Environmental heterogeneity explains contrasting plant species richness between the South African Cape and southwestern Australia

Running title: Heterogeneity and 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 the course of this work.

# Abstract

**Aim:** To assess whether the difference in species richness per unit area between two mediterranean-type biodiversity hotspots is explained by 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 and used to generate environmental heterogeneity and species richness raster layers. Heterogeneity in multiple environmental variables and species richness per unit area, were compared between the two regions at a range of spatial scales. At each scale species richness was also regressed against a major axis of environmental heterogeneity, derived by principal component analysis (PCA), and, using multiple regression, against heterogeneity in individual environmental variables.

**Results:** The GCFR is generally more environmentally-heterogeneous and species-rich than the SWAFR. Species richness per unit area is significantly related to the major axis of heterogeneity across both regions, the latter describing ca. 38–50% of overall heterogeneity, the slope of this relationship differing between the two regions only at the finest spatial scale. Multivariate regressions, and regressions against the first axes of the PCAs (PC1), revealed variations in the dependence of species richness on environmental heterogeneity that differed between the two regions.

**Main conclusions:** We have evidence for a common positive relationship between floristic richness and environmental heterogeneity across the GCFR and SWAFR, dependent on spatial-scale. Though there are region-specific effects, broad, positive associations between various axes of heterogeneity (and the major axis, PC1) and plant species richness hold across the two flora. The generally greater richness per unit area of the GCFR compared to the SWAFR is thus explained by the former’s generally greater environmental heterogeneity and is concordant with its greater levels of floristic turnover.

*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 locally-deterministic, environmental features (e.g. productivity, heterogeneity) that influence species persistence and coexistence (Bøhn & Amundsen, 2004; Ricklefs, 1987, 2004). Since all three effects are potentially influenced by environmental heterogeneity, the latter may be a particularly important driver of regional species richness variation, with physically-heterogeneous regions tending to be more species-rich, as has been repeatedly demonstrated (e.g. Cramer & Verboom, 2016; Kreft & Jetz, 2007; Laliberte et al., 2014; Thuiller et al., 2006; Stein, Gerstner & Kreft, 2014). For example, given that the recruitment success of immigrant lineages into a region is often dictated by the pre-adaptations of those lineages (Ackerly, 2009; Crisp et al., 2009; Donoghue, 2008), a physically-heterogenous environment may promote diversity by admitting a more functionally-diverse array of immigrant lineages. 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 (Sobel, Chen, Watt, & Schemske, 2010; Wiens, 2004a,b). Likewise, in the context of long-term environmental change, physical heterogeneity may offer refugia to a wider array of lineages and so confer a greater level of buffering against lineage extinction (Byrne, 2008). Finally, environmental heterogeneity has repeatedly been shown to facilitate species coexistence at a variety of scales, enhancing regional species richness (Hart et al., 2017). 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, 2012), 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, Milewski, & Newbey, 1994), although the presence of a significant tree component in the SWAFR underpins a striking difference in vegetation physiognomy (Beard et al., 2000; Milewski, 1981). 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 Eocene and Palaeocene (Hopper, 1979; Verboom et al., 2014). There is also evidence of a long history of transoceanic dispersal between them (Bergh & Linder, 2009). 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 (Ackerly, 2009; Beard et al., 2000; Cowling, Rundel, Lamont, Arroyo, & Arianoutsou, 1996; Gioia & Hopper, 2017).

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). Although on-going identification of new species may change these statistics (e.g. Gioia & Hopper, 2017), the overall differences in species km-2 are quite dramatic. One possible 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. Indeed, of the world’s five mediterranean-climate regions, the CFR has the second highest median topographic heterogeneity, being surpassed only by the Mediterranean (Bradshaw & Cowling, 2014). Critically, since the strong relief of the GCFR underlies steep climatic and edaphic gradients (Bradshaw & Cowling, 2014; Jiménez & Ricklefs, 2014), its climatic and edaphic heterogeneity is correspondingly high. Differences in the spatial scale or “grain” of environmental heterogeneity may also explain the differences in richness between these two floras, with SWAFR environments’ likely being “coarser” than those in the GCFR due to the former’s relative topographic uniformity. The central aim of this paper, then, is to test the hypothesis that the observed difference in species richness (per unit area) between the SWAFR and GCFR is a consequence of differences in the physical heterogeneity of these regions. We investigate this across a range of spatial scales, as we expect this to affect the species richness patterns we observe (Hart et al., 2017). 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 at 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 S1). 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, 2004b). The downloaded occurrence data were then cleaned using the “taxize” package (Chamberlain et al., 2016) in R (R Core Team, 2019) (see Supporting Information). Despite spatial variability in collection effort in both regions, we used raw species counts to estimate QDS-scale species richness on the basis that the application of rarefaction techniques severely distorts known richness patterns when applied to the South African flora (Cramer & Verboom, 2016). The final numbers of unique species thus identified as occurring in the GCFR and SWAFR, respectively, were 9,419 and 6,696. We excluded occurrence data (and indeed environmental data, below) originating from coastal pixels at the 0.05° resolution. This excluded coastal/dunal vegetation from our analyses, due to the floristic and environmental dissimilarity of these areas from the “core” sclerophyllous flora under our consideration. 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). Thus, we retained 362 of ca. 449 QDS in the GCFR and 624 of ca. 737 in the SWAFR (ca. 81% and 85% sampling, respectively).

