KOEDOE - African Protected Area Conservation and Science
ISSN: (Online) 2071-0771, (Print) 0075-6458

The spatial distribution of the woodland communities and their associated environmental drivers in the Golden Gate Highlands National Park, South Africa

Authors:
Mahlomola E. Daemane[1,2]
Abel Ramoelo[2,3]
Samuel Adelabu[2]

Affiliations:
[1]Conservation Services Department, Scientific Services, South African National Parks, Kimberley, South Africa

[2]Department of Geography, Faculty of Natural and Agricultural Sciences, University of Free the State, Bloemfontein, South Africa

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

Corresponding author:
Mahlomola Daemane,
ernest.daemane@sanparks.org

Dates:
Received: 31 Jan. 2021
Accepted: 06 July 2021
Published: 23 Sept. 2021

How to cite this article:
Daemane, M.E., Ramoelo, A. & Adelabu, S., 2021, 'The spatial distribution of the woodland communities and their associated environmental drivers in the Golden Gate Highlands National Park, South Africa', *Koedoe* 63(1), a1672. https://doi.org/10.4102/koedoe.v63i1.1672

The extreme variability in the topography, altitude and climatic conditions in the temperate Grassland Mountains of Southern Africa is associated with the complex mosaic of grassland communities with pockets of woodland patches. Understanding the relationships between plant communities and environmental parameters is essential in biodiversity conservation, especially for current and future climate change predictions. This article focused on the spatial distribution of woodland communities and their associated environmental drivers in the Golden Gate Highlands (GGHNP) National Park in South Africa. A generalized linear model (GLM) assuming a binomial distribution, was used to determine the optimal environmental variables influencing the spatial distribution of the woodland communities. The Coefficient of Variation (CV) was relatively higher for the topographic ruggedness index (68.78%), topographic roughness index (68.03), aspect (60.04%), coarse fragments (37.46%) and the topographic wetness index (31.33) whereas soil pH, bulk density, sandy and clay contents had relatively less variation (2.39%, 3.23%, 7.56% and 8.46% respectively). In determining the optimal number of environmental variables influencing the spatial distribution of woodland communities, roughness index, topographic wetness index, soil coarse fragments, soil organic carbon, soil cation exchange capacity and remote-sensing based vegetation condition index were significant ($p < 0.05$) and positively correlated with the woodland communities. Soil nitrogen, clay content, soil pH, fire and elevation were also significant but negatively correlated with the woodland communities. The area under the curve (AUC) of the receiver operating characteristics (ROC) was 0.81. This was indicative of a Parsimonious Model with explanatory predictive power for determination of optimal environmental variables in vegetation ecology.

Conservation implications: The isolated woodland communities are sources of floristic diversity and important biogeographical links between larger forest areas in the wider Drakensberg region. They provide suitable habitats for a larger number of forest species and harbour some of the endemic tree species of South Africa. They also provide watershed protection and other important ecosystem services. Understanding the drivers influencing the spatial distribution and persistence of these woodland communities is therefore key to conservation planning in the area.

Keywords: Afromontane; generalized linear model; environmental parameters; conservation; biodiversity.

Introduction

Several studies across the world have attempted to predict the impact of environmental factors on the spatial distribution of plant communities in different landscapes (Hutyra et al. 2005; Salazar, Nobre & Oyama 2007). The findings from these studies predict changes in dry woodland and Afromontane forests because of climatic changes. The increase in temperatures and other environmental factors are also expected to result in increased fire frequencies and changes in woodland communities (Abiem et al. 2020; Adelabu, Adepoju & Mofokeng 2020; Everand 1986).

Previous studies on structural and floristic composition of the Afromontane forests and other woodland communities in Golden Gate Highlands National Park (GGHNP) were restricted to a smaller area before the expansion of the park to its current footprint (Manfred 1990). Since then, the park has expanded considerably from 11 346 ha to 32 758 ha. Woody communities found in

the study area form an important biogeographical link between larger forest areas in the wider Drakensberg region (White 1978). These woodland communities in the Grassland biome are mostly restricted to deep valleys and drainage lines as compared to the dispersed spatial distribution of trees in the savanna biome. Studies undertaken in the grassland ecosystem suggest that plant communities respond in different ways to the increasing environmental changes and availability of moisture and protection against fire (Adagbasa, Adelabu & Okello 2018; Adagbasa, Adelabu & Okello 2020; Adelabu et al. 2020; Botha, Archibald & Greve 2020; Everand 1986). Other studies in the African continent showed elevation, slope and aspect as the determinants for the spatial and temporal distribution of plant communities and species composition (Sala & Paruelo 1997). Other environmental determinants such as topographic position, physical and chemical properties of the soil were also found to influence plant community types and the associated ecosystem services (Havstad et al. 2007; Sala & Paruelo 1997).

To monitor the ecological status and impact on these woodland communities, vegetation composition and the canopy structural parameters are typically measured *in situ* as indicators of change (Khan et al. 2012; Rahman et al. 2016; Ribichich & Protomastro 1998). However, employing *in situ* sampling in these areas can be time-consuming and impractical because of inaccessible terrains. The basic field measurements and conventional statistics in this instance can also not explain the spatial variation because of multiple interactions amongst state (vegetation, species distribution, understory cover, soil, topography, etc.) and other variables such as climate and human factors (Karahan & Erşahin 2018).

In order to provide a statistical relationship between the spatial distribution of woodland communities and the environmental variables at previously un-sampled locations, the exploratory and predictive models were used to predict probability of occurrence (Guisan, Edwards & Hastie 2002; Li et al. 2014; Pal 2005). Statistical analysis such as regression has been widely used in ecological modelling of the spatial distribution of species and their associated environmental parameters (Franklin 1998; Guisan, Weiss & Weiss 1999; Guisan & Zimmermann 2000; Lenihan 1993; Scott et al. 2002). In this study, we looked at the spatial distribution of the woodland communities and their associated environmental drivers. We used both exploratory statistics and logistic regression model (LRM) (Austin, Nicholls & Margules 1990) to determine the strength of the statistical relationship between a response (e.g. plant species presence) and a suite of environmental variables such as topography, edaphic variables, remotely sensed vegetation condition index (VCI) and fire.

Sampling methods

Study area

This study was undertaken in the GGHNP located in the foothills of the Drakensberg Mountains in eastern Free State Province of South Africa. Golden Gate Highlands National Park lies between latitude 28°30′ and 28°45′ south and longitude 28°30′ and 28°37′ east, on the border between South Africa and Lesotho (Figure 1). The park forms a part of the Maloti Drakensberg Catchment Complex that produces about 50% of the total water supply in South Africa. The area falls under summer rainfall region where the rainy season stretches from September to April with mean annual rainfall of 780 mm. The park is underlain by rock formations representing the upper Karoo Sequence which is intruded by dolerite dykes and sills.

