# Environmental heterogeneity patterns plant species richness and turnover in two hyperdiverse floras

*Ruan van Mazijk, Michael D. Cramer & G. Anthony Verboom*

*2019-05-30*

Department of Biological Sciences, University of Cape Town, Rondebosch, South Africa

Corresponding author: RVM (ruanvmazijk@gmail.com, +27 21 650 3684)

ORCID nos.: RVM: 0000-0003-2659-6909, MDC: 0000-0003-0989-3266, GAV: 0000-0002-1363-9781

**Running title:** Environmental heterogeneity and plant species richness

## Abstract

**Aim:** To quantify the explanatory power of environmental heterogeneity in predicting plant species richness and turnover in two mediterranean-type biodiversity hotpots, the southwestern Cape of South Africa and southwest Australia (SWA), due to habitat diversity's stimulating speciation and supporting diverse assemblages. In each region, we compare different forms of environmental heterogeneity, species turnover and the relative importance of different forms of heterogeneity better predict species richness and turnover. We expected the Cape to be more heterogeneous in most environmental axes, and at a finer grain, such that the consequent high levels of species turnover explain the Cape's greater species richness per unit area. We also conjectured that edaphic heterogeneity would be an important factor in predicting species richness patterns in SWA.

**Location:** The Greater Cape Floristic Region in southwest Africa (the Cape), and the Southwest Australia Floristic Region (SWA)

**Taxon:** Vascular plants

**Methods:** Geospatially explicit floral and environmental data, non-parametric statistics, boosted regression tree modelling

**Results:** The Cape is more environmentally heterogeneous and has higher levels of floristic turnover than SWA. We find that environmental heterogeneity is the main predictor of species richness in the Cape, and somewhat less so for SWA. Edaphic conditions are found to be of more biologically important in the Cape, though this is contingent on the quality of the data modelled.

**Main conclusions:** The Cape was found to have generally more environmentally heterogenous than SWA, though SWA does possess edaphic heterogeneity as great as that in the Cape at coarse spatial scales, and greater levels of floristic turnover in vascular plant communities between grid-cells than SWA. Vascular plant species richness and turnover in the Cape was associated more strongly with environmental heterogeneity than in SWA, and particularly edaphic heterogeneity, than species richness in SWA. In SWA, patterns of species richness and turnover are more strongly associated with climatic axes. The relative strengths of association of different environmental variables with patterns of species richness and turnover was found to vary with spatial scale.

*Keywords:* biodiversity, boosted regression trees, BRTs, environmental heterogeneity, fynbos, Greater Cape Floristic Region, kwongan, macroecology, species richness, species turnover, vascular plants, Southwest Australia Floristic Region

# Acknowledgements

# 1   Introduction

Biodiversity represents the variety of species and the ecological and evolutionary processes that bring about those species (Bøhn & Amundsen, 2004). Studying the distribution of biodiversity in space is a major avenue of biological research (Kerr et al., 2001; Kreft & Jetz, 2007). Region-scale geographic patterns in species richness have long been studied, particularly in biodiversity hotspots (Cook et al., 2015). Indeed, "primary geographic patterns" such as these (sensu Juarez-Barrera et al., 2018) are arguably central when studing the distribution of biota across space. The spatial distribution of species richness can be and often is explained in terms of the physical environment. Certain properties of the environment have been suggested to influence species richness in three ways: (i) available resources and energy, which can determine the number of species able to co-exist in an area (Gaston, 2000; Kreft & Jetz, 2007; Mouchet et al., 2015); (ii) environmental stability through time, which enables species' persistence; and (iii) spatial heterogeneity, which can both stimulate ecological speciation and possible barriers to gene flow and can facilitate greater levels of species' co-existence (Thuiller et al., 2006; Mouchet et al., 2015; Cramer & Verboom, 2016). The physical environment, then, can be used to explain species richness in both a local-deterministic and historical sense (Ricklefs, 1987).

The maintenance of species richness, particularly the coexistence of high numbers of species in biodiversity hotspots, is often regarded as "paradoxical" (Hart et al., 2017), and is a central problem in macro-ecology and biogeography (Ricklefs, 1987; Kreft & Jetz, 2007; Hart et al., 2017). Species richness is constrained by the ability of habitats to support a variety of species—its ecological carrying capacity (Mateo et al., 2017). This is exemplified in modelling approches, wherein species richness is a function of environmental predictors in a correlative framework ("macro-ecological models"; Mateo et al., 2017). Macro-ecological models of species richness implicitly assume that communities are saturated, following species-area and species-energy relationships, and at equilibrium with the environment (Mateo et al., 2017).

A solution to the "paradox" of species co-existence is environmental heterogeneity (EH): a more heterogeneous environment exhibits a larger environmental space, thus facilitating co-existence between more species. Heterogeneity in the physical environment is known to be positively associated with species richness (Rensburg et al., 2002; Hart et al., 2017), and has been demonstrated to do so across many taxa—e.g. Canadian butterflies (Kerr et al., 2001), European vertebrates (Mouchet et al., 2015), South African birds (Rensburg et al., 2002), in communities along marine continental margins (Levin et al., 2010), French scarab beetles (Lobo et al., 2004), and for global terrestrial plants (Kreft & Jetz, 2007). The spatial scale of heterogeneity, or "grain" of the environment, is also important to consider (Hart et al., 2017), as spatial scale in absolute environmental

75  conditions has also been explored (Kerr et al., 2001; Baudena et al., 2015; Mouchet et al., 2015). Species

76  co-existence and biodiversity maintenance is indeed suggested to be scale-dependent (Hart et al., 2017).

77  EH is often under-represented in macro-ecological models of species richness, and has recently been found to

78  explain up to ca. 95% of biome level species richness across South Africa (Cramer & Verboom, 2016). Indeed,

79  models that include EH yield better estimates of the richness of the Cape flora (Thuiller et al., 2006; Cramer &

80  Verboom, 2016). Mediterranean-type terrestrial biodiversity hotspots, such as the Cape flora included in the

81  models by Cramer & Verboom (2016), present interesting study systems in which to investigate the relationship

82  between the environment and species richness. These systems exhibit far greater species richness than

83  predicted by their areas, productivities and latitudes (Cowling et al., 1996; Kreft & Jetz, 2007). There are five

84  Mediterranean biodiversity hotspots on Earth: the California Floristic Province, the Mediterranean Basin, the

85  Chilean Winter Rainfall-Valdivian Forests, the Greater Cape Floristic Region, and the Southwest Australia

86  Floristic Region (Cowling et al., 1996; Hopper & Gioia, 2004; Cook et al., 2015). These ecosystems have

87  regular fire-cycles (Cowling et al., 1996), climatic buffering, and long term stability (Kreft & Jetz, 2007),

88  shrubby, sclerophyllous flora (Hopper & Gioia, 2004). Together, they account for ca. 20% of global vascular

89  plant species, yet only ca. 5% of global land surface areas (Cowling et al., 1996). Various hypotheses have

90  been proposed to explain the high levels of plant species richness in these regions (Cook et al., 2015). The

91  species accumulation hypothesis states that the stability of these regions has allowed many species to accrue.

92  The species co-existence hypothesis states that these hotspots may facilitate greater degrees of species

93  co-existence in smaller spatial areas, due to fine-scale heterogeneity in their environments. Indeed, EH has

94  evolutionary implications too, stimulating ecological speciation across sharp environmental gradients.

95  Both the Southwest Australia Floristic Region (SWA) and the Greater Cape Floristic Region (Cape) are

96  Mediterranean-type biodiversity hotspots, particularly in terms of plant species. Where the Cape (with an area

97  of ca. 189,000 km$^2$) is known to contain about 11,400 plant species (about 0.060 species per km$^2$), SWA (area

98  of ca. 270,000 km$^2$) has about 3,700 species (0.014 species per km$^2$) (Manning et al., 2012). So, the Cape has

99  ca. 4.3 times as many species per km$^2$ as SWA. The Cape and SWA are appropriately often compared, due to

100  the similarities between their environments (e.g. oligotrophic soils, an oceanically buffered moderate climate)

101  and their plants' ecologies (Hopper & Gioia, 2004). These two regions present unique flora out of the five

102  Mediterranean systems, with high levels of endemism (Cowling et al., 1996), and many obligate fire-adapted

103  species (Cowling et al., 1996). Similarities withstanding, SWA is topographically and edaphically distinct from

104  the Cape. The former is topographically rather uniform (i.e. flat)—uniquely so among the world's five

105  Mediterranean-climate regions (Hopper & Gioia, 2004)). SWA possesses a mesoscale chronosequence dune

106  system (Laliberte et al., 2014; Cook et al., 2015), while the Cape is mountainous, topographically

107  heterogeneous, and therefore associated with a large degree of spatial climatic variability, with a fine-scale

108  mosaic of geologies and soils (Cowling et al., 1996; Cramer et al., 2014; Verboom et al., 2017).

109  Both regions have sources of edaphic heterogeneity, but at different scales. This edaphic variability may aid in

110  explaining the species richness in these regions (Beard et al., 2000; Verboom et al., 2017). EH of many forms

111  will likely be important in macro-ecological models in both regions, as both regions have been relatively

112  environmentally stable over evolutionary time-scales (Wardell-Johnson & Horwitz, 1996; Hopper & Gioia,

113  2004; Lambers et al., 2010; Cramer et al., 2014; Laliberte et al., 2014; Cook et al., 2015). For the Cape, high

114  levels of species richness are thought to result from long term climatic stability, and fine grain variation in

115  geology and soils (Cramer et al., 2014). The question thus arises whether heterogeneity is a significant

116  contributor to SWA species richness. In the absence of topographic variability in SWA, it is proposed that the

117  heterogeneity of that region is due to the juxtaposition of soil types (Laliberte et al., 2014; Cook et al., 2015),

118  creating extreme edaphic variation.

