

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

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

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

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

## Abstract

### Aim:

**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 biological importance in the Cape, though this is contingent on the quality of the data modelled.

### Main conclusions:

**Keywords:** biodiversity, environmental heterogeneity, fynbos, Greater Cape Floristic Region, kwongan,

14 macroecology, species richness, species turnover, vascular plants, Southwest Australia Floristic Region

## 15 **Acknowledgements**

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

## 22 **1 Introduction**

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

36 The maintenance of species richness, particularly the coexistence of high numbers of species in biodiversity  
37 hotspots, is often regarded as “paradoxical” (Hart et al., 2017), and is a central problem in macro-ecology and  
38 biogeography (Ricklefs, 1987; Kreft & Jetz, 2007; Hart et al., 2017). Species richness is constrained by the  
39 ability of habitats to support a variety of species—its ecological carrying capacity (Mateo et al., 2017). This is

40 exemplified in modelling approaches, wherein species richness is a function of environmental predictors in a  
 41 correlative framework (“macro-ecological models”; Mateo et al., 2017). Macro-ecological models of species  
 42 richness implicitly assume that communities are saturated, following species-area and species-energy  
 43 relationships, and at equilibrium with the environment (Mateo et al., 2017).

44 A solution to the “paradox” of species co-existence is environmental heterogeneity (EH): a more heterogeneous  
 45 environment exhibits a larger environmental space, thus facilitating co-existence between more species.  
 46 Heterogeneity in the physical environment is known to be positively associated with species richness  
 47 (Rensburg et al., 2002; Hart et al., 2017), and has been demonstrated to do so across many taxa—e.g. Canadian  
 48 butterflies (??), European vertebrates (Mouchet et al., 2015), South African birds (Rensburg et al., 2002), in  
 49 communities along marine continental margins (Levin et al., 2010), French scarab beetles (Lobo et al., 2004),  
 50 and for global terrestrial plants (Kreft & Jetz, 2007). The spatial scale of heterogeneity, or “grain” of the  
 51 environment, is also important to consider (Hart et al., 2017), as spatial scale in absolute environmental  
 52 conditions has also been explored (???; Baudena et al., 2015; Mouchet et al., 2015). Species co-existence and  
 53 biodiversity maintenance is indeed suggested to be scale-dependent (Hart et al., 2017).

54 EH is often under-represented in macro-ecological models of species richness, and has recently been found to  
 55 explain up to ca. 95% of biome level species richness across South Africa (Cramer & Verboom, 2016). Indeed,  
 56 models that include EH yield better estimates of the richness of the Cape flora (Thuiller et al., 2006; Cramer &  
 57 Verboom, 2016). Mediterranean-type terrestrial biodiversity hotspots, such as the Cape flora included in the  
 58 models by Cramer & Verboom (2016), present interesting study systems in which to investigate the relationship  
 59 between the environment and species richness. These systems exhibit far greater species richness than  
 60 predicted by their areas, productivities and latitudes (Cowling et al., 1996; Kreft & Jetz, 2007). There are five  
 61 Mediterranean biodiversity hotspots on Earth: the California Floristic Province, the Mediterranean Basin, the  
 62 Chilean Winter Rainfall-Valdivian Forests, the Greater Cape Floristic Region, and the Southwest Australia  
 63 Floristic Region (Cowling et al., 1996; Hopper & Gioia, 2004; Cook et al., 2015). These ecosystems have  
 64 regular fire-cycles (Cowling et al., 1996), climatic buffering, and long term stability (Kreft & Jetz, 2007),  
 65 shrubby, sclerophyllous flora (Hopper & Gioia, 2004). Together, they account for ca. 20% of global vascular  
 66 plant species, yet only ca. 5% of global land surface areas (Cowling et al., 1996). Various hypotheses have  
 67 been proposed to explain the high levels of plant species richness in these regions (Cook et al., 2015). The  
 68 species accumulation hypothesis states that the stability of these regions has allowed many species to accrue.  
 69 The species co-existence hypothesis states that these hotspots may facilitate greater degrees of species  
 70 co-existence in smaller spatial areas, due to fine-scale heterogeneity in their environments. Indeed, EH has

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

72 Both the Southwest Australia Floristic Region (SWA) and the Greater Cape Floristic Region (Cape) are  
73 Mediterranean-type biodiversity hotspots, particularly in terms of plant species. Where the Cape (with an area  
74 of ca. 189,000 km<sup>2</sup>) is known to contain about 11,400 plant species (about 0.060 species per km<sup>2</sup>), SWA (area  
75 of ca. 270,000 km<sup>2</sup>) has about 3,700 species (0.014 species per km<sup>2</sup>) (???). So, the Cape has ca. 4.3 times as  
76 many species per km<sup>2</sup> as SWA. The Cape and SWA are appropriately often compared, due to the similarities  
77 between their environments (e.g. oligotrophic soils, an oceanically buffered moderate climate) and their plants'  
78 ecologies (Hopper & Gioia, 2004). These two regions present unique flora out of the five Mediterranean  
79 systems, with high levels of endemism (Cowling et al., 1996), and many obligate fire-adapted species (Cowling  
80 et al., 1996). Similarities withstanding, SWA is topographically and edaphically distinct from the Cape. The  
81 former is topographically rather uniform (i.e. flat)—uniquely so among the world's five Mediterranean-climate  
82 regions (Hopper & Gioia, 2004)). SWA possesses a mesoscale chronosequence dune system (Laliberte et al.,  
83 2014; Cook et al., 2015), while the Cape is mountainous, topographically heterogeneous, and therefore  
84 associated with a large degree of spatial climatic variability, with a fine-scale mosaic of geologies and soils  
85 (Cowling et al., 1996; Cramer et al., 2014; Verboom et al., 2017).

86 Both regions have sources of edaphic heterogeneity, but at different scales. This edaphic variability may aid in  
87 explaining the species richness in these regions (Beard et al., 2000; Verboom et al., 2017). EH of many forms  
88 will likely be important in macro-ecological models in both regions, as both regions have been relatively  
89 environmentally stable over evolutionary time-scales (Wardell-Johnson & Horwitz, 1996; Hopper & Gioia,  
90 2004; Lambers et al., 2010; Cramer et al., 2014; Laliberte et al., 2014; Cook et al., 2015). For the Cape, high  
91 levels of species richness are thought to result from long term climatic stability, and fine grain variation in  
92 geology and soils (Cramer et al., 2014). The question thus arises whether heterogeneity is a significant  
93 contributor to SWA species richness. In the absence of topographic variability in SWA, it is proposed that the  
94 heterogeneity of that region is due to the juxtaposition of soil types (Laliberte et al., 2014; Cook et al., 2015),  
95 creating extreme edaphic variation.

96 Our hypotheses concern the Cape and SWA's environments and floras. Our main hypothesis is that the Cape  
97 possesses greater abiotic heterogeneity, and at finer grain, compared to SWA, such as to explain the Cape's  
98 greater species richness per unit area, and proposed greater levels of species turnover between areas. We also  
99 conjecture that the heterogeneity that predicts species richness in SWA will be more pronounced in terms of  
100 edaphic variables. Here we attempt to assess five key predictions of this hypothesis, additionally investigating

101 a seventh prediction to test the conjectured role of edaphic heterogeneity in SWA. Dealing with the two  
102 regions' environments, we assess (i) whether the Cape environment is more heterogeneous than that of SWA  
103 and (ii) whether the Cape environment has more pronounced heterogeneity at finer scales than that of SWA.  
104 Dealing with the distribution of species in the two regions, we assess (iii) whether the Cape exhibits greater  
105 levels of species turnover between areas. Relating each regions' environment and flora, we finally assess (iv)  
106 whether species richness and species turnover are adequately predicted by EH in both regions and whether (v)  
107 species richness and species turnover are better predicted by different forms of EH in either region (e.g. the  
108 importance of edaphic heterogeneity in SWA).

