Similar axes of environmental heterogeneity associated with 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

RvM is grateful to the National Research Foundation and the South African Association of Botanists for bursaries during this work.

# Abstract

**Aim:** To assess whether the difference in species richness per unit area between two mediterranean-type biodiversity hotspots relates to differences in environmental heterogeneity.

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

**Taxon:** Vascular plants (Tracheophytes)

**Methods:** Comparable geospatially-explicit environmental and species occurrence data were obtained for both regions, from which we generated environmental heterogeneity and species richness raster layers. We compare the degrees of various forms of environmental heterogeneity and species richness per unit area between the regions, across a range of spatial scales. We regressed species richness against indices of overall environmental heterogeneity, derived from principal components analyses (PCAs), and in multivariate regressions against the various forms of heterogeneity.

**Results:** The GCFR is generally more environmentally heterogeneous and more species-rich than the SWAFR. The major axes of heterogeneity across both regions (explaining ca. 38–42% variation in heterogeneity, depending on spatial-scale) relates significantly to species richness, with the same slope for each region at broader spatial-scales (*P* < 0.001). Multivariate regressions, and the variation in heterogeneity undescribed by the first axis from the PCAs, reveal axes of environmental heterogeneity associated with species richness with differing strengths in each region.

**Main conclusions:** We have evidence for a common positive relationship between floristic richness and environmental heterogeneity across the GCFR and SWAFR, broadly independent of spatial-scale. Though there are region-specific effects, […]

*Keywords:* biodiversity, environmental heterogeneity, fynbos, Greater Cape Floristic Region, kwongan, macroecology, species richness, species turnover, vascular plants, Southwest Australian Floristic Region

# 1: Introduction

The species richness of a region is a function of its biogeographic context (e.g. proximity to potential immigrant sources), its diversification history and any locally-deterministic, environmental features (e.g. productivity, heterogeneity) that influence species persistence and coexistence (Ricklefs, 1987,2004; Bøhn & Amundsen, 2004). Since all three effects are potentially influenced by environmental heterogeneity, the latter may be a particularly important driver of regional species richness variation (refs), with physically-heterogeneous regions tending to be more 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, enhancing 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, Linder, & Desmet, 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, Linder, Rommerskirchen, & Schefuss, 2011; Hoffmann, Verboom, & Cotterill, 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 old, climatically-buffered infertile landscapes (OCBILs; Hopper, 2009). Owing to these environmental similarities, the SWAFR and GCFR floras are very similar with respect to their plant functional trait spectra (Cowling & Witkowski, 1994), although 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 is home to 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 unit 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 unit 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, Holmern, Prager, Maliti, & Røskaft, 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 each region were obtained from the Global Biodiversity Information Facility (GBIF; 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 (Gioia & Hopper, 2017; Hopper & Gioia, 2004). The downloaded occurrence data were then cleaned using R (ref) and the “taxize” package (Chamberlain et al., 2016) (SI). Despite spatial variability in collection effort in both regions, we treat the species richness data collated here as representative of real patterns in these flora, as concluded by Cramer & Verboom (2016). The final number of unique species in these occurrence data obtained were 8,578 and 6,558 for the GCFR and SWAFR, respectively.

Using R, these cleaned species occurrence record data were collated into QDS, HDS and DS. To compare species richness across equally sized areas, we only made comparisons between squares consisting of all four sub-squares (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) and DS (*S*DS) 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 squares 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 and nominally independent. For example, the inclusion of PDQ in addition to MAP is justified on the basis that, where the latter captures variation in overall rainfall amount, the former measures the intensity of seasonal aridity–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 remote sensing derived layers. Where soil variables were summarised as depth-interval weighted averages, climatic and spectral variables were summarised as annual means, both using the “raster” package for R (Hijmans, 2016). All layers were then projected to a common coordinate reference system (WGS84) using the “rgdal” package (Bivand, Keitt, & Rowlingson, 2017) and resampled to 0.05º resolution using the “resample” function in “raster,” with the “bilinear” method.