119  Our hypotheses concern the Cape and SWA's environments and floras. Our main hypothesis is that the Cape

120  possesses greater abiotic heterogeneity, and at finer grain, compared to SWA, such as to explain the Cape's

121  greater species richness per unit area, and proposed greater levels of species turnover between areas. We also

122  conjecture that the heterogeneity that predicts species richness in SWA will be more pronounced in terms of

123  edaphic variables. Here we attempt to assess five key predictions of this hypothesis, additionally investigating

124  a seventh prediction to test the conjectured role of edaphic heterogeneity in SWA. Dealing with the two

125  regions' environments, we assess (i) whether the Cape environment is more heterogeneous than that of SWA

126  and (ii) whether the Cape environment has more pronounced heterogeneity at finer scales than that of SWA.

127  Dealing with the distribution of species in the two regions, we assess (iii) whether the Cape exhibits greater

128  levels of species turnover between areas. Relating each regions' environment and flora, we finally assess (iv)

129  whether species richness and species turnover are adequately predicted by EH in both regions and whether (v)

130  species richness and species turnover are better predicted by different forms of EH in either region (e.g. the

131  importance of edaphic heterogeneity in SWA).

## 2    Materials and methods

### 2.1    Overview

Our analyses required boundaries for each region and geographically explicit environmental data and vascular plant occurrence records. The environmental variables chosen (Table 1) for this study were intended to cover a reasonable spread of climatic, edaphic, and ecologically relevant environmental axes, and are not intended to be exhaustive. We selected variables describing topography (elevation), productivity (NDVI), soil status and climate and climatic seasonality.

We carried out this investigation at four principal spatial scales: 0.05º x 0.05º squares (the finest common resolution among the environmental data sources used), quarter degree squares (QDS) (Larsen et al., 2009), half degree squares (HDS) (Larsen et al., 2009) and three-quarter degree squares (3QDS). For the Cape, most plant occurrence records are only accurate to QDS level. Thus, analyses involving species occurence data were necessary limited to scales including and above QDS.

Analyses were performed in R v3.4.0–3.5.1 (R Core Team, 2018). Version-numbers of specific R packages used are presented in the bibliography.

### 2.2    Environmental data sources

The GCFR was treated as the area occupied by the Succulent Karoo and Fynbos biomes in the current delineation of South Africa's biome boundaries (Mucina & Rutherford, 2006). The SWAFR was treated as the areas occupied by the Southwest Australia savanna, Swan Coastal Plain Scrub and Woodlands, Jarrah-Karri forest and shrublands, Southwest Australia woodlands, Esperance mallee, and Coolgardie woodlands in the World Wildlife Fund Terrestrial Ecoregions dataset (Olson et al., 2001) in order to closely match the currently delineated SWAFR (Gioia & Hopper, 2017, @Hopper2004). For the sake of readability, we shall refer to the GCFR and SWAFR simply as the Cape and SWA from hereon.

Geospatially-explicit raster layers were acquired for a selection of environmental variables (Table 1), for the regions of interest. Raster data were re-projected to a common coordinate reference: WGS84 (NIMA, 2000), using the "rgdal" (Bivand et al., 2017) package in R (R Core Team, 2018). All data were re-sampled to 0.05º resolution using the "resample" function in the R package "raster" (Hijmans, 2016), with the "bilinear" method.

6

An emphasis was made on using satellite-derived environmental data in this work, in order to minimise

differences in data quality and methodologies between the Cape and SWA. Additionally, satellite-derived data

have been shown to benefit regional-scale species distribution models (Deblauwe et al., 2016), thus motivating

their use in this regional-scale study. The environmental data used in this study were derived from NASA's

SRTM digital elevation model (Farr et al., 2007), NASA's MODIS/Terra spectroradiometric data for land

surface temperature and NDVI, the Climate Hazards Group's CHIRPS rainfall dataset (Funk et al., 2015), and

the International Soil Reference and Information Centre's SoilGrids250m edaphic dataset (Hengl et al., 2017)

(Table 1). SRTM and MODIS are entirely derived from satellite measurements, whereas CHIRPS is

interpolated from weather station data with satellite-derived radiometric measurements. SoilGrids250m is a

machine-learning derived product, based on soil measurements as a function of many covariates, including

MODIS and STRM sources (see Hengl et al., 2017), using random-forests and other classification-tree-based

methods, including gradient-boosting. For the soil data considered here (Table 1), we used depth-interval

weighted average values as the value for a particular soil variable in a given place.

Climatic and spectral data arise from satellites monitoring properties of the Earth's surface through time. We

therefore use the mean annual values for rainfall, surface temperature, and NDVI in each pixel in our analyses.

Pronounced seasonality of rainfall is a known feature of mediterranean systems . We describe this seasonality

by computing computing the precipitation in the driest quarter (PDQ), using methods based on the "biovars"

function in the R package "dismo".

## 2.3   Plant occurrence data

Geospatially-explicit records of vascular plant occurrences were downloaded from the Global Biodiversity

Information Facility (GBIF, Table 1). Queries were made for tracheophyte records from within the borders of

the Cape and SWA as treated here (GBIF,, **???**). Only records with defined species and intra-specific ranks

were kept. Intra-specific occurrences were treated as simply being representative of their species. This resulted

in FIXME unique species names in the Cape, and FIXME in SWA.

We cleaned these data using the R package "taxise" (Scott Chamberlain & Eduard Szocs, 2013; Chamberlain et

al., 2018) to check that these species names had accepted-status among taxonomic databases. We queried two

major taxonomic databases: the Global Name Resolver (GNR), and the Taxonomic Name Resolution Service

(TNRS). Should either one of these services return at least one match for a given name, then that name was

accepted. Those names for which no full binomial matches were found in either database were excluded from

187   the final list of species. The number of species names excluded totalled at FIXME and FIXME for the Cape

188   and SWA respectively. Especially for SWA, these numbers may be deemed appreciably high. But, the

189   occurrence records that would be dropped, as a consequence of these names' removals, appeared randomly

190   distributed in geographic space in both regions. As such, any effect of the loss of these records in this analysis

191   is likely uniform within the two regions.

192   After the unaccepted names were removed, it was important to ensure that a species was not listed under

193   multiple synonyms. Such cases would skew estimates of species richness and turnover in this study. In light of

194   this, the remaining names were queried in the Tropicos and Integrated Taxonomic Information System (ITIS)

195   databases for their known synonyms, again using "taxize". These were collated to produce a nomenclatural

196   "thesaurus" for the Cape and SWA species. This consisted of a list of the accepted species names in a region,

197   each associated with a list of known synonyms. We amended species' names in the GBIF occurrence data, in

198   order ensure species were listed under only one of these synonyms, replacing all appearances of a species'

199   synonyms with the first synonym used in the list.

200   Lastly, We removed any species from both regions that are invasive aliens or non-indigenous. Alien species

201   lists for plants in South Africa and Australia were acquired from the IUCN's Global Invasive Species Database

202   (http://www.iucngisd.org/gisd/).

203   The final total plant species richness in each region was FIXME and FIXME for the Cape and SWA

204   respectively. These final collections of species occurrence records were converted to raster-layers, wherein

205   pixel-values represented the species richness of vascular plants within that pixel. These rasters were produced

206   at QDS, HDS, and 3QDS resolutions.

207   **2.4   Analyses**

208   **2.4.1   Quantifying environmental heterogeneity**

209   In order to assess predictions (i) and (ii), we needed to describe the EH in both regions. Using the R package

210   "raster" (Hijmans, 2016), we used a modified version of the "roughness" index in the "terrain" function. For a

211   three by three neighbourhood **N** of cells, our index of roughness $R$ is the average square-root of the squared

212   difference between each of the $n$ neighbour cells' values $x_i$ and the central focal cell's value $x_{\text{focal}}$:

$$R(\mathbf{N}) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(x_{\text{focal}} - x_i\right)^2} \tag{1}$$

213   This value, notionally equivalent to the standard deviation of values relative to the focal value, is ascribed to

214   the focal cell. Note, in order to use as much data from within regions' borders as possible, roughness was

215   computed if a focal cell had at least one neighbour cell. Using this index, we produced raster layers of each of

216   our nine environmental variable's heterogeneity. We compared the distributions of "roughness" values in each

217   variable in each region with non-parametric Mann-Whitney $U$-tests, as almost all variables were highly

218   non-normal, and could not be normalised by log-transformations. We also compare the effect size of the Cape

219   vs SWA using the "common language effect size" ($CLES$), using the R package "canprot". The $CLES$ is the

220   proportion of all pairwise comparisons between two sample groups' observations where one group's value is

221   greater than the other's. We calculated the $CLES$ as the proportion of pairs where Cape roughness values

222   were greater than that of SWA. This allowed us to assess prediction (i). To compare the spatial scales of

223   heterogeneity (prediction (ii)) between each region, we repeated this analysis at all four spatial scales. This

224   entailed recalculating the roughness layer for each variable after the original layer (0.05 degrees resolution) had

225   been rescaled to each of the coarser resolutions.

### 2.4.2   Quantifying species turnover

227   Regarding prediction (iii), we wished to compare the general degree of species turnover in each region. To

228   compare the extent of species turnover between the Cape and SWA, we determined two metrics of species

229   turnover. The first, computes the mean species turnover as Jaccard distances (Oksanen et al., 2018) between

230   each pair of QDS within each HDS ($\overline{J}_{\text{QDS}}$, based on HDS with $2 \leq n \leq 4$ QDS) in both regions. The second

231   is defined in terms of Whittaker's additive definition of $\beta$-diversity , as follows:

$$\gamma = \alpha + \beta \tag{2}$$

232   Here, we treat species richness at the HDS-scale ($S_{\text{HDS}}$) as $\gamma$-diversity and at the QDS-scale ($\overline{S}_{\text{QDS}}$) as

233   $\alpha$-diversity. Intuitively, the species richness of an area is the result of some combination of the richness of sites

234   within that area and the difference in species complements between those sites. Thus, we partition $\gamma$-diversity

235   as in Equation (2), such that $\beta$-diversity is the difference between $\gamma$- and $\alpha$-diversity. We compare the

236  distributions of $\overline{J}_{\mathrm{QDS}}$ and $T_{\mathrm{HDS}}$ using non-parametric Mann-Whitney $U$-tests, in order to guard against

237  non-normality.

