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 and G. Anthony Verboom

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

Corresponding author: RvM, [ruanvmazijk@gmail.com](mailto:ruanvmazijk@gmail.com)

# Acknowledgements

This work was funded by the South African Department of Science and Technology and the National Research Foundation under the Freestanding Innovation Honours Scholarship and by the South African Association of Botanists Honours Scholarship (both to RvM). Thanks also go to the Department of Biological Sciences, University of Cape Town, for providing a 2TB external hard drive for local GIS data storage.

# Abstract

**Aim:**

**Location:** The Greater Cape Floristic Region, South Africa (GCFR) and Southwest Australian Floristic Region (SWAFR)

**Taxon:** Vascular plants (Tracheophytes)

**Methods:**

**Results:**

**Main conclusions:**

*Keywords:*

# 1 Introduction

The species richness of a region is a function of its biogeographic context (e.g. proximity to potential immigrant source areas), its diversification history (e.g. speciation and extinction history) and any locally-deterministic, environmental features (e.g. environmental productivity, heterogeneity) that influence species persistence and coexistence (Ricklefs 1987, 2004). Since all three effects are potentially influenced by environmental heterogeneity (environmental heterogeneity), the latter may be a particularly important driver of regional species richness variation (refs), with physically-heterogeneous regions being especially prone to be species-rich (refs). For example, given that the recruitment success of immigrant lineages into a region is often dictated by the pre-adaptations of those lineages (Ackerly, Donoghue, Crisp), a physically-heterogenous environment may promote diversity by admitting a more functionally-diverse array of immigrant lineages (ref). In addition, by virtue of its central role in powering adaptive divergence and/or promoting population isolation, environmental heterogeneity is a critical requirement for speciation under most models (Wiens 2004a, b; Sobel et al. 2010; Nosil?). Likewise, in the context of long-term environmental change, physically heterogeneity may offer refugia to a wider array of lineages and so confer a greater level of buffering against lineage extinction (refs Byrne?). Finally, environmental heterogeneity has repeatedly been shown to facilitate species coexistence at a variety of scales, and so enhance regional species richness (refs). Differences in environmental heterogeneity may therefore be critically important in accounting for variation in regional species richness, particularly where the regions under comparison are similar in terms of area, their physical properties, and the time-frames within which their biotas have assembled.

The floristically-rich South Western Australian Floristic Region (SWAFR; Hopper and Gioia 2004) and Greater Cape Floristic Region of South Africa (GCFR; Born et al. 2007) constitute a case in point. Situated on the southwestern corners of their respective continents, the climates of both these regions have been oceanically-moderated at least since the Cretaceous, and both are dominated by a contemporary mediterranean-type climate whose origin can be traced to the Early-Middle (SWAFR: Rundel et al. 2016; Lamont and He 2017) or Late Miocene (GCFR: Dupont et al. 2011; Hoffmann et al 2015). In addition, both regions have been unglaciated since the Permian and are dominated by ancient, weathered landscapes whose soil-nutritional status is amongst the lowest of any landscape on Earth (Stock and Verboom), hence their designation as “OCBILs” (“old, climatically-buffered infertile landscapes”; Hopper 2009). Owing to these environmental similarities, the SWAFR and GCFR floras are very similar in terms of their functional trait spectra (Cowling and Witkowski 1994), though the presence of a significant tree component in the SWAFR underpins a striking difference in vegetation physiognomy (ref). Moreover, the long-term climatic and geological stability of the two regions ensures that the native floras of both reflect long histories of assembly, extending back to the Palaeocene and possibly even earlier (refs), with evidence of a long history of transoceanic dispersal between the two (refs). In this context, it is unsurprising that the two floras show strong taxonomic affinities and that both are species-rich, with high levels of regional endemism (refs).

Notwithstanding these similarities, the SWAFR and GCFR differ markedly in terms of their vascular plant species richness, particularly when considered in relation to geographical area. Where the SWAFR accommodates ~7,380 species in an area of ~302,600 km2 (i.e. 0.024 species km-2; Hopper and Gioia 2004), the GCFR accommodates ~11,430 species in an area of ~189,700 km2 (i.e. 0.060 species km-2; Snijman 2013). One explanation for this striking 2.5-fold species richness difference (per area) relates to differences in the physical heterogeneity of the two regions. Where much of the GCFR, particularly the hyper-diverse (~9,400 species in ~90,800 km2; 0.104 species km-2) “core” Cape Floristic Region (CFR; Goldblatt 1978), is rugged and mountainous, the SWAFR landscape is much more subdued, comprising an ancient, weathered plateau. Since the strong relief of the GCFR underlies steep climatic and edaphic gradients (refs), it is probable that environmental heterogeneity is generally greater. The central aim of this paper, then, is to test the hypothesis that the observed species richness difference (per area) is a consequence of differences in the physical heterogeneity of these regions. Focusing on the quarter-degree square (QDS) and half-degree square (HDS) scales (sensu Larsen et al., 2009), we first compare the distribution of species richness between the two regions, and each region decompose HDS richness into mean QDS richness and between-QDS turnover. Thereafter, we compare environmental heterogeneity between the two regions across a range of spatial scales. Finally, we use linear models to assess whether differences in environmental heterogeneity are sufficient to explain observed differences in species richness between the two regions.