The vegetation of the park is predominantly grassland, with a very small percentage of woodland communities along the rivers and in sheltered places that are protected from fire. Mucina and Rutherford (eds. 2006) identified four dominant vegetation units associated with the study area, namely Northern Drakensberg Highland Grassland (Gd 5), Lesotho Highland Basalt Grassland (Gd 8), Eastern Free State Sandy Grassland (Gm 4), and Basotho Montane Shrubland (Gm 5).

Delineation of forest patches, field data collection, data preparation and analysis

The high-resolution imagery provided via Google Earth has been increasingly used in scientific research (Fortin & Edwards 2001; Madin, Madin & Booth 2011; Pringle 2010) to assist in the selection of field sampling locations (Tang et al. 2010). The woodland polygons were delineated using high-resolution images within Google Earth mainly from Maxar technologies responsible for WorldView 1-3, and Geoeye with less than 1-m resolution from 2015 to 2020. Because the study area is a mountainous environment, time-slide in Google Earth was used to access historical image archives to offset shadow effects. The woodland community polygons were converted from Google Earth files to shapefile and topologically corrected in ArcGIS version 10.5.1. Fieldwork was also conducted to ensure that the woodland community polygons are adjusted and that the boundaries are correctly delineated for further analysis. Purposive sampling was undertaken using the delineated woodland polygons to verify mapped and unmapped woodland communities covered by the mountain shadows. The woodland community polygons data were used as training samples for statistical analysis (Estes et al. 2012).

Environmental data sets

Different environmental data sets were used to determine how they influence the spatial distribution of the woodland communities across the landscape. The environmental variables were downloaded from different databases sources (Table 1). Soil chemical and physical properties were one of the environmental variables used. Soil chemical variables included pH, nitrogen (N), soil organic carbon, cation exchange capacity (CEC), whilst soil physical properties included silt, coarse fragments, bulk density, sand, and clay content (Hengl et al. 2017). Shuttle Radar Topography Mission (SRTM) Digital Elevation Models (DEM) was also acquired (https://www2.jpl.nasa.gov/srtm/). Slope and aspect were computed from DEM using ArcGIS 10.5.1.

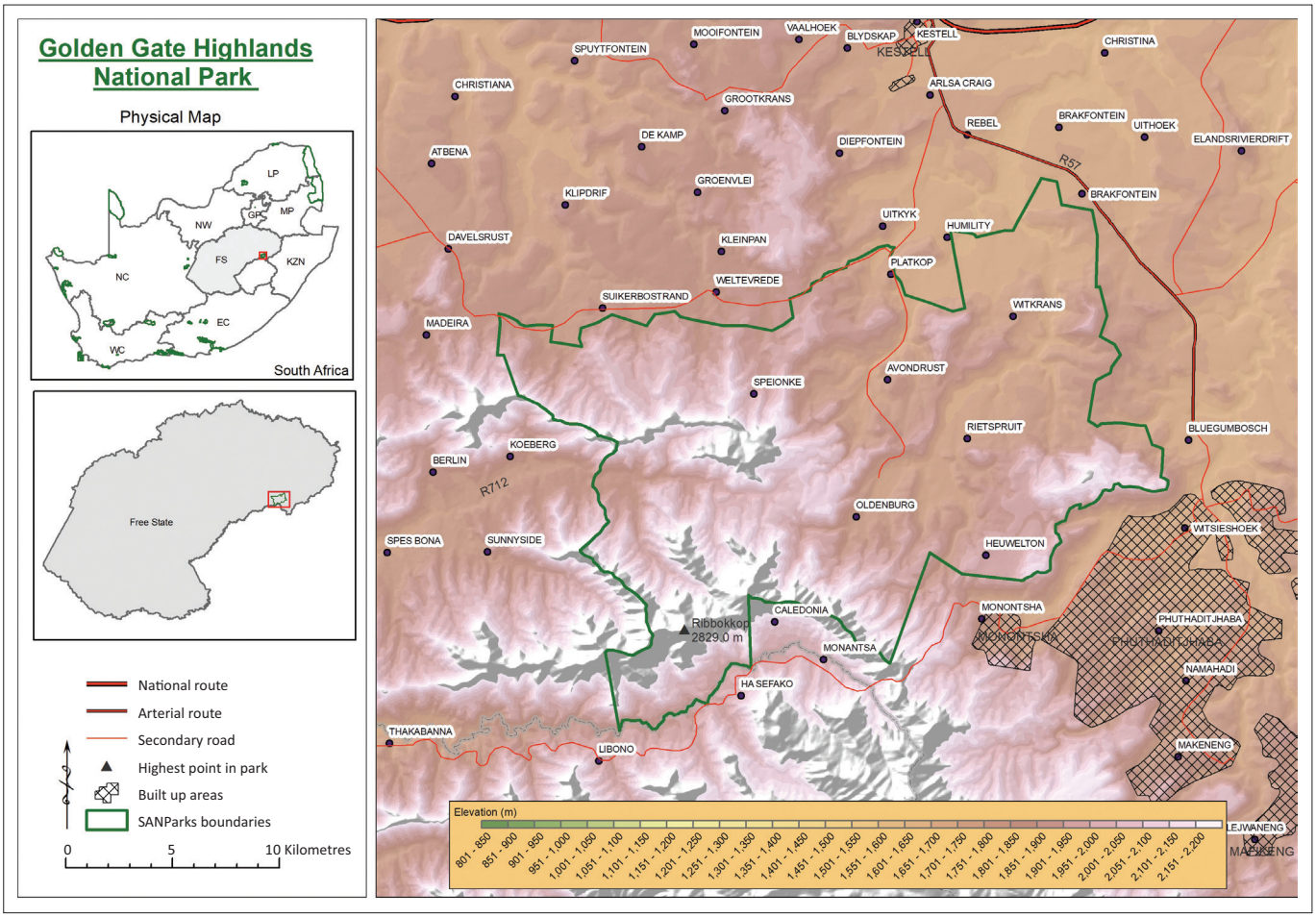


FIGURE 1: Location of the Golden Gate Highlands National Park in the context of the rest of South Africa and the Provinces.

Topographical variables such as slope and aspect were computed using ArcGIS version 10.5.1. Topographic Position Index (TPI), Topographic Wetness Index (TWI), Topographic Ruggedness Index (TRI) and Topographic Roughness were computed using QGIS version 3.0 (Guisan et al. 1999; Mokarram, Roshan & Negahban 2015; Muddarisna et al. 2020; Radułaa, Szymura & Szymura 2018; Riley, DeGloria & Elliot 1999). The TPI computes the differences of elevation at a specific pixel (central pixel) and the average elevation around it within a defined radius (Gallant & Wilson 2000; Weiss 2001). The elevation is an input to the computations of the TPI. The TWI is one of the hydrologically based topographic index. Topographic Wetness Index describes the tendency of a cell to accumulate water (Gruber & Peckham 2009), and can be used as an indicator of soil moisture content. The input variable to calculate TWI is the slope as computed from the elevation data. The TRI was computed according to Riley et al. (1999) providing summary of change in elevation over 3 × 3 pixel window. It computes the terrain heterogeneity measurement for every location, and each pixel contains the difference in elevation from a centre pixel and the eight pixels surrounding it. The extraction of values from various environmental variables were extracted using a tool 'Extracting values to points' embedded in ArcGIS version 10.5.1. The fire frequency was done using Moderate Resolution Imaging Spectroradiometer (MODIS) burn area index, and the VCI for February 2020 was computed from 2001 to 2020 MODIS normalized difference vegetation index (NDVI) (Table 1). Vegetation condition index is a good indicator of how vegetation is affected by drought (Kogan 1990).