## 109 **2 Materials and methods**

### 110 **2.1 Overview**

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

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

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

### 123 **2.2 Environmental data sources**

124 The GCFR was treated as the area occupied by the Succulent Karoo and Fynbos biomes in the current  
125 delineation of South Africa's biome boundaries (Mucina & Rutherford, 2006). The SWAFR was treated as the

126 areas occupied by the Southwest Australia savanna, Swan Coastal Plain Scrub and Woodlands, Jarrah-Karri  
127 forest and shrublands, Southwest Australia woodlands, Esperance mallee, and Coolgardie woodlands in the  
128 World Wildlife Fund Terrestrial Ecoregions dataset (Olson et al., 2001) in order to closely match the currently  
129 delineated SWAFR (Gioia & Hopper, 2017, Hopper & Gioia (2004)). For the sake of readability, we shall refer  
130 to the GCFR and SWAFR simply as the Cape and SWA from hereon.

131 Geospatially-explicit raster layers were acquired for a selection of environmental variables (Table 1), for the  
132 regions of interest. Raster data were re-projected to a common coordinate reference: WGS84 (NIMA, 2000),  
133 using the “rgdal” (???) package in R (R Core Team, 2018). All data were re-sampled to 0.05° resolution using  
134 the “resample” function in the R package “raster” (???), with the “bilinear” method.

135 An emphasis was made on using satellite-derived environmental data in this work, in order to minimise  
136 differences in data quality and methodologies between the Cape and SWA. Additionally, satellite-derived data  
137 have been shown to benefit regional-scale species distribution models (Deblauwe et al., 2016), thus motivating  
138 their use in this regional-scale study. The environmental data used in this study were derived from NASA’s  
139 SRTM digital elevation model (Farr et al., 2007), NASA’s MODIS/Terra spectroradiometric data for land  
140 surface temperature and NDVI, the Climate Hazards Group’s CHIRPS rainfall dataset (Funk et al., 2015), and  
141 the International Soil Reference and Information Centre’s SoilGrids250m edaphic dataset (Hengl et al., 2017)  
142 (Table 1). SRTM and MODIS are entirely derived from satellite measurements, whereas CHIRPS is  
143 interpolated from weather station data with satellite-derived radiometric measurements. SoilGrids250m is a  
144 machine-learning derived product, based on soil measurements as a function of many covariates, including  
145 MODIS and STRM sources (see Hengl et al., 2017), using random-forests and other classification-tree-based  
146 methods, including gradient-boosting. For the soil data considered here (Table 1), we used depth-interval  
147 weighted average values as the value for a particular soil variable in a given place.

148 Climatic and spectral data arise from satellites monitoring properties of the Earth’s surface through time. We  
149 therefore use the mean annual values for rainfall, surface temperature, and NDVI in each pixel in our analyses.  
150 Pronounced seasonality of rainfall is a known feature of mediterranean systems (???). We describe this  
151 seasonality by computing computing the precipitation in the driest quarter (PDQ), using methods based on the  
152 “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, 24 July 2017, GBIF (24 July 2017)). 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 “taxize” (???, (???)) 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 the final list of species. The number of species names excluded totalled at FIXME and FIXME for the Cape and SWA respectively. Especially for SWA, these numbers may be deemed appreciably high. But, the occurrence records that would be dropped, as a consequence of these names’ removals, appeared randomly distributed in geographic space in both regions. As such, any effect of the loss of these records in this analysis is likely uniform within the two regions.

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

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

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

182 at QDS, HDS, and 3QDS resolutions.

## 183 **2.4 Analyses**

### 184 **2.4.1 Quantifying environmental heterogeneity**

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

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

189 This value, notionally equivalent to the standard deviation of values relative to the focal value, is ascribed to  
190 the focal cell. Note, in order to use as much data from within regions’ borders as possible, roughness was  
191 computed if a focal cell had at least one neighbour cell. Using this index, we produced raster layers of each of  
192 our nine environmental variable’s heterogeneity. We compared the distributions of “roughness” values in each  
193 variable in each region with non-parametric Mann-Whitney  $U$ -tests, as almost all variables were highly  
194 non-normal, and could not be normalised by log-transformations. We also compare the effect size of the Cape  
195 vs SWA using the “common language effect size” ( $CLES$ ), using the R package “canprot”. The  $CLES$  is the  
196 proportion of all pairwise comparisons between two sample groups’ observations where one group’s value is  
197 greater than the other’s. We calculated the  $CLES$  as the proportion of pairs where Cape roughness values  
198 were greater than that of SWA. This allowed us to assess prediction (i). To compare the spatial scales of  
199 heterogeneity (prediction (ii)) between each region, we repeated this analysis at all four spatial scales. This  
200 entailed recalculating the roughness layer for each variable after the original layer (0.05 degrees resolution) had  
201 been rescaled to each of the coarser resolutions.

### 202 **2.4.2 Quantifying species turnover**

203 Regarding prediction (iii), we wished to compare the general degree of species turnover in each region. To  
204 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 (???) between each pair of QDS within each HDS ( $\bar{J}_{\text{QDS}}$ , 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  $\beta$ -diversity (???), as follows:

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

Here, we treat species richness at the HDS-scale ( $S_{\text{HDS}}$ ) as  $\gamma$ -diversity and at the QDS-scale ( $\bar{S}_{\text{QDS}}$ ) as  $\alpha$ -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  $\gamma$ -diversity as in Equation (2), such that  $\beta$ -diversity is the difference between  $\gamma$ - and  $\alpha$ -diversity. We compare the distributions of  $\bar{J}_{\text{QDS}}$  and  $T_{\text{HDS}}$  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.

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

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

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

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

241 As recommended by Elith et al. (2008), BRT models were trained on a set of non-collinear predictor variables  
 242 using “gbm.step” in “dismo” (???) and “gbm” (???). Collinear predictor variables can skew the interpretation  
 243 of results, as the relative influence of mutually collinear variables is reduced. Collinearity among the nine  
 244 environmental predictor variables and their respective nine roughness-equivalents was assessed using  
 245 “removeCollinearity” in the R package “virtualspecies” (???) separately for each region, such that variables  
 246 were no more than 80% collinear (Pearson’s  $r \geq 0.80$ ). When faced with a cluster of collinear variables, one  
 247 variable was chosen manually therefrom. Where possible, the roughness-equivalent of an environmental  
 248 variable was included if its absolute-equivalent could also be included. When interpreting the results of BRTs,  
 249 it is important to consider the effects of the variables included as representative of the effect of the excluded  
 250 variables with which it was found to be collinear.

251 In order to select ideal  $lr$  and  $tc$  all models (described below) were trained on the final non-collinear predictor  
 252 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,  
 253  $5 \times 10^{-4}$ ,  $1 \times 10^{-4}$ ). The function “gbm.step” optimises the number of trees ( $nt$ ) using cross-validation during  
 254 model training (Elith et al., 2008) by halting iteration when predictions begin to overfit. For all models, we  
 255 used 10 cross-validation folds (i.e. use 10 different randomly selected training data sets), a tolerance-threshold  
 256 of 0.001, a bagging-fraction of 0.75 (proportion of training data randomly chosen to generate each tree), and

257 trained models starting with 50 trees, with each iterative step adding 50 trees at a time, up to a maximum of  
 258 10,000 trees. Following this iterative parameter optimisation, Gaussian BRT models were constructed with  
 259  $tc = 3$  and  $lr = 0.001$ , along with the other settings described.