Using R, the cleaned species occurrence record data were collated into QDS, HDS and DS (sensu Larsen, Holmern, Prager, Maliti, & Røskaft, 2009; downloaded from <https://mindland.com/wp/projects/quarter-degree-grid-cells/download-qdgc/> [Accessed 27 February, 2020]). In addition, following the additive decomposition (Veech et al., 2002) of Whittaker’s (1960) *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. mean *α* richness), and *T*QDS and *T*HDS represent the residual (i.e. turnover-based) *β* richness, determined as *γ* − *α*. These data were compared between the GCFR and SWAFR using common language effect sizes (*CLES*) using the R package “canprot” (Dick, 2017). The *CLES* is the proportion of all pairwise comparisons of a variable between two categories in which the value in one category exceeds that in the other. Additionally, we tested for differences in regional *α*-values and *β*/*γ*-proportions using two-sided Mann-Whitney *U*-tests as implemented in R, in the face of non-normally distributed data.

## 2.2: Comparing environmental heterogeneity

To compare environmental heterogeneity between the GCFR and SWAFR, we acquired a suite of nine geospatially-explicit environmental variables (Table S1) in the form of raster layers to represent topographic (elevation), climatic (surface temperature (T), mean annual precipitation (MAP), precipitation in the driest quarter (PDQ)), edaphic (clay content, soil carbon (C), pH, cation exchange capacity (CEC)) and vegetational gradients (normalized difference vegetation index; NDVI). As far as possible, these variables were selected to represent environmental axes which are considered regionally important and independent (see Figure S1–3). 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 (Belda et al., 2014). Variable selection was, however, constrained by the availability of suitable raster-layers. Thus, although soil phosphorus concentrations ([P]soil) is probably an important determinant of plant distribution in both the GCFR and SWAFR (Lambers et al., 2006, 2010; Shane et al., 2008), 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 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 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 quantified the environmental heterogeneity of a given square (i.e. 0.10°×0.10°-, QDS-, HDS- and DS-scale) as the variance of the environmental conditions of the four sub-squares (i.e. 0.05°×0.05°-, eighth-degree square-, QDS- and HDS-scale) nested within it. This was done both using the (i) “aggregate” function in the R package “raster” (Hijmans, 2016), with variance set as the aggregation function and (ii) “tidyverse” packages (Wickham et al., 2019). Regarding the latter, data at the QDS-, HDS- and DS-scales was analysed in data-frames with rows labelled with grid-cell codes (sensu Larsen, Holmern, Prager, Maliti, & Røskaft, 2009), to ensure the heterogeneity and richness data were derived from the same grid-cells. Since our index measures within-square heterogeneity at each spatial scale, it can be related directly to species richness at the QDS-, HDS- and DS-scales.

We used principal components analysis (PCA), applied to the nine environmental variables across both regions, to extract a major axis of 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 then run at each spatial scale. The first axis (PC1) thus extracted from each of the four PCAs represents the major axis of heterogeneity across the nine environmental heterogeneity variables considered (see Figure S4).

To compare heterogeneity in the nine environmental variables and in the major axis (PC1) of heterogeneity between the two regions, as with species richness and turnover, we employed *CLES* and two-sided Mann-Whitney *U*-tests in R, in the face of some non-normally distributed forms of heterogeneity. Both analyses were done at all four spatial scales. This enabled us to assess scale-dependence in heterogeneity and to ascertain the spatial scale at which regional environmental heterogeneity differences are most pronounced.

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

We used linear models, in R, to assess the associations between heterogeneity (in the nine selected environmental variables and the major heterogeneity axis represented by PC1) with species richness across the two regions. We did not incorporate spatial autocorrelation into our models of species richness because the aim of our analyses was to expose the associations between different forms of spatial environmental heterogeneity and species richness, rather than to predict species richness in absolute or to partition sources of its variation. The fact that we focussed on heterogeneity between neighbouring pixels is at odds with spatial autocorrelation which occurs when neighbouring pixels are more related to each other than more distant pixels due to their proximity.

We first used analyses of covariance (ANCOVA), at the QDS-, HDS- and DS-scales, to relate species richness (*S*) to the heterogeneity of each environmental variable, and to the major axis of heterogeneity, across the two regions. In order to assess whether these relationships are identical in form across the two regions, we fitted three nested models for each heterogeneity predictor *X*, as follows: 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 “main effect × region” model which includes an interaction between heterogeneity and region (*S* ~ *β*0 + *β*1*X* + *β*2*Region* + *β*3(*X × Region*)). For each of the ten predictors, the best fitting model was determined using Akaike’s information criterion (*AIC*), as the simplest model with *∆AIC* < 2.Where the “main effect only” model describes heterogeneity as having a uniform effect on species richness across the two regions (i.e. a common relationship), the “main effect + region” and “main effect × region” models describe the relationships of species richness to heterogeneity as being region-dependent. Specifically, where the “main effect + region” model describes these relationships as being identical in terms of slope but not intercept, the “main effect × region” describes them as differing in both intercept and slope.

We then used multiple linear regressions to model species richness as a function of region and heterogeneity in all nine environmental variables simultaneously. As with the simple regressions, this was done at all three spatial scales. Starting from a 19-predictor model, including region, the heterogeneities of all nine environmental variables and the interactions of the latter with region, models were simplified using reverse stepwise model selection based on *AIC*-scores. In the final model, the retention of significant heterogeneity-region interaction terms indicates that the dependence of species richness on heterogeneity differs between the two regions. Conversely, where only the main effect is significant the dependence of species richness on heterogeneity is inferred to be uniform across the two regions.