Data preparation for analysis

To prepare data for analysis, delineated woodland polygons were used to extract data from a series of environmental data. The first step was to create a regular point shapefile layer with a spacing of 250 m using QGIS version 3.0. The point data were then overlaid on the delineated woodland polygons to create a binary layer (250 m spacing) indicating the presence (ones) and absence (zeros) of woodland polygons. The 250 m spacing was selected because most of the environmental layers had 250 m spatial resolution. The binary layer was used to extract all the environmental variables associated with the presence and the absence data for statistical analysis (Table 1). There were about 1500 presence points and approximately the same amount of absence points for the park. The binary layer with associated environmental variables was converted into a text file for further statistical analysis. For statistical analysis, the binary field in the latter layer is the dependent variable and all the environmental variables are independent.

TABLE 1: Series of environmental variables acquired from various data sources to determine their influence on the spatial distribution of woodland communities in the Golden Gate Highland National Park.

Variable	Environmental variable	Source	Scale or resolution
Topography	Digital elevation model (DEM) – metres	SRTM, https://www2.jpl.nasa.gov/srtm/	30 m
	Slope	Derived from DEM	30 m
	Aspect		
	Topographic roughness		
	Topographic position index (TPI)		
	Topographic wetness index (TWI)		
	Topographic ruggedness index (TRI)		
Soil Chemical Properties	Nitrogen (cg/kg)	Soilgrids.org (Hengl et al. 2017)	250 m
	pH		
	Soil organic carbon (dg/kg)		
	Cation exchange capacity (mmol(c)/kg)		
Soil Physical Properties	Silt (g/kg)		
	Coarse fragments (cm^3/dm^3)		
	Bulk density (cg/cm^3)		
	Sand (g/kg)		
	Clay content (g/kg)		
Fire	Fire frequency	MODIS Burn Area Index (2000–2017) https://lpdaac.usgs.gov/tools/appeears/	250 m
Vegetation condition index (VCI), Feb 2020	VCI (%) = ([$NDVI_{Feb2020}$ – $NDVI_{min}$]/[$NDVI_{max}$ – $NDVImi_n$]) × 100 (Kogan 1990)	MODIS NDVI (2001–2020), https://lpdaac.usgs.gov/tools/appeears/	250 m

Data analysis and modelling

Descriptive analysis (minimum, maximum, mean, standard deviation and coefficient of variation (standard deviation/mean) × 100) was done to determine the spread and variation of the data sets across the study area, associated with the presence and absence locations. Cross-correlation was also done to determine a level to which explanatory variables are significantly related based on the 95% confidence level.

A generalized linear model (GLM) assuming a binomial distribution (LRM) (Bolker et al. 2009; Hosmer, Lemeshow & Sturdivant 2013), was used to determine optimal environmental variables influencing the spatial distribution of the woodland communities. To minimise multi-collinearity in highly correlated independent variables and over-fitting in training and test datasets, the stepwise logistic model was used:

$$P = \frac{1}{1+e^{-y}} \quad \text{[Eqn 1]}$$

Where P is the probability of occurrence for the forest patches as explained by several environmental variables, and $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots + c$, where β = slope, x = variables and c = intercept:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots + c \quad \text{[Eqn 2]}$$

To determine the optimal number of environmental variables influencing the spatial distribution of woodland communities, several significance tests were done at 95% confidence level ($p < 0.05$). 'Optimal environmental variables' in this article refer to significant variables for explaining the spatial distribution of the woodland communities or a model with the best fit. The optimal variables, and their statistical estimates and confidence levels were also recorded (Table 2). To validate the LRMs used, the data was split into 70% calibration and 30% validation and tested accordingly. The area under-the-curve (AUC) was used as an accuracy indicator for the model (Bradley 1997; Hand 2009). Area under the curve values range from 0 where the model cannot separate the classes (i.e. woodland vs. no woodland) and 1 indicates a good model (two classes clearly distinguished). All the models were run in R statistical programming software using a Rattle Package:

$$AUC = \int_0^1 Pr[TP](v)\,dv \quad \text{[Eqn 3]}$$

Where TP stands for true positive and FP for false positive, and $pr[Tp]$ is a function of $v = pr[Fp]$

The Spearman rank order correlation coefficient, (ρ) was also used to measure the strength and direction of association that exists between two variables (Eqn 3):

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \quad \text{[Eqn 4]}$$

Where d_i = difference in paired ranks and n = number of cases.

Figure 2 provides a summary of methods and procedures followed in data collection and analysis.

Results

Delineation of forest patches

Thirty-two woodland polygons were identified in GGHNP and delineated using the Google Earth programme. The woodland polygons were converted to a shapefile and edited to ensure topology in ArcGIS version 10.5.1. Field verification was done for grouping woodland polygons into communities. Using GGHNP broad vegetation map and classification by Manfred (1990), woodland polygons were grouped into four different categories (Figure 3), namely Afromontane forests (*Olinia*, *Podocarpus*, *Kiggelaria*), *Euclea* woodland, *Leucosidea* woodland, and *Protea* woodland. The largest extent of woodland community, *Leucosidea* woodland was associated with the drainage lines and forest margins, although encroaching the footslopes and midslopes in some areas. Afromontane forests (*Olinia, Podocarpus,* and *Kiggelaria*) was confined to sheltered gorges and deep valleys. The *Euclea* woodland community was found at base of cliffs and shelter of large boulders, and also in small clumps in rocky open grassland. The *Protea* woodland community occurred in open grassland and associated with well drained soils in the footslopes and midslopes.

Exploratory analysis (descriptive)

The descriptive statistics showed relative variability amongst the environmental parameters with appreciably different means (Table 2). For example, variation was relatively higher for TRI (68.78%), topographic roughness (68.03), aspect (60.04%), coarse fragments (37.46%) and TWI (31.33), whereas pH, bulk density, sand and clay content had relatively less variation (2.39%, 3.23%, 7.56% and 8.46%, respectively). However, the topographical position index showed a relatively higher negative value (−818.18%). According to studies on topographic position and landform analysis, negative TPI values signify locations that are lower than their surroundings (gorges, deep valleys), whereas positive TPI values represent locations that are higher than the average of their surroundings (De Reu et al. 2013; Weiss 2001).

