Similar axes of environmental heterogeneity drive plant species richness in two hyper-diverse 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:* biodiversity, environmental heterogeneity, fynbos, Greater Cape Floristic Region, kwongan, macroecology, species richness, species turnover, vascular plants, Southwest Australia Floristic Region

# 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 timeframes within which their biotas have assembled.

The floristically-rich South Western Australian Floristic Region (SWAFR; Hopper & 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 & 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 & Verboom, XXXX), 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 & 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 ca .7,380 species in an area of ca. 302,600 km2 (i.e. 0.024 species km-2; Hopper & Gioia, 2004), the GCFR accommodates ca. 11,430 species in an area of ca. 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 (ca. 9,400 species in ca. 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), half-degree square (HDS) and degree square (DS) scales (sensu Larsen et al., 2009), we first compare the distribution of species richness between the two regions, and in each region decompose broader-scale richness into average finer-scale richness and between-square 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 (ref) 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; ref?) and the Taxonomic Name Resolution Service (TNRS; ref?). 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.” 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/>). Finally, we removed species with fewer than five total collection records in total, in order to discount low-confidence collections [reword]. The final species richness totals thus obtained were 8,578 and 6,558 for the GCFR and SWAFR, respectively.

Using R (ref), these cleaned species occurrence record data were collated into QDS, HDS and DS pixels. To compare species richness across more equally sized areas, we only make comparisons between pixels consisting of all four sub-pixels (e.g. four QDS in an HDS). 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 equations

where QDS and HDS are the average species richness of the four constituent pixels in each HDS and DS respectively (i.e. *α*), and *T*QDS and *T*HDS represent the residual (i.e. turnover-based) richness *β*, determined each as *γ* − *α*.

## 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 nine 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 summarised as depth-interval weighted averages, climatic and spectral variables were summarised 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.

In order 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 nested pixels at various spatial scales. We treated environmental heterogeneity as the variance of the environmental conditions in the four sub-pixels for a given pixel. As such, we calculate heterogeneity based on the twentieth-degree squares’ (0.05°×0.05°), eighth-degree squares’, QDS’ and HDS’ environmental conditions within tenth-degree squares (0.10°×0.10°), QDS, HDS and DS respectively. We implemented this measure of heterogeneity using the R functions “var” and “aggregate” (the latter in the R package “raster” (Hijmans, 2016)). This index only uses neighbouring pixels to describe heterogeneity, similar to indices implemented in the “terrain” function in “raster”. However, our index describes heterogeneity within pixels as opposed to between pixels as in “terrain”. The former is comparable with species richness data and is thus used here.

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 were first log10-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 four PCAs represents the major axis of heterogeneity across the nine environmental variables considered in this study.

To compare the nine forms of environmental heterogeneity and the major axis of heterogeneity between the two regions, we employed common language effect sizes (*CLES*) using the R package “canprot” (ref). The *CLES* describes the proportion of pairwise comparisons, between two categories’ quantitative values, where one category’s values exceeds the other’s. Along with this appealing descriptive statistic, we tested for differences in regions’ heterogeneity values using two-sided Mann-Whitney *U*-tests (ref) in R. We performed the *CLES*-calculations and *U*-tests at the four spatial scales considered. The enabled us to assess scale-dependence in the calculation of heterogeneity and to ascertain the spatial scale (if any) at which environmental heterogeneity is [was?] most pronounced.

## 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 the QDS-, HDS- and DS-scales, we used linear models 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-, HDS- and DS-scales and measures of environmental heterogeneity determined at these same scales. 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, *S*QDS and *S*DS each as functions of environmental heterogeneity.