# 2 Materials and methods

## 2.1 Comparing species richness

To compare vascular plant species richness between the GCFR and SWAFR, geospatially-explicit occurrence records of tracheophytes from within the borders of the GCFR and SWAFR were downloaded from the Global Biodiversity Information Facility (GBIF, see Table 1). For this purpose, the GCFR was treated as the area occupied by the Succulent Karoo and Fynbos Biomes (Mucina & Rutherford 2006), while the SWAFR was treated as the area occupied by Southwest Australia Savanna, Swan Coastal Plain Scrub and Woodlands, Jarrah-Karri Forest and Shrublands, Southwest Australia Woodlands, Esperance Mallee, and Coolgardie Woodlands (Olson et al., 2001) in order to match the current delimitation of the SWAFR (Hopper & Gioia 2004; Gioia & Hopper, 2017). The query results were then cleaned as follows. Firstly, we retained only records identified to the species level, and ignored intraspecific taxa. This resulted in the retention of XXX and XXX unique species names for the GCFR and SWAFR, respectively. The R package “taxize” (Chamberlain & Szocs, 2013; Chamberlain et al., 2018) was then used to query each species name against two taxonomic databases, the Global Name Resolver (GNR) and the Taxonomic Name Resolution Service (TNRS). Where either or both databases returned a match for a name, the name was retained; where not, it was excluded. Although the number of species thus excluded is high (GCFR: XXX; SWAFR: XXX), the geographically-random distribution of the records associated with these names suggests that exclusion of these names will not significantly influence spatial patterns of species richness. In order to ensure that no species was listed under multiple synonyms, the retained names were then queried against the Tropicos and Integrated Taxonomic Information System (ITIS) for known synonyms, again using “taxize.” Finally, we removed all records of species identified as non-native, using lists of invasive plants for South Africa and Australia from the IUCN’s Global Invasive Species Database (<http://www.iucngisd.org/gisd/>). The final species richness totals thus obtained were XXX and XXX for the GCFR and SWAFR, respectively. Using R, the cleaned species occurrence record data were converted to QDS- and HDS-resolution raster-layers for use in subsequent analyses. In addition, following Whittaker’s (ref) original additive decomposition of *γ*-diversity, we decomposed the species richness of each HDS (*S*HDS) into its *α* (“plot” richness) and *β* (turnover) components, using the equation

(Equation 1)

where QDS is the average species richness of the four constituent QDS, and *T*QDS is the residual (i.e. turnover-based) richness, determined as *S­*HDS ­– QDS.

## 2.2 Comparing environmental heterogeneity

To compare environmental heterogeneity between the GCFR and SWAFR, we acquired a broad suite of geospatially-explicit environmental data in the form of raster layers. For the purpose of analysis, we then selected a subset of ten variables (Table 1) to represent topographic (elevation), climatic (surface T, MAP, PDQ), edaphic (clay content, soil C, pH, CEC) and vegetational (NDVI) gradients. As far as possible, these variables were selected to represent environmental axes which are considered regionally important. For example, the inclusion of PDQ in addition to MAP is justified on the basis that, where the letter captures variation in overall rainfall amount, the former measures the intensity of seasonal aridity which is a key feature of mediterranean-type climates (ref). Variable selection was, however, constrained by the availability of suitable raster-layers. Thus, although soil [P] is probably an important determinant of plant distribution in both the GCFR and SWAFR (refs), this variable could not be included owing to a lack of comparable data layers for the two regions. Indeed, wherever possible, we made use of satellite layers\*\*\*. Where soil variables were summarized as depth-interval weighted averages, climatic and spectral variables were summarized as annual means using the “raster” package for R (Hijmans, 2016). All layers were then projected to a common coordinate reference system (WGS84; ref) using the “rgdal” package (Bivand et al., 2017) and resampled to 0.05º resolution using the “resample” function in “raster,” with the “bilinear” method.

To quantify heterogeneity in the variables under study, we developed an index that would account for the spatial configuration of different environmental conditions at a range of scales. Our index, based on raster data, employs a moving-window of nine pixels, comprising a focal pixel and its eight neighbouring pixels. The heterogeneity (*EH*) of a pixel is determined as the mean magnitude of the difference between the value of the focal pixel (*x*\*) and those of its neighbour pixels (*x*1, *x*2, … *x*8) as follows:

(Equation 2)

We implemented this measure of heterogeneity using the “focal” function in the R package “raster” (Hijmans, 2016), applying it only to pixels having at least one neighbour pixel.

Since spatial scale is an important consideration in ecological studies (refs), environmental heterogeneity was considered at five spatial scales: 0.05º x 0.05º squares (the finest common resolution among the environmental data sources used), eighth degree squares (EDS), QDS, HDS and three-quarter degree squares (3QDS) (sensu Larsen et al., 2009). To do this, the absolute environmental raster data were first aggregated to broader spatial scales from 0.05º x 0.05º, using the “aggregate” function in “raster” (Hijmans, 2016), with the “mean” method. We then converted these to rasters of environmental heterogeneity using Equation 2.