## 2.4: Species richness hotspots

To identify hotspots of exceptional richness, i.e. squares whose species richness exceeds that expected on the basis of their environmental heterogeneity, we used the residuals associated with the PC1-based ANCOVA models and the multiple regression models, at all three spatial scales, to identify outlier points. For this purpose, outliers were defined as any points with residual species richness more than two standard deviations from the mean predicted species richness at a given level of environmental heterogeneity (in both the PC1-based and multiple regression models). We also used *F*-tests to assess whether the variances of the residuals associated with ANCOVA and multiple regressions differed between the GCFR and SWAFR. Finally, to assess whether the exceptional richness of hotspots is best explained by factors other than environmental heterogeneity, and to assess the sensitivity of results to these hotspots, we repeated the ANCOVA and multiple regression analyses with hotspots omitted and compared the coefficients of determination from each.

# 3: Results

## 3.1: Comparing species richness

Vascular plant species richness varies spatially in both the GCFR and SWAFR, with both regions possessing a hotspot of exceptional richness (the Kogelberg Centre in the GCFR; Greater Perth in the SWAFR) and declining species richness towards its interior margin (Figure 1a,b). Comparisons of species richness between the regions using two-sided Mann-Whitney *U*-tests reveal that species richness per unit area is similar at the QDS- (Figure 2c; *P* = 0.624, *CLES* = 0.490), somewhat greater in the GCFR at HDS- (Figure 2b; *P* = 0.061, *CLES* = 0.579) and significantly greater in the GCFR at DS-scales (Figure 2a; *P* = 0.050, *CLES* = 0.717).

Partitioning *S*HDS into its *α*- and *β*-components (QDS and *T*QDS respectively; Figure 2g), we found that the average proportional contribution of floristic turnover (i.e. *T*QDS/*S*HDS; Figure 2e) is greater in the GCFR (0.64) than in the SWAFR (0.60) (two-sided Mann-Whitney *U*-test, *P* < 0.001; *CLES* = 0.750). This is also the case at the DS-scale (Figure 2d; *T*HDS/*S*DS, GCFR:0.57, SWAFR: 0.49; two-sided Mann-Whitney *U* test, *P* < 0.001, *CLES* = 0.967). Summarily, at both of these scales, the plant species richness of squares in the GCFR is somewhat more attributable to turnover between sub-squares (points further above 1:1-line in Figure 2f,g) than for squares in the SWAFR (points closer to 1:1-line in Figure 2f,g).

## 3.2: Comparing environmental heterogeneity

*CLES* comparisons revealed the GCFR to be more heterogeneous than the SWAFR in all nine environmental variables, and across the full range of spatial scales (Figure 3), with a few exceptions at the DS-scale (i.e. MAP, NDVI, CEC, clay and soil C). The same was true for the major axis of heterogeneity described by PC1 (Figure 3j), which accounted for between ca. 38% (at the 0.10°×0.10°-scale) and ca. 50% (at the DS-scale) of the variance in all nine heterogeneity variables. The greater overall heterogeneity of the GCFR is striking when compared visually (e.g. at the HDS-scale: Figure 1c vs d). In general, the disparity in heterogeneity between the two regions seems greater for topographic (as expected) and climatic (*CLES* between ca. 0.60 and 1.00; Figure 3a–d) than edaphic variables (*CLES* between ca. 0.60 and 0.80; Figure 3f–i).

Regressions indicate that the degree to which the GCFR is more environmentally heterogeneous than the SWAFR is positively scale-dependent (i.e. more heterogeneous at broad scales), with the notable exceptions of MAP and clay (Figure 3b,h), in which the GCFR is disproportionately more heterogeneous at fine scales. The major axis of heterogeneity (PC1) displays scale-independence of most forms of heterogeneity, with its *CLES* being more or less uniform across spatial scales (Figure 3j).

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

The univariate ANCOVA results (Table 1) show that heterogeneity in each of the nine environmental heterogeneity variables, as well as the main axis of heterogeneity (PC1), influence species richness in a consistently positive manner across the two study regions (exceptions being DS-scale elevation and CEC; Table 1c). Notably, “main effect × region” models were optimal for some variables. Although, the interaction terms associated with these either describe a form of heterogeneity as having a positive effect on richness in both regions or as having a positive effect in one region and negligible effect in the other (e.g. HDS-scale heterogeneity in clay; Table 1b). In addition, at the HDS- and DS-scales, the effect of the major axis of heterogeneity (PC1) on species richness is uniform across the two regions (i.e. “main effect only” model favoured: Table 1b,c; Figure 4b,c). At the QDS-scale, though the “main effect + region” model was favoured, the difference in intercepts between the GCFR and SWAFR (92.5 species; Table 1a; Figure 4a) is small relative to the spread of residual species richness at the QDS-scale within each region (*SD*GCFR = 343.5; *SD*SWAFR = 245.8; Table S3a). This indicates that regional differences in the form of the QDS-scale species richness-heterogeneity relationship, while statistically significant, are subtle. Quadratic forms of these best fitting PC1-models were also compared, with the above linear forms of the “main effect only” models favoured at both the HDS- and DS-scales (*∆AIClinear* < 2). At the QDS-scale, however, a quadratic model (i.e., of the form *S* ~ *β*0 + *β*1*X + β*2*X*2) was favoured (*∆AIClinear* = 23.78), with positive effects of PC1 on richness (*β*1 = 69.02, *β*2 = 9.99; both *P* < 0.001) describing an upward-opening parabolic relationship.