For the univariate regressions, we fitted three nested models for each of the nine axes of environmental heterogeneity and the major axis of heterogeneity. For each predictor variable *X*, we fit: a “main effect only” model (*S* ~ *β*0 + *β*1*X*), a “main effect + region” model (*S* ~ *β*0 + *β*1*X* + *β*2*Region*) and a model with an interaction term for region (“main effect × region”; (*S* ~ *β*0 + *β*1*X* + *β*2*Region* + *β*3(*X × Region*)). The best fitting of these three models for each of the ten predictor variables was determined using Akaike’s information criterion (*AIC*; ref), such that the selected model was the simplest model with *∆AIC* < 2 (ref).Using this ANCOVA-like approach, we assess when each form of heterogeneity poorly predicts species richness across the two regions (i.e. when a “main effect + region” model is best fitting but with little support for the main effect), when there is a common relationship between a form of heterogeneity in both regions (i.e. when a “main effect only” or a “main effect + region” model is best fitting) or when species richness in each region relates differently to a form of heterogeneity (i.e. when a “main effect × region” model is best fitting). [reword this explanation/para!]

While 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. Species richness (log10-transformed as above) was fit as a function of all forms of environmental heterogeneity and the interaction of each with region [reword] at all three spatial scales. Each of these three starting models was simplified using reverse stepwise regression model selection, based on *AIC*-scores in R. The resulting three models represent the best-fitting models describing forms of heterogeneity with significant covariance with species richness. In addition, forms of heterogeneity that relate to richness differently across the two regions are identified when the interaction-term between that variable and region is retained during model simplification (akin to the ANCOVA-like approach used in the univariate models of species richness, described above). [reword this explanation/para!]

# 3: Results

## 3.1: Comparing species richness

Our vascular plant species richness datasets for the GCFR and SWAFR (Figure 5a,b) are consistent with the knowledge that species richness is unevenly distributed. Comparisons of the QDS- and HDS-scale vascular plant species richness of the GCFR and SWAFR confirms the greater richness of the former (Figure 2a,b; *P* < 0.05 at all scales, two-sided Mann-Whitney *U*-tests). Although, there are QDS in the SWAFR as rich, if not richer than in GCFR (and likewise for HDS). Calculating the *CLES* of GCFR vs SWAFR richness demonstrates that the GCFR is pronouncedly richer than the SWAFR at coarser spatial scales (for *S*QDS: *CLES* = 0.516, *S*HDS: 0.542, *S*DS: 0.658 (see SI)).

We partitioned *S*HDS into its - and -components (QDS and *T*QDS respectively) and demonstrate that most HDS and DS in both the GCFR and SWAFR are composed of QDS and HDS, respectively, that account for about 60% of *S*HDS (Figure 2c) and SDS (see SI). At the HDS-scale, few QDS can account for more than 50% of *S*HDS (Figure 2c). After accounting for the generally greater *S*HDS in the GCFR (Figure 2b), *S*HDS is more attributable to floristic turnover in the GCFR than it is in the SWAFR (Figure 2d; for *T*QDS / QDS: *CLES* = 0.696, *T*HDS / HDS: 0.741 (see SI)).

## 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 (Figure 1). The same was true for the major axis of heterogeneity, described by PC1 (Figure 5c,d), which accounted for between 38.19% (at the tenth-degree scale) and 42.44% (at the QDS-scale) of the variance in all nine variables across the four spatial scales (Figure 1j). Edaphic axes were more heterogeneous in the GCFR, though less severely than other forms of environmental heterogeneity considered (Figure 1f–i).

The degree to which the GCFR is more environmentally heterogeneous than the SWAFR is largely scale-independent, with the notable exceptions of MAP (Figure 1b), in which the GCFR and SWAFR are more similarly heterogeneous at coarser spatial scales, and NDVI and clay (Figure 1e,g), which are more pronouncedly heterogeneous in the GCFR than in SWAFR at coarser spatial scales. Indeed, some heterogeneity variables (MAP, NDVI and CEC; Figure 1b,e,f) did not differ significantly between the two regions at DS-scales (*P* > 0.05; two-sided Mann-Whitney *U*-tests). The scale-independence of environmental heterogeneity, despite the scale-dependence in a few variables, is reflected in the major axis of heterogeneity (PC1): the *CLES* of GCFR vs SWAFR PC1 is relatively constant across spatial scales scale (Figure 1j).