260 The optimum configuration of  $lr$  and  $tc$  for the final model is a trade-off between model fit (e.g. pseudo- $R^2$ ;  
 261 Equation (4)) and complexity ( $nt$ ). A  $tc$  of 5 was chosen for the final model. This follows the  
 262 recommendations of Elith et al. (2008), where  $lr$  and  $tc$  are advised to be adjusted inversely. This was chosen  
 263 in order to account for the complex interactions determining species richness. To avoid overfitting, an  
 264 intermediate  $lr$  of 0.001 was chosen.

#### 265 **2.4.4 Assessing BRT-predictions' fit**

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

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

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

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

278 In order to assess the reliability of the conclusions drawn from these models, we randomly permuted the  
 279 response data ( $S_{\text{QDS}}$ ,  $S_{\text{HDS}}$  and  $\bar{J}_{\text{QDS}}$ ) with respect to the environmental and heterogeneity data, and reran all  
 280 six BRT-models 999 times (with the final non-collinear predictor sets and preconfigurations above). This also  
 281 allows us to remove any effect of spatial autocorrelation in generating the observed correlations between

patterns of species occurrence and environment (???), 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)),  $R^2_{\text{E-O}}$ ) 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

309 between roughness in mean annual rainfall (R MAP) and land surface temperature (R Surface T) in the Cape  
 310 and SWA is less pronounced when considered at coarse spatial scales (Figure ??b). At all spatial scales  
 311 considered, roughness in rainfall seasonality (R PDQ), however, is equally more heterogeneous in the Cape  
 312 than SWA. Biological productivity, as measured by NDVI, is fairly similarly heterogeneous in the Cape and  
 313 SWA ( $CLES < 0.60$ , Figure 1). Concerning edaphic variables, the Cape and SWA are similarly  
 314 heterogeneous at coarser scales, particularly in terms of CEC and Soil C ( $CLES \approx 0.50$ , Figure 1b).

### 315 **3.2 Comparing species turnover in the two regions**

316 Following calculations of  $\bar{J}_{QDS}$  and  $T_{HDS}$  for each HDS-cell in each region, we also used non-parametric  
 317 Mann-Whitney  $U$ -tests to compare the distributions of values in the Cape and SWA. The Cape possesses  
 318 generally greater floristic turnover than SWA, for both measures of turnover defined here ( $P < 0.0001$ , Figure  
 319 2a,b).  $\bar{J}_{QDS}$  measures the average pairwise Jaccard distance between QDS-cells in each HDS-cell.  $T_{HDS}$ ,  
 320 however, represents the inferred  $\beta$  component of  $\gamma$ -diversity. As  $\gamma$ -diversity ( $= S_{HDS}$ ) in the Cape has a  
 321 greater  $\beta$ -diversity component ( $= T_{HDS}$ ) than SWA, the complement is necessarily true:  $\gamma$ -diversity in the  
 322 Cape has a lesser  $\alpha$ -diversity component ( $= \bar{S}_{QDS}$ ) than SWA.

### 323 **3.3 Predicting richness and turnover with environmental heterogeneity**

324 We found vascular plant species richness and turnover both to be predicted primarily by environmental  
 325 heterogeneity in the Cape (Figure 3a–c) and at least in-part by environmental heterogeneity in SWA (Figure  
 326 3d–f). Our six BRT-models performed adequately, and detected relationships between patterns of species  
 327 occurrence and the environment that do not occur in the permuted datasets (Figures 4 and 3, Table 2).

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

332 Across our BRT-models of species richness and turnover, the importance of different environmental variables  
 333 in predictions differed substantially between the Cape and SWA. Additionally, the relative importance of  
 334 absolute and heterogeneity variables also differs between the Cape and SWA (Figure 3). Most obviously,  
 335 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 ( $= \bar{S}_{QDS}$ ), and to a lesser extent at HDS-scales ( $= S_{HDS}$ ), 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 select as representative of itself, NDVI and soil C.

**In the equations below, collinear groups of variables are listed as predictors enclosed within braces.**

**Cape:**

$$\begin{aligned}\bar{S}_{QDS} &\sim pH + R_{SurfT} + \left\{ \begin{matrix} MAP \\ NDVI \\ SurfT \\ SoilC \end{matrix} \right\} + \left\{ \begin{matrix} R_{Clay} \\ R_{pH} \end{matrix} \right\} + RElev \\ S_{HDS} &\sim R_{SurfT} + RElev + R_{CEC} + \left\{ \begin{matrix} R_{MAP} \\ R_{PDQ} \\ R_{NDVI} \end{matrix} \right\} + \left\{ \begin{matrix} MAP \\ NDVI \\ SurfT \\ SoilC \end{matrix} \right\} \\ T_{HDS} &\sim RElev + R_{SurfT} + \left\{ \begin{matrix} R_{MAP} \\ R_{PDQ} \\ R_{NDVI} \end{matrix} \right\} + Elev + R_{CEC}\end{aligned}$$

$$\begin{aligned}\bar{S}_{\text{QDS}} &\sim \left\{ \begin{array}{c} MAP \\ NDVI \\ SoilC \end{array} \right\} + PDQ + RMAP \\ S_{\text{HDS}} &\sim \left\{ \begin{array}{c} MAP \\ NDVI \\ SoilC \end{array} \right\} + CEC + RElev \\ T_{\text{HDS}} &\sim \left\{ \begin{array}{c} MAP \\ NDVI \\ SoilC \end{array} \right\} + RElev + CEC\end{aligned}$$

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

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

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

367 The positive association of heterogeneity variables in the Cape as opposed to SWA (Figure 6a,b vs d,e) concurs  
 368 with their greater importance in BRT-model predictions (Figure 3).

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

variables as species richness, but with opposite signs to its relationships (Figure 6c,f), is indicative of the richness-dependence of the measure of floristic turnover used here (Jaccard distances) to quantify turnover at the HDS-scale.

## 4 Discussion

The Cape is generally more environmentally heterogeneous than SWA, though SWA does possess edaphic heterogeneity as great as that in the Cape at coarse spatial scales.

The Cape has greater levels of floristic turnover in vascular plant communities between grid-cells than SWA.

Vascular plant species richness and turnover in the Cape is associated more strongly with environmental heterogeneity, 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 and natures of associations of different environmental variables with patterns of species richness and turnover varies with spatial scale.

Jaccard distances, as used here as  $\bar{J}_{QDS}$ , are highly richness dependent. Although they represent the proportional floristic turnover between cells, this proportion itself is sensitive to the richness of sites under comparison. For example, the turnover between two low richness sites has is likely to be greater by chance, due to the absence or presence of a few species, than the turnover between two high richness sites, where the absence or presence of a few species does not greatly affect the turnover calculated.

We have support for the hypothesis that the difference in plant species richness between the Cape and SWA is accounted for by the greater abiotic heterogeneity in the Cape. As expected, the Cape is shown to possess (i) a quantifiably more heterogeneous environment that is (ii) generally heterogeneous at a finer spatial scale than SWA. We have shown that vascular plant species richness (iii) can be explained in terms of environmental conditions including environmental heterogeneity in both the Cape and SWA. Also, we have shown that (iv) the sets of environmental axes that explain plant species richness differ between the Cape and SWA. These findings contribute towards an understanding of the ecological conditions associated with high levels of species co-existence in these two regions.