Finally, we used principal components analysis (PCA), applied to the nine environmental variables across both regions, to derive a measure of overall environmental heterogeneity. For this purpose, the layers describing heterogeneity in the nine environmental variables at each spatial scale first log(*x* + 1) transformed to ensure normality and then subjected to PCA. A separate PCA was done for each spatial scale. The first axis (PC1) extracted from each of the five PCAs represents the major axis of heterogeneity across the nine environmental variables considered in this study.

We need here an explanation of the method used to compare heterogeneity of the nine variables and the major axis of heterogeneity between the two regions.

## 2.3 Environmental heterogeneity as an explanation of species richness

In the absence of strong non-linearity of the relationships between environmental heterogeneity (in the nine selected variables, and the major heterogeneity axis represented by PC1) and species richness at either the QDS- or HDS-scale, we used linear models, at both the QDS- and HDS-scales, to assess the explanatory power of environmental heterogeneity as a determinant of species richness across the two regions. These analyses made use of the species richness data collated at the QDS- and HDS-scales and measures of environmental heterogeneity determined at these same scales. Since the moving-window approach to determining heterogeneity (outlined above) quantifies the mean difference of a particular pixel (e.g. QDS or HDS) from its eight neighbours, rather than its internal heterogeneity, we developed a different measure of heterogeneity for the purpose of relating pixel species richness to internal pixel heterogeneity. Under this measure, each pixel is treated as comprising four subpixels (e.g. four QDS in an HDS; four EDS in a QDS) and the environmental heterogeneity (*EH*) is determined as the mean absolute difference between the value of each constituent subpixel and the other three subpixels within the more inclusive pixel. Specifically,

(Equation 3)

As for the moving-window measure of heterogeneity, the heterogeneities of the nine selected environmental variables were log(*x* + 1)-transformed and subjected to PCA to yield a major axis of environmental heterogeneity (PC1).

To test the dependence of species richness on environmental heterogeneity and to assess whether the form of this dependence is identical across the two regions, we fitted simple and multiple linear regression models specifying *S*HDS or *S*QDS as a function of environmental heterogeneity. In each instance, we fitted three models, an

In order to understand the relationships between environmental heterogeneity and species richness differ between the GCFR and SWAFR, we fit both simple and multiple linear regression models as functions of environmental heterogeneity values. The rationale of the univariate models was to describe empirical patterns of covariance between each axis of environmental heterogeneity and species richness. The multivariate models allow us to account for differences in richness across multiple axes of environmental heterogeneity simultaneously.

# 3 Results

## 3.1 Comparing species richness

Comparison of QDS- and HDS-scale species richness between the GCFR and SWAFR confirms the greater richness of the former,

Using Whittaker’s definition of additive turnover (ref), we partitioned *S*HDS into its - and -components (QDS and *T*QDS respectively). Using this method, we can see that almost all HDS in both the GCFR and SWAFR are composed of QDS that only account for no more than ca. 50% of *S*HDS (Figure 2a). After accounting for the generally greater *S*HDS in the GCFR (Figure 2b), *S*HDS is more attributable to floristic turnover in in the GCFR than it is in the SWAFR (Figure 2c).

## 3.2 Comparing environmental heterogeneity

Regressions of *CLES* against spatial scale identified the GCFR as being consistently more heterogeneous than the SWAFR for all nine environmental variables, across the full range of spatial scales studied (Fig. 1; Two-sided t test (mean *CLES* > 0.5): t = \*\*\*, df = \*\*\*, P < 0.001) [Across the five spatial scales, all *CLES*-values differed significantly from 0.5 following two-sided *t*-tests (*P* < 0.001).]. The same was true for the major axis of heterogeneity, described by PC1, which accounted for between 43.64% and 46.40% of the variance in all nine variables across the five spatial scales. In addition, for MAP, Surface T, CEC, Soil C and PC1, *CLES* was negatively related to spatial scale, indicating that the greater heterogeneity of the GCFR was especially pronounced at small spatial scales. The opposite was true of elevation and Clay, in which the greater heterogeneity of the GCFR was more pronounced at large spatial scales. NDVI and pH showed no scale-dependence.

## 3.3 Environmental heterogeneity as an explanation of species richness

We regressed vascular plant species richness (*S*) against each axis of environmental heterogeneity separately (Table 2, Figure 3) and in multivariate models (Figure 4) at both HDS- (Table 2a, Figure 4a) and QDS-scales (Table 2b, Figure 4b). Heterogeneity in elevation and mean annual surface temperature were consistently positively “main effect only” across scales when considered in univariate models (Table 2). This pattern is mirrored by the major axis of environmental heterogeneity: PC1 (Figure 3), though there is scale-dependence in the slope of that relationship. In a univariate context, all other environmental heterogeneity-variables showed varying degrees of covariance with *S* (Table 2).