## 3.3: Environmental heterogeneity as an explanation of species richness

We regressed vascular plant species richness against each axis of environmental heterogeneity (Table 2) and the major axis of heterogeneity (PC1) separately (Table 2, Figure 3). Insofar as PC1 describes much of the variation in environmental heterogeneity (ca. 39–42%), we found strong support for a common relationship between species richness and environmental heterogeneity at HDS- and DS-scales (Figure 3b,c). The GCFR and SWAFR pixels occupy different areas along this relationship [reword?]. At QDS-scales, however, there was evidence for differences in the slopes and intercepts between the GCFR and SWAFR for the regressions against PC1 (Figure 3a). It is noteworthy that the differences in these slopes is small (24.61 species per unit PC1, *P* = 0.034) relative to the variation in species richness observed across QDS (*SD* = 304.75 species).

Considering each axis of environmental heterogeneity separately, we found strong support for a common set of axes of environmental heterogeneity positively associated with species richness across the GCFR and SWAFR at HDS- and DS-scales (Table 2b,c). This reflects the results for the regressions against PC1. Indeed, at QDS-scales, there was evidence for differences in the slopes and intercepts for the regressions of different forms heterogeneity between the GCFR and SWAFR (Table 2a).

We also regressed vascular plant species richness against each axis of environmental heterogeneity in multivariate models (Figure 4). Many of the effects of the different axes of heterogeneity on species richness have the same signs across spatial scales (Figure 4a–c, from QDS- to DS-scale respectively). The two regions’ partial effects for each form of heterogeneity, where they are distinct model terms, are either: always of the same sign (e.g. heterogeneity in MAP; Figure 4a,b), or one region has a significant effect where the other does not (e.g. heterogeneity in PDQ; Figure 4a–c). Thus, regions’ various partial effects were not found to have opposing signs.

From these multivariate models, the estimated difference in species richness between GCFR and SWAFR pixels (“SWAFR”-term; Figure 4) is conditional on equality in all forms of environmental heterogeneity in that model. At the QDS-scale, the SWAFR is more species rich for a given level of heterogeneity across all the axes in that model (estimated 104.07 more species, *P* < 0.001; Figure 4a), while at the DS-scale the GCFR is more rich for given levels of heterogeneity (estimated 1112.54 more species, *P* = 0.002 Figure 4c). At the HDS-scale, however, the regions are similarly species rich, holding all else constant (*P* = 0.58; Figure 4b).

Both the regressions against PC1 and the multivariate regressions underpredict species richness in areas of observed high richness and overpredict in areas that are relatively species poor (Figure 5e–h), failing to explain the most of variation of species richness (PC1 regressions: *R*­2 = 0.14–0.28; multivariate regressions: 0.24–0.61)—with the exception of the multivariate model at the DS-scale (*R*2 = 0.61; Figure 4c). Notably, the PC1 regressions and multivariate regressions are remarkably similar in their predictions of species richness (Table 3). This further supports PC1 as a meaningful index of environmental heterogeneity and the common relationship between heterogeneity and species richness.

# 4: Discussion

Outline:

1. Species richness:
   * GCFR > SWAFR on average
   * GCFR >>> SWAFR broader spatial scales
     + Likely due to greater floristic turnover between GCFR pixels
   * Though, there is great variation in pixels’ richness over space! (ref map)
2. Environmental heterogeneity:
   * GCFR > SWAFR, generally
   * GCFR >/≈ SWAFR at broader spatial scales
   * GCFR >/≈ SWAFR along edaphic axes
3. SR ~ EH:
   * Scale-independent relationship, generally
   * GCFR and SWAFR have broad similarities—a “common relationship” …
4. … though there are regional “idiosyncrasies”—region-specific effecrs
   * …
5. PC1 is *good* (Re: correlation of residuals etc. between PC1-models
6. Heterogeneity-determinism-unexplained richness = history
   * Cf. absolute environmental conditions (ref map)
   * Re: drought refugia?