We fit multiple regression models to describe the effects of the different heterogeneity variables on species richness in concert. The partial effects of heterogeneity predictors retained in the optimal multiple regression models (Figure 5) are usually positive or neutral in the GCFR and positive or neutral in the SWAFR, with exceptions being heterogeneity in elevation at the DS-scale, PDQ at QDS- and DS-scales, CEC at the QDS- and DS-scales and pH at all spatial scales. This adds nuance to the generally positive influence of heterogeneity on species richness displayed by the PC1-based ANCOVAs (Figure 4) and in most forms of heterogeneity when compared in isolation (Table 1). Interestingly, DS-scale heterogeneity in elevation and CEC have both negative univariate (Table 1c) and partial (Figure 5c) effects. Heterogeneity in MAP has the greatest explanatory power in the QDS- and HDS-scale models (Table S2a,b). Like the optimal ANCOVA models, multiple regression models suggest region-dependence in the relationships of species dependence to environmental heterogeneity at all three spatial scales (Figure 5). Particularly at the DS-scale (Figure 5c), the coefficients associated with the partial effects of heterogeneity in SWAFR relative to the GCFR effectively cancel out the effects of these variables estimated in the GCFR. This indicates weaker or absent partial relationships between richness and the different forms of heterogeneity in the SWAFR.

The coefficients of determination associated with the optimum regression models (Figure 5: *R*2 = 0.27, 0.43 and 0.85) are consistently greater than those associated with the optimal ANCOVA models based on PC1 (Figure 4: *R*2 = 0.13, 0.22 and 0.49). This is because PC1 does not capture all of the variation in environmental heterogeneity relevant to the prediction of species richness.

## 3.4: Species richness hotspots

Residual-based outlier identification yielded very similar results for the PC1-based ANCOVA and multiple regression models at all spatial scales (Table 2; Figure 1e–h, S9e–h, S10e–h). For both the GCFR and SWAFR, outliers are geographically clustered, in areas corresponding to recognized diversity centres (Figure S5, S6). At the QDS-scale, for example, GCFR outliers are concentrated in the Kogelberg-Hottentots-Holland and Cederberg areas, while SWAFR outliers are concentrated in the Mt Lesueur, Perth, Stirling-Albany and Fitzgerald River areas (Figure S5a,b; S6a,b; S9e–h). Fewer hotspots are necessarily resolved at the HDS-scale (Figure S5c,d; S6c,d; 1e–h) and only one at the DS-scale when considering the PC1-based ANCOVAs’ residuals (the Hottentots Holland area in the GCFR). Omission of outliers from the PC1-based ANCOVA (Figure S7) and multiple regressions (Figure S8) yielded qualitatively similar models as before with improved coefficients of determination. At the broader DS-scale, the GCFR and SWAFR do not differ significantly in their residual standard deviations (following both the PC1-based ANCOVA and multiple regressions) (Table S3c). This highlights that the commonality of the richness-heterogeneity relationship is recovered more accurately at broader spatial scales after the omission of hotspots.

# 4: Discussion

Consistent with a recent meta-analysis identifying environmental heterogeneity as a universal driver of species richness (Stein et al., 2014), we found heterogeneity to have a consistently positive influence on species richness in the GCFR and SWAFR. Most significant partial effects associated with environmental heterogeneity terms in our multiple regression models and the overall effect of environmental heterogeneity (PC1) in our ANCOVA results are positive. Thus, we find no evidence for the hump-backed response of species richness to heterogeneity anticipated by some authors (Allouche et al., 2012; Carnicer et al., 2013), at least at the scales considered here. A quadratic relationship between richness and heterogeneity was only favoured at the QDS-scale, and even then not representing a unimodal effect but an upward-opening parabola. Additionally, and also consistent with (Stein et al., 2014), we find the strength of the heterogeneity-species richness relationship to associate positively with spatial scale (grain), as evidenced by that fact that the coefficients of determination associated with our ANCOVA and multiple regression models were greatest at the DS-scale and smallest at the QDS-scale. One possible explanation of this effect is the fact that larger areas accommodate more environmental variability (Wüest et al., 2019), and so facilitate stronger heterogeneity-species richness relationships (Van Rensburg, Chown, & Gaston, 2002). In speciation hotspots (see below) like the GCFR or the SWAFR, however, an important additional consideration is the spatial scale of speciation (Stein et al., 2014). Stronger coupling of species richness to heterogeneity at the DS-scale possibly arising because the DS, in contrast to the QDS and HDS, is sufficiently large to capture differences in richness as a result of allopatric speciation processes.

The observation that species richness responds to environmental heterogeneity in a uniform manner across the GCFR and SWAFR or at least, where this is not the case, that regional differences in the form of the richness-heterogeneity relationship are subtle, suggests that the greater vascular plant species richness of the GCFR is partly attributable to the greater physiographic heterogeneity of this region. As noted by Cowling et al. (2015) the lower heterogeneity of SWAFR, both now and during the Tertiary, would have constrained opportunities for radiation, thereby producing a flora that is generally less diverse. In addition, the greater heterogeneity of the GCFR might facilitate denser species packing, through the provision of greater niche diversity (Cramer et al., 2014; Hart et al., 2017). Our data reveal that, despite its significantly greater DS-scale species richness, the GCFR does not have significantly greater QDS- and HDS-scale species richness than the SWAFR. This pattern, which reflects higher rates of QDS- and HDS-scale species turnover in the GCFR, implies a greater role for dispersal limitation and local species differentiation in driving high DS-scale richness in the Cape. This interpretation is consistent with existing evidence for a much higher frequency of single-site endemic taxa in the Cape than in the Australian flora (Linder, 2019) and evidence of such in our species occurrence dataset (Figure S11).