### 2.4.3  Predicting richness and turnover with environmental heterogeneity

239  Regarding prediction (iii), we wished to compare the general degree of species turnover in each region. For (iv)

240  and (v) we modelled species richness ($S$) and turnover as a function of various combinations of environmental

241  and environmental heterogeneity variables in both regions using boosted regression-tree (BRT) modelling

242  techniques. This allowed us to explore which axes of environmental heterogeneity are most influential on

243  vascular plant species richness and turnover, and the differences in the importance of such axes between the

244  Cape and SWA.

245  BRTs are a flexible machine learning-based model of response variables and do so without involving normal

246  null-hypothesis significance testing (Elith et al., 2008), and have been employed previously to model species

247  richness (see Mouchet et al., 2015; Thuiller et al., 2006; Cramer & Verboom, 2016) as macro-ecological

248  models. BRTs are developed through the iterative generation of non-linear regression trees. BRTs are an

249  ensemble-approach, in which a prediction $\widehat{y}_i$ is based on the weighted sum of the predictions of progressively

250  "less important" regression trees ($t_k$), as opposed to the predictions of one tree (Elith et al., 2008). For $k \to nt$

251  number of trees, where each tree is itself a function of the matrix $\mathbf{X}$ of $j$ predictor variables ($t_k = f(x_{ij})$), a

252  BRT-model can be represented as follows:

$$\widehat{y}_i = \sum_{k=1}^{nt} w_k t_k \tag{3}$$

253  BRTs have two major meta-parameters over which users have control (Friedman, 2001): the learning rate ($lr$,

254  the rate at which iterative trees reduce predictive deviance during model-training, controlling the contribution

255  of each tree to the final model) and tree complexity ($tc$, the number of nodes on a given regression-tree, i.e. the

256  maximum interaction depth the model is permitted to fit).

257  BRTs were implemented here to predict both vascular plant species richness and turnover in each HDS, as a

258  function of environmental variables and environmental roughness values in those cells, as Gaussian responses,

259  thus resulting in two BRT-models for each region. We treated richness as $S_{HDS}$ and turnover as $\overline{J}_{\mathrm{QDS}}$. The

260  natural logarithm of species richness was used, in order to satisfy the assumptions of a Gaussian response.

261  Note, this is not strictly because BRTs have any parametric assumptions concerning the distribution of the

262  response variable, but rather to aid in applying the Gaussian-family of BRT algorithms to the richness data

263  available. Additionally, BRTs were implemented to predict vascular plant species richness at the QDS-scale

264  ($S_{QDS}$), thus resulting in a total of six BRT-models presented here.

265  As recommended by Elith et al. (2008), BRT models were trained on a set of non-collinear predictor variables

266  using "gbm.step" in "dismo" (Hijmans et al., 2017) and "gbm" . Collinear predictor variables can skew the

267  interpretation of results, as the relatively influence of mutually collinear variables is reduced. Collinearity

268  among the nine environmental predictor variables and their respective nine roughness-equivalents was assessed

269  using "removeCollinearity" in the R package "virtualspecies" (Leroy et al., 2015) separately for each region,

270  such that variables were no more than 80% collinear (Pearson's $r \geq 0.80$). When faced with a cluster of

271  collinear variables, one variable was chosen manually therefrom. Where possible, the roughness-equivalent of

272  an environmental variable was included if its absolute-equivalent could also be included. When interpreting the

273  results of BRTs, it is important to consider the effects of the variables included as representative of the effect of

274  the excluded variables with which it was found to be collinear.

275  In order to select ideal $lr$ and $tc$ all models (described below) were trained on the final non-collinear predictor

276  sets iteratively for 25 combinations of a range of $tc$ values (1 to 5) and a range of $lr$ values (0.01, 0.005, 0.001,

277  $5 \times 10^{-4}$, $1 \times 10^{-4}$). The function "gbm.step" optimises the number of trees ($nt$) using cross-validation during

278  model training (Elith et al., 2008) by halting iteration when predictions begin to overfit. For all models, we

279  used 10 cross-validation folds (i.e. use 10 different randomly selected training data sets), a tolerance-threshold

280  of 0.001, a bagging-fraction of 0.75 (proportion of training data randomly chosen to generate each tree), and

281  trained models starting with 50 trees, with each iterative step adding 50 trees at a time, up to a maximum of

282  10,000 trees. Following this iterative parameter optimisation, Gaussian BRT models were constructed with

283  $tc = 3$ and $lr = 0.001$, along with the other settings described.

284  The optimum configuration of $lr$ and $tc$ for the final model is a trade-off between model fit (e.g. pseudo-$R^2$;

285  Equation (4)) and complexity ($nt$). A $tc$ of 5 was chosen for the final model. This follows the

286  recommendations of Elith et al. (2008), where $lr$ and $tc$ are advised to be adjusted inversely. This was chosen

287  in order to account for the complex interactions determining species richness. To avoid overfitting, an

288  intermediate $lr$ of 0.001 was chosen.

### 2.4.4 Assessing BRT-predictions' fit

BRT-model performance can be described by measuring the variance in a dataset a BRT-model has explained, quantified here by $R^2_{\text{pseudo}}$, which is the proportion of null deviance $D_{null}$ explained by some model $i$. Formally, it is defined as follows:

$$R^2_{\text{pseudo}} = 1 - \frac{D_i}{D_{\text{null}}} \tag{4}$$

The derivation of this metric is not easy to interpret, as it is not immediately clear what model deviance is. Alternatively, comparing expected (i.e. model-predicted) and observed data has more heuristic appeal. We employed this metric of BRT-model performance too. We regressed expected against observed richness and turnover, and calculated the $R^2$-value for those regressions (hereafter $R^2_{\text{E}-\text{O}}$).

The BRT-model fitting algorithm contains intrinsic stochasticity, due to the random partitions made in a dataset during cross-validation. Though this randomness is usually negligible (e.g. variables' contributions vary from run-to-run by a few decimal places), we reran each of the six BRT-models (see above) 1000 times in order to account for this stochasticity. Where indicated, we either present the averages of these replicate-models' results or the results of a representative model from each set of replicates.

In order to assess the reliability of the conclusions drawn from these models, we randomly permuted the response data ($S_{\text{QDS}}$, $S_{\text{HDS}}$ and $\overline{J}_{\text{QDS}}$) with respect to the environmental and heterogeneity data, and reran all six BRT-models 999 times (with the final non-collinear predictor sets and preconfigurations above). This also allows us to remove any effect of spatial autocorrelation in generating the observed correlations between patterns of species occurrence and environment (Cramer & Verboom, 2016), and to allow us to assess the significance of our results relative to a random null. Notably, as the predictor variables themselves are likely spatially autocorrelated, correlation structure in model residuals is accounted for by the correlation structure in the environmental data. Nonetheless, we wished to demonstrate our results more robustly and thus carried out these permutation tests. For all six models, the majority of the 999 permuted models failed to find associations between the response and predictor variables. The results of those that succeeded to fit a model to completion (usually ca. 200 out of 999) are presented. The replicate and permuted BRT-models were compared using various measures of model performance (above; $nt$, $R^2_{\text{pseudo}}$ (Equation (4)), $R^2_{\text{E}-\text{O}}$) and the ranks of these values for each replicate BRT-model relative to the 999 permuted models for that region/scope.

# 3 Results

## 3.1 Describing environmental heterogeneity across scales

Across all variables considered, the Cape is more environmentally heterogeneous in the majority of pairwise comparisons of grid-cells ($CLES > 0.50$, Mann-Whitney $U$-test: $P < 0.05$, Figure 1). The degree to which the Cape is more heterogeneous varies between environmental variables. These effects also vary somewhat across spatial scales. In some variables, the differentiation between Cape and SWA heterogeneity lessens at coarser scales (Figure 1b). Indeed, when comparing the overall ranking and medians of Cape vs SWA roughness values for each variable, we only find non-significant differences at the 3QDS scale (Mann-Whitney $U$ tests, $P > 0.05$, Figure 1b).

Most obviously, and as expected, topographic heterogeneity is generally greater in the Cape than in SWA (Figure 1). Though SWA has a slightly wider distribution of elevational roughness values at coarse scales compared to fine scales than the Cape, the relative difference in heterogeneity between the two regions seems invariant with spatial scale ($CLES \approx 0.95$, Figure 1b). This concurs with our expectations, as the Cape is mountainous and known to have steep elevational gradients , while SWA is much more topographically uniform. Elevational roughness differs between the two regions in the manner we expected, giving us more confidence in the environmental reality of our other results here.

Climatic heterogeneity is less differentiated between the Cape and SWA than with topographic heterogeneity (Figure 1a), though the Cape is indeed more climatically heterogeneous (Figure **??**b). Notably, the difference between roughness in mean annual rainfall (R MAP) and land surface temperature (R Surface T) in the Cape and SWA is less pronounced when considered at coarse spatial scales (Figure **??**b). At all spatial scales considered, roughness in rainfall seasonality (R PDQ), however, is equally more heterogeneous in the Cape than SWA. Biological productivity, as measured by NDVI, is fairly similarly heterogeneous in the Cape and SWA ($CLES < 0.60$, Figure 1). Concerning edaphic variables, the Cape and SWA are similarly heterogeneous at coarser scales, particularly in terms of CEC and Soil C ($CLES \approx 0.50$, Figure 1b).