The Spearman Rank Correlation showed no significant correlation between most of the environmental parameters (Figure 4). However, significant and positive correlations were found between a few variables such as Topographic Roughness versus TRI (r = 0.985), CEC versus Soil Organic Carbon (r = 0.823); Silt versus CEC (r = 0.783); Elevation versus Soil Organic Carbon (r = 0.757), Clay Content versus Soil Organic Carbon (r = 0.64) etc., whilst negative and significant correlations were found between variables such as Sand versus Silt (r = −0.879); Sand versus Clay Content

TABLE 2: The descriptive statistics for the environmental variables associated with sampling points in the Golden Gate Highlands National Park.

Variables	Environmental variable	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)
Topography	Digital elevation model (DEM) – metres	1662.00	2757.00	1957.07	175.921	8.90
	Aspect	0.00	359.53	182.87	109.79	60.04
	Topographic roughness	1.00	145.00	23.08	15.70	68.03
	Topographic position index (TPI)	-36.38	35.87	-0.33	2.70	-818.18
	Topographic wetness index (TWI)	2.93	17.06	6.64	2.08	31.33
	Topographic ruggedness index (TRI)	0.38	51.87	7.02	4.83	68.78
Soil chemical properties	Nitrogen (cg/kg)	183.00	461.00	254.40	37.43	14.71
	pH	55.00	63.00	58.64	1.41	2.39
	Soil organic carbon (dg/kg)	13.00	70.00	23.98	7.01	29.24
	Cation exchange capacity (mmol(c)/kg)	11.00	31.00	17.67	4.48	25.33
Soil physical properties	Silt (g/kg)	16.00	31.00	22.24	2.57	11.49
	Coarse fragments (cm^3/dm^3)	1.00	20.00	9.19	3.44	37.46
	Bulk density (cg/cm^3)	1137.00	1520.00	1398.88	45.23	3.23
	Sand (g/kg)	41.00	62.00	52.91	4.01	7.56
	Clay Content (g/kg)	20.00	33.00	24.87	2.10	8.46
Fire (2000–2017)	Fire frequency	1.00	7.00	3.79	1.02	26.59
Vegetation condition index (VCI), Feb 2020	VCI (%)	0.15	0.99	0.72	0.14	19.69

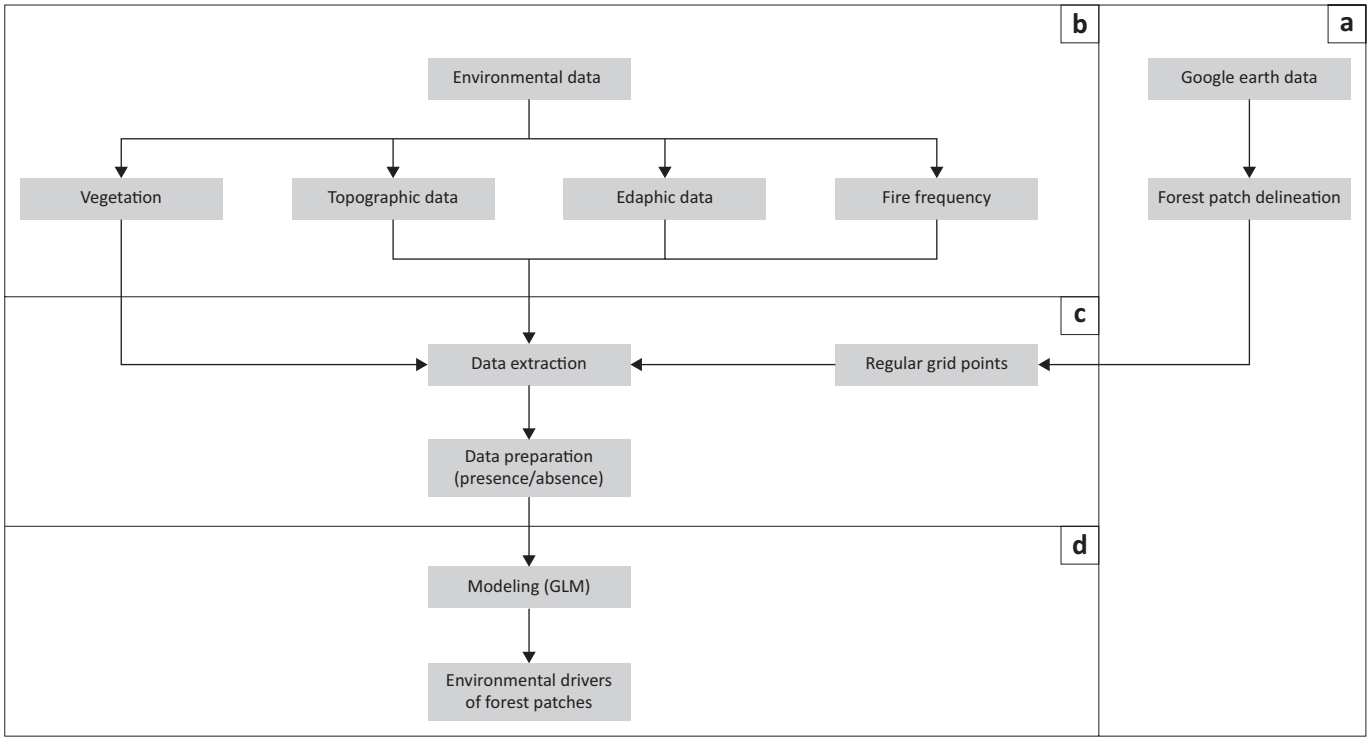


FIGURE 2: Synopsis of the methodology reflecting different stages of the data mining, input and outputs.