Considering the multivariate models of *S*HDS (Figure 4a) and *S*QDS (Figure 4b), the importance of different axes of environmental heterogeneity varies between the HDS- and QDS-scales (Table 3). At the HDS-scale, the GCFR and SWAFR share no “common effects” of environmental heterogeneity on *S*, while at the QDS-scale the relationships between *S* and heterogeneity in elevation, MAP and CEC are common to the two regions.

At the HDS-scale (i.e. for *S*HDS; Table 2a), there is evidence for a difference in the slopes of the GCFR and SWAFR's relationships with heterogeneity in MAP. Heterogeneity in NDVI and clay only present evidence for the same slope in each region, but differing intercepts. Heterogeneity in CEC and pH have non-significant slopes and significant region-effects­­­—suggesting that these variables values' have weak relationships with *S*HDS, and that the region-effect explains more of the variance. Other variables (heterogeneity in elevation, PDQ, surface T and soil C) only present evidence for a continuous effect of that heterogeneity, explaining the difference in the regions' *S*HDS in terms of the roughness values themselves, without the need to invoke a region term. Think of it this way:

* If there is no need for any information concerning the region a cell belongs to, then the environmental roughness "rule" is followed well across the two regions in a similar way.
* If the region-effect is significant, but not the roughness effect, then that roughness axis isn't doing a very good job of explaining anything, and must defer to the region-effect.
* When both the region- and roughness-effect are significant, this represents a softer version of the above, where the roughness axis can explain some variance, but not all.
* When there is a significant interaction between region and roughness, then each region is playing a whole new game with that axes in terms of how richness is being driven.

At the QDS-scale (i.e. for *S*QDS; Table 2b), it is noteworthy that all axes best-supported to have an additive region term only also had non-significant roughness-effects [expand?].

I also regressed against PC1. Like heterogeneity in elevation and surface T, PC1 was the only explanatory variable "needed" in regressions for *S*HDS (also see Figure 3) and *S*QDS Figure 3 shows quite nicely how, in general, the GCFR and SWAFR are following the same "rule" (species richness increases with increasing environmental heterogeneity (PC1)) but occupy different areas along that relationship (the GCFR being more rich and more rough than the SWAFR).

# 4 Discussion

…

# Tables

Table 1: Georeferenced environmental data and vascular plant species occurrence data sources used in this study. Data were acquired for the Greater Cape Floristic Region and Southwest Australian Floristic Region, with the temporal extent of data products used described where applicable.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset(s) | Source | Temporal extent | Citation(s) |
| Plant species occurrences | GBIF |  | GBIF (2017a,b) |
| Elevation | SRTM (v2.0) |  | Farr et al. (2007) |
| NDVI | MODIS (MOD13C2) | Feb. 2000 to Apr. 2017 | NASA (2017a) |
| Surface T | MODIS (MOD11C3) | Feb. 2000 to Apr. 2017 | NASA (2017b) |
| MAP, PDQ | CHIRPS (v2.0) | Jan. 1981 to Feb. 2017 | Funk et al. (2015) |
| CEC, clay, soil C, pH | SoilGrids250m |  | Hengl et al. (2017) |

Abbreviations are as follows: NDVI, normalized difference vegetation index; T, temperature; MAP, mean annual precipitation; PDQ, precipitation in the driest quarter; CEC, cation exchange capacity; C, carbon.

Table 2 (next page): Results of univariate regressions of vascular plant species richness against different axes of environmental heterogeneity and overall environmental heterogeneity (PC1) across the Greater Cape Floristic Region and Southwest Australian Floristic Region (SWAFR), at both (a) HDS- and (b) QDS-scale. For each axis of environmental heterogeneity, we fit three univariate models: *S* as a function of environmental heterogeneity, *S* as a function of environmental heterogeneity with an additive term describing region and *S* as a of environmental heterogeneity with an interaction term for region. We used Akaike’s information criterion (*AIC*; ref) to select which of these three model types fit best for each environmental heterogeneity predictor variable. In each case, the best-fitting model (those presented) was selected as the simplest model with *∆AIC* < 2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Model term (signs and significances) | | | | | | |
| Response | Model type | Predictor | Slope |  | SWAFR effect |  | Slope:SWAFR |  |
| (a) *S*HDS | Main effect × region | MAP | + | \*\*\* | + |  | – | \*\* |
|  | Main effect + region | Clay | + | \* | – | \*\* |  |  |
|  |  | NDVI | + | \*\*\* | – | \* |  |  |
|  | Main effect only | Elevation | + | \*\*\* |  |  |  |  |
|  |  | PDQ | + | \*\*\* |  |  |  |  |
|  |  | Soil C | + | \*\*\* |  |  |  |  |
|  |  | Surface T | + | \*\*\* |  |  |  |  |
|  |  | PC1 | + | \*\*\* |  |  |  |  |
|  | Region only | CEC | – |  | – | \*\* |  |  |
|  |  | pH | + |  | – | \*\* |  |  |
| (b) *S*QDS | Main effect × region | NDVI | + | \*\*\* | – | \*\* | – | \*\*\* |
|  |  | PDQ | + | \*\*\* | + |  | + | \*\*\* |
|  |  | Soil C | + | \*\*\* | – |  | – | \*\* |
|  | Main effect only | Elevation | + | \*\*\* |  |  |  |  |
|  |  | MAP | + | \*\*\* |  |  |  |  |
|  |  | Surface T | + | \*\*\* |  |  |  |  |
|  |  | PC1 | + | \*\*\* |  |  |  |  |
|  | Region only | CEC | – |  | – | \*\*\* |  |  |
|  |  | Clay | + |  | – | \*\*\* |  |  |
|  |  | pH | – |  | – | \*\*\* |  |  |

Abbreviations are as in Table 1.