## 3.2 Comparing species turnover in the two regions

Following calculations of $\overline{J}_{\mathrm{QDS}}$ and $T_{\mathrm{HDS}}$ for each HDS-cell in each region, we also used non-parametric Mann-Whitney $U$-tests to compare the distributions of values in the Cape and SWA. The Cape possesses

generally greater floristic turnover than SWA, for both measures of turnover defined here ($P < 0.0001$, Figure 2a,b). $\overline{J}_{\mathrm{QDS}}$ measures the average pairwise Jaccard distance between QDS-cells in each HDS-cell. $T_{HDS}$, however, represents the inferred $\beta$ component of $\gamma$-diversity. As $\gamma$-diversity ($= S_{HDS}$) in the Cape has a greater $\beta$-diversity component ($= T_{HDS}$) than SWA, the complement is necessarily true: $\gamma$-diversity in the Cape has a lesser $\alpha$-diversity component ($= \overline{S}_{\mathrm{QDS}}$) than SWA.

### 3.3 Predicting richness and turnover with environmental heterogeneity

We found vascular plant species richness and turnover both to be predicted primarily by environmental heterogeneity in the Cape (Figure 3a–c) and at least in-part by environmental heterogeneity in SWA (Figure 3d–f). Our six BRT-models performed adequately, and detected relationships between patterns of species occurrence and the environment that do not occur in the permuted datasets (Figures 4 and 3, Table 2).

BRT-models of species richness at the QDS-scale in each region generally performed best, as these models had fit greater number of trees ($nt$, Figure 4a), and possess higher $R^2$-values (Figure 4b,c). SWA models of species richness and turnover at the HDS-scale out-performed Cape models. At the QDS-scale, the Cape and SWA models performed equally well (Figure 4, Table 3).

Across our BRT-models of species richness and turnover, the importance of different environmental variables in predictions differed substantially between the Cape and SWA. Additonally, the relative importance of absolute and heterogeneity variables also differs between the Cape and SWA (Figure 3). Most obviously, species richness and turnover in the Cape are predicted mostly by environmental heterogeneity, which is not the case in SWA (Figure 3). Species richness and turnover in the Cape are predicted by a broad suite of environmental variables, with no individual variable contributing more than ca. 20% to any model prediction (Figure 3a–c). The SWA models' predictions, however, are largely determined by MAP (Figure 3d–f).

Species richness at QDS-scales ($= \overline{S}_{\mathrm{QDS}}$), and to a lesser extent at HDS-scales ($= S_{\mathrm{HDS}}$), in the Cape is predicted largely edaphic conditions (Figure 3a,b). Species richness in SWA, at both scales, is mostly predicted by MAP and other climatic variables (Figure 3d,e). Interestingly, topograhphic heterogeneity did not feature as highly in contributing to Cape predictions as we expected (Figure 3a–c).

It is important to consider variables not included formally in these BRT-models that were found to be collinear with some of the variables included (see SI). Here, we interpret the effects of variables excluded from the analyses as well as those included, as the forms and importances of these relationships are likely similar. In the

370   Cape (concerning clusters of collinear variables relevant to relevant to those retained during BRT-model

371   fitting), MAP was included in the BRT-analyses as representative of a cluster of collinear variables consisting

372   of itself, NDVI, surface T and soil C at the. Roughness in soil clay content represented itself, roughness in soil

373   pH and roughness in NDVI. In SWA, MAP was select as representative of itself, NDVI and soil C.

374   Our BRT-models of species richness in both regions rank environmental variables somewhat differently at

375   QDS- and HDS-scales (Figure 3a,b,d,e). These differences in rankings are similar to those between two

376   unrelated lists ($P_{1-2} > 0.05$, Figure 5). This suggests some scale-dependence of different environmental

377   variables' associations with species richness.

378   It is noteworthy that BRT-models of species turnover ($= \overline{J}_{\mathrm{QDS}}$, at HDS-scales) (Figure 3c,f) rank variables

379   similarly to models of richness at HDS-scales ($P_{2-3} \leq 0.005$, Figure 5). This is likely due to the fact that

380   proportional floristic turnover covaries with species richness. As such, though the signs of relationships

381   determining turnover may differ from those determining richness, the importances of different variables would

382   be similar.

383   In addition to different variables being more strongly associated with species richness and turnover in the Cape

384   compared to SWA (Figure 3), the forms of those relationships vary (Figure 6). We found MAP, and roughness

385   therein, to relate positively with species richness in both regions at both scales (Figure 6a,b,d,e). As MAP is

386   collinear with NDVI and soil C in both regions (and surface T in the Cape), this can be interpreted as the signal

387   of a biological productivity and resource availability associating with high levels of species richness.

388   The positive association of heterogeneity variables in the Cape as opposed to SWA (Figure 6a,b vs d,e) concurs

389   with their greater importance in BRT-model predictions (Figure 3).

390   The fact that species turnover ($T_{HDS} = \overline{J}_{QDS}$) in the Cape and SWA is largely predicted by the same

391   variables as species richness, but with opposite signs to its relationships (Figure 6c,f), is indicative of the

392   richness-dependence of the measure of floristic turnover used here (Jaccard distances) to quantify turnover at

393   the HDS-scale.


## 4   Discussion

395   Here we have provided support for the hypothesis that the difference in plant species richness between the

396   Cape and SWA is accounted for by the greater abiotic heterogeneity in the Cape. As expected, the Cape is

15

397   shown to possess (i) a quantifiably more heterogeneous environment that is (ii) generally heterogeneous at a

398   finer spatial scale than SWA, with (iii) greater levels of floristic turnover. We have shown that vascular plant

399   species richness (iv) can be explained in terms of environmental conditions including environmental

400   heterogeneity in both the Cape and SWA. Also, (v) the sets of environmental axes that explain plant species

401   richness differ between the Cape and SWA. These findings contribute towards an understanding of the

402   ecological conditions associated with high levels of species co-existence in these two regions, and strengthens

403   the generality of a positive relationship between EH and biodiversity.

404   Cook et al. (2015) summarised the suites of hypotheses for the high levels of plant diversity in

405   mediterranean-type ecosystems. Broadly, these biodiversity hotspots are thought to arise due to greater levels

406   of evolutionary diversification, ecological co-existence and time for species accumulation by either

407   immigration or in situ evolution. Our study concerns the dual role of EH in facilitating ecological speciation

408   along environmental gradients and a diversity of habitats in wich greater numbers of species may co-exist. We

409   have correlative evidence for these processes, in that EH was positively associated with high levels of species

410   richness and turnover in both regions. Further research should be aimed at distinguishing the roles of ecological

411   speciation and species co-existence associated with EH in mediterranean-type ecosystems. The role of

412   environmental stability is relevant here (Cook et al., 2015, Pinto-Ledezma et al., 2018), both by allowing more

413   continuous species accumulation when habitats persists through geologic time and by allowing environmental

414   gradients to be stable long enough for ecological speciation to occur (Pinto-Ledezma et al., 2018). In the

415   contexts of the Cape and SWA, this long term environmental stability, and the stability of EH through time, is

416   indeed expected to contribute to these regions biodiversity (Hopper, 1979; Cowling et al., 1996).

417   *Ed: I'd like to include more information on the landscape/geologic history of the regions here, and link this*

418   *to the notion of environmental stability and "species accumulation", sensu Cook et al. (2015) and*

419   *Pinto-Ledezma et al. (2018).*

420   The Cape and SWA regions present differentiable environmental spaces, each with varying degrees of

421   heterogeneity across spatial scales and environmental axes. The clear distinction of the regions' topographic

422   features is as expected. The Cape region has been found previously to have the second highest median

423   topographic heterogeneity of the five Mediterranean-climate regions (Bradshaw & Cowling, 2014). In the

424   models developed by Cramer & Verboom (2016) for South Africa, roughness in topography was largely

425   superseded as an important predictor of species richness by other roughness variables. This is the case with our

426   results too, likely as topographic complexity can be considered a proxy for gradients in other environmental

16

axes, such as rainfall or temperature. In line with our hypotheses, a combination of absolute and heterogeneity variables are associated with species richness and turnover. These effects are region-specific, particularly in that patterns in the Cape are more strongly determined by edaphic conditions and EH while SWA patterns are more strongly determined by absolute environmetal conditions, particularly climate. Kreft & Jetz (2007) modelled global terrestrial vascular plant species richness, which focussed primarily on using absolute environmental variables, underestimated the richness of the Cape flora. Though Kreft & Jetz (2007) did include topographic heterogeneity in their predictor set, topography is often a proxy for more biologically meaningful variables (Cramer & Verboom, 2016). This explains why the inclusion of these variables (e.g. roughness in mean annual precipitation) yields more accurate predictions of species richness. Indeed, Thuiller et al. (2006) also included topographic heterogeneity.

We also detected strongly non-linear relationships between environmental and heterogeneity variables and species richness and turnover. We expected this, as non-linear relationships are common in ecology (Zanne et al. 2018), making our use of machine-learning methods appropriate here.

Comparing our models to those fit by Cramer & Verboom (2016), we did not directly find spatial heterogeneity in NDVI to be an important determinant of species richness and turnover patterns in the Cape. Rather absolute MAP, which is collinear with absolute NDVI, was found to be important. This speaks to the relationship between vegetation productivity and climate, and how suitably productive habitats can foster more species . An additional interpretation of spatial heterogeneity (and indeed absolute conditions) in NDVI is that of changing canopy light-contexts (Cramer & Verboom, 2016). Power et al. (2017) found that floristic turnover between biomes in South Africa to be associated with differences in biomes' leaf area indices. We found NDVI to be more heterogeneous across the Cape than SWA (Figures 1). NDVI is an integrating variable, which captures information about productivity, light availability, and soil nutrients (Power et al., 2017). Absolute NDVI and its correlates can be interpreted as contributing to predicting species richness in the Cape. This demonstrates the role of ecological productivity in facilitating the co-existence diverse species assemblages. EH, then, must be considered alongside resource- and energy-availability axes when intepreting the drivers of species richness. In so much as a diverse environmental space supports more species, the materials and productivity required for biota to thrive are also needed to support species (Gaston, 2000; Kerr et al., 2001; Bøhn & Amundsen, 2004; Kreft & Jetz, 2007). As such, our findings, along with those of previous studies (Rensburg et al., 2002; Thuiller et al., 2006; Kreft & Jetz, 2007; Cramer & Verboom, 2016), suggest that there are ecological and evolutionary consequence to both resource availability and EH.