These two regions present differentiable environmental spaces, each with varying degrees of heterogeneity



396 across spatial scale. The clear separation of the regions' topographic features is as expected (Figures ??A, ??),  
397 particularly regarding topography. The Cape region has been found previously to have the second highest  
398 median topographic heterogeneity of the five Mediterranean-climate regions (Bradshaw & Cowling, 2014).  
399 The Cape has a much wider range of scales exhibited in the heterogeneity across its environmental axes.  
400 Notably, each region has finer scale heterogeneity in some variables, and coarser scale in others—neither  
401 region is more fine or coarse than the other over all environmental axes. BRT-models of species richness in  
402 both regions reveal species richness to depend on those environmental axes that differentiate the two regions  
403 (Figures ??), ??). The importance of variables is also shown to vary with spatial scale (Figure ??), as  
404 previously suggested may be the case when modelling geographic patterns of biodiversity (Baudena et al.,  
405 2015). Indeed, as Cowling et al. (1996) describes differing patterns of species richness across spatial scales, so  
406 do the predictors of those patterns vary with scale (Hart et al., 2017).

407 The fact that a combination of absolute and roughness variables is also as predicted by the hypothesis in this  
408 study. In the models developed by Cramer & Verboom (2016) for South Africa, roughness in topography was  
409 largely superseded as an important predictor of species richness by other roughness variables.

410 ...

411 The determinants of vascular plant species are shown to be region specific (Figures ??, ??, ??).

412 ...

413 NDVI is more heterogeneous across the Cape than SWA (Figures ??A). The fact that thermal variables tend to  
414 be more rough in the Cape (Figure ??A) is likely due to possible covariance of the MODIS/Terra products with  
415 topography, as MODIS data used here describes land surface temperature. As the Cape is topographically  
416 rugged, the roughness of NDVI may arise from this. Despite this, NDVI is an integrating variable, which  
417 captures information about productivity, light availability, and soil nutrients (Power et al., 2017). The fact that  
418 absolute NDVI contributes to predicting species richness in the Cape, especially at finer spatial scales (Figure  
419 ??E) demonstrates the role of ecological productivity in facilitating the coexistence diverse species  
420 assemblages. Environmental heterogeneity, then, is integral to explaining patterns of species richness, but must  
421 be considered along with resource- and energy-availability axes. In so much as a diverse environmental space  
422 supports more species, the materials and productivity required for biota to thrive are also needed to support  
423 species (???; Gaston, 2000; Bøhn & Amundsen, 2004; Kreft & Jetz, 2007). As such, my findings, along with  
424 those of previous studies (Rensburg et al., 2002; Thuiller et al., 2006; Kreft & Jetz, 2007; Cramer & Verboom,

2016), suggest that there is ecological and evolutionary consequence to resource availability *and* environmental heterogeneity, in that they tend to be positively associated with species richness.

The combined BRT-model of species richness for both regions reveals soil clay content as an important predictor, at coarse spatial scales, despite this variable not being particularly important within each region separately (Figure ??). Though this model does not strictly consider the regions as separate, this finding may indicate that the relationship between clay content and species richness differs between the regions. So far as clay content can be used to predict species richness, it matters more to those predictions when applied to large sections (i.e. coarse scales) of each regions.

Kreft & Jetz (2007) modelled global terrestrial vascular plant species richness, which focussed on primarily absolute environmental values, 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. @Cramer2016 described 68% of species richness at the QDS scale across South Africa. Regarding the Cape, depending on whether one consults pseudo- $R^2$  (Table 3), the ratio of mean predicted to observed richness per grid-cell (Table 5), or the distributions of predicted vs. observed richness values per grid-cell (Figure ??), I have achieved a similarly suitable level of predictive accuracy. There is, though, still unexplained species richness in light of my models. As Cramer & Verboom (2016), Rensburg et al. (2002), Thuiller et al. (2006), and Mouchet et al. (2015) have done, these macro-ecological models are a-historical. Evolutionary considerations of species richness in geographic space are worthwhile, especially in regions with environments stable over evolutionary time.

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

455 those with similar post-fire responses.

456 Though the Cape was correctly predicted to have, on average, more species per grid-cell at HDS and 3QDS  
457 scales than SWA, this was not the case for QDS grid-cells (Table 5). This demonstrates that the Cape is indeed  
458 overall more rich in plant species than SWA, but a given HDS in SWA contains fewer species than a given Cape  
459 HDS. Thus, the greater richness in the Cape is a product of greater turnover in species at spatial scales no more  
460 coarse than the HDS. Species turnover is an interesting aspect to species richness studies, as it species turnover  
461 is implicit to species-area and co-existence-area relationships (Hart et al., 2017). One could expect patterns of  
462 endemism and species turnover to concur with patterns in environmental heterogeneity to some degree.

463 Following from the understanding that functionally diverse assemblages, which are more likely to be more  
464 species rich, are likely to arise and/or occur in areas with diverse ecological pressures (Molina-Venegas et al.,  
465 2015), one would expect, then, heterogeneous habitats such as those in Mediterranean-type biodiversity  
466 hotspots to exhibit high levels functional beta diversity along steep environmental gradients (Molina-Venegas  
467 et al., 2015). If the niches concerning these functions are phylogenetically conserved among those biota, then  
468 one would also expect high levels of species and phylogenetic beta diversity along these gradients  
469 (Molina-Venegas et al., 2015). This concurs with the notion put forward by Power et al. (2017), wherein  
470 megadiverse systems such as these represent the results of “phylogenetic niche conservatism on a  
471 heterogeneous landscape”. Thus, species and phylogenetic turnover should covary with environmental  
472 heterogeneity in some way. Indeed, endemism, at certain scales, could also follow this pattern. Thuiller et al.  
473 (2006) demonstrated that there is phylogenetic and biome related determinants of species richness. This makes  
474 sense, in light of the difficulty of crossing biome boundaries in Mediterranean systems (Power et al., 2017).  
475 NDVI and light availability, and the heterogeneity therein, are associated with high levels of floristic turnover  
476 (Power et al., 2017). This may be indicative of ecological specialisation precluding species from crossing these  
477 boundaries, thus increasing the level of endemism within a region, while also increasing the level of turnover,  
478 and thus likely species richness, along environmental gradients. Although, this may be debated. Beard et al.  
479 (2000) state that the high levels of endemism in SWA are function of habitat specialisation to soil mosaics. Cf.  
480 Laliberte et al. (2014), who say that this endemism is likely due to environmental filtering along these soil  
481 turnover sequences, as opposed to the juxtaposition of specialised species along soil gradients.

482 Summarily, we have demonstrated support for the generality of as a meaningful predictor of patterns of species  
483 richness and turnover in mediterranean-type ecosystems. In SWA and the Cape, high levels of are also likely  
484 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.

#### 4.1 Future studies (a.k.a. “to do after first review maybe”)

- 3QDS scale BRTs
- $S_{\text{HDS}} \sim \bar{S}_{\text{QDS}} + \bar{\delta}_{ij}??$  (= “ $\gamma = \alpha + \beta$ ”-analysis) (see `explore-turnover-metrics.pdf` note, where  $\bar{\delta}_{ij}(\mathbf{N})?? = \bar{\alpha}_i(\mathbf{N}) \times \bar{\beta}_{ij}(\mathbf{N}) = \bar{S}_{\text{QDS}} \times \bar{J}_{\text{QDS}}$ )
- Expanding the analyses to other/all mediterranean-type regions
- Investigating environmental stability through time (e.g. WorldClim LGM vs today)
- Phylogenetics/mega-trees in different flora?  $\beta_{\text{phyllo}}$ ?
- Teasing apart co-existence and evolution
- Teasing about turnover into turnover-sensu-stricto and nestedness
- Functional diversity  $\sim$  EH? Functional turnover?