Table 3: Interpretation of region-specific scale-dependencies of the effects of different axes of environmental heterogeneity on vascular plant species richness (*S*) in the Greater Cape Floristic Region (GCFR) and Southwest Australian Floristic Region (SWAFR) (Figure 4). Here, positive scale-dependence (+) represents a greater magnitude of effect on *S* at broader spatial scales (i.e. at the HDS-scale), while negative scale-dependence (–) represents a greater magnitude at finer spatial scales (i.e. at the QDS-scale).

|  |  |  |
| --- | --- | --- |
|  | Axes of heterogeneity | |
| Scale-dependence | GCFR | SWAFR |
| + | Clay, pH |  |
| None |  | CEC |
| – | NDVI, soil C | PDQ |

Abbreviations of variables are as in Tables 1 and 2.

# Figures

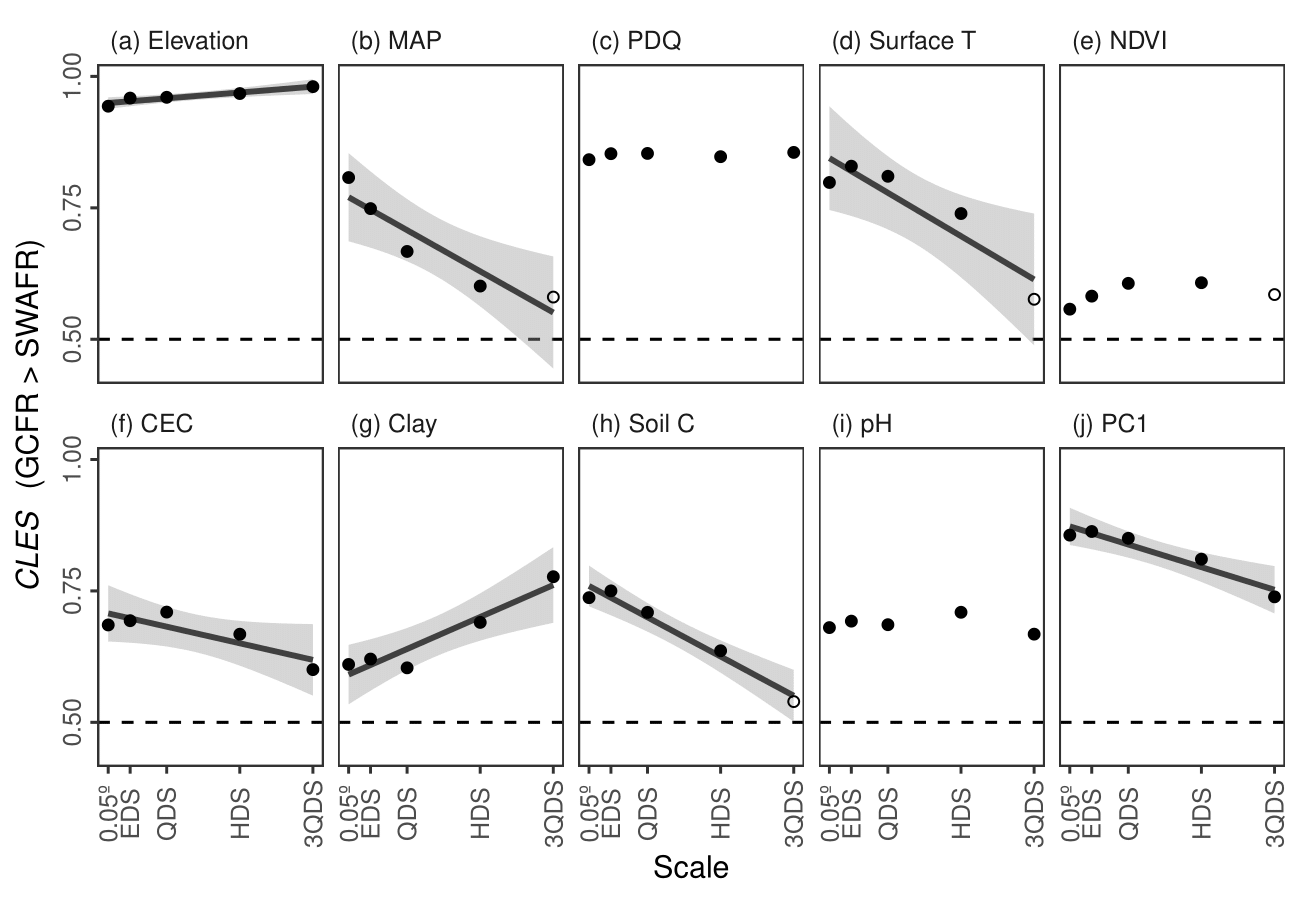


Figure 1: Simple linear regressions of the common language effect size (*CLES*; ref) of (a–i) various forms of environmental heterogeneity (*EH*; Equation 2) and (j) the first principal component of EH (PC1) in the Greater Cape Floristic Region (GCFR) and Southwest Australian Floristic Region (SWAFR). The *CLES* here is treated as the effect of GCFR relative to SWAFR values. Filled points represent comparisons where the GCFR and SWAFR significantly differed in *EH* (*P* ≤ 0.05, Mann-Whitney *U*-tests), while unfilled points represent those that were not significant (*P* > 0.05). Only significant (*P* ≤ 0.05) regression lines are plotted, with the exception of the fit for CEC, which was plotted in light of its marginal significance (*P* = 0.06). Grey bands denote 95% confidence intervals about the fitted lines. PC1 accounted for between 43.64 and 46.40% of the variation in *EH* across the five spatial scales at which it was calculated. Abbreviations of variables are as in Tables 1–3.