# Tables

**Table 1:** Georeferenced environmental data1 and vascular plant species occurrence data sources used in this study. Data were acquired for the GCFR and SWAFR, 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) |

1 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:** Summarised results of univariate regressions of vascular plant species richness against different axes of environmental heterogeneity1 (log10-transformed) and overall environmental heterogeneity (PC1) across the GCFR and SWAFR at the (a) QDS-, (b) HDS- and (c) DS-scales. The signs (+, −) of the heterogeneity variables’ slope terms and the SWAFR term (where applicable) are presented alongside their significances2.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Response | Model type | Predictor | Main effect | | SWAFR effect | | SWAFR interaction | |
| (a) *S*QDS | Main effect × region | MAP | + | \*\*\* | + | \*\* | − | \*\*\* |
|  |  | NDVI | + | \*\*\* | − | \* | − | \*\*\* |
|  |  | Soil C | + | \*\*\* | − |  | − | \*\* |
|  |  | PC1 | + | \*\*\* | + | \*\*\* | − | \* |
|  | Main effect + region | Elevation | + | \*\*\* | + | \*\* |  |  |
|  |  | PDQ | + | \*\*\* | + | \* |  |  |
|  |  | CEC | + | \* | − | \* |  |  |
|  |  | Clay | + | \*\* | − | \* |  |  |
|  | Main effect only | Surface T | + | \*\*\* |  |  |  |  |
|  | Region only | pH | + | ~ | − | \* |  |  |
| (b) *S*HDS | Main effect only | Elevation | + | \*\* |  |  |  |  |
|  |  | MAP | + | \*\*\* |  |  |  |  |
|  |  | PDQ | + | \*\*\* |  |  |  |  |
|  |  | Surface T | + | \*\*\* |  |  |  |  |
|  |  | NDVI | + | \*\*\* |  |  |  |  |
|  |  | Clay | + | \*\*\* |  |  |  |  |
|  |  | Soil C | + | \*\*\* |  |  |  |  |
|  |  | pH | + |  |  |  |  |  |
|  |  | PC1 | + | \*\*\* |  |  |  |  |
|  | Region only | CEC | − |  | − | \* |  |  |
| (c) *S*DS | Main effect only | Elevation | + | \*\* |  |  |  |  |
|  |  | MAP | + | \*\*\* |  |  |  |  |
|  |  | PDQ | + | \*\*\* |  |  |  |  |
|  |  | NDVI | + | \*\* |  |  |  |  |
|  |  | Clay | + | \*\*\* |  |  |  |  |
|  |  | PC1 | + | \*\*\* |  |  |  |  |
|  | Region only | Surface T | + | ~ | − | \* |  |  |
|  |  | CEC | − |  | − | \* |  |  |
|  |  | Soil C | + |  | − | \* |  |  |
|  |  | pH | + |  | − | \* |  |  |