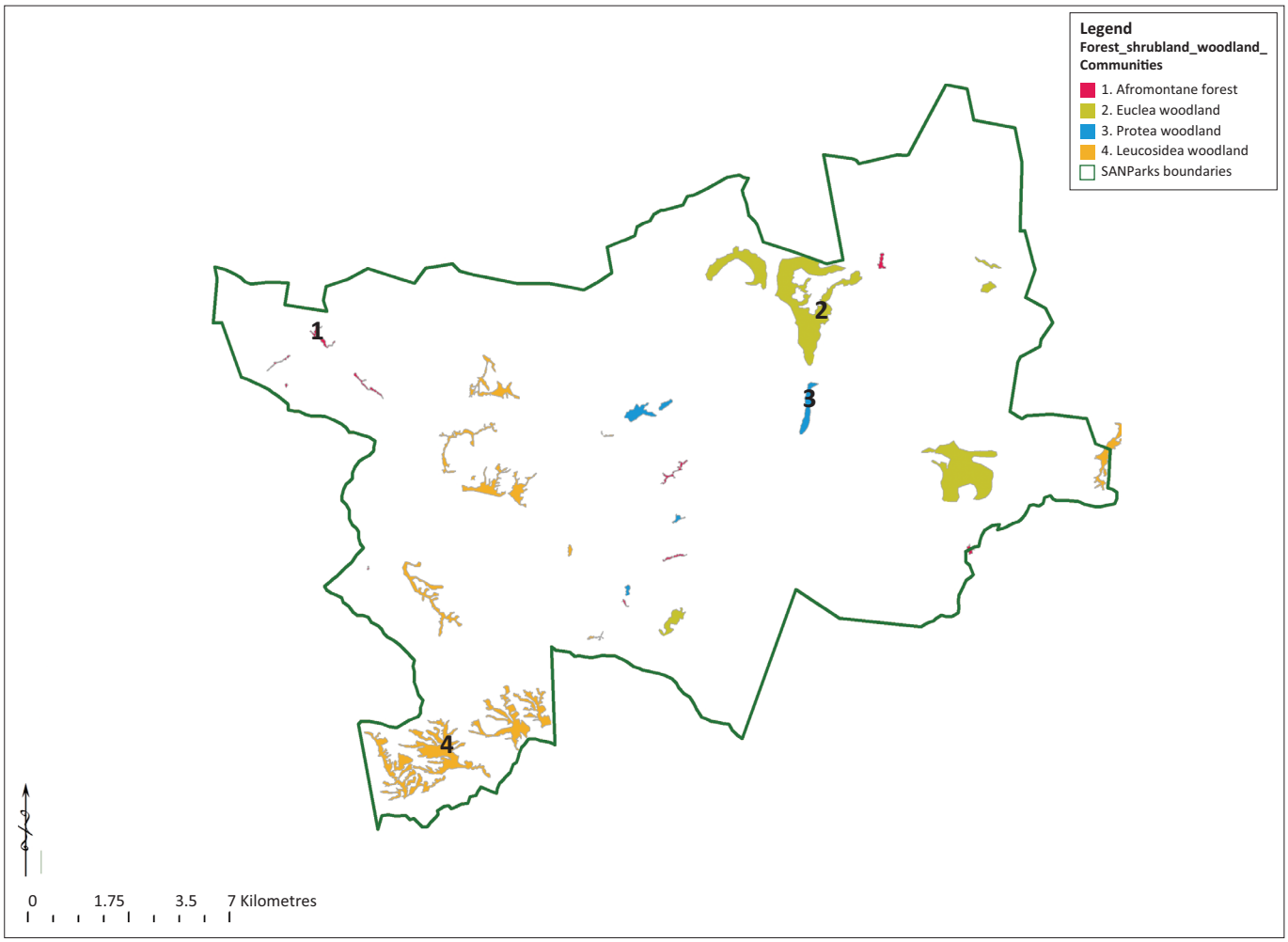
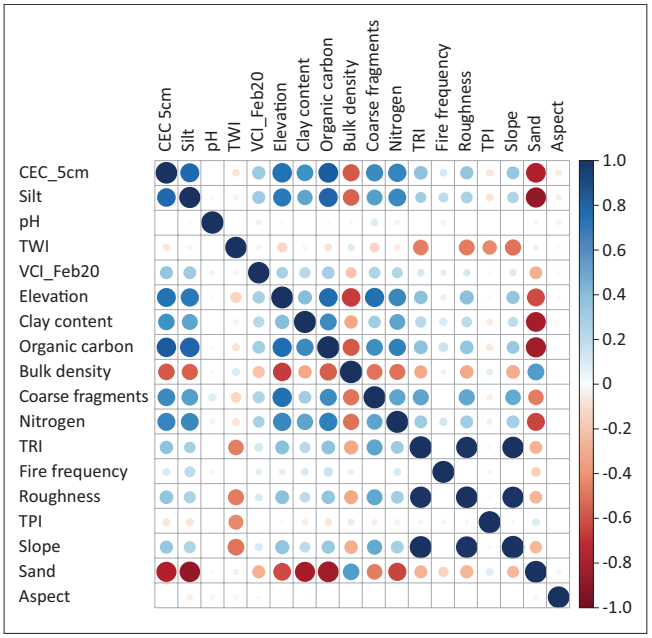


FIGURE 3: The spatial distribution of woodland communities in the Golden Gate Highlands National Park landscape.



CEC, Cation Exchange Capacity; TWI, Topographic Wetness Index; VCI, Vegetation Condition Index; TRI, Topographic Ruggedness Index; TPI , Topographic Position Index.

FIGURE 4: Correlogram showing relationship between environmental variables associated with the woodland communities in the Golden Gate Highlands National Park.

(*r* = –0.825); Soil Organic Carbon versus Sand (*r* = –0.832); CEC versus Sand (*r* = –0.799).

Optimal environmental parameters influencing woodland communities

In determining the optimal number of environmental variables influencing the spatial distribution of woodland communities, topography, edaphic factors (soil chemical and physical properties), fire frequency and remote sensing-based VCI were found to be significant at 95% confidence level (Table 3). The environmental variables such as topographic roughness, TWI, soil organic carbon, CEC and coarse fragments were significant and correlated positively with the woodland assemblages (Table 3). Elevation, soil nitrogen, soil pH, clay content, and fire were also significant but negatively correlated with the woodland assemblages.

Model validation

The AUC (= 0.81) of the receiver operating characteristic (ROC) was used in evaluating the GLM model for predicting the accuracy of the distributions of the woodland assemblages (Figure 5). The ROC curve was generated by plotting the True Positive Rate (TPR) against

the False Positive Rate (FPR). The TPR shows the proportion of positive samples that are correctly predicted, whilst FPR shows the proportion of positive samples that are incorrectly predicted (Wang & Zheng 2013). The AUC indicated that the optimal environmental variables selected as determinants of the spatial distribution of woodland assemblages were based on the good or parsimonious model.

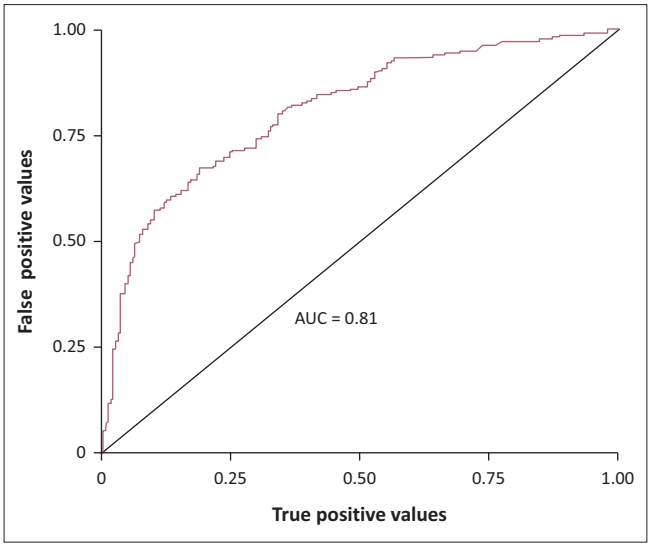
Discussion

The study assessed the spatial distribution of the woodland communities and their environmental drivers in the GGHNP. The use of high-resolution imagery from Google Earth and ESRI Base maps, and field verification data was effective in delineating the woodland in the current study. Georeferenced data sourced from different databases were also found to be of great value in describing the spatial distribution of the woodland and their associated environmental parameters. Our GLM model analysis identified topography, edaphic factors (soil chemical and physical properties), fire frequency and remote sensing-based VCI as the most important environmental variables influencing the spatial distribution of the woodland communities. Three topographic factors, namely elevation, topographic roughness and TWI influenced the occurrence of the woodland communities. Although temperature and precipitation were not included as environmental variables in the study, they are critical variables influencing the state of the vegetation in the mountain ecosystems (Basist, Bell & Meentemeyer 1994). However, local climatic conditions strongly influence the relationship between topography and the spatial distribution of precipitation (Basist et al. 1994); hence, temperature and precipitation were excluded in the study. This suggests that one can estimate the spatial distribution of mean annual precipitation using topographically based regression equations (Basist et al. 1994). In addition, we also used VCI, which is an index for assessing drought impact on vegetation.