457  We detected a strong set of links between various edaphic properties and vascular plant species richness and

458  turnover in the Cape. The importance of absolute pH speaks to the association of many Cape communities with

459  nutrient impoversished, acidic soils . We did not, however, find edaphic conditions to be as important as we

460  expected in SWA. This may be due to a real lack of importance of edaphic features in driving assemblage

461  patterns in SWA. ALternatively, there is a risk that the quality of soil data provided to our models, from

462  SoilGrids250m , is insufficient for our purposes here. There are two potential reasons for this. Firstly, teh

463  SoilGrids250m data product is acknowledged to be less accurate at the scales we investigate here , such that the

464  true variation in soil properties in SWA is not represented in our dataset. Secondly, it could be that the link

465  between soils and assemblage patterns in SWA exists, but simply at a finer scale than that at which we

466  performed our analyses here. These two reasons are not mutually exclusive, in that SoilGrids250m may be

467  both inaccurate at the scales we consider here and that our analysis considers scales greater than those

468  important in SWA plant community assemblage patterns. This is particularly relevant to our study, as the Cape

469  has a much wider range of scales exhibited in the heterogeneity across its environmental axes. Notably, each

470  region has finer scale heterogeneity in some variables, and coarser scale in others—neither region is more fine

471  or coarse than the other over all environmental axes.

472  *Ed: I'd obviously like to include more information on the landscape/geologic history of the regions here*

473  *too, and with reference to particular soil variables (Mike?).*

474  Another issue related to spatial scales, as with any ecological study, is that of spatially uniform and

475  representative plant species occurence data. Efforts were made herein to minimise differences in data-quality

476  between SWA and Cape by using remote-sensing derived environmental data. Species occurence data,

477  however, prove difficult to obtain with uniform methodology. South African biodiversity data is stored only to

478  QDS-level accuracy. As such, our analyses were limited to that as the finest spatial scale, such that our

479  environmental datasets necessarily, then, were also limited to the QDS-scale. This may complicate the

480  interpretation of our analyses, as these data may be too coarse in scale to capture patterns of species richness

481  and turnover and their associations with the environmenta in reality in empirically heterogeneous environments

482  like the Cape and SWA.

483  In addition to the scale of our data, spatial bias in species occurence data must also be considered. GBIF data

484  are sourced from both herbaria and plot-monitoring datasets. The relative contribution of each of these to GBIF

485  data may vary between countries and regions—i.e. between the Cape and SWA. As recently noted by Guerin

486  (2018), herbarium records and plot-based inventories describe regional flora with varying accuracy, especially

487  with respect to documenting high levels of floristic turnover. We emphasise here our concerns that there the

488  floral occurence data and soils information use in this study are especially at risk of spatial bias and/or

489  inaccuracies.

490  ...

491  Jaccard distances, as used here as $\overline{J}_{\mathrm{QDS}}$, are highly richness dependent. Although they represent the

492  proportional floristic turnover between cells, this proportion itself is sensitive to the richness of sites under

493  comparison. For example, the turnover between two low richness sites has is likely to be greater by chance,

494  due to the absence or presence of a few species, than the turnover between two high richness sites, where the

495  absence or presence of a few species does not greatly affect the turnover calculated.

496  ...

497  Our findings here are correlative. There are, however, many proposed mechanisms to explain the correlative

498  signals demonstrated here. Our findings support the hypothesis that mediterranean systems' plant species

499  richness is a function of spatial variability in environmental conditions. This can stimulate diversification, and

500  maintain that diversity by providing a range of habitats for species co-existence. Oligtrophic soils can stimulate

501  an increase in functional diversity, through the evolution of diverse nutrient acquisition strategies (Lambers et

502  al., 2010; Verboom et al., 2017)—e.g. sclerophylly (Cramer et al., 2014; Cook et al., 2015). An aspect of the

503  environment I have neglected to consider is fire, shown to also contribute to predictions here in the Cape

504  (Cramer & Verboom, 2016). Cardillo (2012) have shown the structuring forces behind species co-occurrence

505  patterns, and thus likely species richness,differ between species-pairs with different post-fire responses and

506  those with similar post-fire responses.

507  Following from the understanding that functionally diverse assemblages, which are more likely to be more

508  species rich, are likely to arise and/or occur in areas with diverse ecological pressures (Molina-Venegas et al.,

509  2015), one would expect, then, heterogeneous habitats such as those in Mediterranean-type biodiversity

510  hotspots to exhibit high levels functional beta diversity along steep environmental gradients (Molina-Venegas

511  et al., 2015). If the niches concerning these functions are phylogenetically conserved among those biota, then

512  one would also expect high levels of species and phylogenetic beta diversity along these gradients

513  (Molina-Venegas et al., 2015). This concurs with the notion put forward by Power et al. (2017), wherein

514  megadiverse systems such as these represent the results of "phylogenetic niche conservatism on a

515  heterogeneous landscape". Thus, species and phylogenetic turnover should covary with environmental

19

heterogeneity in some way. Indeed, endemism, at certain scales, could also follow this pattern. Thuiller et al. (2006) demonstrated that there are phylogenetic and biome-related determinants of species richness. This makes sense in light of the low probability of lineages crossing biome boundaries in Mediterranean systems (Power et al., 2017). NDVI and light availability, and the heterogeneity therein, are associated with high levels of floristic turnover (Power et al., 2017). This may be indicative of ecological specialisation precluding species from crossing these boundaries, thus increasing the level of endemism within a region, while also increasing the level of turnover, and thus likely species richness, along environmental gradients. Although, this may be debated. Beard et al. (2000) state that the high levels of endemism in SWA are function of habitat specialisation to soil mosaics. Cf. Laliberte et al. (2014), who say that this endemism is likely due to environmental filtering along these soil turnover sequences, as opposed to the juxtaposition of specialised species along soil gradients.

*Ed: Is it worth expanding this discussion to functional diversity and phylogenetic diversity metrics too?*

Summarily, we have demonstrated support for the generality of EH as a meaningful predictor of patterns of species richness and turnover in mediterranean-type ecosystems. In SWA and the Cape, high levels of are also likely the results of long-term landscape and climatic stability (Hopper, 1979). Thus, the roles of environmental variability through space that persists through deep time are two important ways in which the environment relates to biodiversity in these regions. The efficacy of environmental stability through geological time should be explored further, comparing the Cape and SWA.

**Tables**

Table 1: Georeferenced vascular plant species occurence and environmental data sources used in this study. Data were acquired for the Cape and SWA regions, with the temporal extent of data products used described where applicable. Abbreviations are as follows: MAP, mean annual precipitation; PDQ, precipitation in the driest quarter; CEC, cation exchange capacity.

| Variable | Source | Temporal extent | Citation |
| --- | --- | --- | --- |
| Plant species occurrences | GBIF | TODO | @GBIFCape, @GBIFSWA |
| Elevation | SRTM v2.0 | | @Farr2007 |
| NDVI | MODIS (MOD13C2) | Feb. 2000 to Apr. 2017 | @MOD13C2 |
| **Climatic variables** | | | |
| Surface temperature | MODIS (MOD11C3) | Feb. 2000 to Apr. 2017 | @MOD11C3 |
| MAP | CHIRPS v2.0 | Jan. 1981 to Feb. 2017 | @Funk2015 |
| PDQ | CHIRPS v2.0 | Jan. 1981 to Feb. 2017 | @Funk2015 |
| **Soil variables** | | | |
| CEC | SoilGrids250m (CECSOL M 250m) | | @Hengl2017 |
| Clay | SoilGrids250m (CLYPPT M 250m) | | |
| Soil C | SoilGrids250m (OCDENS M 250m) | | |
| pH | SoilGrids250m (PHIKCL M 250m) | | |

Table 2: Average proportional-ranks for BRT-model performance measures ($nt$, $R^2_{\mathrm{pseudo}}$ (Equation (4)), $R^2_{\mathrm{E-O}}$ (see text)) of the 1000 replicate BRT-models relative to 999 BRT-models fit to permuted datasets. Each of the 1000 replicate BRT-mdels was ranked against the 999 permuted BRT-models. The average rank of each, as a proportion, is presented.

| Model | $nt$ | $R^2_{pseudo}$ | $R^2_{E-O}$ |
|---|---|---|---|
| GCFR | 1.000 | 1.000 | 1.000 |
| SWAFR | 1.000 | 1.000 | 1.000 |
| GCFR | 0.987 | 1.000 | 0.988 |
| SWAFR | 1.000 | 1.000 | 1.000 |
| GCFR | 0.977 | 0.992 | 0.979 |
| SWAFR | 0.997 | 1.000 | 1.000 |

Table 3: Estimated differences between replicate Cape and SWA BRT-models' performance measures ($nt$, $R^2_{\text{pseudo}}$ (Equation (4)), $R^2_{\text{E}-\text{O}}$ (see text)) following two-sided $t$-tests. Positive values indicate that the Cape models had greater values. In all cases, the Cape and SWA had highly significantly different values for these quality measures ($P < 0.0001$).

