

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

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

2019-02-27

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 hotspots, 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

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

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

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

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

33 **Acknowledgements**

34 This work was funded by the South African Department of Science and Technology (DST) and
35 the National Research Foundation (NRF) under the DST-NRF Freestanding Innovation Honours
36 Scholarship (to RVM), and by the South African Association of Botanists (SAAB) Honours
37 Scholarship (to RVM). Thanks go to the Department of Biological Sciences, University of Cape
38 Town, for providing a 2TB external hard drive for local GIS data storage. Many computations
39 were performed using facilities provided by the University of Cape Town's ICTS High
40 Performance Computing team (<hpc.uct.ac.za >).

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 studying 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 approaches, 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

71 vertebrates (Mouchet et al., 2015), South African birds (Rensburg et al., 2002), in communities
72 along marine continental margins (Levin et al., 2010), French scarab beetles (Lobo et al., 2004),
73 and for global terrestrial plants (Kreft & Jetz, 2007). The spatial scale of heterogeneity, or
74 “grain” of the environment, is also important to consider (Hart et al., 2017), as spatial scale in
75 absolute environmental conditions has also been explored (Kerr et al., 2001; Baudena et al., 2015;
76 Mouchet et al., 2015). Species co-existence and biodiversity maintenance is indeed suggested to
77 be scale-dependent (Hart et al., 2017).

78 EH is often under-represented in macro-ecological models of species richness, and has recently
79 been found to explain up to ca. 95% of biome level species richness across South Africa (Cramer
80 & Verboom, 2016). Indeed, models that include EH yield better estimates of the richness of the
81 Cape flora (Thuiller et al., 2006; Cramer & Verboom, 2016). Mediterranean-type terrestrial
82 biodiversity hotspots, such as the Cape flora included in the models by Cramer & Verboom
83 (2016), present interesting study systems in which to investigate the relationship between the
84 environment and species richness. These systems exhibit far greater species richness than
85 predicted by their areas, productivities and latitudes (Cowling et al., 1996; Kreft & Jetz, 2007).
86 There are five Mediterranean biodiversity hotspots on Earth: the California Floristic Province,
87 the Mediterranean Basin, the Chilean Winter Rainfall-Valdivian Forests, the Greater Cape
88 Floristic Region, and the Southwest Australia Floristic Region (Cowling et al., 1996; Hopper &
89 Gioia, 2004; Cook et al., 2015). These ecosystems have regular fire-cycles (Cowling et al., 1996),
90 climatic buffering, and long term stability (Kreft & Jetz, 2007), shrubby, sclerophyllous flora
91 (Hopper & Gioia, 2004). Together, they account for ca. 20% of global vascular plant species, yet
92 only ca. 5% of global land surface areas (Cowling et al., 1996). Various hypotheses have been
93 proposed to explain the high levels of plant species richness in these regions (Cook et al., 2015).
94 The species accumulation hypothesis states that the stability of these regions has allowed many
95 species to accrue. The species co-existence hypothesis states that these hotspots may facilitate
96 greater degrees of species co-existence in smaller spatial areas, due to fine-scale heterogeneity in
97 their environments. Indeed, EH has evolutionary implications too, stimulating ecological
98 speciation across sharp environmental gradients.

99 Both the Southwest Australia Floristic Region (SWA) and the Greater Cape Floristic Region
100 (Cape) are Mediterranean-type biodiversity hotspots, particularly in terms of plant species.
101 Where the Cape (with an area of ca. 189,000 km²) is known to contain about 11,400 plant species

102 (about 0.060 species per km²), SWA (area of ca. 270,000 km²) has about 3,700 species (0.014
103 species per km²) (Manning et al., 2012). So, the Cape has ca. 4.3 times as many species per km²
104 as SWA. The Cape and SWA are appropriately often compared, due to the similarities between
105 their environments (e.g. oligotrophic soils, an oceanically buffered moderate climate) and their
106 plants' ecologies (Hopper & Gioia, 2004). These two regions present unique flora out of the five
107 Mediterranean systems, with high levels of endemism (Cowling et al., 1996), and many obligate
108 fire-adapted species (Cowling et al., 1996). Similarities withstanding, SWA is topographically and
109 edaphically distinct from the Cape. The former is topographically rather uniform
110 (i.e. flat)—uniquely so among the world's five Mediterranean-climate regions (Hopper & Gioia,
111 2004)). SWA possesses a mesoscale chronosequence dune system (Laliberte et al., 2014; Cook et
112 al., 2015), while the Cape is mountainous, topographically heterogeneous, and therefore
113 associated with a large degree of spatial climatic variability, with a fine-scale mosaic of geologies
114 and soils (Cowling et al., 1996; Cramer et al., 2014; Verboom et al., 2017).

115 Both regions have sources of edaphic heterogeneity, but at different scales. This edaphic
116 variability may aid in explaining the species richness in these regions (Beard et al., 2000;
117 Verboom et al., 2017). EH of many forms will likely be important in macro-ecological models in
118 both regions, as both regions have been relatively environmentally stable over evolutionary
119 time-scales (Wardell-Johnson & Horwitz, 1996; Hopper & Gioia, 2004; Lambers et al., 2010;
120 Cramer et al., 2014; Laliberte et al., 2014; Cook et al., 2015). For the Cape, high levels of species
121 richness are thought to result from long term climatic stability, and fine grain variation in geology
122 and soils (Cramer et al., 2014). The question thus arises whether heterogeneity is a significant
123 contributor to SWA species richness. In the absence of topographic variability in SWA, it is
124 proposed that the heterogeneity of that region is due to the juxtaposition of soil types (Laliberte
125 et al., 2014; Cook et al., 2015), creating extreme edaphic variation.

126 Our hypotheses concern the Cape and SWA's environments and floras. Our main hypothesis is
127 that the Cape possesses greater abiotic heterogeneity, and at finer grain, compared to SWA, such
128 as to explain the Cape's greater species richness per unit area, and proposed greater levels of
129 species turnover between areas. We also conjecture that the heterogeneity that predicts species
130 richness in SWA will be more pronounced in terms of edaphic variables. Here we attempt to assess
131 five key predictions of this hypothesis, additionally investigating a seventh prediction to test the
132 conjectured role of edaphic heterogeneity in SWA. Dealing with the two regions' environments, we

133 assess (i) whether the Cape environment is more heterogeneous than that of SWA and (ii)
134 whether the Cape environment has more pronounced heterogeneity at finer scales than that of
135 SWA. Dealing with the distribution of species in the two regions, we assess (iii) whether the Cape
136 exhibits greater levels of species turnover between areas. Relating each regions' environment and
137 flora, we finally assess (iv) whether species richness and species turnover are adequately predicted
138 by EH in both regions and whether (v) species richness and species turnover are better predicted
139 by different forms of EH in either region (e.g. the importance of edaphic heterogeneity in SWA).

140 **2 Materials and methods**

141 **2.1 Overview**

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

147 We carried out this investigation at four principal spatial scales: $0.05^{\circ} \times 0.05^{\circ}$ squares (the finest
148 common resolution among the environmental data sources used), quarter degree squares (QDS)
149 (Larsen et al., 2009), half degree squares (HDS) (Larsen et al., 2009) and three-quarter degree
150 squares (3QDS). For the Cape, most plant occurrence records are only accurate to QDS level.
151 Thus, analyses involving species occurrence data were necessary limited to scales including and
152 above QDS.

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

155 **2.2 Environmental data sources**

156 The GCFR was treated as the area occupied by the Succulent Karoo and Fynbos biomes in the
157 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, Hopper & Gioia (2004)). 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 0.05° resolution using the “resample” function in the R package “raster” (Hijmans, 2016), with the “bilinear” method.

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

188 in the driest quarter (PDQ), using methods based on the “biovars” function in the R package
189 “dismo”.

190 **2.3 Plant occurrence data**

191 Geospatially-explicit records of vascular plant occurrences were downloaded from the Global
192 Biodiversity Information Facility (GBIF, Table 1). Queries were made for tracheophyte records
193 from within the borders of the Cape and SWA as treated here (GBIF, 24 July 2017, ???). Only
194 records with defined species and intra-specific ranks were kept. Intra-specific occurrences were
195 treated as simply being representative of their species. This resulted in FIXME unique species
196 names in the Cape, and FIXME in SWA.