TABLE 3: The optimal environmental variables, statistical estimates and confidence levels associated with the spatial distribution of woodland communities in the Golden Gate Highlands National Park.

Variables	Estimates or slope	SD errors	*Z*-value	*P*-value ($p < 0.05$)
Topography				
Elevation	-0.0088	0.0007	-12.184	< 0.05000
Topographic roughness	0.0425	0.0047	8.964	< 0.05000
Topographic wetness index	0.0921	0.0281	3.279	0.00104
Soil chemical properties				
Nitrogen	-0.0062	0.0022	-2.880	0.00392
pH	-0.1213	0.0410	-2.960	0.00311
Organic carbon	0.0805	0.0158	5.090	< 0.05000
Cation exchange capacity	0.1634	0.0248	6.590	< 0.05000
Soil physical properties				
Coarse fragments	0.3081	0.0275	11.190	< 0.05000
Clay content	-0.2254	0.0352	-6.390	< 0.05000
FIRE (0017)				
Fire frequency_0017 (2000–2017)	-0.2317	0.0541	-4.290	< 0.05000
Vegetation condition index				
Vegetation condition index_feb20	0.9034	0.4062	2.220	0.02619
Model intercept	22.4436	2.9502	7.608	< 0.05000

SD, Standard deviation.



AUC, areas under the curve.

FIGURE 5: Areas under the curve indicating the performance of the model.

Elevation in the GGHNP was found to be one of the environmental parameters limiting the distribution of the woodland communities. Elevation was significant but negatively correlated with the woodland communities. Woodland communities were absent above 2757 m and below 1662 m above sea level. In the Natal Drakensberg, the woodland communities limit lies in the montane and subalpine vegetation belts just below 2500 m above sea level (Chawla et al. 2008; Killick 1963). Topographic wetness index and topographic roughness derived from digital elevation model also had significant and positive correlation, implying that the soil moisture content was an important driver in the spatial distribution of the four different types of woodland communities (Li & McCarty 2019).

Although soil clay content was significantly and negatively associated with the woodland communities, relatively low variation in soil clay content did occur (coefficient variation = 8.46%). The woodland communities associated with the gorges, deep valleys and drainage lines (i.e. Afromontane forests and *Leucosidea* shrublands) have relative high water retention capacity because of the clay content (Groen et al. 2008). These communities are also associated with high soil organic content because of positive correlation between soil organic content versus clay content ($r = 0.64$). The *Euclea* woodland is mostly associated with sheltered boulders and shallow soils whereas *Protea* woodland is associated with well drained sandy soils in the footslopes and midlopes of GGHNP. The rocky and sandy habitats are poor in soil organic matter and have low water retention capacity and decreased water infiltration (Chow et al. 2007). This association is supported by negative and significant correlation found between soil organic carbon versus sand ($r = -0832$), and sand versus clay content ($r = -0.825$).

Soil nitrogen and soil pH were also found to be significant but negatively correlated with the woodland communities.

The *Protea* and *Euclea* woodland communities are associated with low soil pH in steeper and drier slopes because of drained nutrients (Woldemariam et al. 2008). Studies in the Maloti-Drakensberg Park also showed steep topography to be associated with leached nutrients and low pH soils (Carbutt & Edward 2015). However, despite intrinsically high levels of total soil nitrogen, the soil economy of the Drakensberg Mountains (> 1800 m above sea level) is characterised by low inorganic soil nitrogen availability because of the effect of low temperatures on soil nitrogen mineralisation (Carbutt et al. 2013).

Fire frequency was significant but also negatively correlated with the woodland communities. Nearly 26.9% spatial variation in fire frequency can be ascribed to different habitats associated with the woodland communities. The Afromontane forests and the *Leucosidea* woodland communities occurring in gorges and deep valleys are protected from fire (Manfred 1990). The *Euclea* woodland community remain relatively self-protected from burning because of rocky habitat and the lack of grass biomass required for intense frequent fires (Van Langevelde et al. 2003). Positive correlation was also found between satellite-based VCI and woodland communities. Approximately 19.69% of the spatial variations in VCI can be alluded to evergreen woodland communities occurring in areas of deep valley, drainage lines and gorges (Afromontane forest and *Leucosidea* woodland) and deciduous communities such as *Euclea* woodland community.

Conclusion

The GGHNP forms part of the Maluti Drakensberg Transfrontier region characterised by diverse topography and steep altitudinal gradients. It is an ideal region in which to study possible vegetation changes or shifts as a consequence of environmental changes (Hill 1996).

The analysis from the environmental data in GGHNP found topography, soil properties (chemical and physical), fire, and VCI to be the most influential factors in the spatial distribution of the woodland communities. Topographical variables such as roughness and topographical wetness index were found to be the most significant variables positively correlated with the woodland communities. Soil organic carbon, coarse fragments, CEC and VCI were also significant and positively correlated with the spatial distribution of the plant communities. Elevation, nitrogen, pH, fire frequency and clay content were also significant but negatively correlated with the woodland communities.

The data derived from this study will be of great conservation value as the woodland communities found in the area are the remnants of the Afromontane forests in the Maloti Drakensberg region (Kotze & Lawes 2007; White 1978). These woodland communities act as the nutrients pump and seedbank to the surrounding ecosystems, whilst maintaining genetic and floristic diversity. Understanding the relationships between these woodland communities and environmental parameters is also essential in the biodiversity conservation of the grassland biome and their associated woodland communities.

Acknowledgements

The authors would like to thank Lufuno Munyai and Abel Matsapula for assisting with field work and GIS maps.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

M.E.D., A.R. and S.A. developed the conceptual framework. M.E.D. designed the ecological sampling methods and undertook ecological surveys. A.R. was responsible for the statistical models and analysis. M.E.D. took the lead in writing the manuscript with inputs from all the authors.

Ethical considerations

This article followed all ethical standards for research without direct contact with human or animal subjects.

Funding information

Financial support provided by South African National Parks to undertake the field work is acknowledged. University of Free State covered tuition fees for the main author during the duration of the study.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and/or its supplementary materials.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the author.

