows that the machine-learning models for biodiversity mapping may not require satellite sensors with high spatial and spectral resolution.

A wide range of relationships between satellite-based VIs and vegetation characteristics have been established, with low to moderate predictive performance (Haboudane, 2004). Our research corroborates studies that propose using VIs in addition to spectral bands to improve the predictiveness of species diversity metrics (Rocchini et al., 2016). This is mainly because when used alone, VIs yield poor predictive performance and correlations despite the enhanced spatial resolution of a satellite sensor; hence, our study elucidates the use of machine learning and relevant spectral bands, and VIs improve the estimation of species richness and diversity. Even

TABLE 3 Random forest regression for predicted species richness, diversity, and remote-sensing data sets.

	Species richness				Species diversity			
	Landsat 8		Sentinel-2- MSI		Landsat 8		Sentinel-2-MSI	
	Training	Test	Training	Test	Training	Test	Training	Test
Number of plots	99	43	99	43	99	43	99	43
January								
R^2	0.90	0.88	0.90	0.80	0.87	0.86	0.85	0.86
RMSE	1.477	1.558	1.614	1.78	0.182	0.197	0.201	0.201
February								
R^2	0.88	0.87	0.84	0.92	0.89	0.90	0.89	0.89
RMSE	1.626	1.722	0.178	1.646	0.189	0.178	0.152	0.161
March								
R^2	0.87	0.88	0.91	0.91	0.90	0.88	0.91	0.91
RMSE	1.704	1.871	1.544	1.539	0.186	0.209	0.167	0.175
January–March								
R^2	0.81	0.88	0.88	0.85	0.79	0.85	0.82	0.8343
RMSE	1.712	2.008	1.788	1.746	0.203	0.219	0.195	0.188

Abbreviations: MSI; RMSE, root-mean-squared error.

so, the differences in predictive accuracy between the Sentinel-2 and Landsat 8 sensors was not substantial; this is not surprising since the two sensors are not too distinct spectrally and spatially. This finding aligns with studies that postulate multispectral sensors with relatively moderate to high spatial resolution could be good candidates for biodiversity mapping (Rocchini et al., 2016).

The relationship between spectral bands, VIs, and species diversity indices somewhat depends on the species diversity metric (Oldeland et al., 2010; Rocchini et al., 2016). However, our study demonstrated little to no effect in using different diversity metrics; opposing studies suggesting that using the Shannon–Wiener index improves the predictive performance by three-fold compared to species richness (Oldeland et al., 2010; Rocchini et al., 2016). Nonetheless, the prediction accuracy was highest in both January (richness) and March (diversity) for Landsat and March (richness and diversity for Sentinel-2). The estimation accuracies increased with deteriorating phenology for Sentinel-2, rendering it beneficial for mapping grass species diversity in senescence. This is because of its strategically positioned spectral bands, especially the inclusion of red-edge bands, making it helpful in studying vegetation characteristics (Thenkabail et al., 2004).

- This research explored satellite remote sensing as a primary tool for identifying biodiversity hotspots in South Africa's mountainous grasslands and predicting changes. The RF remote-sensing model predicted species richness and diversity with relatively high accuracy. These models present an opportunity for plant species monitoring using remote sensing, which has always been associated with many challenges concerning species diversity monitoring (Rocchini et al., 2016). Remote sensing is a cost-effective and less labor-intensive tool for biodiversity management, and its development is imperative for monitoring the inevitable consequences of global environmental change. Previously, species richness at local scales was studied using SVHs. However, the RF models in this study provide better estimates of plant species richness than the proposed SVH (Rocchini et al., 2007, 2018). In contrast to SVH, the RF predictive models do not require remote-sensing sensors with high spatial and spectral resolution. SVH starts with a heterogeneity map correlated with field sampling data for estimation models.

6 | CONCLUSION AND RECOMMENDATIONS

Determining species to establish relationships between remote-sensing data and species may be difficult because of subtle differences in spectral signature measures among species. The type of sensor and modeling algorithms has always limited remote-sensing approaches; however, it is demonstrated in this research that advancement in technology could enable species quantification and monitoring efficiently. NIR, the selected spectral band for predicting species richness and diversity, remains the ideal band for vegetation monitoring using remote sensing. This selection also augments NIR as the spectral band that allows species discrimination related to species traits, especially chlorophyll, which can also be measured using NIR-based vegetation indices and, despite having relatively low spatial and spectral resolution, Landsat 8 bands yielded impressive modeling accuracy virtually comparable to those of Sentinel-2. For future research, we suggest testing remote-sensing images with very high spectral and spatial resolution and special uncrewed aerial vehicles for species diversity mapping and incorporating terrain in the geospatial models, especially in areas such as GGHNP with complex mountainous terrain.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Google Earth Engine at https://code.earthengine.google.com/0a7251d85e04c56d261069189cbc17ff.