In order to quantify heterogeneity in these environmental variables, we developed an index that would account for the spatial configuration of environmental conditions. Our index, based on raster data, employs nested squares at various spatial scales. We treated environmental heterogeneity as the variance of the environmental conditions in the four sub-squares for a given square. We calculated heterogeneity at the tenth-degree-square- (0.10°×0.10°), QDS-, HDS- and DS-scales (thus based on the twentieth-degree squares’ (0.05°×0.05°), eighth-degree squares’, QDS’ and HDS’ environmental conditions). We implemented this measure of heterogeneity using the “aggregate” function in the R package “raster” (Hijmans, 2016) with variance as the aggregation function. This index only uses neighbouring squares to describe heterogeneity, similar to indices implemented in the “terrain” function in “raster”. However, our index describes heterogeneity within squares, as opposed to between squares as in “terrain”. Our species richness dataset is based on within-square data (species occurrences), making it comparable with this heterogeneity data.

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. 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 heterogeneity variables considered.

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’ values where one category’s values exceed the other’s. Additionally, 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. This enabled us to assess scale-dependence in heterogeneity and to ascertain the spatial scale (if any) at which environmental heterogeneity is 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 from the PCA to describe empirical patterns of covariance between axes of environmental heterogeneity and species richness. For each heterogeneity 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 assessed when each form of heterogeneity poorly predicted 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).

The multiple linear regressions allowed us to account for differences in richness across multiple axes of environmental heterogeneity simultaneously. Species richness was regressed against the nine separate forms of environmental heterogeneity, including the interactions of each with a region-term. This was repeated at the three spatial scales. These three models were then simplified using reverse stepwise regression model selection (ref), based on *AIC*-scores in R. The final regressions represent the best-fitting models describing forms of heterogeneity with significant associations with species richness. Additionally, forms of heterogeneity that relate to richness differently across the two regions are identified when the interaction-term between that variable and the region-term was retained during model simplification.

# 3: Results

## 3.1: Comparing species richness

Vascular plant species richness is spatially variable in both of the GCFR and SWAFR (Figure 5a,b). Comparisons of the QDS- and HDS-scale species richness of each region using two-sided Mann-Whitney *U*-tests show similar species richness per unit area in both regions at the QDS- (*P* = 0.402, *CLES* = 0.516; Figure 2a) and HDS-scales (*P* = 0.275, *CLES* = 0.542; Figure 2b), while the GCFR is significantly more species-rich than the SWAFR at the DS-scale (*P* = 0.038, *CLES* = 0.658; SI, Figure S1a).

We partitioned *S*HDS into its *α*- and *β*-components (QDS and *T*QDS respectively; Figure 2c) and demonstrate that *T*QDS in both regions accounts for at least 60% of *S*HDS in most HDS (Figure 2c,d). *S*HDS is more attributable to floristic turnover in the GCFR than it is in the SWAFR (*P* < 0.001; *CLES* = 0.696; Figure 2d). Likewise, at the DS-scale (SI, Figure S1b), floristic turnover (*T*HDS) accounts for ca. 55% and ca. 48% of *S*DS on average in the GCFR and SWAFR respectively. with GCFR *S*DS being significantly more turnover-driven than in the SWAFR in this regard (*P* = 0.001, *CLES* = 0.741; SI, Figure S1b).

## 3.2: Comparing environmental heterogeneity

Regressions of *CLES* against spatial scale identified the GCFR as being consistently more (or at least equally) 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 1j; 5c,d), which accounted for between 38% (at the tenth-degree scale) and 42% (at the QDS-scale) of the variance in all nine heterogeneity variables. There is a greater disparity in topographic and climatic heterogeneity between the GCFR and SWAFR (all *CLES* > 0.60; Figure 1a–d) than there is in edaphic heterogeneity (all *CLES* < 0.75; 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 somewhat more 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 major axis of heterogeneity (PC1) reflects the scale-independence of most forms of heterogeneity—the *CLES* of GCFR vs SWAFR PC1 is relatively constant across spatial scales scale (Figure 1j)—despite the scale-dependence exhibited by a few variables.