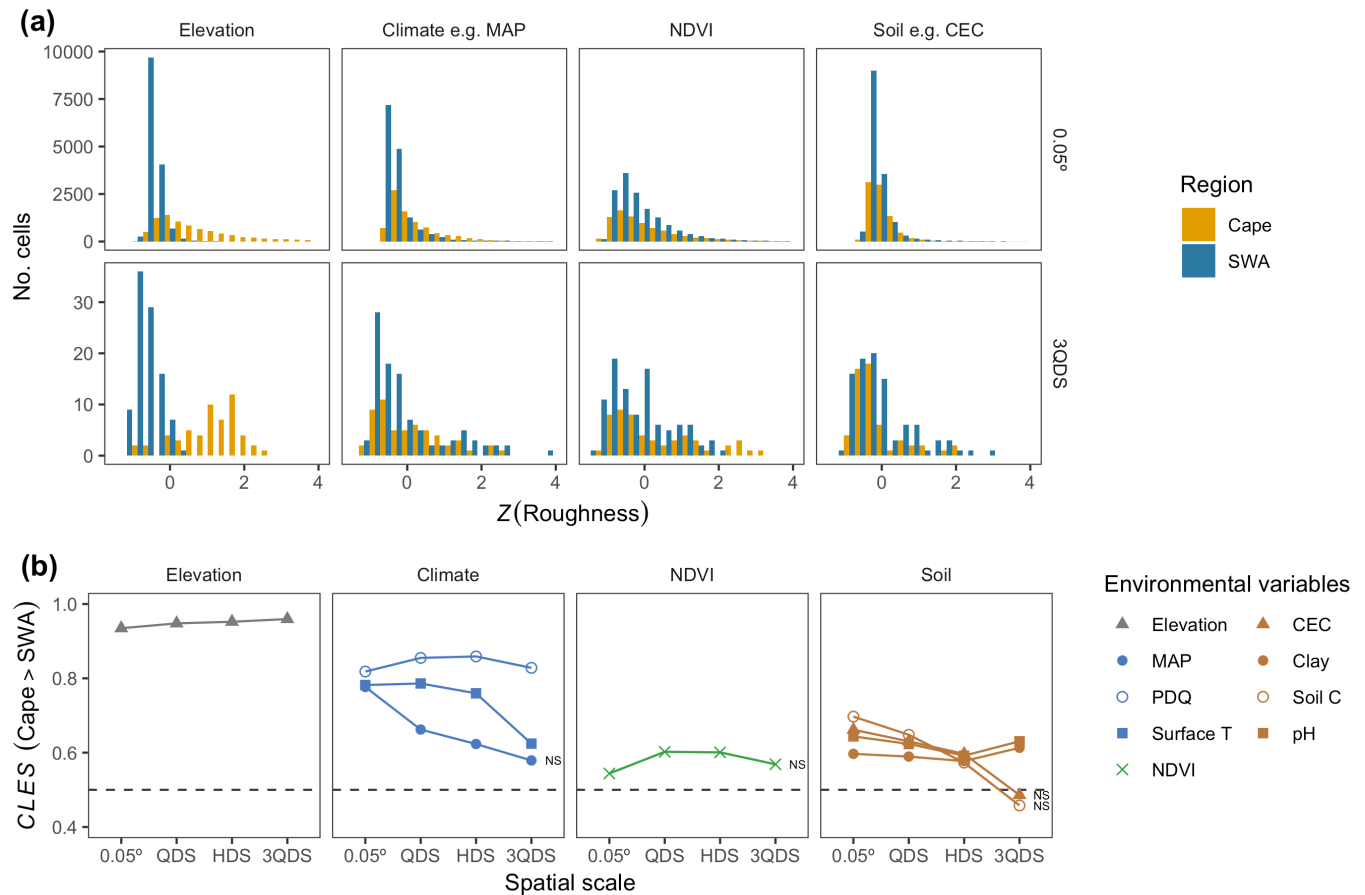| Model | $nt$ | $R^2_{\text{pseudo}}$ | $R^2_{\text{E}-\text{O}}$ |
|---|---|---|---|
| QDS-richness | 542.938 | 0.063 | -0.005 |
| HDS-richness | -808.994 | -0.064 | -0.233 |
| HDS-turnover | -997.045 | -0.052 | -0.296 |

**Figures**



Figure 1: Comparisons of different types of environmental heterogeneity between the the Greater Cape Floristic Region (Cape) and the Southwest Australia Floristic Region (SWA). We present (a) distributions of roughness values (Equation (1)) for example variables from each broad category of the environment concerned. (b) The common language effect size ($CLES$; see text) of Cape versus SWA roughness values is shown for all variables, grouped by broad categories of the environment, describing differences in the distributions of Cape and SWA roughness values. We used Mann-Whitney $U$-tests to assess differences in these distributions. Non-significant differences ($P_U > 0.05$) are denoted as such ("NS"). Note, $U$-tests were performed using only a random set of 5000 cells at the 0.05-degree-scale, as the $U$-test as implemented in R cannot handle more than that many values to compare.
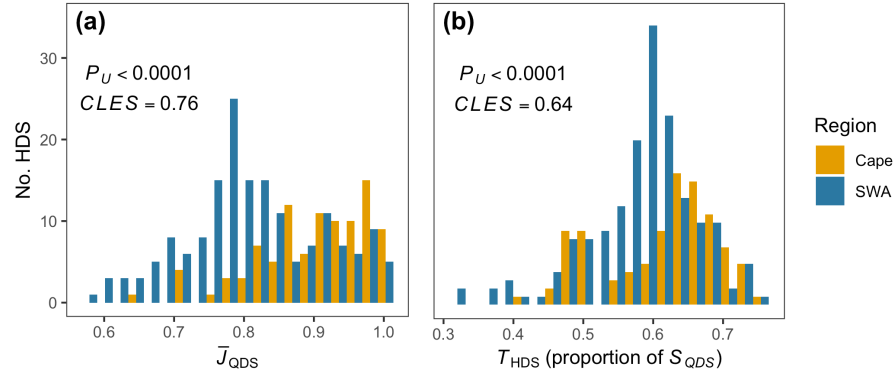
Figure 2: Species turnover, described in two forms ((a) mean Jaccard distance between QDS in each HDS ($\overline{J}_{\mathrm{QDS}}$), (b) additively defined turnover ($T_{\mathrm{HDS}}$, Equation (2)) as a proportion of HDS richness ($S_{\mathrm{HDS}}$)), compared between the Cape and SWA. Mann-Whitney $U$-tests between the Cape and SWA distributions of $\overline{J}_{\mathrm{QDS}}$ and $T_{\mathrm{HDS}}$ yielded significant differences (see $P$-values and common language effect sizes ($CLES$) inset).
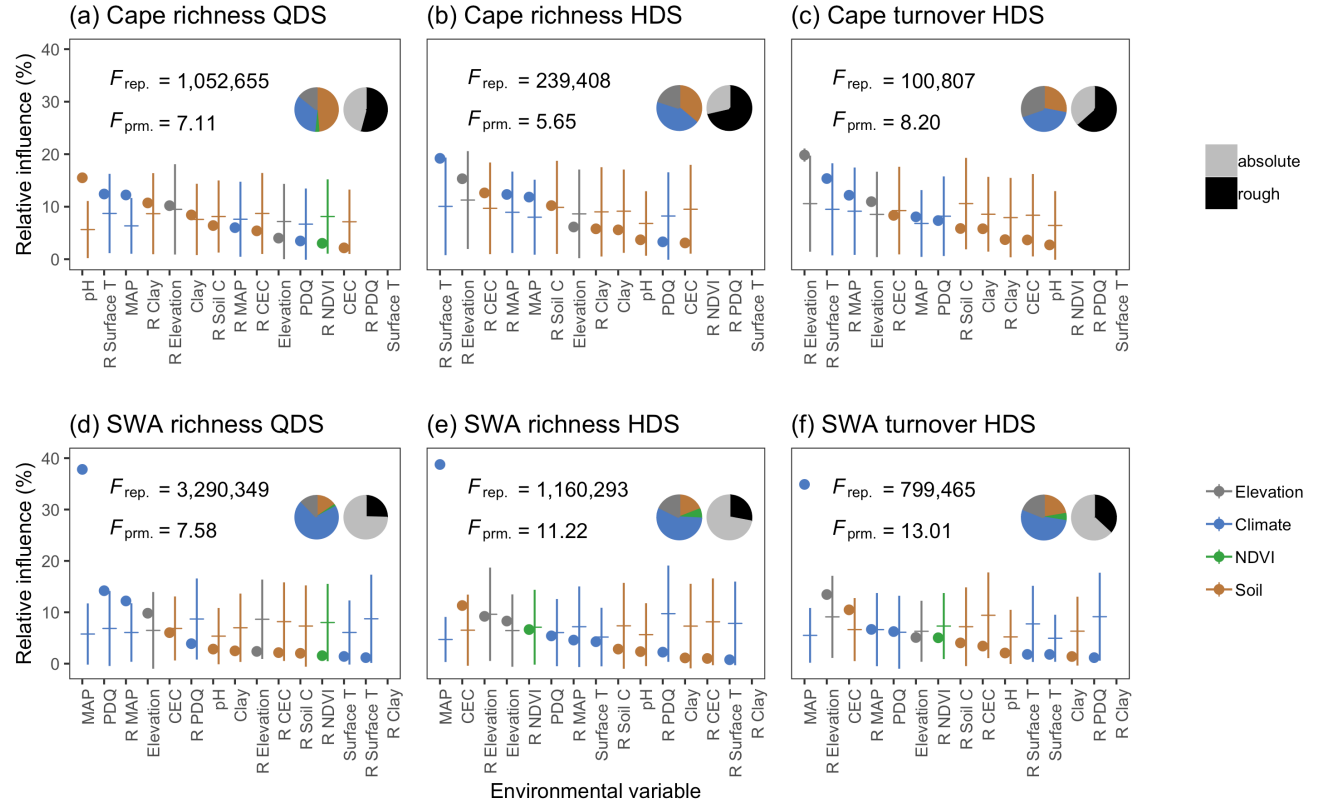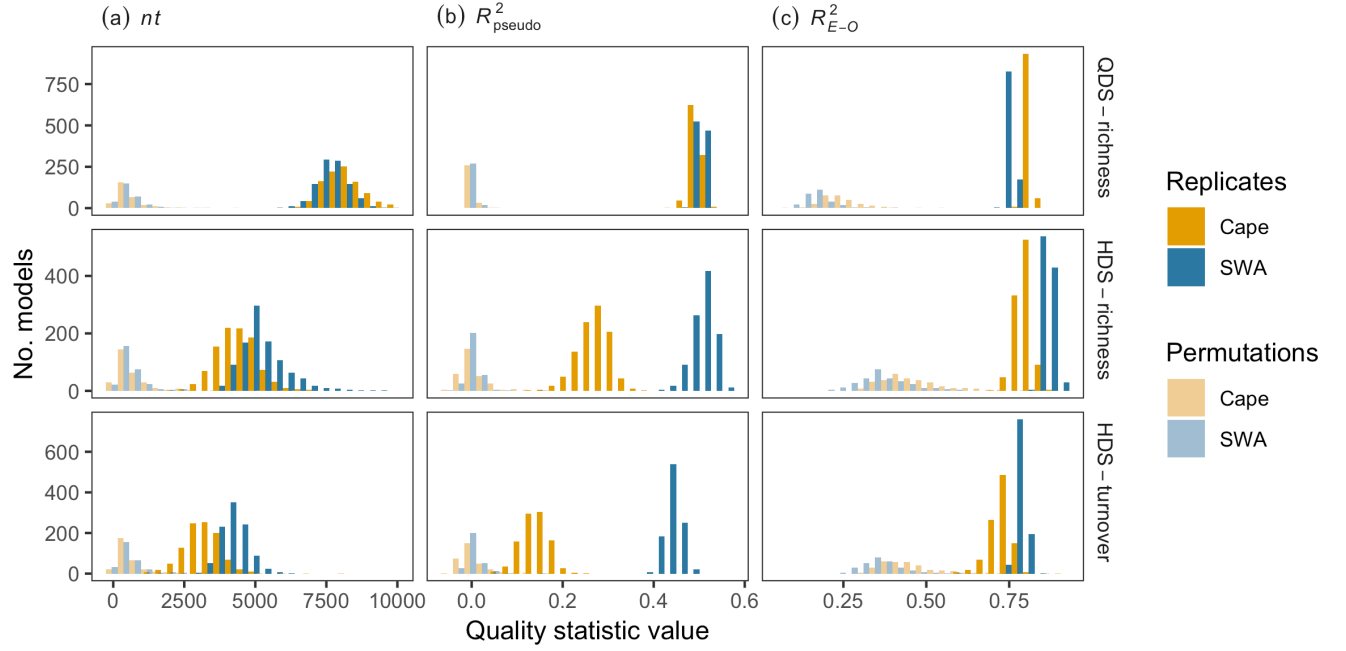
Figure 3: Relative influence of environmental variables (including heterogeneity variables—prefixed with "R") in boosted regression tree (BRT) model predictions in the Greater Cape Floristic Region (Cape, a–c) and the Southwest Australia Floristic Region (SWA, b–d) of vascular plant species richness at the (b,e) QDS-scale ($= \overline{S}_{QDS}$), (a,d) HDS-scale ($= S_{HDS}$) and (c,f) turnover ($= \overline{J}_{QDS}$). All BRT-models were permitted to fit three-way interactions between environmental variables. Points denote the mean contribution of an environmental variable to model-predictions across the 1000 replicate BRT-models for that region/scope. Horizontal ticks denote the mean for the 999 permuted BRT-models. Standard deviations above and below these means are shown with vertical lines. Note, in the case of the replicate, standard deviations are so small such that the vertical lines are obscured by the points. Colours represent the general category of the environment to which a variable belongs (keyed), as in Figure 1b. Left-most piecharts inset in each panel display the same information. Right-most piecharts group contributions according to whether a variable was absolute or roughness-transformed (keyed). $F$-statistics inset are for one-way ANOVAs of differences in variables' relative influences—for both the replicate ($F_{rep.}$) and permuted ($F_{prm.}$) BRT-models.