Although our data confirm a broad dependence of GCFR and SWAFR plant species richness on environmental heterogeneity, as defined by the nine environmental variables examined, the generally low coefficients of determination associated with these relationships indicate a role for other factors. Firstly, species richness is almost certainly influenced by heterogeneity in multiple other environmental variables, some of which may vary at spatial scales beyond the fine-scale resolution of available environmental layers. Cramer, Wootton, van Mazijk, & Verboom (2019), for example, recently highlighted the superiority of locally-modelled soil layers, which also include aspects of soil chemistry, as predictors of vegetation type in the GCFR, compared with the globally-modelled layers used in this study. Unfortunately, we were unable to make use of these layers as comparable data are lacking between the GCFR and SWAFR. Despite the obvious importance of soil variables as determinants of plant distribution, their inclusion in broad-scale spatial modelling exercises remains problematic since the spatial scale at which soils vary is typically much finer than the spatial resolution of modelled layers (Figueiredo et al., 2018). Similarly, the association of many point-endemics in the GCFR with highly-localized bogs, whose distributions are geomorphologically- rather than climatically-determined, presents challenges for species distribution modelling on account of their small size (Born & Linder, 2018). The same may well be true for the SWAFR flora where phylogenetically-relictual species typically inhabit waterlogged situations (Hopper & Gioia, 2004).

A second factor affecting the explanatory power of our models is the inclusion of only the heterogeneities of environmental variables and not their absolute values. Although the absolute values of certain environmental variables, particularly those influencing biological productivity (i.e. energy-water theory), have often been found to correlate positively with species richness at broad scales (Currie, 1991; Hawkins et al., 2003; Kreft & Jetz, 2007), we elected to omit these variables from our analyses because we wished explicitly to assess the relative ability of various forms of heterogeneity to account for species richness variation across and between the GCFR and SWAFR. Besides, as noted by Cowling, Bradshaw, Colville, & Forest (2017), energy-water theory does not appear to hold in the GCFR where species richness is maximized in cool, nutrient-impoverished areas.

A third factor underpinning the unexplained variance in our models is spatial variation in collection effort and its consequences for species richness estimates. Although some authors have argued for the application of sample- (Gotelli & Colwell, 2001) or coverage-based rarefaction techniques (Chao & Jost, 2012) to correct for variable collection effort (Engemann et al., 2015), the application of these approaches to systematically-biased collection data is potentially problematic. These techniques assume that the relationship between true species richness and sampling effort is weak. In contrast, herbarium collectors commonly focus their efforts on rare species, and many herbaria do not accept material indiscriminately, being reluctant to accept multiple accessions of the same species from a single area. Moreover, rarefaction-methods, while able to account for spatially non-uniform sampling effort, are not ideal for non-uniformities in sampling that arise from targetted collection efforts (e.g. collection trips targetting a specific taxon and disregarding others in a given locality). We have desisted from applying these techniques (as in Gioia & Hopper, 2017) because, at least for the South African flora, these methods severely distort known species richness (Cramer & Verboom, 2016). Use of rarefaction may artificially inflate species richness (relative to areas of greater richness) in areas of low collection, where collections are low not because of inaccessibility, but because species richness is low. Indeed, Cramer & Verboom (2016) found that the cumulative species richness provided a reasonable estimate of the biome-level floral richness, and that these numbers were consistent with results from an exhaustive atlassing project (i.e. Protea Atlas; Rebelo, 2001) with low collecting bias (Merow et al., 2013).

A fourth and final factor potentially compromising the strength of the species richness-heterogeneity relationship is the existence of diversity hotspots whose high species richness is not directly linked to physiographic heterogeneity. We suspect, for example, that the exceptional species richness of the Perth QDS/HDS reported in this study partly reflects the introduction of plants from elsewhere in the SWAFR into the Greater Perth metropolitan area. Indeed, our species occurrence data seem to show a non-trivial proportion of species with outlier-records (relative to the rest of their ranges) in the vicinity of Perth: ca. 7% (based on a random sample of 300 of the 2,944 species with occurrences near Perth; see Figure S12). Bearing in mind such artefacts, species richness in both the GCFR and SWAFR is often concentrated in hotspots (cf. Gioia & Hopper, 2017; Oliver, Linder, & Rourke, 1983) whose exceptional richness may be a consequence of their long-term climatic and/or hydrological stability. In the Cape Floristic Region (CFR; sensu Goldblatt, 1978) for example, higher species richness in the west has been attributed to reduced rates of extinction, associated with greater climatic stability through the Pleistocene (Cowling & Lombard, 2002; Cowling et al., 2017). In this study, we used the residuals derived from PC1-based ANCOVA and multiple regression models to identify hotspots whose species richness exceeds that expected on the basis of their underlying heterogeneities. Though spatially autocorrelated, at least some of the hotspots thus identified correspond to centres of long-term environmental stability. In the GCFR, for example, the southwestern mountains (Kogelberg-Hottentots Holland) have been identified as a long-term climatic and hydrological refugium, especially for moisture-loving species which inhabit the numerous bogs and seeps found there (Linder, 2019; Wüest et al., 2019). Similarly, pollen and midden-based isotope data provide evidence of relatively muted Pleistocene climate change in the Cederberg (Meadows, Chase, & Seliane, 2010; Meadows & Sugden, 1991, 1993). The biota of the SWAFR, too, show evidence of climatically-forced range contraction in the Pleistocene (Byrne, 2008; Byrne & Hines, 2004). Putative refugia include the Sterling and Porungurup ranges (Keppel et al., 2017), areas relatively elevated and topographically complex in the SWAFR, supporting greater levels of species richness and endemism. Additionally, maps in Byrne (2008; Figure 2) identify one refugium in the vicinity of Perth and a second in the vicinity of Fitzgerald River.