References

Abiem, I., Arellano, G., Kenfack, D. & Chapman, H., 2020, 'Afromontane forest diversity and the role of grassland-forest transition in tree species distribution', *Diversity* 12(1), 30. https://doi.org/10.3390/d12010030

Adagbasa, E.G., Adelabu, S.A. & Okello, T.W., 2020, 'Development of post-fire vegetation response-ability model in grassland mountainous ecosystem using GIS and remote sensing', *ISPRS Journal of Photogrammetry* 164, 173–183. https://doi.org/10.1016/j.isprsjprs.2020.04.006

Adagbasa, G., Adelabu, S. & Okello, T., 2018, 'Spatio-Temporal Assessment of Fire Severity in a Protected and Mountainous Ecosystem', IGARSS 2018 - 2018 IEEE *International Geoscience and Remote Sensing Symposium*, 22–27 July 2018, Valencia, Spain, pp. 6572–6575. https://doi.org/10.1109/IGARSS.2018.8518268

Adelabu, S.A., Adepoju, K.A. & Mofokeng, O.D., 2020, 'Estimation of fire potential index in mountainous protected region using remote sensing', *Geocarto International* 35(1), 29–46. https://doi.org/10.1080/10106049.2018.1499818

Austin, M.P., Nicholls, A.O. & Margules, C.R., 1990, 'Measurement of the realized qualitative niche: Environmental niche of five Eucalyptus species', *Ecological Monographs* 60(2), 161–177. https://doi.org/10.2307/1943043

Basist, A., Bell, G.D. & Meentemeyer, V., 1994, 'Statistical relationships between topography and precipitation patterns', *Journal of Climate* 7(9), 1305–1315. https://doi.org/10.1175/1520-0442(1994)007<1305:SRBTAP>2.0.CO;2

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Henry, M. et al., 2009, 'Generalized linear mixed models: A practical guide for ecology and evolution', *Trends in Ecology & Evolution* 24(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Botha, M., Archibald, S. & Greve, M., 2020, 'What drives grassland forest boundaries? Assessing fire and frost effects on tree seedling survival and architecture', *Ecology and Evolution* 10(19), 10719–10734. https://doi.org/10.1002/ece3.6730

Bradley, A.P., 1997, 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition* 30(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Carbutt, C. & Edwards, T., 2015, 'Reconciling ecological and phytogeographical spatial boundaries to clarify the limits of the montane and alpine regions of sub-Sahelian Africa', *South African Journal of Botany* 98, 64–75. https://doi.org/10.1016/j.sajb.2015.01.014

Carbutt, C., Edwards, T.J., Fynn, R.W.S. & Beckett, R.P., 2013, 'Evidence for temperature limitation of nitrogen mineralisation in the Drakensberg Alpine Centre', *South African Journal of Botany* 88, 447–454. https://doi.org/10.1016/j.sajb.2013.09.001

Chawla, A., Rajkumar, S., Singh, K., Lal, B., Singh, R. & Thukral, A., 2008, 'Plant species diversity along an altitudinal gradient of Bhabha Valley in western Himalaya', *Journal of Mountain Science* 5, 157–177. https://doi.org/10.1007/s11629-008-0079-y

Chow, T.H., Rees, H.W., Monteith, J.O. & Toner, P., 2007, 'Effects of coarse fragment contenton soil physical properties, soil erosion and potato production', *Canadian Journal of Soil Science* 87(5), 565–577.

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P. et al., 2013, 'Application of the topographic position index to heterogeneous landscapes', *Geomorphology* 186, 39–49 https://doi.org/10.1016/j.geomorph.2012.12.015

Estes, A.B., Kuemmerle, T., Kushnir, H., Radeloff, V.C. & Shugart, H.H., 2012, 'Landcover change and human population trends in the Serengeti ecosystem from 1984–2003', *Biological Conservation* 147(1), 255–263. https://doi.org/10.1016/j.biocon.2012.01.010

Everand, D.A., 1986, 'The effects of fire on the *Podocarpus latifolius* forests of the Royal Natal National Park, Natal Drakensberg', South *African Journal of Botany* 52(1), 60–66.

Fortin M.J. & Edwards G., 2001, 'Delineation and Analysis of Vegetation Boundaries', in C.T. Hunsaker, M.F. Goodchild, M.A. Friedl & T.J. Case (eds.), *Spatial Uncertainty in Ecology*, pp. 158–174, Springer, New York, NY. https://doi.org/10.1007/978-1-4613-0209-4_8

Franklin, J., 1998, 'Predicting the distribution of shrub species in southern California from climate and terrain-derived variables', *Journal of Vegetation Science* 9(5), 733–748. https://doi.org/10.2307/3237291

Gallant, J.C. & Wilson, J.P., 2000, 'Primary topographic attributes', in J.P. Wilson, J.C. & Gallant (eds.), *Terrain analysis: Principles and applications*, pp. 51–85, Wiley, New York, NY.

Groen, T.A., Van Langevelve, F., Van deVijvert, C.A.D.M., Govender, N. & Prins, H.H.T., 2008, 'Soil clay content and fire frequency affect clustering in trees in South African savannas', *Journal of Tropical Ecology* 24(3), 269–279. https://doi.org/10.1017/S0266467408004872

Gruber, S. & Peckham, S., 2009, 'Land-surface parameters and objects in hydrology', *Developments in Soil Science* 33, 171–194. https://doi.org/10.1016/S0166-2481(08)00007-X

Guisan, A., Edwards, T.C. & Hastie, T., 2002, 'Generalized linear and generalized additive models in studies of species distributions: Setting the scene', *Ecological Modeling* 157(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guisan, A., Weiss, S.B. & Weiss, A.D., 1999, 'GLMs versus CCA spatial modeling of plant species distribution', *Plant Ecology* 143, 107–122. https://doi.org/10.1023/A:1009841519580

Guisan, A. & Zimmermann, N.E., 2000, 'Predictive habitat distribution models in ecology', *Ecological Modelling* 135(2–3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hand, D.J., 2009, 'Measuring classifier performance: A coherent alternative to the area under the ROC curve', *Machine Learning* 77, 103–123. https://doi.org/10.1007/s10994-009-5119-5

Havstad, K.M., Peters, D.P.C., Skaggs, R., Brown, J., Bestelmeyer, B.T., Fedrickson, E. et al., 2007, 'Ecological services to and from rangelands of the United States', *Ecological Economics* 64(2), 261–268. https://doi.org/10.1016/j.ecolecon.2007.08.005

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A. et al., 2017, 'SoilGrids250m: Global gridded soil information based on machine learning', *PLoS One* 12(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hill, T.R., 1996, 'Description, classification and ordination of the dominant vegetation communities, Cathedral Peak, KwaZulu-Natal Drakensberg', *South African Journal of Botany* 62(5), 263–269. https://doi.org/10.1016/S0254-6299(15)30655-4

Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X., 2013, *Applied logistic regression*, John Wiley & Sons, Hoboken, NJ.