197 We cleaned these data using the R package “taxise” (Scott Chamberlain & Eduard Szocs, 2013;
198 Chamberlain et al., 2018) to check that these species names had accepted-status among
199 taxonomic databases. We queried two major taxonomic databases: the Global Name Resolver
200 (GNR), and the Taxonomic Name Resolution Service (TNRS). Should either one of these services
201 return at least one match for a given name, then that name was accepted. Those names for which
202 no full binomial matches were found in either database were excluded from the final list of
203 species. The number of species names excluded totalled at FIXME and FIXME for the Cape and
204 SWA respectively. Especially for SWA, these numbers may be deemed appreciably high. But, the
205 occurrence records that would be dropped, as a consequence of these names’ removals, appeared
206 randomly distributed in geographic space in both regions. As such, any effect of the loss of these
207 records in this analysis is likely uniform within the two regions.

208 After the unaccepted names were removed, it was important to ensure that a species was not
209 listed under multiple synonyms. Such cases would skew estimates of species richness and turnover
210 in this study. In light of this, the remaining names were queried in the Tropicos and Integrated
211 Taxonomic Information System (ITIS) databases for their known synonyms, again using “taxize”.
212 These were collated to produce a nomenclatural “thesaurus” for the Cape and SWA species. This
213 consisted of a list of the accepted species names in a region, each associated with a list of known
214 synonyms. We amended species’ names in the GBIF occurrence data, in order ensure species were
215 listed under only one of these synonyms, replacing all appearances of a species’ synonyms with
216 the first synonym used in the list.

217 Lastly, We removed any species from both regions that are invasive aliens or non-indigenous.
 218 Alien species lists for plants in South Africa and Australia were acquired from the IUCN’s Global
 219 Invasive Species Database (<http://www.iucngisd.org/gisd/>).

220 The final total plant species richness in each region was FIXME and FIXME for the Cape and
 221 SWA respectively. These final collections of species occurrence records were converted to
 222 raster-layers, wherein pixel-values represented the species richness of vascular plants within that
 223 pixel. These rasters were produced at QDS, HDS, and 3QDS resolutions.

224 **2.4 Analyses**

225 **2.4.1 Quantifying environmental heterogeneity**

226 In order to assess predictions (i) and (ii), we needed to describe the EH in both regions. Using
 227 the R package “raster” (Hijmans, 2016), we used a modified version of the “roughness” index in
 228 the “terrain” function. For a three by three neighbourhood \mathbf{N} of cells, our index of roughness R is
 229 the average square-root of the squared difference between each of the n neighbour cells’ values x_i
 230 and the central focal cell’s value x_{focal} :

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

231 This value, notionally equivalent to the standard deviation of values relative to the focal value, is
 232 ascribed to the focal cell. Note, in order to use as much data from within regions’ borders as
 233 possible, roughness was computed if a focal cell had at least one neighbour cell. Using this index,
 234 we produced raster layers of each of our nine environmental variable’s heterogeneity. We
 235 compared the distributions of “roughness” values in each variable in each region with
 236 non-parametric Mann-Whitney U -tests, as almost all variables were highly non-normal, and could
 237 not be normalised by log-transformations. We also compare the effect size of the Cape vs SWA
 238 using the “common language effect size” ($CLES$), using the R package “canprot”. The $CLES$ is
 239 the proportion of all pairwise comparisons between two sample groups’ observations where one
 240 group’s value is greater than the other’s. We calculated the $CLES$ as the proportion of pairs
 241 where Cape roughness values were greater than that of SWA. This allowed us to assess prediction

(i). To compare the spatial scales of heterogeneity (prediction (ii)) between each region, we repeated this analysis at all four spatial scales. This entailed recalculating the roughness layer for each variable after the original layer (0.05 degrees resolution) had been rescaled to each of the coarser resolutions.

2.4.2 Quantifying species turnover

Regarding prediction (iii), we wished to compare the general degree of species turnover in each region. To compare the extent of species turnover between the Cape and SWA, we determined two metrics of species turnover. The first, computes the mean species turnover as Jaccard distances (Oksanen et al., 2018) between each pair of QDS within each HDS (\bar{J}_{QDS} , based on HDS with $2 \leq n \leq 4$ QDS) in both regions. The second is defined in terms of Whittaker's additive definition of β -diversity, as follows:

$$\gamma = \alpha + \beta \quad (2)$$

Here, we treat species richness at the HDS-scale (S_{HDS}) as γ -diversity and at the QDS-scale (\bar{S}_{QDS}) as α -diversity. Intuitively, the species richness of an area is the result of some combination of the richness of sites within that area and the difference in species complements between those sites. Thus, we partition γ -diversity as in Equation (2), such that β -diversity is the difference between γ - and α -diversity. We compare the distributions of \bar{J}_{QDS} and T_{HDS} using non-parametric Mann-Whitney U -tests, in order to guard against non-normality.

2.4.3 Predicting richness and turnover with environmental heterogeneity

Regarding prediction (iii), we wished to compare the general degree of species turnover in each region. For (iv) and (v) we modelled species richness (S) and turnover as a function of various combinations of environmental and environmental heterogeneity variables in both regions using boosted regression-tree (BRT) modelling techniques. This allowed us to explore which axes of environmental heterogeneity are most influential on vascular plant species richness and turnover, and the differences in the importance of such axes between the Cape and SWA.

266 BRTs are a flexible machine learning-based model of response variables and do so without
 267 involving normal null-hypothesis significance testing (Elith et al., 2008), and have been employed
 268 previously to model species richness (Thuiller et al., 2006; see Mouchet et al., 2015; Cramer &
 269 Verboom, 2016) as macro-ecological models. BRTs are developed through the iterative generation
 270 of non-linear regression trees. BRTs are an ensemble-approach, in which a prediction \hat{y}_i is based
 271 on the weighted sum of the predictions of progressively “less important” regression trees (t_k), as
 272 opposed to the predictions of one tree (Elith et al., 2008). For $k \rightarrow nt$ number of trees, where
 273 each tree is itself a function of the matrix \mathbf{X} of j predictor variables ($t_k = f(x_{ij})$), a BRT-model
 274 can be represented as follows:

$$\hat{y}_i = \sum_{k=1}^{nt} w_k t_k \quad (3)$$

275 BRTs have two major meta-parameters over which users have control (Friedman, 2001): the
 276 learning rate (lr , the rate at which iterative trees reduce predictive deviance during
 277 model-training, controlling the contribution of each tree to the final model) and tree complexity
 278 (tc , the number of nodes on a given regression-tree, i.e. the maximum interaction depth the model
 279 is permitted to fit).

280 BRTs were implemented here to predict both vascular plant species richness and turnover in each
 281 HDS, as a function of environmental variables and environmental roughness values in those cells,
 282 as Gaussian responses, thus resulting in two BRT-models for each region. We treated richness as
 283 S_{HDS} and turnover as \bar{J}_{QDS} . The natural logarithm of species richness was used, in order to
 284 satisfy the assumptions of a Gaussian response. Note, this is not strictly because BRTs have any
 285 parametric assumptions concerning the distribution of the response variable, but rather to aid in
 286 applying the Gaussian-family of BRT algorithms to the richness data available. Additionally,
 287 BRTs were implemented to predict vascular plant species richness at the QDS-scale (S_{QDS}), thus
 288 resulting in a total of six BRT-models presented here.

289 As recommended by Elith et al. (2008), BRT models were trained on a set of non-collinear
 290 predictor variables using “gbm.step” in “dismo” (Hijmans et al., 2017) and “gbm”. Collinear
 291 predictor variables can skew the interpretation of results, as the relative influence of mutually
 292 collinear variables is reduced. Collinearity among the nine environmental predictor variables and

their respective nine roughness-equivalents was assessed using “removeCollinearity” in the R package “virtualspecies” (Leroy et al., 2015) separately for each region, such that variables were no more than 80% collinear (Pearson’s $r \geq 0.80$). When faced with a cluster of collinear variables, one variable was chosen manually therefrom. Where possible, the roughness-equivalent of an environmental variable was included if its absolute-equivalent could also be included. When interpreting the results of BRTs, it is important to consider the effects of the variables included as representative of the effect of the excluded variables with which it was found to be collinear.