Given their importance for long-term species persistence (McLaughlin et al., 2017), the accurate identification and effective protection of climatically- and/or hydrologically-stable hotspots must be an important conservation objective, particularly in the face of contemporary climatic deterioration. In the context of the GCFR, therefore, it is important to establish whether the high species richness in the west is in fact attributable to a broad longitudinal effect, as implied by some authors (Cowling & Lombard, 2002; Cowling et al., 2017; Verboom et al., 2014), or whether it is tied to the presence of hotspots whose exceptional richness is attributable to effects that are more local in nature. Our data support the latter interpretation, with the Kogelberg-Hottentots Holland area emerging as a particularly important refugium for Cape plant diversity. Given that much of the locally-endemic diversity of this area is associated with bog and seepage habitats (Linder, 2019), recent proposals to abstract water from the Table Mountain Group aquifer which feeds these bogs, present a serious threat to the continued persistence of Cape floristic diversity, particularly since the hydrological consequences of such abstraction remain poorly understood (Slingsby et al., 2018). The long-term integrity of groundwater-dependent terrestrial vegetation is similarly threatened by groundwater abstraction in the Perth area of the SWAFR (Barron et al., 2014; Froend & Sommer, 2010).

Although the plant species richness of the GCFR has been identified as being globally anomalous (Kreft & Jetz, 2007), our data reveal that species richness in the GCFR and SWAFR varies broadly as a common function of environmental heterogeneity, the latter promoting diversification and facilitating species coexistence. This corroborates the suggestion (Cowling et al., 2015) that the greater richness of the GCFR, relative to that of the SWAFR, is a product of its greater environmental heterogeneity. Importantly, since environmental heterogeneity invariably is spatially-configured, the greater richness of the GCFR flora is associated with higher rates of species turnover in space. The species richness-environmental heterogeneity relationship is punctuated by the existence of local hotspots, whose exceptional richness may be a consequence of historical factors. Though these regions’ hotspots may be important for species persistence (e.g. in the face of climate change), we find evidence for a unified conceptual framework for thinking about the species richness in of these two mediterranean floras.

# Tables

**Table 1:** Values and significances1 of coefficients from univariate ANCOVAs 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. The SWAFR interaction terms describe the difference in slope between the GCFR and SWAFR lines for that model.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Response | Model type | Predictor | Main effect | | SWAFR effect | | SWAFR interaction | |
| (a) *S*QDS | Main effect × region | Elevation | 140.2 | \*\*\* | 124.4 | \*\*\* | -56.1 | \* |
|  |  | MAP | 172.0 | \*\*\* | 54.5 | \*\* | -54.1 | \*\* |
|  |  | PDQ | 73.4 | \*\*\* | 55.6 | \* | 61.9 | \*\* |
|  |  | NDVI | 154.9 | \*\*\* | -7.8 |  | -102.2 | \*\*\* |
|  | Main effect + region | PC1 | 67.0 | \*\*\* | 92.5 | \*\*\* |  |  |
|  | Main effect only | Surface T | 62.1 | \*\*\* |  |  |  |  |
|  |  | CEC | 14.7 |  |  |  |  |  |
|  |  | Clay | 42.1 | \*\*\* |  |  |  |  |
|  |  | Soil C | 62.9 | \*\*\* |  |  |  |  |
|  |  | pH | 21.9 | \* |  |  |  |  |
| (b) *S*HDS | Main effect × region | MAP | 399.0 | \*\*\* | -41.5 |  | -185.0 | \*\* |
|  |  | Clay | -12.8 |  | -216.1 | \*\* | 173.6 | \* |
|  | Main effect only | Elevation | 163.7 | \*\*\* |  |  |  |  |
|  |  | PDQ | 226.3 | \*\*\* |  |  |  |  |
|  |  | Surface T | 135.9 | \*\*\* |  |  |  |  |
|  |  | NDVI | 246.6 | \*\*\* |  |  |  |  |
|  |  | Soil C | 159.4 | \*\*\* |  |  |  |  |
|  |  | PC1 | 123.1 | \*\*\* |  |  |  |  |
|  | Region only | CEC | -26.3 |  | -251.9 | \*\* |  |  |
|  |  | pH | 53.8 |  | -193.0 | \* |  |  |
| (c) *S*DS | Main effect × region | Elevation | -1455.9 | \* | -2278.4 | \*\* | 1668.5 | \* |
|  |  | MAP | 683.3 | \*\*\* | -519.1 | \*\* | -382.1 | \* |
|  |  | CEC | -933.3 | \*\* | -1043.4 | \*\*\* | 837.1 | \* |
|  | Main effect + region | Clay | 273.0 | \* | -542.8 | \* |  |  |
|  |  | Soil C | 246.5 | \* | -615.4 | \* |  |  |
|  | Main effect only | PDQ | 363.1 | \*\* |  |  |  |  |
|  |  | Surface T | 336.7 | \*\* |  |  |  |  |
|  |  | NDVI | 475.3 | \*\*\* |  |  |  |  |
|  |  | pH | 448.4 | \*\*\* |  |  |  |  |
|  |  | PC1 | 231.1 | \*\*\* |  |  |  |  |