#### 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 ( $nt$ ,  $R_{\text{pseudo}}^2$  (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 ( $nt$ ,  $R_{\text{pseudo}}^2$  (Equation (4)),  $R_{\text{E-O}}^2$  (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

511 quality measures ( $P < 0.0001$ ).

## 512 **Figure captions**

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

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

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

528 Figure 3: Relative influence of environmental variables (including heterogeneity variables—prefixed with “R”)   
529 in boosted regression tree (BRT) model predictions in the Greater Cape Floristic Region (Cape, a–c) and the  
530 Southwest Australia Floristic Region (SWA, b–d) of vascular plant species richness at the (b,e) QDS-scale  
531 ( $= \bar{S}_{QDS}$ ), (a,d) HDS-scale ( $= S_{HDS}$ ) and (c,f) turnover ( $= \bar{J}_{QDS}$ ). All BRT-models were permitted to fit  
532 three-way interactions between environmental variables. Points denote the mean contribution of an  
533 environmental variable to model-predictions across the 1000 replicate BRT-models for that region/scope.  
534 Horizontal ticks denote the mean for the 999 permuted BRT-models. Standard deviations above and below  
535 these means are shown with vertical lines. Note, in the case of the replicate, standard deviations are so small  
536 such that the vertical lines are obscured by the points. Colours represent the general category of the  
537 environment to which a variable belongs (keyed), as in Figure 1b. Left-most piecharts inset in each panel  
538 display the same information. Right-most piecharts group contributions according to whether a variable was

539 absolute or roughness-transformed (keyed).  $F$ -statistics inset are for one-way ANOVAs of differences in  
540 variables' relative influences—for both the replicate ( $F_{\text{rep.}}$ ) and permuted ( $F_{\text{prm.}}$ ) BRT-models.

541 Figure 4: Distributions of three measures of boosted regression tree (BRT) model performance: (a) the number  
542 of trees in the model  $nt$ , (b)  $R^2_{\text{pseudo}}$  (Equation (4)), (c)  $R^2_{\text{E-O}}$  (see text). These measures are presented for the  
543 six sets of permuted (pale bars) and six sets of replicate BRT-models (dark bars) as in Figure 3, coloured  
544 according to the region of interest as in Figures 1a and 2. In all cases, replicate BRT-models almost entirely  
545 out-rank the permuted models in terms of performance (Table 2) and the Greater Cape Floristic Region (Cape)  
546 and Southwest Australia Floristic Region (SWA) models had significantly different values for each metric  
547 (Table 3). Note, the actual differences between Cape and the SWA models' values is not realistically significant  
548 in some cases (e.g. the difference in  $nt$  between the Cape and SWA QDS richness models is statistically  
549 significant, but are observedly so similar as not to affect interpretation).

550 Figure 5: Differences in the rankings of environmental variables' (including heterogeneity variables) relative  
551 influences on boosted regression tree (BRT) model predictions of vascular plant species richness and turnover  
552 in (a) the Greater Cape Floristic Region (Cape) and (b) Southwest Australia Floristic Region (SWA) (as in  
553 Figure 3). Each point represents an environmental variable's rank in BRT-model importance, decreasing in  
554 importance from left to right. Rankings used here are the same as that of the average relative influence for  
555 variables across replicate BRT-models, presented in Figure 3. Coloured lines connect points representing the  
556 same environmental variable. Points' outlines are coloured according to the general category of the  
557 environment (keyed) to which a variable belongs, as in Figures 1b and 3, while points' centres are coloured  
558 according to whether a variable was roughness-transformed or not. The comparisons of variables' rankings of  
559 interest are between QDS- and HDS-scale richness ( $= \bar{S}_{\text{QDS}}$  and  $S_{\text{HDS}}$  respectively; rows nos. 1 and 2) and  
560 between HDS-scale richness and turnover ( $= \bar{J}_{\text{QDS}}$ ) (rows nos. 2 and 3). Statistics ( $\Delta$ - and  $P$ -values) inset at  
561 the top and bottom of each panel refer to these comparisons respectively.  $\Delta$ -values represent the average  
562 absolute difference in ranks across variables between two models' rankings. The associate  $P$ -value results from  
563 ranking the observed  $\Delta$ -values against 999  $\Delta$ -values based on random permutations of variables' rankings  
564 (SI1), such that more significant  $P$ -values denote rankings more similar than would be expected by chance.

565 Figure 6: Marginal effects of environmental conditions and heterogeneity on vascular plant species richness at  
566 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  
567 the Greater Cape Floristic Region (Cape; a–c) and Southwest Australia Floristic Region (SWA; d–f) following  
568 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.

## 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.
- 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.
- 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., Stévant, 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.

- 598 Elith, J., Leathwick, J.R., & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
- 599 Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S.,  
600 Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007) The shuttle radar topography mission.  
601 *Reviews of Geophysics*, **45**, 1–33.
- 602 Funk, C.C., Peterson, P.J., Landsfeld, M., Pedreros, D.H., Verdin, J., Shukla, S., Husak, G., Rowland, J.D., Harrison, L., Hoell, A., &  
603 Michaelsen, J. (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring  
604 extremes. *Scientific Data*, **2**, 150066.
- 605 Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.
- 606 GBIF (24 July 2017) GBIF Occurrence Download..
- 607 GBIF (24 July 2017) GBIF Occurrence Download..
- 608 Gioia, P. & Hopper, S.D. (2017) A new phytogeographic map for the Southwest Australian Floristic Region after an exceptional decade  
609 of collection and discovery. *Botanical Journal of the Linnean Society*, **184**, 1–15.
- 610 Hart, S.P., Usinowicz, J., & Levine, J.M. (2017) The spatial scales of species coexistence. *Nature Ecology & Evolution*, **1**, 1066–1073.
- 611 Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti?, A., Shangguan, W., Wright, M.N.,  
612 Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E.,  
613 Wheeler, I., Mantel, S., & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning.  
614 *PLoS ONE*, **12**, e0169748.
- 615 Hopper, S.D. (1979) Biogeographical Aspects of Speciation in the Southwest Australian Flora. *Annual Review of Ecology and*  
616 *Systematics*, **10**, 399–422.
- 617 Hopper, S.D. & Gioia, P. (2004) The Southwest Australian Floristic Region: Evolution and Conservation of a Global Hot Spot of  
618 Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 623–650.
- 619 Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of*  
620 *Sciences*, **104**, 5925–5930.
- 621 Laliberte, E., Zemunik, G., & Turner, B.L. (2014) Environmental filtering explains variation in plant diversity along resource gradients.  
622 *Science*, **345**, 1602–1605.
- 623 Lambers, H., Brundrett, M.C., Raven, J.A., & Hopper, S.D. (2010) Plant mineral nutrition in ancient landscapes: high plant species  
624 diversity on infertile soils is linked to functional diversity for nutritional strategies. *Plant and Soil*, **334**, 11–31.
- 625 Larsen, R., Holmern, T., Prager, S.D., Maliti, H., & Røskaft, E. (2009) Using the extended quarter degree grid cell system to unify  
626 mapping and sharing of biodiversity data. *African Journal of Ecology*, **47**, 382–392.
- 627 Levin, L.A., Sibuet, M., Gooday, A.J., Smith, C.R., & Vanreusel, A. (2010) The roles of habitat heterogeneity in generating and