Hutyra, L.R., Munger, J.W., Nobre, C.A., Saleska, S.R., Vieira, S.A. & Wofsy, S.C., 2005, 'Climatic variability and vegetation vulnerability in Amazonia', *Geophysical Research Letters* 32, L24712. https://doi.org/10.1029/2005GL024981

Karahan, G. & Erşahin, S., 2018, 'Geostatistics in characterizing spatial variability of forest ecosystems', *Eurasian Journal of Forest Science* 6(1), 9–22.

Khan, S.M., Page, S., Ahmad, H., Shaheen, H. & Harper, D., 2012, 'Vegetation dynamics in the Western Himalayas, diversity indices and climate change', *Science, Technology and Development* 31, 232–243.

Killick, D.J.B., 1963, 'An account of the plant ecology of the Cathedral Peak Area of the Natal Drakensberg', *Memoirs of the Botanical Survey of South Africa* 34, 1–178.

Kogan, F.N., 1990, 'Remote sensing of weather impacts on vegetation in non-homogeneous areas', *International Journal of Remote Sensing* 11(8), 1405–1419. https://doi.org/10.1080/01431169008955102

Kotze, D.J. & Lawes, M.J., 2007, 'Viability of ecological processes in small Afromontane forest patches in South Africa', *Journal of Austral Ecology* 32(3), 294–304. https://doi.org/10.1111/j.1442-9993.2007.01694.x

Lenihan, J.M., 1993, 'Ecological responses surfaces for North American tree species and their use in forest classification', *Journal of Vegetation Science* 4(5), 667–680. https://doi.org/10.2307/3236132

Li, C., Wang, J., Wang, L., Hu, L. & Gong, P., 2014, 'Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper Imagery', *Remote Sensing* 6(2), 964–983. https://doi.org/10.3390/rs6020964

Li, X. & McCarty, G.W., 2019, 'Application of Topographic Analyses for Mapping Spatial Patterns of Soil Properties', in A. Pepe & Q. Zhao (eds.), *Geospatial Analyses of Earth Observation (EO) data*, Books on Demand, IntechOpen, London. https://doi.org/10.5772/intechopen.86109

Madin, E.M.P., Madin, J.S. & Booth, D.J., 2011, 'Landscape of fear visible from space', *Scientific Reports* 1, 14. https://doi.org/10.1038/srep00014

Manfred, D.M., 1990, 'The structure and floristic composition of Afromontane forests in the Golden Gate Highlands National Park', B.Sc. Honours thesis, Dept. of Botany, University of Johannesburg, Johannesburg.

Mokarram, M., Roshan, G. & Negahban, S., 2015, 'Landform classification using topography position index (case study: Salt dome of Korsia-Darab plain, Iran)', *Modeling Earth Systems and Environment* 1, 40. https://doi.org/10.1007/s40808-015-0055-9

Mucina, L. & Rutherford, M.C. (eds.), 2006, *The vegetation of South Africa, Lesotho and Swaziland*, Strelitzia 19, South African National Biodiversity Institute, Pretoria.

Muddarisna, N., Yuniwati, E.D., Masruroh, H. & Oktaviansyah, A.R., 2020, 'An automated approach using Topographic Position Index (TPI) for landform mapping (Case Study: Gede Watershed, Malang Regency, East Java, Indonesia)', *IOP Conference Series, Earth and Environmental Science* 412, 012027. https://doi.org/10.1088/1755-1315/412/1/012027

Pal, M., 2005, 'Random forest classifier for remote sensing classification', *International Journal of Remote Sensing* 26(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pringle, H., 2010, 'Google Earth shows clandestine worlds', *Science* 329(5995), 1008–1009. https://doi.org/10.1126/science.329.5995.1008

Radułaa, M.W., Szymura, T.H. & Szymura, M., 2018, 'Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values', *Ecological Indicators* 85, 172–179. https://doi.org/10.1016/j.ecolind.2017.10.011

Rahman, A.U., Khan, S.M., Khan, S., Hussain, A., Rahman, I.U., Iqbal, Z. et al., 2016, 'Ecological assessment of plant communities and associated edaphic and topographic variables in the Peochar Valley of the Hindu Kush Mountains', *Mountain Research and Development* 36(3), 332–341. https://doi.org/10.1659/MRD-JOURNAL-D-14-00100.1

Ribichich, A.M. & Protomastro, J., 1998, 'Woody vegetation structure of xeric forest stands under different edaphic site conditions and disturbance histories in the Biosphere Reserve "Parque Costero del Sur," Argentina', *Plant Ecology* 139, 189–201. https://doi.org/10.1023/A:1009718819857

Riley, S.J., DeGloria, S.D. & Elliot, R., 1999, 'A terrain ruggedness index that quantifies topographic heterogeneity', *Intermountains Journal of Science* 5(1–4), 23–27.

Sala, O.E. & Paruelo, J.M., 1997, 'Ecosystem Services in Grasslands', in G.C. Daily (ed.), *Nature's Services: Societal Dependence on Natural Ecosystems*, pp. 237–252, Island Press, Washington, DC.

Salazar, L.F., Nobre, C.A. & Oyama, M.D., 2007, 'Climate change consequences on the biome distribution in tropical South America', *Geophysical Research Letters* 34, L09708. https://doi.org/10.1029/2007GL029695

Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M. et al., 2002, *Predicting species occurrences: Issues of accuracy and scale*, Island Press, Covelo, CA.

Tang, L., Shao, G., Piao, Z., Dai, L., Jenkins, M.A., Wang, S. et al., 2010, 'Forest degradation deepens around and within protected areas in East Asia', *Biological Conservation* 143(5), 1295–1298. https://doi.org/10.1016/j.biocon.2010.01.024

Van Langevelde, F., Van De Vijver, C.A., Kumar, L., Van De Koppel, J., De Ridder, N., Van Andel, J. et al., 2003, 'Effects of fire and herbivory on the stability of savanna ecosystems', *Ecology* 84(2), 337–350. https://doi.org/10.1890/0012-9658(2003)084[0337:EOFAHO]2.0.CO;2

Wang, H. & Zheng, H., 2013, 'True Positive Rate', in W. Dubitzky, O. Wolkenhauer, K.H. Cho & H. Yokota (eds.), *Encyclopedia of Systems Biology*, pp. 136–141, Springer, New York, NY, https://doi.org/10.1007/978-1-4419-9863-7-255

Weiss, A.D., 2001, 'Topographic position and landforms analysis', Ecoregional Data Management Team, The Nature Conservancy, Northwest Division, Seattle, Washington, Poster Presentation, *ESRI Users Conference*, San Diego, California.

White, F., 1978, 'The afromontane region', in M.J.A. Werger (ed.), *Biogeography and Ecology of Southern Africa*, pp. 463–513, Springer, Berlin/Heidelberg. https://doi:10.1007/978-94-009-9951-0_11

Woldemariam, G.T., Borsch, T., Denich, M. & Teketay, D., 2008, 'Floristic composition and environmental factors characterizing coffee forests in south-west Ethiopia', *Forest Ecology Management* 255(7), 2138–2150. https://doi.org/10.1016/j.foreco.2007.12.028