In order to select ideal lr and tc all models (described below) were trained on the final non-collinear predictor 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, 5×10^{-4} , 1×10^{-4}). The function “gbm.step” optimises the number of trees (nt) using cross-validation during model training (Elith et al., 2008) by halting iteration when predictions begin to overfit. For all models, we used 10 cross-validation folds (i.e. use 10 different randomly selected training data sets), a tolerance-threshold of 0.001, a bagging-fraction of 0.75 (proportion of training data randomly chosen to generate each tree), and trained models starting with 50 trees, with each iterative step adding 50 trees at a time, up to a maximum of 10,000 trees. Following this iterative parameter optimisation, Gaussian BRT models were constructed with $tc = 3$ and $lr = 0.001$, along with the other settings described.

The optimum configuration of lr and tc for the final model is a trade-off between model fit (e.g. pseudo- R^2 ; Equation (4)) and complexity (nt). A tc of 5 was chosen for the final model. This follows the recommendations of Elith et al. (2008), where lr and tc are advised to be adjusted inversely. This was chosen in order to account for the complex interactions determining species richness. To avoid overfitting, an 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_{pseudo} , which is the proportion of null deviance D_{null} explained by some model i . Formally, it is defined as follows:

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

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

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

330 In order to assess the reliability of the conclusions drawn from these models, we randomly
331 permuted the response data (S_{QDS} , S_{HDS} and \bar{J}_{QDS}) with respect to the environmental and
332 heterogeneity data, and reran all six BRT-models 999 times (with the final non-collinear predictor
333 sets and preconfigurations above). This also allows us to remove any effect of spatial
334 autocorrelation in generating the observed correlations between patterns of species occurrence and
335 environment (Cramer & Verboom, 2016), and to allow us to assess the significance of our results
336 relative to a random null. Notably, as the predictor variables themselves are likely spatially
337 autocorrelated, correlation structure in model residuals is accounted for by the correlation
338 structure in the environmental data. Nonetheless, we wished to demonstrate our results more
339 robustly and thus carried out these permutation tests. For all six models, the majority of the 999
340 permuted models failed to find associations between the response and predictor variables. The
341 results of those that succeeded to fit a model to completion (usually ca. 200 out of 999) are
342 presented. The replicate and permuted BRT-models were compared using various measures of
343 model performance (above; nt , R_{pseudo}^2 (Equation (4)), R_{E-O}^2) and the ranks of these values for
344 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 \bar{J}_{QDS} and T_{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). \bar{J}_{QDS} measures the average pairwise Jaccard distance between QDS-cells in each HDS-cell. T_{HDS} , however, represents the inferred β component of γ -diversity. As γ -diversity ($= S_{\text{HDS}}$) in the Cape has a greater β -diversity component ($= T_{\text{HDS}}$) than SWA, the complement is necessarily true: γ -diversity in the Cape has a lesser α -diversity component ($= \bar{S}_{\text{QDS}}$) 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. Additionally, 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 ($= \bar{S}_{\text{QDS}}$), and to a lesser extent at HDS-scales ($= S_{\text{HDS}}$), 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, topographic 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 Cape (concerning clusters of collinear variables relevant to those retained during BRT-model fitting), MAP was included in the BRT-analyses as representative of a cluster of collinear variables consisting of itself, NDVI, surface T and soil C at the. Roughness in soil clay content represented itself, roughness in soil pH and roughness in NDVI. In SWA, MAP was selected as representative of itself, NDVI and soil C.

Our BRT-models of species richness in both regions rank environmental variables somewhat differently at QDS- and HDS-scales (Figure 3a,b,d,e). These differences in rankings are similar to those between two unrelated lists ($P_{1-2} > 0.05$, Figure 5). This suggests some scale-dependence of different environmental variables' associations with species richness.

It is noteworthy that BRT-models of species turnover ($= \bar{J}_{\text{QDS}}$, at HDS-scales) (Figure 3c,f) rank variables similarly to models of richness at HDS-scales ($P_{2-3} \leq 0.005$, Figure 5). This is likely due to the fact that proportional floristic turnover covaries with species richness. As such, though the signs of relationships determining turnover may differ from those determining richness, the importances of different variables would be similar.

In addition to different variables being more strongly associated with species richness and turnover in the Cape compared to SWA (Figure 3), the forms of those relationships vary (Figure 6). We found MAP, and roughness therein, to relate positively with species richness in both regions at both scales (Figure 6a,b,d,e). As MAP is collinear with NDVI and soil C in both regions (and surface T in the Cape), this can be interpreted as the signal of a biological productivity and resource availability associating with high levels of species richness.

The positive association of heterogeneity variables in the Cape as opposed to SWA (Figure 6a,b

429 vs d,e) concurs with their greater importance in BRT-model predictions (Figure 3).

430 The fact that species turnover ($T_{HDS} = \bar{J}_{QDS}$) in the Cape and SWA is largely predicted by the
431 same variables as species richness, but with opposite signs to its relationships (Figure 6c,f), is
432 indicative of the richness-dependence of the measure of floristic turnover used here (Jaccard
433 distances) to quantify turnover at the HDS-scale.

434 4 Discussion

435 Here we have provided support for the hypothesis that the difference in plant species richness
436 between the Cape and SWA is accounted for by the greater abiotic heterogeneity in the Cape. As
437 expected, the Cape is shown to possess (i) a quantifiably more heterogeneous environment that is
438 (ii) generally heterogeneous at a finer spatial scale than SWA, with (iii) greater levels of floristic
439 turnover. We have shown that vascular plant species richness (iv) can be explained in terms of
440 environmental conditions including environmental heterogeneity in both the Cape and SWA.
441 Also, (v) the sets of environmental axes that explain plant species richness differ between the
442 Cape and SWA. These findings contribute towards an understanding of the ecological conditions
443 associated with high levels of species co-existence in these two regions, and strengthens the
444 generality of a positive relationship between EH and biodiversity.

445 Cook et al. (2015) summarised the suites of hypotheses for the high levels of plant diversity in
446 mediterranean-type ecosystems. Broadly, these biodiversity hotspots are thought to arise due to
447 greater levels of evolutionary diversification, ecological co-existence and time for species
448 accumulation by either immigration or in situ evolution. Our study concerns the dual role of EH
449 in facilitating ecological speciation along environmental gradients and a diversity of habitats in
450 wich greater numbers of species may co-exist. We have correlative evidence for these processes, in
451 that EH was positively associated with high levels of species richness and turnover in both
452 regions. Further research should be aimed at distinguishing the roles of ecological speciation and
453 species co-existence associated with EH in mediterranean-type ecosystems. The role of
454 environmental stability is relevant here (Cook et al., 2015, Pinto-Ledezma et al., 2018), both by
455 allowing more continuous species accumulation when habitats persists through geologic time and
456 by allowing environmental gradients to be stable long enough for ecological speciation to occur

457 (Pinto-Ledezma et al., 2018). In the contexts of the Cape and SWA, this long term environmental
458 stability, and the stability of EH through time, is indeed expected to contribute to these regions
459 biodiversity (Hopper, 1979; Cowling et al., 1996).

460 *Ed: I'd like to include more information on the landscape/geologic history of the regions here, and*
461 *link this to the notion of environmental stability and "species accumulation", sensu Cook et al.*
462 *(2015) and Pinto-Ledezma et al. (2018).*

463 The Cape and SWA regions present differentiable environmental spaces, each with varying
464 degrees of heterogeneity across spatial scales and environmental axes. The clear distinction of the
465 regions' topographic features is as expected. The Cape region has been found previously to have
466 the second highest median topographic heterogeneity of the five Mediterranean-climate regions
467 (Bradshaw & Cowling, 2014). In the models developed by Cramer & Verboom (2016) for South
468 Africa, roughness in topography was largely superseded as an important predictor of species
469 richness by other roughness variables. This is the case with our results too, likely as topographic
470 complexity can be considered a proxy for gradients in other environmental axes, such as rainfall
471 or temperature. In line with our hypotheses, a combination of absolute and heterogeneity
472 variables are associated with species richness and turnover. These effects are region-specific,
473 particularly in that patterns in the Cape are more strongly determined by edaphic conditions and
474 EH while SWA patterns are more strongly determined by absolute environmental conditions,
475 particularly climate. Kreft & Jetz (2007) modelled global terrestrial vascular plant species
476 richness, which focussed primarily on using absolute environmental variables, underestimated the
477 richness of the Cape flora. Though Kreft & Jetz (2007) did include topographic heterogeneity in
478 their predictor set, topography is often a proxy for more biologically meaningful variables
479 (Cramer & Verboom, 2016). This explains why the inclusion of these variables (e.g. roughness in
480 mean annual precipitation) yields more accurate predictions of species richness. Indeed, Thuiller
481 et al. (2006) also included topographic heterogeneity.