1 Represented as follows: \*\*\*, *P* < 0.001; \*\*, *P* < 0.01; \*, *P* < 0.05; 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 2:** Pearson’s *r* correlation coefficients comparing the predicted and residual species richness (*S*) between the PC1-based ANCOVAs (Table 1, Figure 4) and the multivariate regressions (Figure 5) of vascular plant species richness across the GCFR and SWAFR. All correlation coefficients were significant (*P* < 0.001; two-sided *t*-test).

|  |  |  |
| --- | --- | --- |
|  | Correlation | |
| Spatial scale | Predicted *S* | Residual *S* |
| QDS | 0.680 | 0.908 |
| HDS | 0.699 | 0.834 |
| DS | 0.723 | 0.369 |

# Figures

A close up of a map

Description automatically generated

**Figure 1 (previous page):** HDS-scale maps for the GCFR and SWAFR of **(a,b)** vascular plant species richness (*S*HDS), **(c,d)** values of the major axis of environmental heterogeneity (PC1) from the PCA of nine forms of environmental heterogeneity (log10-transformed; Table S1) and residual species richness following regressions against **(e,f)** PC1 (see also Figure 4b) and **(g,h)** the multivariate (MV) regression model (see also Figure 5b). Map projection used: WGS84. Colour versions of these maps, and the QDS- and DS-scale equivalents (Figure S9, S10 respectively) are available in the online version.



**Figure 2:** Frequency distributions of vascular plant species richness (*S*) in the GCFR and SWAFR at the **(a)** DS-, **(b)** HDS-, and **(c)** QDS-scales, and of the proportional contributions of floristic turnover (i.e. *T*/*S*) to **(d)** *S*DS (i.e. *T*HDS / *S*DS) and **(e)** *S*HDS. (i.e. *T*QDS / *S*HDS). Frequencies are scaled as the proportions of cells within each region. The derivations of *T*DS and *T*HDS are demonstrated with scatter plots of mean **(f)** HDS- and **(g)** QDS-scale richness (i.e. HDS and QDS, respectively) against turnover (*T*HDS and *T*QDS, respectively) with contour lines denoting the *S*DS and *S*HDS that arise as their sums.