## 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; Table 2, Figure 3) separately. At QDS-scales, however, there was evidence for differences in the slopes and intercepts of this relationship between the regions (Figure 3a). Although, the difference 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). Insofar as PC1 describes much of the variation in environmental heterogeneity (ca. 38–42%), we found evidence for a common relationship between species richness and environmental heterogeneity at HDS- and DS-scales (Figure 3b,c), with GCFR and SWAFR squares simply occupying different areas along this relationship. This covariation of richness with overall heterogeneity is consistent with results considering each axis of environmental heterogeneity separately (Table 2). At QDS-scales, there was also evidence for differences in regions’ slopes and intercepts for these relationships (Table 2a), while we found a common set of axes of environmental heterogeneity that were positively associated with species richness in both the GCFR and SWAFR at HDS- and DS-scales (Table 2b,c).

We also regressed vascular plant species richness against each axis of environmental heterogeneity in multiple linear regressions (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). In cases where the partial effect of a form of heterogeneity was represented in a model separately for each region, the coefficients are always of the same sign (e.g. heterogeneity in MAP; Figure 4a,b) or one region’s coefficient is not significant where the other’s is (e.g. heterogeneity in PDQ; Figure 4a–c). Thus, regions’ various partial effects were never found to have opposing signs. From these multivariate models, the estimated difference in species richness between GCFR and SWAFR squares (“SWAFR”-term; Figure 4) is conditional on all forms of environmental heterogeneity used in that model being constant. At the QDS-scale, for given levels of heterogeneity, the SWAFR is more species-rich (estimated 104.07 more species, *P* < 0.001; Figure 4a) while the partial-differences in richness estimated at other scales reflect observed differences in species richness: similarly rich at HDS-scales (*P* = 0.58; Figure 4b); GCFR more rich at DS-scales (estimated 1112.54 more species, *P* = 0.002; Figure 4c).

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 most of the variation of species richness (PC1 regressions: *R*­2 = 0.14–0.28; multivariate regressions: *R*­2 = 0.24–0.61; Figure 3,4)—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 (Figure 5e–h; Table 3). This further supports PC1 as a meaningful index of environmental heterogeneity and the common relationship between heterogeneity and species richness across the GCFR and SWAFR.

# 4: Discussion

## Paragraph 1: Interpretation & generalization

Broadly, we found support for the hypothesis that differences in the observed species richness (per unit area) between the GCFR and SWAFR are associated with differences in these regions’ environmental heterogeneity, over a variety of environmental axes. We thus conclude there to be an underlying common relationship between species richness and heterogeneity in these two floras.

Extent vs grain of *S* ~ heterogeneity rule: does it break down beyond MTEs? Kreft & Jetz study was larger grain AND larger extent? How do these patterns (S~H) play out over global scales (in terms of both extent & grain)?

## Paragraph 2: Idiosyncrasies & scale

There are, however, regional idiosyncrasies to the richness-heterogeneity relationship (Figure 4), including evidence for a small difference in the slope between species richness and the major axis of environmental heterogeneity (PC1) at finer spatial-scales (Figure 3a).

As spatial-scale is increased from fine to broad (i.e. from QDS- to DS-scales), the GCFR

The greater species richness per unit area in the GCFR at broader spatial-scales is attributable to the greater turnover-partition of squares

The greater floristic turnover exhibited within GCFR squares (compared to the SWAFR) at HDS- ands DS-scales makes sense in light of the greater richness of these same squares and the greater heterogeneity in the GCFR across scales.

[Greater disparity in topographic and climatic heterogeneity than in edaphic heterogeneity between the regions]

[SWAFR richness > than GCFR | heterogeneity (Figure 4)]

[…]

## Paragraph 3: History & outliers

Heterogeneity-determinism-unexplained richness = history

* Cf. absolute environmental conditions (map?)
* Re: drought refugia?
* Re: Linder 2009 Restionaceae paper
  + Distance from the Kogelberg
    - S ~ DK vs ResPC1 ~ DK
  + Is there a SWAFR equivalent? Likely not…

## Paragraph 4: [To discuss after the previous 3 are written]

Ideas for now:

* Biodiversity = richness AND…
  + Composition,
  + Functional diversity,
  + Phylogenetic diversity (Re: Felix’ map in the fynbos book)
* Conservation
  + Habitat diversity = more species (Re: European reserve study Tony mentioned)