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

486 Comparing our models to those fit by Cramer & Verboom (2016), we did not directly find spatial

487 heterogeneity in NDVI to be an important determinant of species richness and turnover patterns
 488 in the Cape. Rather absolute MAP, which is collinear with absolute NDVI, was found to be
 489 important. This speaks to the relationship between vegetation productivity and climate, and how
 490 suitably productive habitats can foster more species . An additional interpretation of spatial
 491 heterogeneity (and indeed absolute conditions) in NDVI is that of changing canopy light-contexts
 492 (Cramer & Verboom, 2016). Power et al. (2017) found that floristic turnover between biomes in
 493 South Africa to be associated with differences in biomes' leaf area indices. We found NDVI to be
 494 more heterogeneous across the Cape than SWA (Figures 1). NDVI is an integrating variable,
 495 which captures information about productivity, light availability, and soil nutrients (Power et al.,
 496 2017). Absolute NDVI and its correlates can be interpreted as contributing to predicting species
 497 richness in the Cape. This demonstrates the role of ecological productivity in facilitating the
 498 co-existence diverse species assemblages. EH, then, must be considered alongside resource- and
 499 energy-availability axes when interpreting the drivers of species richness. In so much as a diverse
 500 environmental space supports more species, the materials and productivity required for biota to
 501 thrive are also needed to support species (Gaston, 2000; Kerr et al., 2001; Bøhn & Amundsen,
 502 2004; Kreft & Jetz, 2007). As such, our findings, along with those of previous studies (Rensburg
 503 et al., 2002; Thuiller et al., 2006; Kreft & Jetz, 2007; Cramer & Verboom, 2016), suggest that
 504 there are ecological and evolutionary consequence to both resource availability and EH.

505 We detected a strong set of links between various edaphic properties and vascular plant species
 506 richness and turnover in the Cape. The importance of absolute pH speaks to the association of
 507 many Cape communities with nutrient impoverished, acidic soils . We did not, however, find
 508 edaphic conditions to be as important as we expected in SWA. This may be due to a real lack of
 509 importance of edaphic features in driving assemblage patterns in SWA. Alternatively, there is a
 510 risk that the quality of soil data provided to our models, from SoilGrids250m , is insufficient for
 511 our purposes here. There are two potential reasons for this. Firstly, the SoilGrids250m data
 512 product is acknowledged to be less accurate at the scales we investigate here , such that the true
 513 variation in soil properties in SWA is not represented in our dataset. Secondly, it could be that
 514 the link between soils and assemblage patterns in SWA exists, but simply at a finer scale than
 515 that at which we performed our analyses here. These two reasons are not mutually exclusive, in
 516 that SoilGrids250m may be both inaccurate at the scales we consider here and that our analysis
 517 considers scales greater than those important in SWA plant community assemblage patterns. This

518 is particularly relevant to our study, as the Cape has a much wider range of scales exhibited in
519 the heterogeneity across its environmental axes. Notably, each region has finer scale heterogeneity
520 in some variables, and coarser scale in others—neither region is more fine or coarse than the other
521 over all environmental axes.

522 *Ed: I'd obviously like to include more information on the landscape/geologic history of the regions*
523 *here too, and with reference to particular soil variables (Mike?).*

524 Another issue related to spatial scales, as with any ecological study, is that of spatially uniform
525 and representative plant species occurrence data. Efforts were made herein to minimise differences
526 in data-quality between SWA and Cape by using remote-sensing derived environmental data.
527 Species occurrence data, however, prove difficult to obtain with uniform methodology. South
528 African biodiversity data is stored only to QDS-level accuracy. As such, our analyses were limited
529 to that as the finest spatial scale, such that our environmental datasets necessarily, then, were
530 also limited to the QDS-scale. This may complicate the interpretation of our analyses, as these
531 data may be too coarse in scale to capture patterns of species richness and turnover and their
532 associations with the environment in reality in empirically heterogeneous environments like the
533 Cape and SWA.

534 In addition to the scale of our data, spatial bias in species occurrence data must also be
535 considered. GBIF data are sourced from both herbaria and plot-monitoring datasets. The relative
536 contribution of each of these to GBIF data may vary between countries and regions—i.e. between
537 the Cape and SWA. As recently noted by Guerin (2018), herbarium records and plot-based
538 inventories describe regional flora with varying accuracy, especially with respect to documenting
539 high levels of floristic turnover. We emphasise here our concerns that there the floral occurrence
540 data and soils information use in this study are especially at risk of spatial bias and/or
541 inaccuracies.

542 ...

543 Jaccard distances, as used here as \bar{J}_{QDS} , are highly richness dependent. Although they represent
544 the proportional floristic turnover between cells, this proportion itself is sensitive to the richness
545 of sites under comparison. For example, the turnover between two low richness sites has is likely
546 to be greater by chance, due to the absence or presence of a few species, than the turnover

547 between two high richness sites, where the absence or presence of a few species does not greatly
548 affect the turnover calculated.

549 ...

550 Our findings here are correlative. There are, however, many proposed mechanisms to explain the
551 correlative signals demonstrated here. Our findings support the hypothesis that mediterranean
552 systems' plant species richness is a function of spatial variability in environmental conditions.
553 This can stimulate diversification, and maintain that diversity by providing a range of habitats for
554 species co-existence. Oligotrophic soils can stimulate an increase in functional diversity, through
555 the evolution of diverse nutrient acquisition strategies (Lambers et al., 2010; Verboom et al.,
556 2017)—e.g. sclerophylly (Cramer et al., 2014; Cook et al., 2015). An aspect of the environment I
557 have neglected to consider is fire, shown to also contribute to predictions here in the Cape
558 (Cramer & Verboom, 2016). Cardillo (2012) have shown the structuring forces behind species
559 co-occurrence patterns, and thus likely species richness, differ between species-pairs with different
560 post-fire responses and those with similar post-fire responses.

561 Following from the understanding that functionally diverse assemblages, which are more likely to
562 be more species rich, are likely to arise and/or occur in areas with diverse ecological pressures
563 (Molina-Venegas et al., 2015), one would expect, then, heterogeneous habitats such as those in
564 Mediterranean-type biodiversity hotspots to exhibit high levels functional beta diversity along
565 steep environmental gradients (Molina-Venegas et al., 2015). If the niches concerning these
566 functions are phylogenetically conserved among those biota, then one would also expect high
567 levels of species and phylogenetic beta diversity along these gradients (Molina-Venegas et al.,
568 2015). This concurs with the notion put forward by Power et al. (2017), wherein megadiverse
569 systems such as these represent the results of “phylogenetic niche conservatism on a
570 heterogeneous landscape”. Thus, species and phylogenetic turnover should covary with
571 environmental heterogeneity in some way. Indeed, endemism, at certain scales, could also follow
572 this pattern. Thuiller et al. (2006) demonstrated that there are phylogenetic and biome-related
573 determinants of species richness. This makes sense in light of the low probability of lineages
574 crossing biome boundaries in Mediterranean systems (Power et al., 2017). NDVI and light
575 availability, and the heterogeneity therein, are associated with high levels of floristic turnover
576 (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.

Table captions

Captions are also repeated alongside their respective tables for readability.

Table 1: Georeferenced vascular plant species occurrence 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.