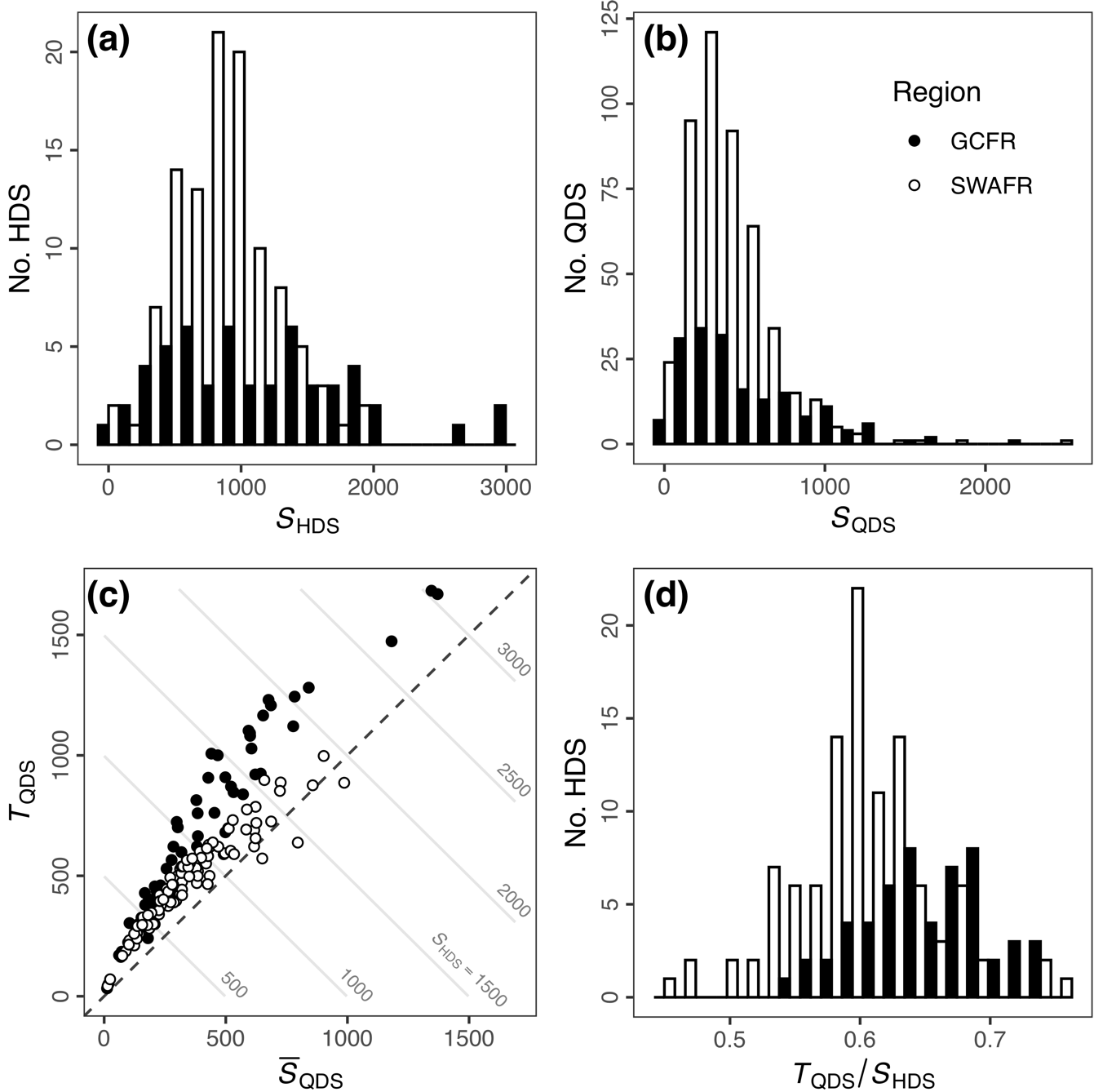


Figure 2: Distributions of (a) HDS- and (b) QDS-scale vascular plant species richness in the Greater Cape Floristic Region (GCFR) and Southwest Australian Floristic Region (SWAFR). (c) Scatter plot of mean QDS-scale richness (QDS) and turnover (*T*QDS; Equation 1) with contour lines denoting the *S*HDS that arises as their sum (i.e. increasing from lower-left to upper-right). (d) The distribution of the turnover partition of *S*HDS, expressed as a proportion (*T*QDS / *S*HDS). *P*-values inset (a,b,d) are from two-sided Mann-Whitney *U*-tests. Common language effect size (*CLES*) values inset (a,b,d) are for comparisons where GCFR values are greater than SWAFR values, as in Figure 1.