- 628 maintaining biodiversity on continental margins: an introduction. *Marine Ecology*, **31**, 1–5.
- 629 Lobo, J.M., Jay-robert, P., Lumaret, J.-p., Lobo, J.M., Jay-robert, P., & Lumaret, J.-p. (2004) Modelling the Species Richness  
630 Distribution for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography*, **27**, 145–156.
- 631 Mateo, R.G., Mokany, K., & Guisan, A. (2017) Biodiversity Models: What If Unsaturaton Is the Rule? *Trends in Ecology &*  
632 *Evolution*, **32**, 556–566.
- 633 Molina-Venegas, R., Aparicio, A., Slingsby, J.A., Lavergne, S., & Arroyo, J. (2015) Investigating the evolutionary assembly of a  
634 Mediterranean biodiversity hotspot: Deep phylogenetic signal in the distribution of eudicots across elevational belts. *Journal*  
635 *of Biogeography*, **42**, 507–518.
- 636 Mouchet, M., Levers, C., Zupan, L., Kuemmerle, T., Plutzar, C., Erb, K., Lavorel, S., Thuiller, W., & Haberl, H. (2015) Testing the  
637 effectiveness of environmental variables to explain European terrestrial vertebrate species richness across biogeographical  
638 scales. *PLoS ONE*, **10**, 1–16.
- 639 Mucina, L. & Rutherford, M.C. (2006) *The vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity  
640 Institute,
- 641 NIMA (2000) Amendment 1. 3 January 2000. Department of Defense World Geodetic System 1984. Its Definition and Relationships  
642 with Local Geodetic Systems. 1–3.
- 643 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,  
644 H.E., Morrison, J.C., & Others (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of  
645 terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, **51**, 933–938.
- 646 Power, S.C., Verboom, G.A., Bond, W.J., & Cramer, M.D. (2017) Environmental correlates of biome-level floristic turnover in South  
647 Africa. *Journal of Biogeography*, **44**, 1745–1757.
- 648 R Core Team (2018) *R: A Language and Environment for Statistical Computing. Version 3.5.0*. R Foundation for Statistical  
649 Computing, Vienna, Austria.
- 650 Rensburg, B.J. van, Chown, S.L., & Gaston, K.J. (2002) Species Richness, Environmental Correlates, and Spatial Scale: A Test Using  
651 South African Birds. *The American Naturalist*, **159**, 566–577.
- 652 Ricklefs, R.E. (1987) Community diversity: relative roles of local and regional processes. *Science, New Series*, **235**, 167–171.
- 653 Thuiller, W., Midgley, G.F., Rouget, M., Cowling, R.M., F. Midgley, G., Rougeti, M., & M. Cowling, R. (2006) Predicting patterns of  
654 plant species richness in megadiverse South Africa. *Ecography*, **29**, 733–744.
- 655 Verboom, G.A., Stock, W.D., & Cramer, M.D. (2017) Specialization to extremely low-nutrient soils limits the nutritional adaptability  
656 of plant lineages. *The American Naturalist*, **In press**.
- 657 Wardell-Johnson, G. & Horwitz, P. (1996) Conserving biodiversity and the recognition of heterogeneity in ancient landscapes: a case



## 659 **Biosketches**

660 **Ruan van Mazijk** is a Masters student interested in phylogenetic systematics, macro-ecology, comparative  
661 work and plant functional ecology.

662 **Michael D. Cramer**

663 **G. Anthony Verboom**

## 664 **Author contributions**

665 MDC and GAV conceived the study question, which RVM investigated under their supervision for his BSc  
666 Hons project. The analyses and programming work were largely devised by RVM, with input from the other  
667 authors, and was carried out by RVM. RVM wrote the first draft of the manuscript and all authors contributed  
668 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	??
<b>Climatic variables</b>			
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	??
<b>Soil variables</b>			
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 ( $nt$ ,  $R^2_{\text{pseudo}}$  (Equation (4)),  $R^2_{\text{E-O}}$  (see text)) of the 1000 replicate BRT-models relative to 999 BRT-models fit to permuted datasets. Each of the 1000 replicate BRT-models was ranked against the 999 permuted BRT-models. The average rank of each, as a proportion, is presented.

Model	$nt$	$R^2_{\text{pseudo}}$	$R^2_{\text{E-O}}$
<b>QDS-richness</b>			
GCFR	1.000	1.000	1.000
SWAFR	1.000	1.000	1.000
<b>HDS-richness</b>			
GCFR	0.987	1.000	0.988
SWAFR	1.000	1.000	1.000
<b>HDS-turnover</b>			
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)),  $R^2_{\text{E-O}}$  (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	$nt$	$R^2_{\text{pseudo}}$	$R^2_{\text{E-O}}$
QDS-richness	542.938	0.063	-0.005
HDS-richness	-808.994	-0.064	-0.233
HDS-turnover	-997.045	-0.052	-0.296