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

Table 3: Estimated differences between replicate Cape and SWA BRT-models' performance measures (nt , R_{pseudo}^2 (Equation (4)), $R_{\text{E-O}}^2$ (see text)) following two-sided t -tests. Positive values

604 indicate that the Cape models had greater values. In all cases, the Cape and SWA had highly
605 significantly different values for these quality measures ($P < 0.0001$).

606 **Figure captions**

607 Captions are also repeated alongside their respective figures for readability.

608 Figure 1: Comparisons of different types of environmental heterogeneity between the the Greater
609 Cape Floristic Region (Cape) and the Southwest Australia Floristic Region (SWA). We present
610 (a) distributions of roughness values (Equation (1)) for example variables from each broad
611 category of the environment concerned. (b) The common language effect size (*CLES*; see text) of
612 Cape versus SWA roughness values is shown for all variables, grouped by broad categories of the
613 environment, describing differences in the distributions of Cape and SWA roughness values. We
614 used Mann-Whitney *U*-tests to assess differences in these distributions. Non-significant differences
615 ($P_U > 0.05$) are denoted as such (“NS”). Note, *U*-tests were performed using only a random set of
616 5000 cells at the 0.05-degree-scale, as the *U*-test as implemented in R cannot handle more than
617 that many values to compare.

618 Figure 2: Species turnover, described in two forms ((a) mean Jaccard distance between QDS in
619 each HDS (\bar{J}_{QDS}), (b) additively defined turnover (T_{HDS} , Equation (2)) as a proportion of HDS
620 richness (S_{HDS})), compared between the Cape and SWA. Mann-Whitney *U*-tests between the
621 Cape and SWA distributions of \bar{J}_{QDS} and T_{HDS} yielded significant differences (see *P*-values and
622 common language effect sizes (*CLES*) inset).

623 Figure 3: Relative influence of environmental variables (including heterogeneity
624 variables—prefixed with “R”) in boosted regression tree (BRT) model predictions in the Greater
625 Cape Floristic Region (Cape, a–c) and the Southwest Australia Floristic Region (SWA, b–d) of
626 vascular plant species richness at the (b,e) QDS-scale ($= \bar{S}_{\text{QDS}}$), (a,d) HDS-scale ($= S_{\text{HDS}}$) and
627 (c,f) turnover ($= \bar{J}_{\text{QDS}}$). All BRT-models were permitted to fit three-way interactions between
628 environmental variables. Points denote the mean contribution of an environmental variable to
629 model-predictions across the 1000 replicate BRT-models for that region/scope. Horizontal ticks
630 denote the mean for the 999 permuted BRT-models. Standard deviations above and below these
631 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 ($F_{\text{rep.}}$) and permuted ($F_{\text{prm.}}$) 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_{pseudo} (Equation (4)), (c) $R^2_{\text{E-O}}$ (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 Figures 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 ($= \bar{S}_{\text{QDS}}$ and S_{HDS} respectively; rows nos. 1 and 2) and between HDS-scale richness and turnover ($= \bar{J}_{\text{QDS}}$) (rows nos. 2 and 3). Statistics (Δ - and P -values) inset at the top and bottom of each panel refer to these comparisons respectively. Δ -values represent the average absolute difference in ranks across variables between two models' rankings. The associate P -value results from ranking the observed

663 Δ -values against 999 Δ -values based on random permutations of variables' rankings (SI1), such
664 that more significant P -values denote rankings more similar than would be expected by chance.

665 Figure 6: Marginal effects of environmental conditions and heterogeneity on vascular plant species
666 richness at the QDS-scale ($= \bar{S}_{\text{QDS}}$; a,d), HDS-scale ($= S_{\text{HDS}}$; b,e) and turnover ($= \bar{J}_{\text{QDS}}$; c,f) in
667 response variables in the Greater Cape Floristic Region (Cape; a–c) and Southwest Australia
668 Floristic Region (SWA; d–f) following boosted regression tree (BRT) modelling. Marginal effect
669 functions presented are derived from a representative BRT-model from the set of replicate
670 BRT-models (for each of the six modelling cases) (see SI regarding how representative
671 BRT-models were selected). Marginal effects represent the effect of a predictor variable when all
672 other predictors are set at their means. Marginal effect functions are shown for environmental
673 variables that contributed $\geq 10\%$ to a model's predictions. Functions are coloured as keyed, with
674 solid lines representing absolute environmental variables and dotted representing heterogeneity
675 variables ("rough"). Environmental variables were all rescaled here such as to be centred on zero
676 (i.e. Z -transformed), facilitating comparison of functions' forms.

677 References

- 678 Baudena, M., Sánchez, A., Georg, C.-P., Ruiz-Benito, P., Rodríguez, M.Á., Zavala, M.A., & Rietkerk, M. (2015)
679 Revealing patterns of local species richness along environmental gradients with a novel network tool.
680 *Scientific Reports*, **5**, 11561.
- 681 Beard, J.S., Chapman, A.R., & Gioia, P. (2000) Species richness and endemism in the Western Australian flora.
682 *Journal of Biogeography*, **27**, 1257–1268.
- 683 Bivand, R., Keitt, T., & Rowlingson, B. (2017) rgdal: Bindings for the Geospatial Data Abstraction Library. R
684 package version 1.2-7..
- 685 Bradshaw, P.L. & Cowling, R.M. (2014) Landscapes, rock types, and climate of the Greater Cape Floristic Region.
686 *Fynbos: Ecology, evolution and conservation of a megadiverse region* (ed. by N. Allsopp, J.F. Colville, and
687 G.A. Verboom), pp. 26–46. Oxford University Press, Oxford.
- 688 Bøhn, T. & Amundsen, P.-A. (2004) Ecological Interactions and Evolution: Forgotten Parts of Biodiversity?
689 *BioScience*, **54**, 804.
- 690 Cardillo, M. (2012) The phylogenetic signal of species co-occurrence in high-diversity shrublands: different patterns
691 for fire-killed and fire-resistant species. *BMC Ecology*, **12**, 21.

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

695 Cook, L.G., Hardy, N.B., & Crisp, M.D. (2015) Three explanations for biodiversity hotspots: small range size,
696 geographical overlap and time for species accumulation. An Australian case study. *New Phytologist*, **207**,
697 390–400.

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

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

702 Cramer, M.D., West, A.G., Power, S.C., Skelton, R., & Stock, W.D. (2014) Plant ecophysiological diversity. *Fynbos:*
703 *Ecology, evolution and conservation of a megadiverse region* pp. 248–272. Oxford University Press, Oxford.

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

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

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

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

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

716 GBIF (24 July 2017) GBIF Occurrence Download..

717 GBIF (24 July 2017) GBIF Occurrence Download..

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

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

- 722 Hart, S.P., Usinowicz, J., & Levine, J.M. (2017) The spatial scales of species coexistence. *Nature Ecology &*
723 *Evolution*, **1**, 1066–1073.
- 724 Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti?, A., Shangguan,
725 W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes,
726 N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017) SoilGrids250m: Global
727 gridded soil information based on machine learning. *PLoS ONE*, **12**, e0169748.
- 728 Hijmans, R.J. (2016) raster: Geographic Data Analysis and Modeling. R package version 2.5-8..
- 729 Hijmans, R.J., Phillips, S., Leathwick, J., & Elith, J. (2017) dismo: Species Distribution Modeling. R package
730 version 1.1-4..
- 731 Hopper, S.D. (1979) Biogeographical Aspects of Speciation in the Southwest Australian Flora. *Annual Review of*
732 *Ecology and Systematics*, **10**, 399–422.
- 733 Hopper, S.D. & Gioia, P. (2004) The Southwest Australian Floristic Region: Evolution and Conservation of a
734 Global Hot Spot of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 623–650.
- 735 Kerr, J.T., Southwood, T.R., & Cihlar, J. (2001) Remotely sensed habitat diversity predicts butterfly species
736 richness and community similarity in Canada. *Proceedings of the National Academy of Sciences of the*
737 *United States of America*, **98**, 11365–11370.
- 738 Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the*
739 *National Academy of Sciences*, **104**, 5925–5930.
- 740 Laliberte, E., Zemunik, G., & Turner, B.L. (2014) Environmental filtering explains variation in plant diversity along
741 resource gradients. *Science*, **345**, 1602–1605.
- 742 Lambers, H., Brundrett, M.C., Raven, J.A., & Hopper, S.D. (2010) Plant mineral nutrition in ancient landscapes:
743 high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. *Plant*
744 *and Soil*, **334**, 11–31.
- 745 Larsen, R., Holmern, T., Prager, S.D., Maliti, H., & Røskft, E. (2009) Using the extended quarter degree grid cell
746 system to unify mapping and sharing of biodiversity data. *African Journal of Ecology*, **47**, 382–392.
- 747 Leroy, B., Meynard, C.N., Bellard, C., & Courchamp, F. (2015) Virtualspecies, an r package to generate virtual
748 species distributions. *Ecography*,.
- 749 Levin, L.A., Sibuet, M., Gooday, A.J., Smith, C.R., & Vanreusel, A. (2010) The roles of habitat heterogeneity in
750 generating and maintaining biodiversity on continental margins: an introduction. *Marine Ecology*, **31**, 1–5.
- 751 Lobo, J.M., Jay-robert, P., Lumaret, J.-p., Lobo, J.M., Jay-robert, P., & Lumaret, J.-p. (2004) Modelling the
752 Species Richness Distribution for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography*, **27**, 145–156.