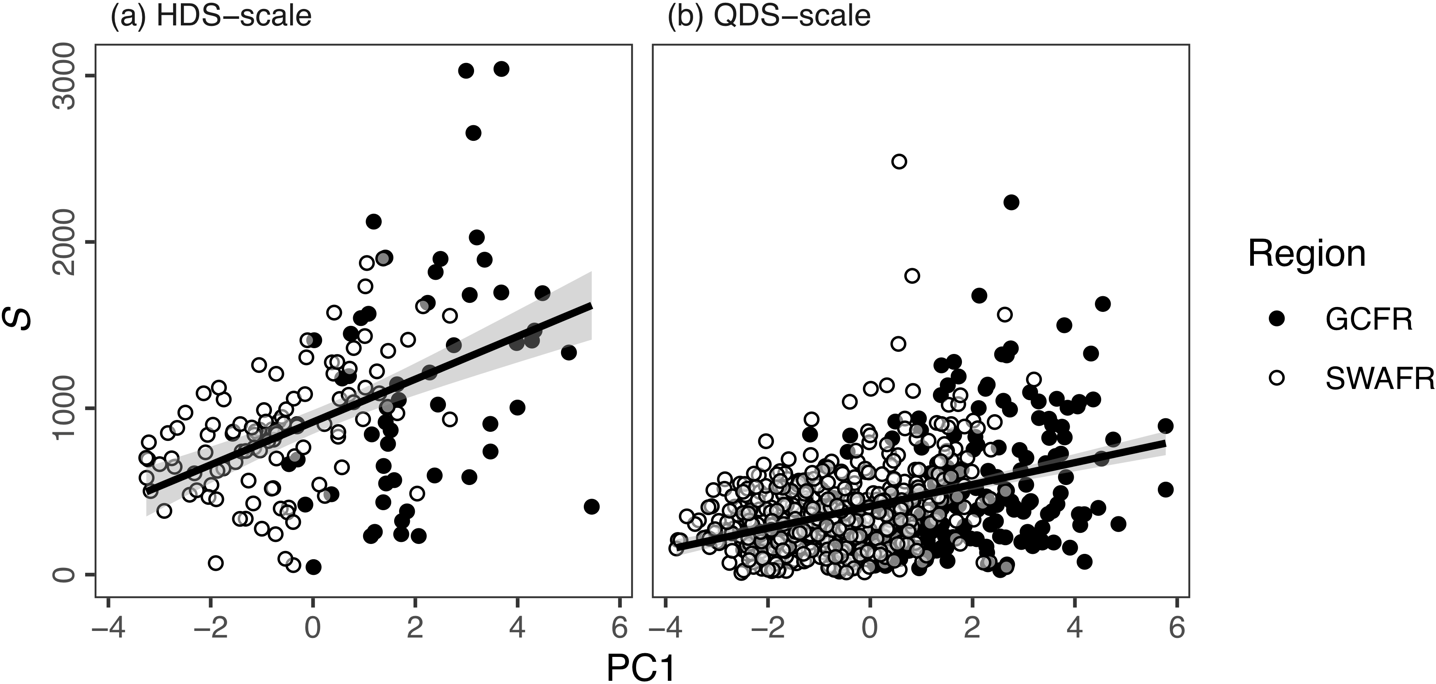


Figure 3: Simple linear regressions of vascular plant species richness as (a) *S*HDS (*R*2 = 0.23) and (b) *S*QDS (*R*2 = 0.15) against each respective scale’s first principle component (PC1) of environmental heterogeneity (*EH*; Equation 3) in the Greater Cape Floristic Region (GCFR) and Southwest Australian Floristic Region (SWAFR). Grey bands denote 95% confidence intervals about the fitted lines. When calculated at the QDS-scale, PC1 explained 39.86% of the variation in EH, while at the HDS-scale PC1 explained 41.55% of the variation in *EH*.