1 Abbreviations are as in Table 1.

2 Represented as follows: \*\*\*, *P* < 0.001; \*\*, *P* < 0.01; \*, *P* < 0.05; ~, *P* < 0.1; blank, NS.

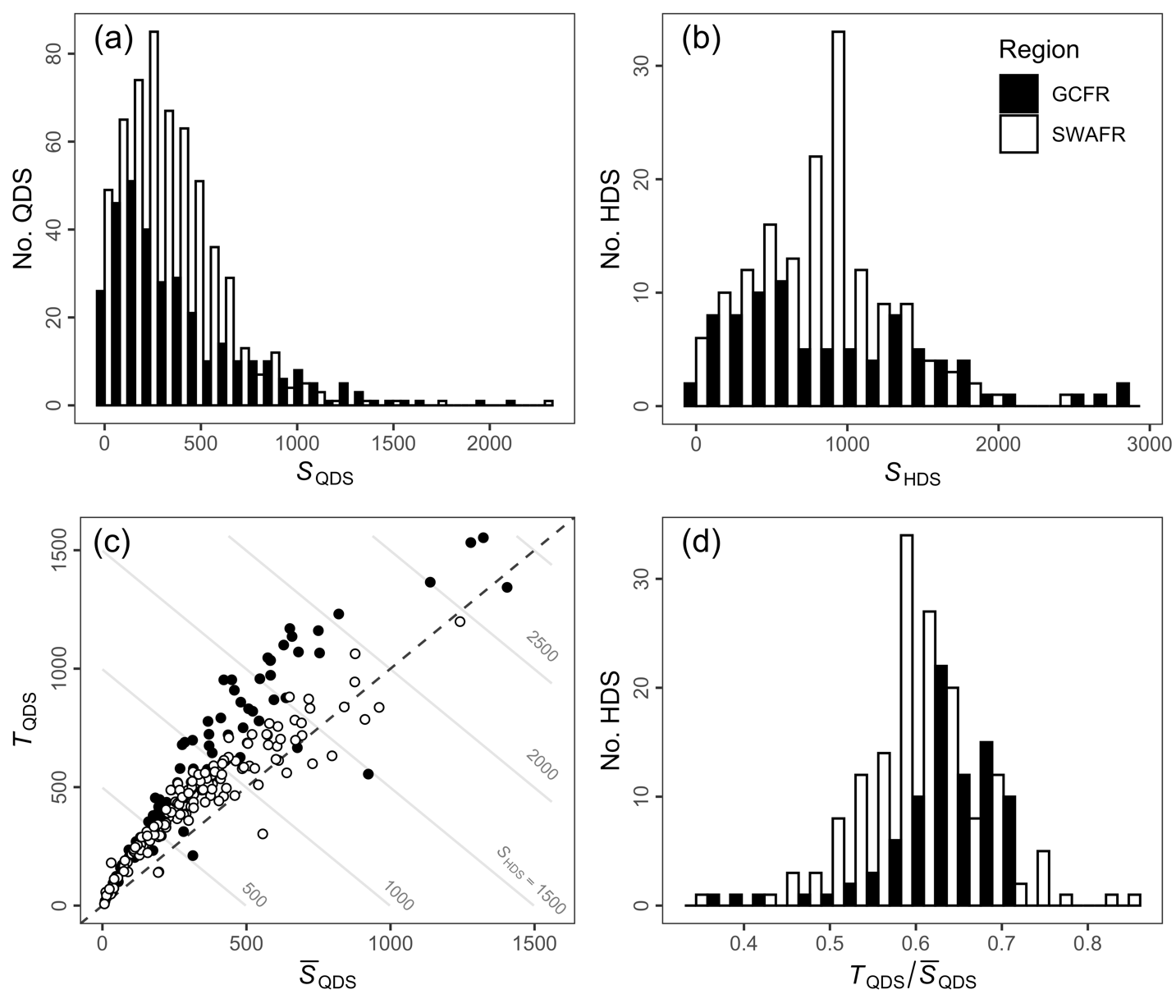
**Table 3:** Pearson’s *r* (ref) correlation coefficients comparing the predicted and residual species richness (*S*) between regressions of *S* against the major axis of environmental heterogeneity (PC1; Table 2; Figure 3) and the multivariate regressions (Figure 4). All correlation coefficients were significant (*P* << 0.001; two-sided *t*-test).

|  |  |  |
| --- | --- | --- |
|  | Correlation | |
| Spatial scale | Predicted *S* | Residual *S* |
| QDS | 0.743 | 0.934 |
| HDS | 0.711 | 0.878 |
| DS | 0.638 | 0.656 |