- 753 Manning, J., Goldblatt, P., & Others (2012) *Plants of the Greater Cape Floristic Region. 1: The Core Cape flora.*
 754 South African National Biodiversity Institute,
- 755 Mateo, R.G., Mokany, K., & Guisan, A. (2017) Biodiversity Models: What If Unsaturation Is the Rule? *Trends in*
 756 *Ecology & Evolution*, **32**, 556–566.
- 757 Molina-Venegas, R., Aparicio, A., Slingsby, J.A., Lavergne, S., & Arroyo, J. (2015) Investigating the evolutionary
 758 assembly of a Mediterranean biodiversity hotspot: Deep phylogenetic signal in the distribution of eudicots
 759 across elevational belts. *Journal of Biogeography*, **42**, 507–518.
- 760 Mouchet, M., Levers, C., Zupan, L., Kuemmerle, T., Plutzar, C., Erb, K., Lavorel, S., Thuiller, W., & Haberl, H.
 761 (2015) Testing the effectiveness of environmental variables to explain European terrestrial vertebrate species
 762 richness across biogeographical scales. *PLoS ONE*, **10**, 1–16.
- 763 Mucina, L. & Rutherford, M.C. (2006) *The vegetation of South Africa, Lesotho and Swaziland.* South African
 764 National Biodiversity Institute,
- 765 NIMA (2000) Amendment 1. 3 January 2000. Department of Defense World Geodetic System 1984. Its Definition
 766 and Relationships with Local Geodetic Systems. 1–3.
- 767 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O’Hara, R.B.,
 768 Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., & Wagner, H. (2018) *Vegan: Community ecology*
 769 *package.*
- 770 Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D’amico,
 771 J.A., Itoua, I., Strand, H.E., Morrison, J.C., & Others (2001) Terrestrial Ecoregions of the World: A New
 772 Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving
 773 biodiversity. *BioScience*, **51**, 933–938.
- 774 Power, S.C., Verboom, G.A., Bond, W.J., & Cramer, M.D. (2017) Environmental correlates of biome-level floristic
 775 turnover in South Africa. *Journal of Biogeography*, **44**, 1745–1757.
- 776 R Core Team (2018) *R: A Language and Environment for Statistical Computing. Version 3.5.0.* R Foundation for
 777 Statistical Computing, Vienna, Austria.
- 778 Rensburg, B.J. van, Chown, S.L., & Gaston, K.J. (2002) Species Richness, Environmental Correlates, and Spatial
 779 Scale: A Test Using South African Birds. *The American Naturalist*, **159**, 566–577.
- 780 Ricklefs, R.E. (1987) Community diversity: relative roles of local and regional processes. *Science, New Series*, **235**,
 781 167–171.
- 782 Scott Chamberlain & Eduard Szocs (2013) Taxize - taxonomic search and retrieval in r. *F1000Research*,.
- 783 Thuiller, W., Midgley, G.F., Rouget, M., Cowling, R.M., F. Midgley, G., Rougeti, M., & M. Cowling, R. (2006)

- 784 Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, **29**, 733–744.
- 785 Verboom, G.A., Stock, W.D., & Cramer, M.D. (2017) Specialization to extremely low-nutrient soils limits the
786 nutritional adaptability of plant lineages. *The American Naturalist*, **In press**,.
- 787 Wardell-Johnson, G. & Horwitz, P. (1996) Conserving biodiversity and the recognition of heterogeneity in ancient
788 landscapes: a case study from south-western Australia. *Forest Ecology and Management*, **85**, 219–238.

789 **Biosketches**

790 **Ruan van Mazijk**

791 **Michael D. Cramer**

792 **G. Anthony Verboom**

793 **Author contributions**

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

Table 1: Georeferenced vascular plant species occurrence 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	??, ??
Elevation	SRTM v2.0		??
NDVI	MODIS (MOD13C2)	Feb. 2000 to Apr. 2017	??
Climatic variables			
Surface temperature	MODIS (MOD11C3)	Feb. 2000 to Apr. 2017	??
MAP	CHIRPS v2.0	Jan. 1981 to Feb. 2017	??
PDQ	CHIRPS v2.0	Jan. 1981 to Feb. 2017	??
Soil variables			
CEC	SoilGrids250m (CECSOL M 250m)		??
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_{pseudo}^2 (Equation (4)), R_{E-O}^2 (see text)) of the 1000 replicate BRT-models relative to 999 BRT-models fit to permuted datasets. Each of the 1000 replicate BRT-models was ranked against the 999 permuted BRT-models. The average rank of each, as a proportion, is presented.

Model	nt	R_{pseudo}^2	R_{E-O}^2
QDS-richness			
GCFR	1.000	1.000	1.000
SWAFR	1.000	1.000	1.000
HDS-richness			
GCFR	0.987	1.000	0.988
SWAFR	1.000	1.000	1.000
HDS-turnover			
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_{pseudo}^2 (Equation (4)), R_{E-O}^2 (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_{pseudo}^2	R_{E-O}^2
QDS-richness	542.938	0.063	-0.005
HDS-richness	-808.994	-0.064	-0.233
HDS-turnover	-997.045	-0.052	-0.296