ORCID

Katlego Mashiane https://orcid.org/0000-0001-5676-9554

REFERENCES

Adjorlolo, C. & Botha, J.O. (2015) Integrating remote sensing and conventional models for modeling grazing/browsing capacity in southern African savannas. *Journal of Applied Remote Sensing*, 9(1), 96041. Available from: https://doi.org/10.1117/1.JRS.9.096041

Brown, L.R., du Preez, P.J., Bezuidenhout, H., Bredenkamp, G.J., Mostert, T.H.C. & Collins, N.B. (2013) Guidelines for phytosociological classifications and descriptions of vegetation in southern Africa. *Koedoe*, 55(1), 1103. Available from: https://doi.org/10.4102/koedoe.v55i1.1103

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P. et al. (2012) Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59–67. Available from: https://doi.org/10.1038/nature11148

Cardinale, B.J., Gonzalez, A., Allington, G.R.H. & Loreau, M. (2018) Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. *Biological Conservation*, 219, 175–183. Available from: https://doi.org/10.1016/j.biocon.2017.12.021

Ferreira, S., Deacon, A., Sithole, H., Bezuidenhout, H., Daemane, M. & Herbst, M. (2011) From numbers to ecosystems and biodiversity: a mechanistic approach to monitoring. *Koedoe*, 53(2), 1–12. Available from: https://doi.org/10.4102/koedoe.v53i2.998

Gessner, U., Machwitz, M., Esch, T., Tillack, A., Naeimi, V., Kuenzer, C. et al. (2015) Multi-sensor mapping of west African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data. *Remote Sensing of Environment*, 164, 282–297. Available from: https://doi.org/10.1016/j.rse.2015.03.029

Haboudane, D. (2004) Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. Available from: https://doi.org/10.1016/j.rse.2003.12.013

Hautier, Y., Isbell, F., Borer, E.T., Seabloom, E.W., Harpole, W.S., Lind, E.M. et al. (2018) Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. *Nature Ecology & Evolution*, 2(1), 50–56. Available from: https://doi.org/10.1038/s41559-017-0395-0

Ingram, J.C., Dawson, T.P. & Whittaker, R.J. (2005) Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, 94(4), 491–507. Available from: https://doi.org/10.1016/j.rse.2004.12.001

Lausch, A., Bastian, O., Klotz, S., Leitão, P.J., Jung, A., Rocchini, D. et al. (2018) Understanding and assessing vegetation health by in situ species and remote-sensing approaches. *Methods in Ecology and Evolution*, 9(8), 1799–1809. Available from: https://doi.org/10.1111/2041-210X.13025

Lyon, J.G. & Huete, A. (2016) In: Thenkabail, P.S. & Lyon, J.G. (Eds.) *Hyperspectral remote sensing of vegetation*. Boca Raton: CRC Press. Available from: https://doi.org/10.1201/b11222

Mucina, L. & Rutherford, M.C. (2006) The vegetation of South Africa, Lesotho and Swaziland. *Strelitzia*, 19, 1–30.

Naeem, S., Thompson, L.J., Lawler, S.P., Lawton, J.H. & Woodfin, R.M. (1995) Empirical evidence that declining species diversity may Alter the performance of terrestrial ecosystems. *Philosophical Transactions: Biological Sciences*, 347, 55946.

Nagendra, H., Rocchini, D., Ghate, R., Sharma, B. & Pareeth, S. (2010) Assessing plant diversity in a dry tropical forest: comparing the utility of landsat and ikonos satellite images. *Remote Sensing*, 2(2), 478–496. Available from: https://doi.org/10.3390/rs2020478

Oksanen, J. (2017) Vegan: ecological diversity. R Package Version 2.4-4, 11.