# Figures



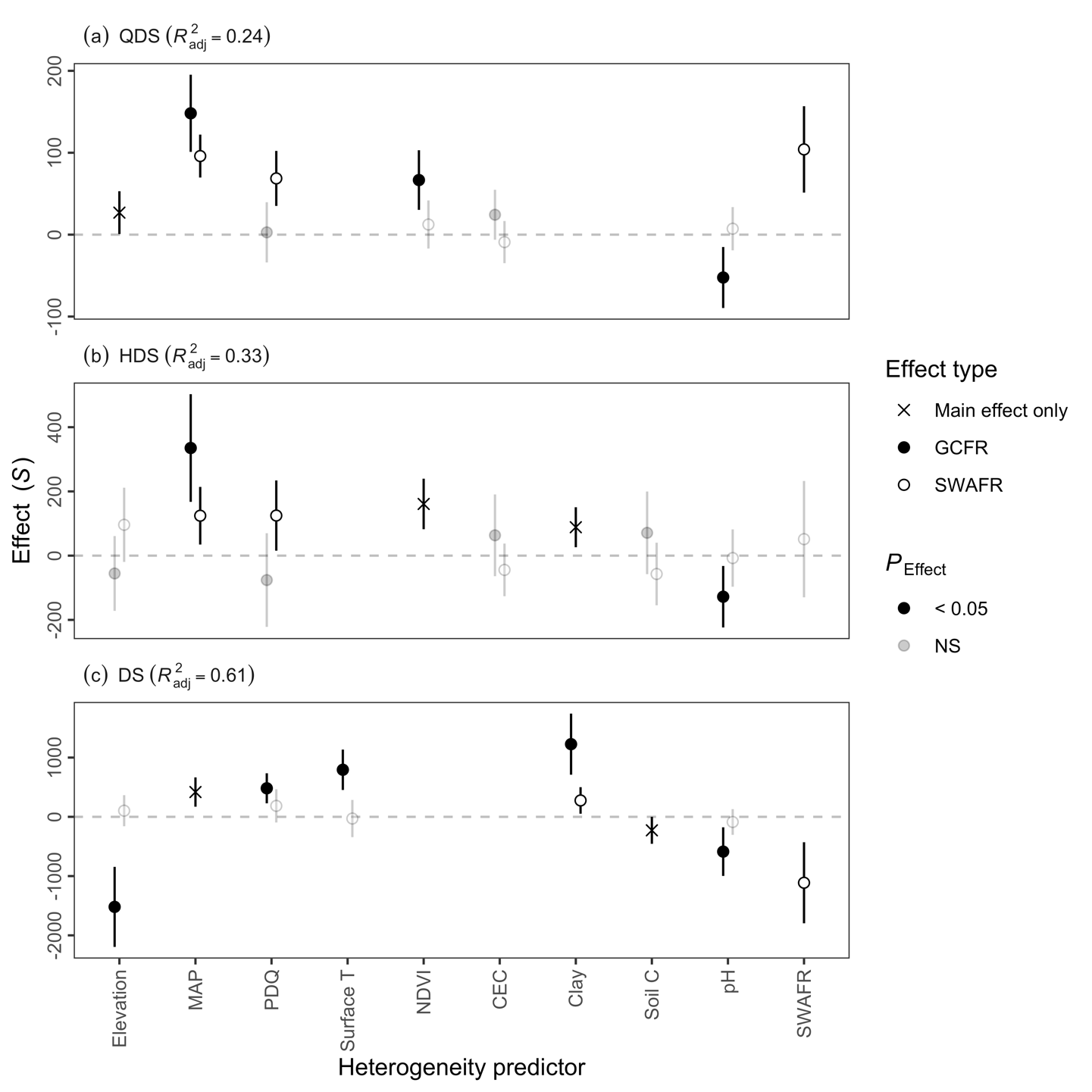
**Figure 1:** The common language effect size (*CLES*) of (a–i) various forms of environmental heterogeneity (log10-transformed) and (j) the major axis thereof (PC1) in the GCFR and SWAFR. *CLES* here is treated as the effect of GCFR relative to SWAFR values. Filled points represent comparisons where the GCFR and SWAFR significantly differed in heterogeneity (*P* < 0.05; two-sided Mann-Whitney *U*-tests), while unfilled points represent those that were not significant. Following simple linear regressions of *CLES* against scale, we found some evidence for negative relationships (depicted by lines) for MAP (slope = –0.157; *P* = 0.098), NDVI (slope = 0.116; *P* < 0.001) and Clay (slope = 0.158, *P* = 0.037). Abbreviations are as in Tables 1–2.