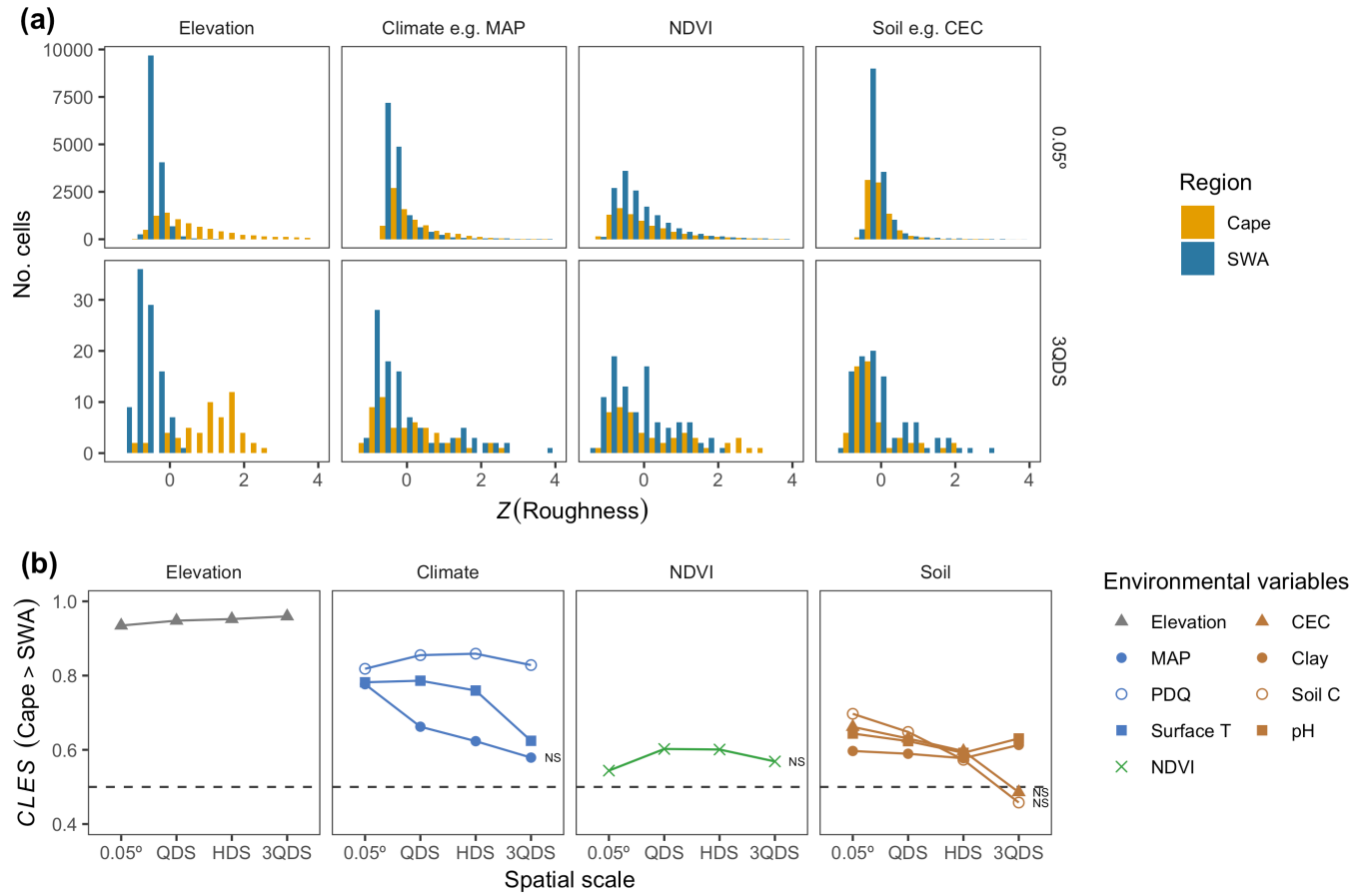


Figure 1: Comparisons of different types of environmental heterogeneity between the the Greater Cape Floristic Region (Cape) and the Southwest Australia Floristic Region (SWA). We present (a) distributions of roughness values (Equation (1)) for example variables from each broad category of the environment 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.

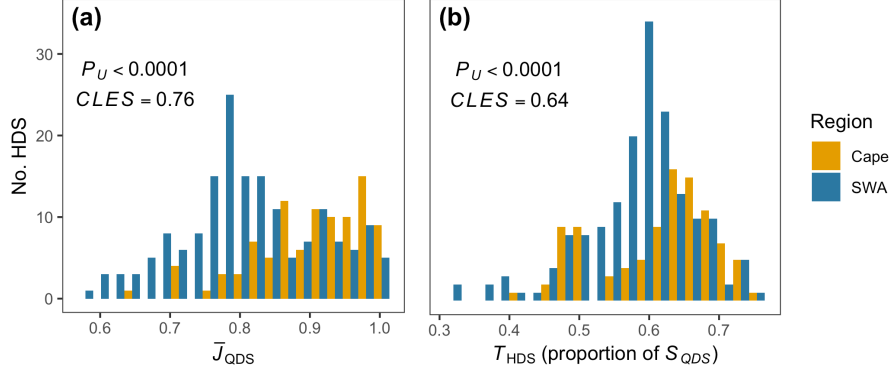


Figure 2: Species turnover, described in two forms ((a) mean Jaccard distance between QDS in each HDS (\bar{J}_{QDS}), (b) additively defined turnover (T_{HDS} , Equation (2)) as a proportion of HDS richness (S_{HDS})), compared between the Cape and SWA. Mann-Whitney U -tests between the Cape and SWA distributions of \bar{J}_{QDS} and T_{HDS} yielded significant differences (see P -values and common language effect sizes ($CLES$) inset).

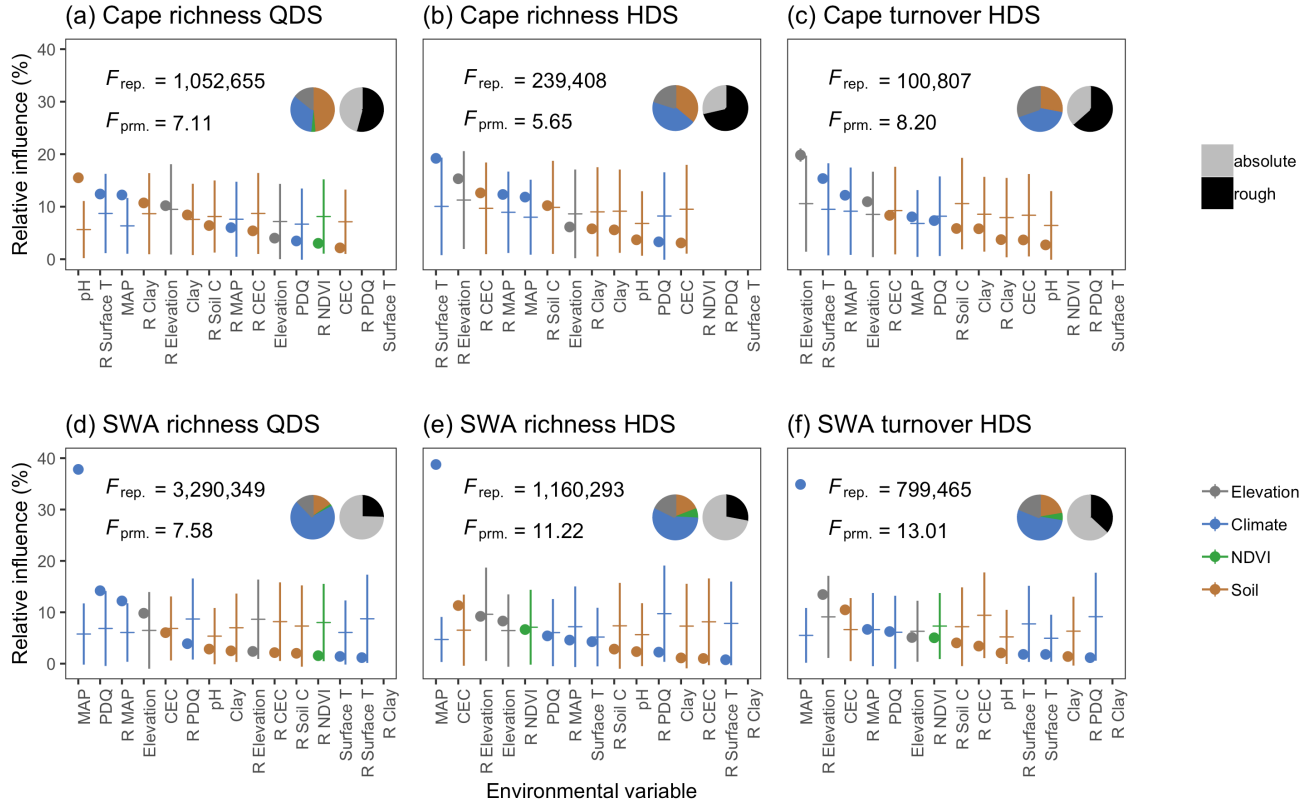


Figure 3: Relative influence of environmental variables (including heterogeneity variables—prefixed with “R”) in boosted regression tree (BRT) model predictions in the Greater Cape Floristic Region (Cape, a–c) and the Southwest Australia Floristic Region (SWA, b–d) of vascular plant species richness at the (b,e) QDS-scale ($= \bar{S}_{QDS}$), (a,d) HDS-scale ($= S_{HDS}$) and (c,f) turnover ($= \bar{J}_{QDS}$). 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.