Figure 4: Distributions of three measures of boosted regression tree (BRT) model performance: (a) the number of trees in the model $nt$, (b) $R^2_{\text{pseudo}}$ (Equation (4)), (c) $R^2_{\text{E}-\text{O}}$ (see text). These measures are presented for the six sets of permuted (pale bars) and six sets of replicate BRT-models (dark bars) as in Figure 3, coloured according to the region of interest as in Figures 1a and 2. In all cases, replicate BRT-models almost entirely out-rank the permuted models in terms of performance (Table 2) and the Greater Cape Floristic Region (Cape) and Southwest Australia Floristic Region (SWA) models had significantly different values for each metric (Table 3). Note, the actual differences between Cape and the SWA models' values is not realistically significant in some cases (e.g. the difference in $nt$ between the Cape and SWA QDS richness models is statistically significant, but are observedly so similar as not to affect interpretation).
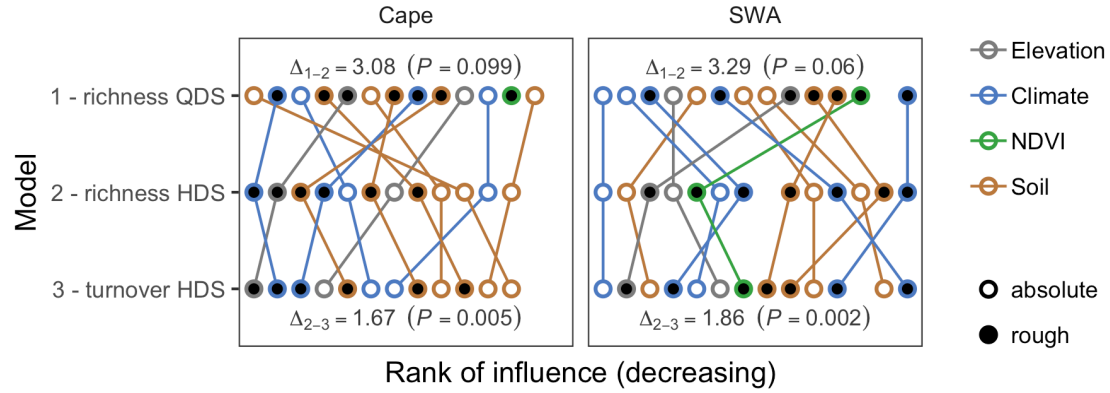
Figure 5: Differences in the rankings of environmental variables' (including heterogeneity variables) relative influences on boosted regression tree (BRT) model predictions of vascular plant species richness and turnover in (a) the Greater Cape Floristic Region (Cape) and (b) Southwest Australia Floristic Region (SWA) (as in Figure 3). Each point represents an environmental variable's rank in BRT-model importance, decreasing in importance from left to right. Rankings used here are the same as that of the average relative influence for variables across replicate BRT-models, presented in Figure 3. Coloured lines connect points representing the same environmental variable. Points' outlines are coloured according to the general category of the environment (keyed) to which a variable belongs, as in Figuress 1b and 3, while points' centres are coloured according to whether a variable was roughness-transformed or not. The comparisons of variables' rankings of interest are between QDS- and HDS-scale richness ($= \overline{S}_{QDS}$ and $S_{HDS}$ respectively; rows nos. 1 and 2) and between HDS-scale richness and turnover ($= \overline{J}_{QDS}$) (rows nos. 2 and 3). Statistics ($\Delta$- and $P$-values) inset at the top and bottom of each panel refer to these comparisons respectively. $\Delta$-values represent the average absolute difference in ranks across variables between two models' rankings. The associate $P$-value results from ranking the observed $\Delta$-values against 999 $\Delta$-values based on random permutations of variables' rankings (SI1), such that more significant $P$-values denote rankings more similar than would be expected by chance.
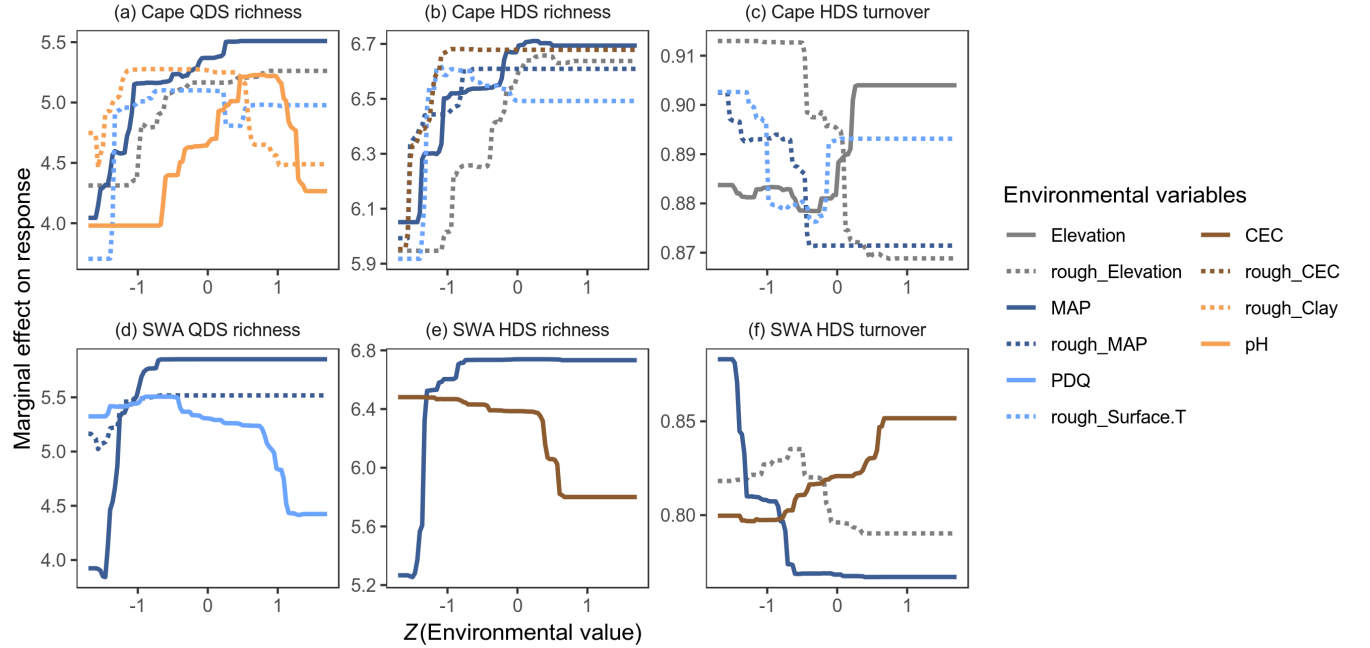
Figure 6: Marginal effects of environmental conditions and heterogeneity on vascular plant species richness at the QDS-scale (= $\overline{S}_{QDS}$; a,d), HDS-scale (= $S_{HDS}$; b,e) and turnover (= $\overline{J}_{QDS}$; c,f) in response variables in the Greater Cape Floristic Region (Cape; a–c) and Southwest Australia Floristic Region (SWA; d–f) following boosted regression tree (BRT) modelling. Marginal effect functions presented are derived from a representative BRT-model from the set of replicate BRT-models (for each of the six modelling cases) (see SI regarding how representative BRT-models were selected). Marginal effects represent the effect of a predictor variable when all other predictors are set at their means. Marginal effect functions are shown for environmental variables that contributed $\geq 10\%$ to a model's predictions. Functions are coloured as keyed, with solid lines representing absolute environmental variables and dotted representing heterogeneity variables ("rough"). Environmental variables were all rescaled here such as to be centred on zero (i.e. $Z$-transformed), facilitating comparison of functions' forms.