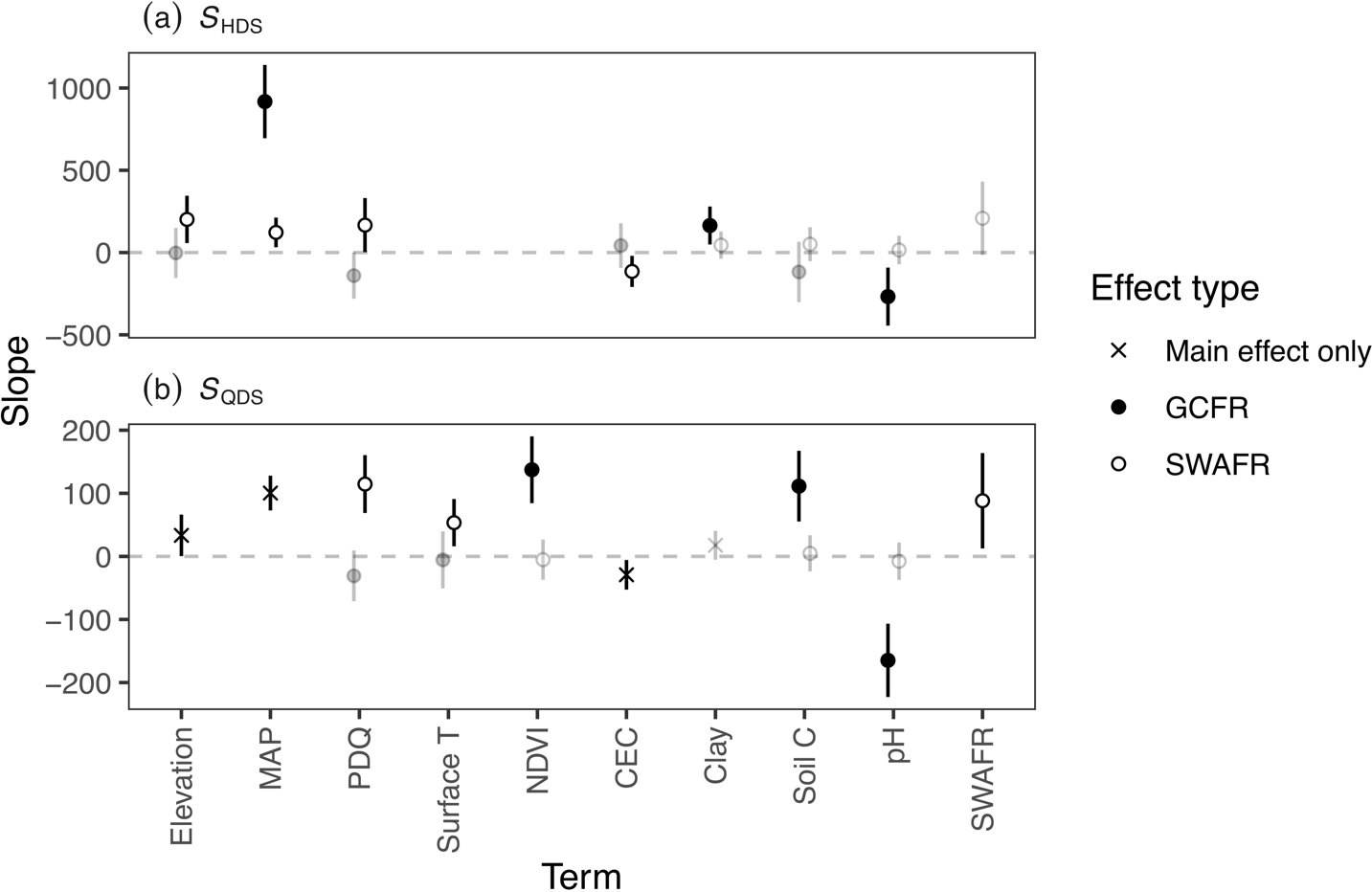


Figure 4: Slope estimates from multiple linear regressions of vascular plant species richness as (a) *S*HDS (*R*2adj = 0.49) and (b) *S*QDS (*R*2adj = 0.33) against the various forms of environmental heterogeneity across the Greater Cape Floristic Region (GCFR) and Southwest Australian Floristic Region (SWAFR). Each model was simplified, from a starting model with all predictors and their interactions with region, using reverse stepwise regression model selection based on *AIC*-scores in R. Points with error bars denote slope estimates and their 95% confidence intervals. Estimates illustrated in black were significant (*P* < 0.05), while those in grey were not, but still retained during stepwise model selection. Abbreviations of variables are as in Tables 1–3 and Figure 1.

# Data availability statement

…

# References

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

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

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

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

GBIF (24 July 2017a) GBIF Occurrence Download.

GBIF (24 July 2017b) GBIF Occurrence Download.

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

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

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

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

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

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

NASA (?) Vegetation indices monthly l3 global 0.05Deg cmg (mod13c2) v[Version]. *NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota*,.

NASA (?) Land surface temperature/emissivity monthly l3 global 0.05Deg cmg (mod11c3) v[Version]. *NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota*,.

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

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

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

# Biosketches

…

# Author contributions

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

# ORCID

RvM: 0000-0003-2659-6909

MDC: 0000-0003-0989-3266

GAV: 0000-0002-1363-9781