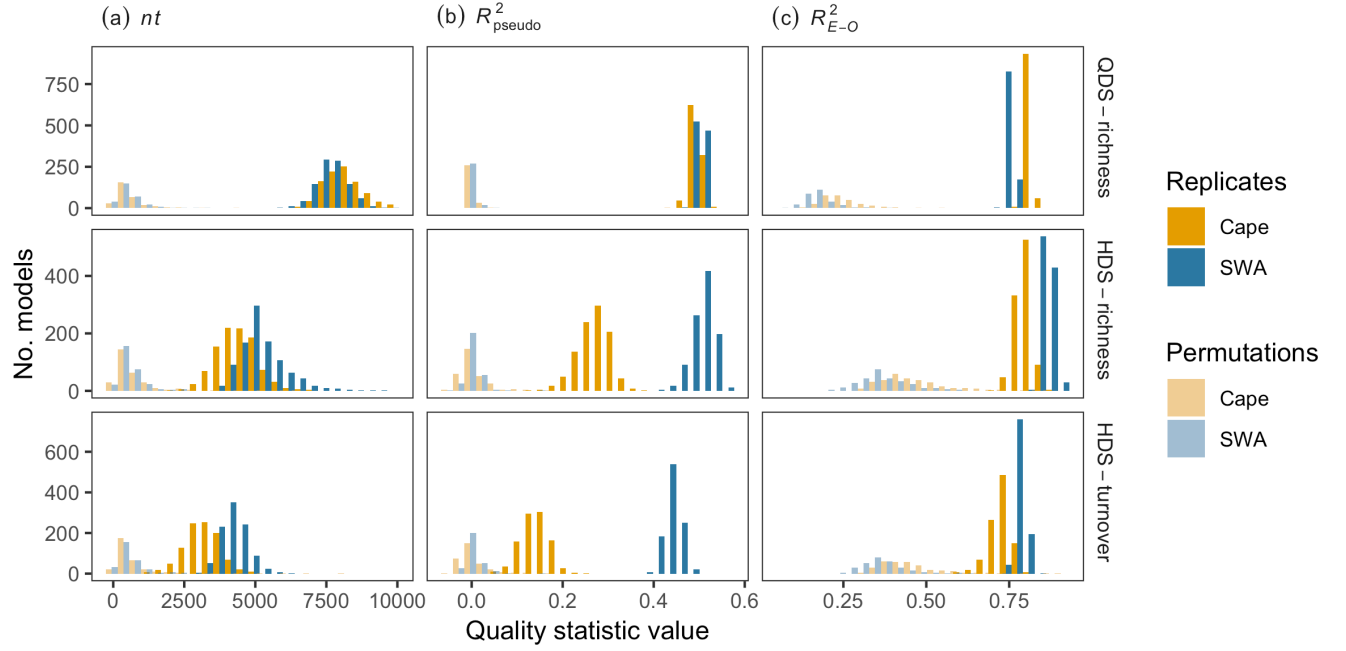


Figure 4: Distributions of three measures of boosted regression tree (BRT) model performance: (a) the number of trees in the model nt , (b) R^2_{pseudo} (Equation (4)), (c) R^2_{E-O} (see text). These measures are presented for the six sets of permuted (pale bars) and six sets of replicate BRT-models (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).

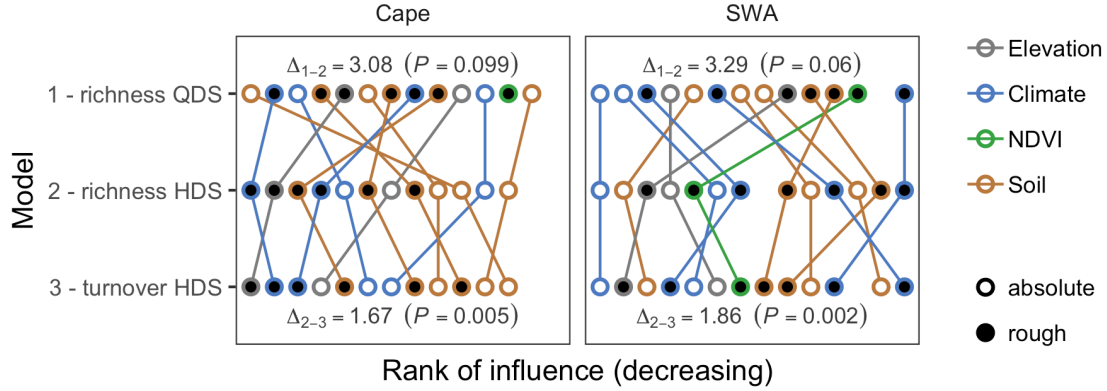


Figure 5: Differences in the rankings of environmental variables' (including heterogeneity variables) relative influences on boosted regression tree (BRT) model predictions of vascular plant species richness and turnover in (a) the Greater Cape Floristic Region (Cape) and (b) Southwest Australia Floristic 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 Figures 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 ($= \bar{S}_{\text{QDS}}$ and S_{HDS} respectively; rows nos. 1 and 2) and between HDS-scale richness and turnover ($= \bar{J}_{\text{QDS}}$) (rows nos. 2 and 3). Statistics (Δ - and P -values) inset at the top and bottom of each panel refer to these comparisons respectively. Δ -values represent the average absolute difference in ranks across variables between two models' rankings. The associate P -value results from ranking the observed Δ -values against 999 Δ -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.

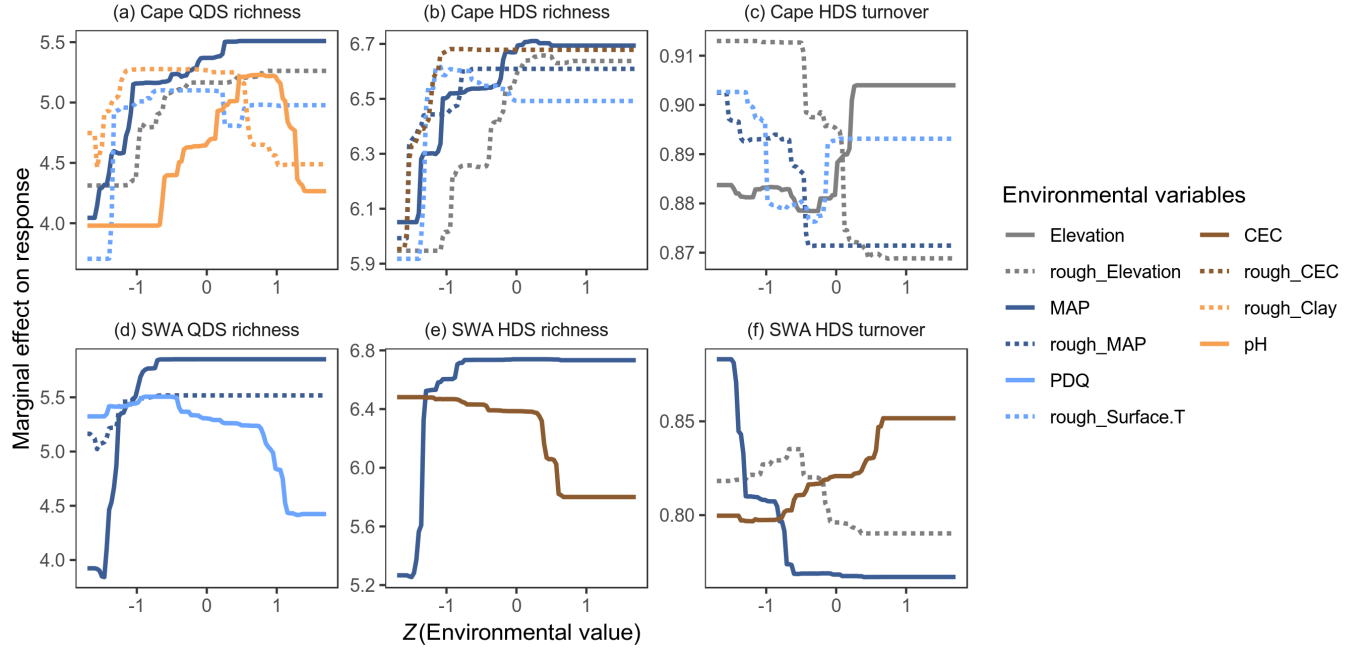


Figure 6: Marginal effects of environmental conditions and heterogeneity on vascular plant species richness at the QDS-scale ($= \bar{S}_{\text{QDS}}$; a,d), HDS-scale ($= S_{\text{HDS}}$; b,e) and turnover ($= \bar{J}_{\text{QDS}}$; c,f) in response variables in the Greater Cape Floristic Region (Cape; 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.