Oldeland, J., Dorigo, W., Lieckfeld, L., Lucieer, A. & Jürgens, N. (2010) Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery. *Remote Sensing of Environment*, 114(6), 1155–1166. Available from: https://doi.org/10.1016/j.rse.2010.01.003

Oliver, T.H., Heard, M.S., Isaac, N.J.B., Roy, D.B., Procter, D., Eigenbrod, F. et al. (2015) Biodiversity and resilience of ecosystem functions. *Trends in Ecology & Evolution*, 30(11), 673–684. Available from: https://doi.org/10.1016/j.tree.2015.08.009

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J. et al. (2013) Essential biodiversity variables. *Science*, 339(6117), 277–278. Available from: https://doi.org/10.1126/science.1229931

Puletti, N., Chianucci, F. & Castaldi, C. (2018) Use of Sentinel-2 for forest classification in Mediterranean environments. *Annals of Silvicultural Research*, 42(1), 32–38. Available from: https://doi.org/10.12899/ASR-1463

Reddy, C.S. (2021) Remote sensing of biodiversity: what to measure and monitor from space to species? *Biodiversity and Conservation*, 30(10), 2617–2631. Available from: https://doi.org/10.1007/s10531-021-02216-5

Richter, R., Reu, B., Wirth, C., Doktor, D. & Vohland, M. (2016) The use of airborne hyperspectral data for tree species classification in a species-rich central European forest area. *International Journal of Applied Earth Observation and Geoinformation*, 52, 464–474. Available from: https://doi.org/10.1016/j.jag.2016.07.018

Rocchini, D., Bacaro, G., Chirici, G., Da Re, D., Feilhauer, H., Foody, G.M. et al. (2018) Remotely sensed spatial heterogeneity as an exploratory tool for taxonomic and functional diversity study. *Ecological Indicators*, 85(December 2017), 983–990. Available from: https://doi.org/10.1016/j.ecolind.2017.09.055

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S. et al. (2010) Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. *Ecological Informatics*, 5(5), 318–329. Available from: https://doi.org/10.1016/j.ecoinf.2010.06.001

Rocchini, D., Boyd, D.S., Féret, J., Foody, G.M., He, K.S., Lausch, A. et al. (2016) Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation*, 2(1), 25–36. Available from: https://doi.org/10.1002/rse2.9

Rocchini, D., Ricotta, C. & Chiarucci, A. (2007) Using satellite imagery to assess plant species richness: the role of multispectral systems. *Applied Vegetation Science*, 10(3), 325–331. Available from: https://doi.org/10.1111/j.1654-109X.2007.tb00431.x

SANParks. (2020) Golden gate highlands national park park management plan. In: *South African National Parks*. Pretoria: SANParks.

Schmidtlein, S. & Fassnacht, F.E. (2017) The spectral variability hypothesis does not hold across landscapes. *Remote Sensing of Environment*, 192, 114–125. Available from: https://doi.org/10.1016/j.rse.2017.01.036

Symstad, A.J. & Jonas, J.L. (2011) Incorporating biodiversity into rangeland health: plant species richness and diversity in Great Plains grasslands. *Rangeland Ecology & Management*, 64(6), 555–572. Available from: https://doi.org/10.2111/REM-D-10-00136.1

Thenkabail, P.S., Enclona, E.A., Ashton, M.S. & Van Der Meer, B. (2004) Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91(3–4), 354–376. Available from: https://doi.org/10.1016/j.rse.2004.03.013

Torresani, M., Rocchini, D., Sonnenschein, R., Zebisch, M., Marcantonio, M., Ricotta, C. et al. (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. Available from: https://doi.org/10.1016/j.ecoinf.2019.04.001

Vihervaara, P., Auvinen, A.P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S. et al. (2017) How essential biodiversity variables and remote sensing can help national biodiversity monitoring. *Global Ecology and Conservation*, 10, 43–59. Available from: https://doi.org/10.1016/j.gecco.2017.01.007

Xulu, S., Mbatha, N. & Peerbhay, K. (2021) Burned area mapping over the southern cape forestry region, South Africa using sentinel data within gee cloud platform. *ISPRS International Journal of Geo-Information*, 10(8), 511. Available from: https://doi.org/10.3390/ijgi10080511

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Random forest predicted species richness for January (top left), February (top right), March (bottom left), and January–February (bottom right) using Landsat 8 optimal variables.

Appendix S2. Random forest predicted species diversity for January (top left), February (top right), March (bottom left), and January–February (bottom right) using Landsat 8 optimal variables.

Appendix S3. Random forest predicted species richness for January (top left), February (top right), March (bottom left), and January–February (bottom right) using Sentinel-2 optimal variables.

Appendix S4. Random forest predicted species diversity for January (top left), February (top right), March (bottom left), and January–February (bottom right) using Sentinel-2 optimal variables.

How to cite this article: Mashiane, K., Ramoelo, A. & Adelabu, S. (2024) Prediction of species richness and diversity in sub-alpine grasslands using satellite remote sensing and random forest machine-learning algorithm. *Applied Vegetation Science*, 27, e12778. Available from: https://doi.org/10.1111/avsc.12778