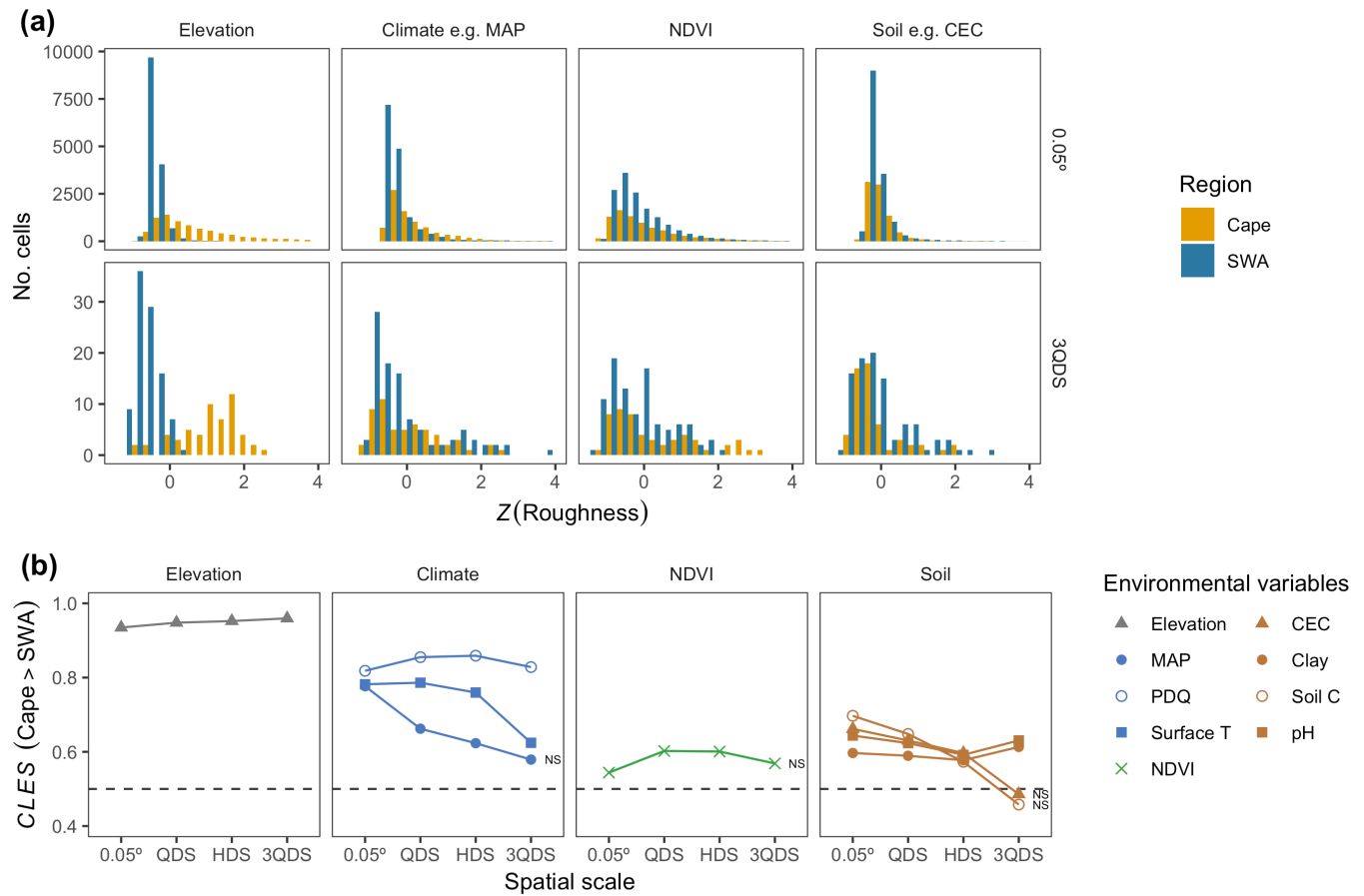


Figure 1: Comparisons of different types of environmental heterogeneity between the the Greater Cape Floristic Region (Cape) and the Southwest Australia Floristic Region (SWA). We present (a) distributions of roughness values (Equation (1)) for example variables from each broad category of the environment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

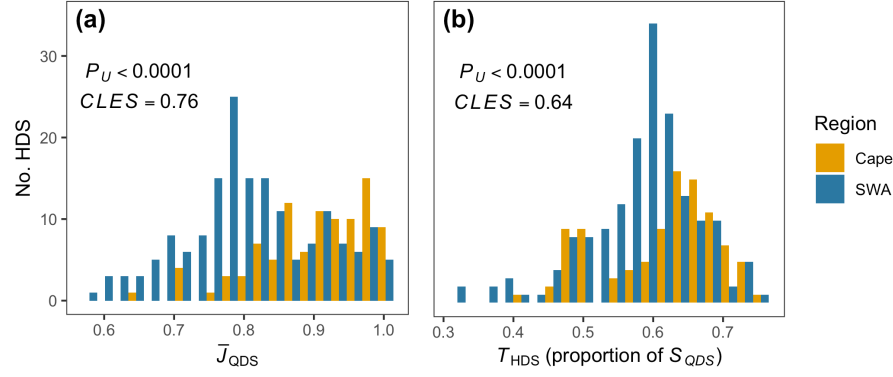


Figure 2: Species turnover, described in two forms ((a) mean Jaccard distance between QDS in each HDS ( $\bar{J}_{QDS}$ ), (b) additively defined turnover ( $T_{HDS}$ , Equation (2)) as a proportion of HDS richness ( $S_{HDS}$ )), compared between the Cape and SWA. Mann-Whitney  $U$ -tests between the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

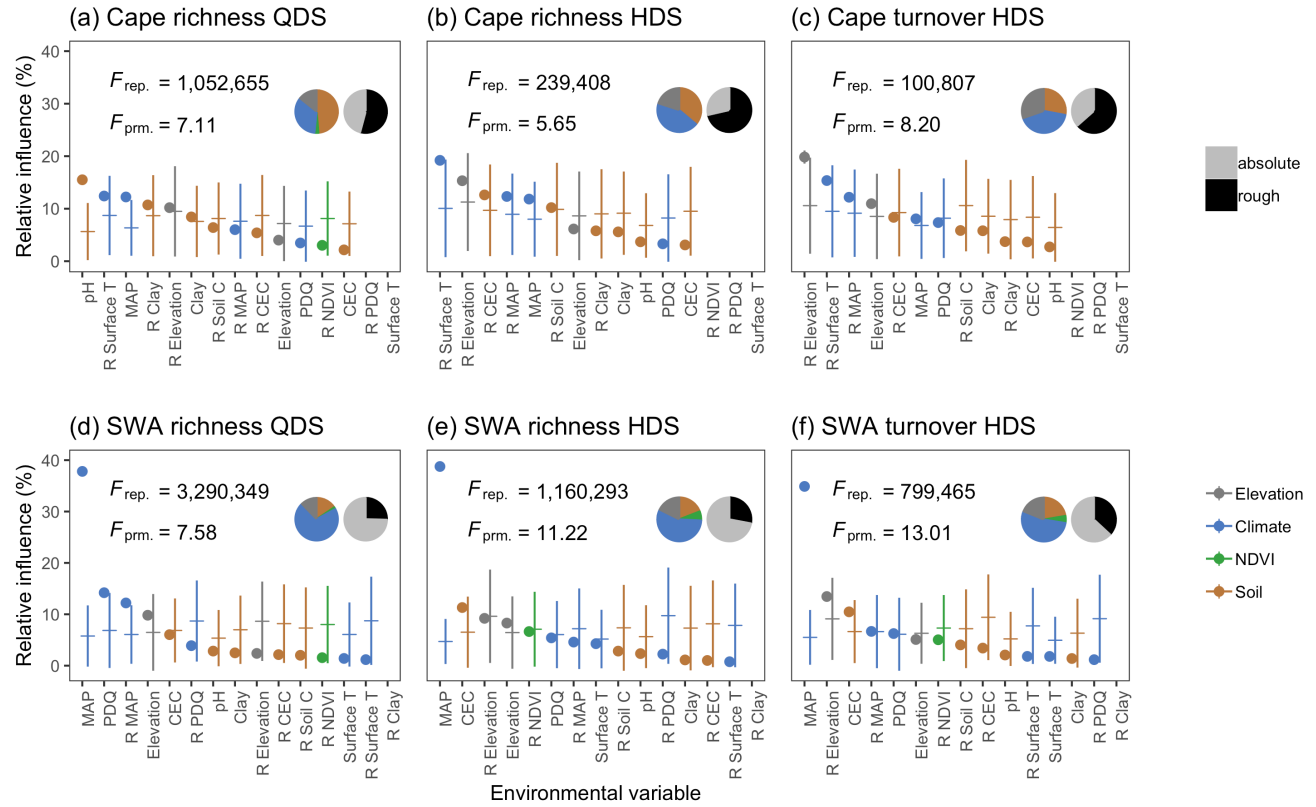


Figure 3: Relative influence of environmental variables (including heterogeneity variables—prefixed with “R”) in boosted regression tree (BRT) model predictions in the Greater Cape Floristic Region (Cape, a–c) and the Southwest Australia Floristic Region (SWA, b–d) of vascular plant species richness 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.