## Possible weaknesses to our study [not a paragraph]

1. What about absolute environment variables?
   * Deal w/ post-review
   * And emphasise throughout that that is not what we are here to do
2. Species-occurrence data & collection effort?
   * Address outliers, and compare to Gioia & Hopper 2007 (Re: rarefaction etc.) to show that, indeed, our data are flawed.
   * Also note Cramer & Verboom 2017 managed in-spite of it (Re: rarefaction etc.)!
3. Spatial autocorrelation?
   * Our study is *necessarily* correlative
   * Spatial arranged response AND predictors “cancel each other out”

# Tables



**Table 2:** Signs and significances1 of coefficients from univariate regressions of vascular plant species richness against different axes of environmental heterogeneity2 (log10-transformed) and overall environmental heterogeneity (PC1) across the GCFR and SWAFR at the (a) QDS-, (b) HDS- and (c) DS-scales.

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

2 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 3:** Pearson’s *r* correlation coefficients comparing the predictions and residuals of regressions of species richness (*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 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 Table 2.



**Figure 2:** Frequency 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 expressed as a proportion (*T*QDS / *S*HDS).



**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 Table 2 and Figure 1.

A screenshot of a map

Description automatically generated

**Figure 5:** 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 multivariate (MV) model (Figure 4b). Map projection used: WGS84. QDS- and DS-scale equivalents of these maps are available in the online version (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. Retrieved from https://cran.r-project.org/package=rgdal

Born, J., Linder, H. P., & Desmet, P. (2007). The Greater Cape Floristic Region. *Journal of Biogeography*, *34*(1), 147–162. https://doi.org/10.1111/j.1365-2699.2006.01595.x

Chamberlain, S., Szocs, E., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., … O’Donnell, J. (2016). taxize: Taxonomic information from around the web. R package version 0.7.8. Retrieved from https://github.com/ropensci/taxize

Cramer, M. D., & Verboom, G. A. (2016). Measures of biologically relevant environmental heterogeneity improve prediction of regional plant species richness. *Journal of Biogeography*, *44*(3), 1–13. https://doi.org/10.1111/jbi.12911

Dupont, L. M., Linder, H. P., Rommerskirchen, F., & Schefuss, E. (2011). Climate-driven rampant speciation of the Cape flora. *Journal of Biogeography*, *38*(6), 1059–1068. https://doi.org/10.1111/j.1365-2699.2011.02476.x

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), 1–15. https://doi.org/10.1093/botlinnean/box010

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. R package version 2.5-8. Retrieved from https://cran.r-project.org/package=raster

Hoffmann, V., Verboom, G. A., & Cotterill, F. P. D. (2015). Dated plant phylogenies resolve Neogene climate and landscape evolution in the cape floristic region. *PLoS ONE*, *10*(9), 1–25. https://doi.org/10.1371/journal.pone.0137847

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*(1), 623–650. https://doi.org/10.1146/annurev.ecolsys.35.112202.130201

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*(3), 382–392. https://doi.org/10.1111/j.1365-2028.2008.00997.x

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

Rundel, P. W., Arroyo, M. T. K., Cowling, R. M., Keeley, J. E., Lamont, B. B., & Vargas, P. (2016). Mediterranean Biomes: Evolution of Their Vegetation, Floras, and Climate. *Annual Review of Ecology, Evolution, and Systematics*, *47*(1), 383–407. https://doi.org/10.1146/annurev-ecolsys-121415-032330

# Biosketches

**Ruan van Mazijk** is […]

**Michael D. Cramer** is an ecophysiologist interested in physiological 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.

# Supplementary Information

## Species occurrence data cleaning

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 et al., 2016) was then used to query each species name against two major 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; ref?) 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].

## Supplementary tables

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| , Surface T | (v[Version]) |  | .b |
|  |  |  |  |
|  |  |  |  |

## Supplementary figures

## Supplementary references

Chamberlain, S., Szocs, E., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., … O’Donnell, J. (2016). taxize: Taxonomic information from around the web. R package version 0.7.8. Retrieved from <https://github.com/ropensci/taxize>

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. DOI: […]

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. DOI: […]

GBIF (2017a). GBIF Occurrence Download. DOI: […]

GBIF (2017b). GBIF Occurrence Download. DOI: […]

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. DOI: […]

NASA (2017a). 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, U.S.A. DOI: […]

NASA (2017b). 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, U.S.A. DOI: […]