**Figure 2:** Distributions of (a) QDS- and (b) HDS-scale vascular plant species richness in the GCFR and SWAFR. (c) Scatter plot of mean QDS-scale richness (QDS) and turnover (*T*QDS) with contour lines denoting the *S*HDS that arises as their sum. (d) The distribution of the turnover partition of *S*HDS (*T*QDS; in c) expressed as a proportion (*T*QDS / QDS).



**Figure 3:** Simple linear regressions of vascular plant species richness as (a) *S*QDS, (b) *S*HDS and (c) *S*DS against each respective scale’s major axis of environmental heterogeneity (PC1) across the GCFR and SWAFR. These three linear models are presented in Table 2, all with highly significant slopes (*P* < 0.001). For *S*QDS, the separate fits for the GCFR (grey) and SWAFR (black) are presented, following the best fitting model at that scale (see Table 2a). The *R*2-values of each model and the variation in environmental heterogeneity explained by PC1 from each of the three PCAs are noted in parentheses in the panel and horizontal axis headings respectively.



**Figure 4:** Slope estimates from multiple linear regressions of vascular plant species richness as (a) *S*QDS, (b) *S*HDS and (c) *S*DS against the various forms of environmental heterogeneity (log10-transformed and scaled) across GCFR and SWAFR. Points with error bars denote partial effect estimates and their 95% confidence intervals. Filled and empty points represent effect estimates for the GCFR and SWAFR respectively when region-interaction terms were retained during stepwise model selection, while crosses represent main effects (i.e. no region-interaction term retained). Estimates illustrated in black were significant (*P* < 0.05), while those in grey were not, but still retained during stepwise model selection. The multiple adjusted *R*2-values of each model are noted in parentheses in the panel headings. Abbreviations of variables are as in Tables 1–2 and Figure 1.

A screenshot of a map

Description automatically generated

**Figure 5 (previous page):** HDS-scale maps for the GCFR and SWAFR of (a,b) vascular plant species richness, (c,d) the major axis of environmental heterogeneity (PC1) from the PCA of nine forms of environmental heterogeneity (log10-transformed), residuals from regressions of species richness against (e,f) PC1 (Figure 3b) and (g,h) the simplified multivariate (MV) model (Figure 4b). Map projection used: WGS84 (ref). QDS- and DS-scale equivalents of these maps are available in the online version (see SI).

# Data availability statement

…

# ORCID

Ruan van Mazijk: <https://orcid.org/0000-0003-2659-6909>

Michael D. Cramer: <https://orcid.org/0000-0003-0989-3266>

G. Anthony Verboom: <https://orcid.org/0000-0002-1363-9781>

# 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

**Ruan van Mazijk** is […]

**Michael D. Cramer** is an ecophysiologist interested in ecophysiological specialization in the hyper‐diverse Cape flora and the link between nutrient‐impoverished soils and species richness.

**G. Anthony Verboom** works on the assembly of the hyper‐diverse Cape flora, its vegetation organization and the role of speciation and extinction in its radiation.

# Author contributions

MDC and GAV conceived the study question, which RvM investigated and developed under their supervision for his BSc Hons project. RvM collated the data and carried out the GIS work. All authors contributed to the the analyses, which were then carried out by RvM, who wrote the first draft of the manuscript. All authors contributed equally to the writing thereafter.