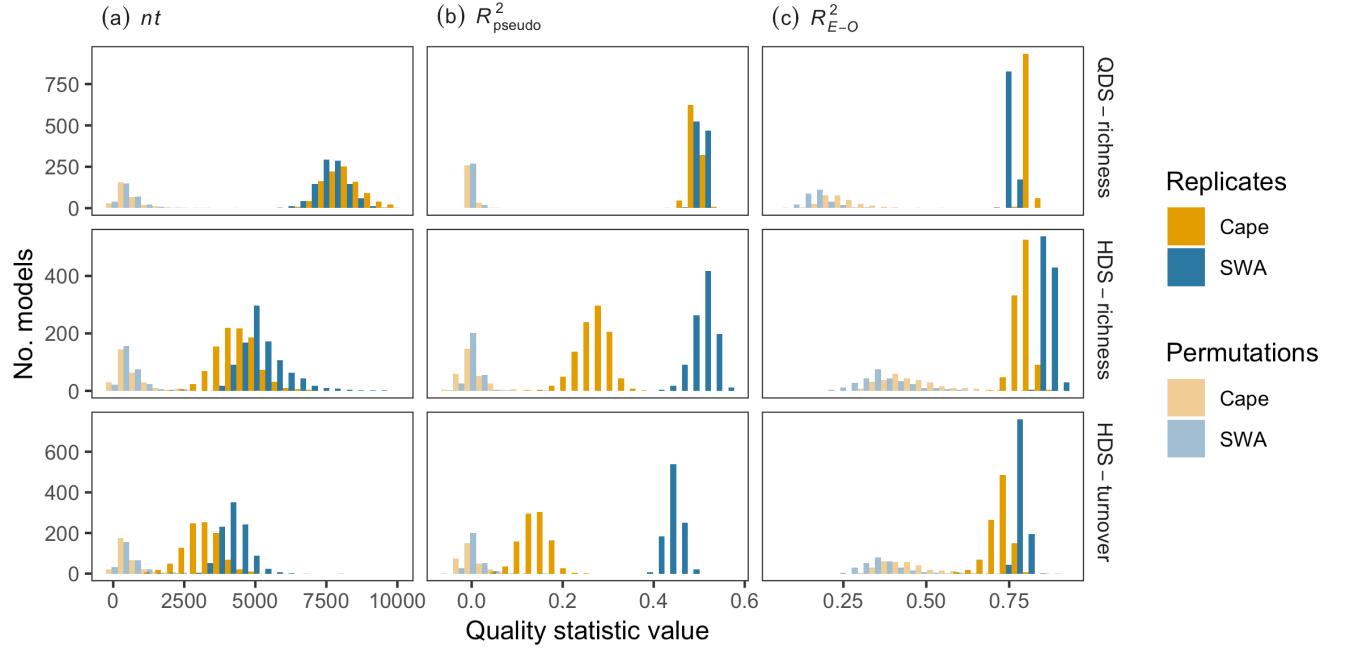


Figure 4: Distributions of three measures of boosted regression tree (BRT) model performance: (a) the number of trees in the model  $nt$ , (b)  $R^2_{\text{pseudo}}$  (Equation (4)), (c)  $R^2_{E-O}$  (see text). These measures are presented for the six sets of permuted (pale bars) and six sets of replicate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

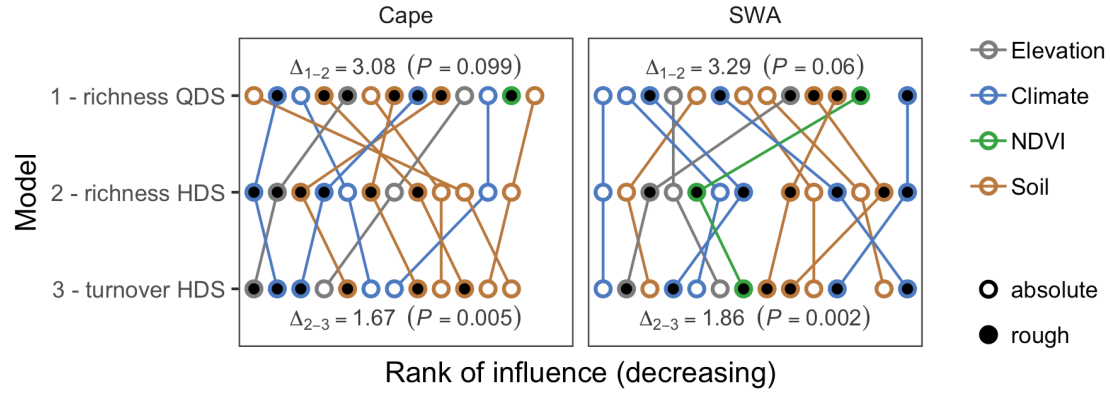


Figure 5: Differences in the rankings of environmental variables' (including heterogeneity variables) relative influences on boosted regression tree (BRT) model predictions of vascular plant species richness and turnover in (a) the Greater Cape Floristic Region (Cape) and (b) Southwest Australia Floristic 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 ( $= \bar{S}_{QDS}$  and  $S_{HDS}$  respectively; rows nos. 1 and 2) and between HDS-scale richness and turnover ( $= \bar{J}_{QDS}$ ) (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.

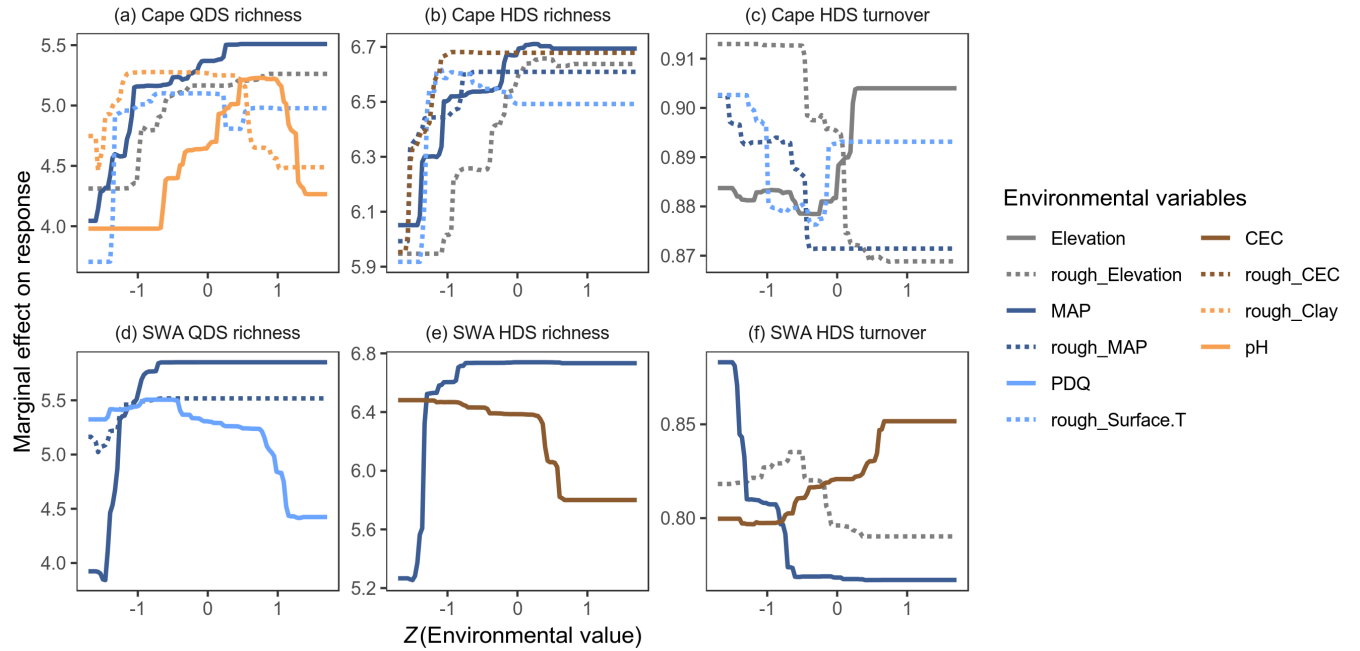


Figure 6: Marginal effects of environmental conditions and heterogeneity on vascular plant species richness at the QDS-scale ( $= \bar{S}_{QDS}$ ; a,d), HDS-scale ( $= S_{HDS}$ ; b,e) and turnover ( $= \bar{J}_{QDS}$ ; c,f) in response variables in the Greater Cape Floristic Region (Cape; a–c) and Southwest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