## Biosketches

**Ruan van Mazijk**

**Michael D. Cramer**

**G. Anthony Verboom**

## Author contributions

MDC and GAV conceived the study question, which RVM investigated under their supervision for his BSc Hons project. The analyses and programming work were largely devised by RVM, with input from the other authors, and was carried out by RVM. RVM wrote the first draft of the manuscript and all authors contributed equally thereafter.

## References

Baudena, M., Sánchez, A., Georg, C.-P., Ruiz-Benito, P., Rodríguez, M.Á., Zavala, M.A., & Rietkerk, M. (2015) Revealing patterns of local species richness along environmental gradients with a novel network tool. *Scientific Reports*, **5**, 11561.

Beard, J.S., Chapman, A.R., & Gioia, P. (2000) Species richness and endemism in the Western Australian flora. *Journal of Biogeography*, **27**, 1257–1268.

Bivand, R., Keitt, T., & Rowlingson, B. (2017) rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-7..

Bradshaw, P.L. & Cowling, R.M. (2014) Landscapes, rock types, and climate of the Greater Cape Floristic Region. *Fynbos: Ecology, evolution and conservation of a megadiverse region* (ed. by N. Allsopp, J.F. Colville, and G.A. Verboom), pp. 26–46. Oxford University Press, Oxford.

Bøhn, T. & Amundsen, P.-A. (2004) Ecological Interactions and Evolution: Forgotten Parts of Biodiversity? *BioScience*, **54**, 804.

Cardillo, M. (2012) The phylogenetic signal of species co-occurrence in high-diversity shrublands: different patterns for fire-killed and fire-resistant species. *BMC Ecology*, **12**, 21.

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B.G., Marchand, P., & Tran, V. (2018) *Taxize: Taxonomic information from around the web.*

Cook, L.G., Hardy, N.B., & Crisp, M.D. (2015) Three explanations for biodiversity hotspots: small range size, geographical overlap and time for species accumulation. An Australian case study. *New Phytologist*, **207**, 390–400.

Cowling, R.M., Rundel, P.W., Lamont, B.B., Arroyo, M.K., & Arianoutsou, M. (1996) Plant diversity in mediterranean-climate regions. *Trends in Ecology and Evolution*, **11**, 362–366.

Cramer, M.D. & Verboom, G.A. (2016) Measures of biologically relevant environmental heterogeneity improve prediction of regional plant species richness. *Journal of Biogeography*, 1–13.

Cramer, M.D., West, A.G., Power, S.C., Skelton, R., & Stock, W.D. (2014) Plant ecophysiological diversity.

*Fynbos: Ecology, evolution and conservation of a megadiverse region* pp. 248–272. Oxford University Press, Oxford.

Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.C., Wieringa, J.J., Ramesh, B.R., Stévart, T., & Couvreur, T.L.P. (2016) Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, **25**, 443–454.

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

Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007) The shuttle radar topography mission. *Reviews of Geophysics*, **45**, 1–33.

Funk, C.C., Peterson, P.J., Landsfeld, M., Pedreros, D.H., Verdin, J., Shukla, S., Husak, G., Rowland, J.D., Harrison, L., Hoell, A., & Michaelsen, J. (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, **2**, 150066.

Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.

GBIF GBIF Occurrence Download..

GBIF GBIF Occurrence Download..

Gioia, P. & Hopper, S.D. (2017) A new phytogeographic map for the Southwest Australian Floristic Region after an exceptional decade of collection and discovery. *Botanical Journal of the Linnean Society*, **184**, 1–15.

Guerin, G. et a. (2018) When macroecological transitions are a fiction of sampling: Comparing herbarium records to plot-based species inventory data. *Ecography*, 1864–1875.

Hart, S.P., Usinowicz, J., & Levine, J.M. (2017) The spatial scales of species coexistence. *Nature Ecology & Evolution*, **1**, 1066–1073.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti?, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, **12**, e0169748.

596     Hijmans, R.J. (2016) raster: Geographic Data Analysis and Modeling. R package version 2.5-8..

597     Hijmans, R.J., Phillips, S., Leathwick, J., & Elith, J. (2017) dismo: Species Distribution Modeling. R package
598     version 1.1-4..

599     Hopper, S.D. (1979) Biogeographical Aspects of Speciation in the Southwest Australian Flora. *Annual Review*
600     *of Ecology and Systematics*, **10**, 399–422.

601     Hopper, S.D. & Gioia, P. (2004) The Southwest Australian Floristic Region: Evolution and Conservation of a
602     Global Hot Spot of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 623–650.

603     Kerr, J.T., Southwood, T.R., & Cihlar, J. (2001) Remotely sensed habitat diversity predicts butterfly species
604     richness and community similarity in Canada. *Proceedings of the National Academy of Sciences of the*
605     *United States of America*, **98**, 11365–11370.

606     Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the*
607     *National Academy of Sciences*, **104**, 5925–5930.

608     Laliberte, E., Zemunik, G., & Turner, B.L. (2014) Environmental filtering explains variation in plant diversity
609     along resource gradients. *Science*, **345**, 1602–1605.

610     Lambers, H., Brundrett, M.C., Raven, J.A., & Hopper, S.D. (2010) Plant mineral nutrition in ancient
611     landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional
612     strategies. *Plant and Soil*, **334**, 11–31.

613     Larsen, R., Holmern, T., Prager, S.D., Maliti, H., & Røskaft, E. (2009) Using the extended quarter degree grid
614     cell system to unify mapping and sharing of biodiversity data. *African Journal of Ecology*, **47**, 382–392.

615     Leroy, B., Meynard, C.N., Bellard, C., & Courchamp, F. (2015) Virtualspecies, an r package to generate virtual
616     species distributions. *Ecography*,.

617     Levin, L.A., Sibuet, M., Gooday, A.J., Smith, C.R., & Vanreusel, A. (2010) The roles of habitat heterogeneity
618     in generating and maintaining biodiversity on continental margins: an introduction. *Marine Ecology*, **31**, 1–5.

619     Lobo, J.M., Jay-robert, P., Lumaret, J.-p., Lobo, J.M., Jay-robert, P., & Lumaret, J.-p. (2004) Modelling the
620     Species Richness Distribution for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography*, **27**, 145–156.

621     Manning, J., Goldblatt, P., & Others (2012) *Plants of the Greater Cape Floristic Region. 1: The Core Cape*

*flora.* South African National Biodiversity Institute,

Mateo, R.G., Mokany, K., & Guisan, A. (2017) Biodiversity Models: What If Unsaturation Is the Rule? *Trends in Ecology & Evolution*, **32**, 556–566.

Molina-Venegas, R., Aparicio, A., Slingsby, J.A., Lavergne, S., & Arroyo, J. (2015) Investigating the evolutionary assembly of a Mediterranean biodiversity hotspot: Deep phylogenetic signal in the distribution of eudicots across elevational belts. *Journal of Biogeography*, **42**, 507–518.

Mouchet, M., Levers, C., Zupan, L., Kuemmerle, T., Plutzar, C., Erb, K., Lavorel, S., Thuiller, W., & Haberl, H. (2015) Testing the effectiveness of environmental variables to explain European terrestrial vertebrate species richness across biogeographical scales. *PLoS ONE*, **10**, 1–16.

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

NIMA (2000) Amendment 1. 3 January 2000. Department of Defense World Geodetic System 1984. Its Definition and Relationships with Local Geodetic Systems. 1–3.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., & Wagner, H. (2018) *Vegan: Community ecology package.*

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., & Others (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, **51**, 933–938.

Power, S.C., Verboom, G.A., Bond, W.J., & Cramer, M.D. (2017) Environmental correlates of biome-level floristic turnover in South Africa. *Journal of Biogeography*, **44**, 1745–1757.

R Core Team (2018) *R: A Language and Environment for Statistical Computing. Version 3.5.0.* R Foundation for Statistical Computing, Vienna, Austria.

Rensburg, B.J. van, Chown, S.L., & Gaston, K.J. (2002) Species Richness, Environmental Correlates, and Spatial Scale: A Test Using South African Birds. *The American Naturalist*, **159**, 566–577.

648    Ricklefs, R.E. (1987) Community diversity: relative roles of local and regional processes. *Science, New*

649    *Series*, **235**, 167–171.

650    Scott Chamberlain & Eduard Szocs (2013) Taxize - taxonomic search and retrieval in r. *F1000Research*,.

651    Thuiller, W., Midgley, G.F., Rouget, M., Cowling, R.M., F. Midgley, G., Rougeti, M., & M. Cowling, R. (2006)

652    Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, **29**, 733–744.

653    Verboom, G.A., Stock, W.D., & Cramer, M.D. (2017) Specialization to extremely low-nutrient soils limits the

654    nutritional adaptability of plant lineages. *The American Naturalist*, **In press**,.

655    Wardell-Johnson, G. & Horwitz, P. (1996) Conserving biodiversity and the recognition of heterogeneity in

656    ancient landscapes: a case study from south-western Australia. *Forest Ecology and Management*, **85**,

657    219–238.