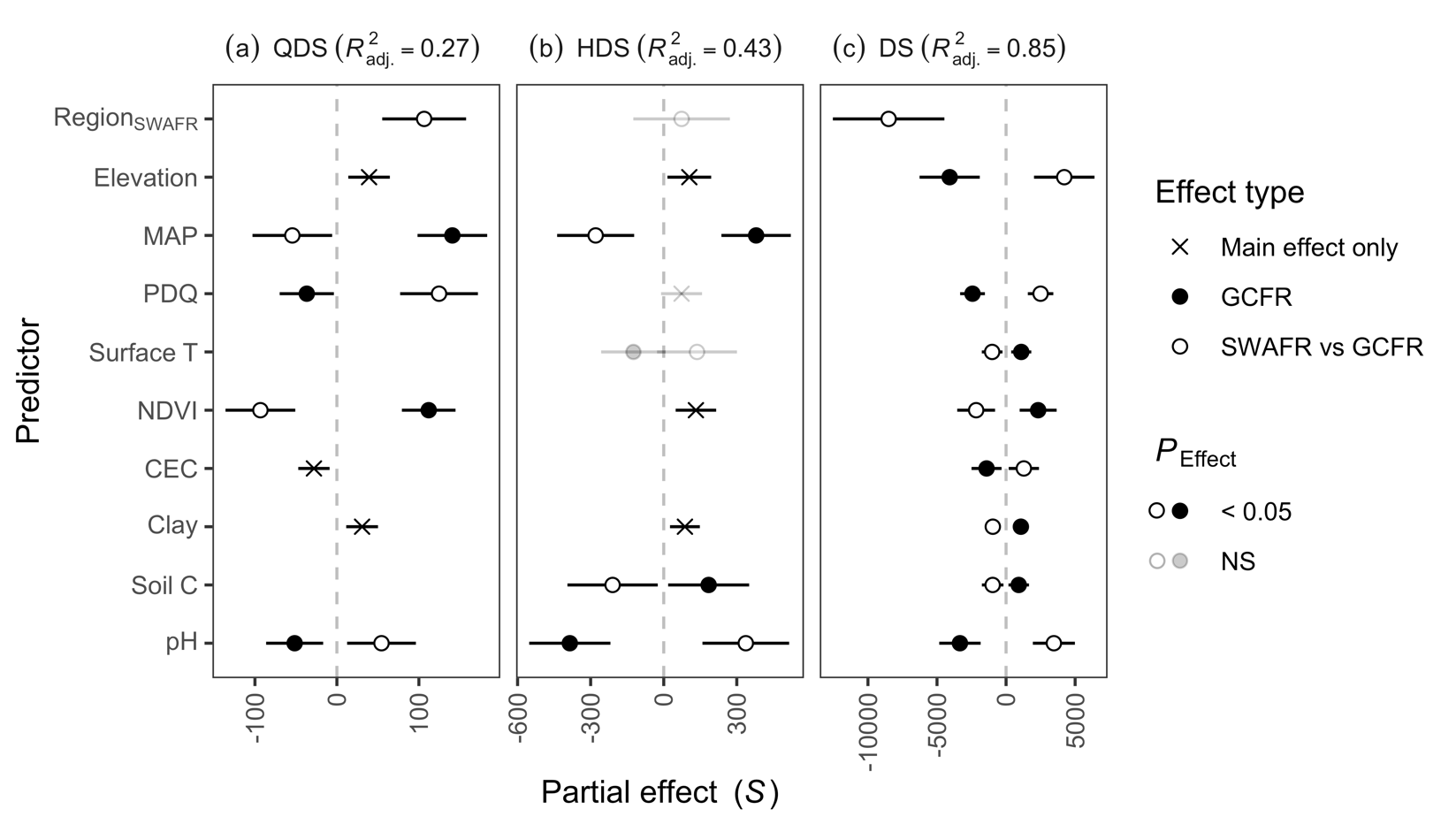
**A close up of a map

Description automatically generated**

**Figure 3:** 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 evidence for positive scale-dependence (depicted by lines) for most variables. Both strongly significant (*P* < 0.05; a–c,g,i) and marginally significant (*P* < 0.1; d,e,h) have been plotted, the latter for illustration. Note, only two forms of heterogeneity were found to be negatively scale-dependent **(b,h)**. Abbreviations are as in Table 1.

****

**Figure 4:** 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. Grid-cells representing hotspots of species richness (identified as having residual *S* greater than two standard deviations from the mean) are identified by triangles (black for the GCFR, white for SWAFR). These three linear models are presented in Table 1, all with highly significant slopes (*P* < 0.001) and were fitted including species richness hotspots. For *S*QDS, the separate fits for the GCFR (black) and SWAFR (grey) are presented following the best fitting model at that scale: a “main effect + region” model (*S* ~ *β*0 + *β*1*X* + *β*2*Region*; Table 1a). The *R*2-values of each model and the percentage 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. The equivalent results when hotpots were excluded is available in the online version (Figure S7).

****

**Figure 5:** Partial effect 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 re-scaled) across the GCFR and SWAFR. Points with error bars denote slopes and their 95% confidence intervals. Filled and empty points represent effect estimates for the GCFR and the SWAFR relative to the GCFR, respectively (i.e. region-interaction terms were retained during stepwise model selection). 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 1 and Figure 3. The equivalent results when species richness hotpots were excluded is available in the online version (Figure S8).

# Data availability statement

The raw vascular plant occurrence records used are available from GBIF (GCFR: <https://doi.org/10.15468/dl.n6u6n0>; SWAFR: <https://doi.org/10.15468/dl.46okua>). Analyses in the form of R-scripts, cleaned species lists, raster-layers of the nine forms of environmental heterogeneity, the major axis of heterogeneity (PC1) (at each of the four spatial scales) and species richness (at the three spatial scales) are available in the DRYAD Digital Repository: [URL].

# References

Ackerly, D. D. (2009). Evolution, origin and age of lineages in the Californian and Mediterranean floras. *Journal of Biogeography*, *36*(7), 1221–1233. https://doi.org/10.1111/j.1365-2699.2009.02097.x

Allouche, O., Kalyuzhny, M., Moreno-Rueda, G., Pizarro, M., & Kadmon, R. (2012). Area-heterogeneity tradeoff and the diversity of ecological communities. *Proceedings of the National Academy of Sciences*, *109*(43), 17495–17500.

Barron, O., Froend, R., Hodgson, G., Ali, R., Dawes, W., Davies, P., & McFarlane, D. (2014). Projected risks to groundwater-dependent terrestrial vegetation caused by changing climate and groundwater abstraction in the Central Perth Basin, Western Australia. *Hydrological Processes*, *28*(22), 5513–5529.

Beard, J. S., Chapman, A. R., & Gioia, P. (2000). Species richness and endemism in the Western Australian flora. *Journal of Biogeography*, *27*(6), 1257–1268. https://doi.org/10.1046/j.1365-2699.2000.00509.x

Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: from Köppen to Trewartha. *Climate Research*, *59*(1), 1–13.

Bergh, N. G., & Linder, H. P. (2009). Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies (Asteraceae--Gnaphalieae). *Molecular Phylogenetics and Evolution*, *51*(1), 5–18.

Bivand, R., Keitt, T., & Rowlingson, B. (2017). *rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-7*. https://cran.r-project.org/package=rgdal

Bøhn, T., & Amundsen, P.-A. (2004). Ecological interactions and evolution: forgotten parts of biodiversity? *BioScience*, *54*(9), 804–805.

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

Born, Julia, & Linder, H. P. (2018). Water availability, fundamental niches and realized niches: A case study from the Cape flora. *Austral Ecology*, *43*(6), 696–705. https://doi.org/10.1111/aec.12616

Bradshaw, P. L., & Cowling, R. M. (2014). Landscapes, rock types, and climate of the Greater Cape Floristic Region. In N. Allsopp, J. F. Colville, & G. A. Verboom (Eds.), *Fynbos: Ecology, Evolution and Conservation of a Megadiverse Region* (pp. 26–46). Oxford University Press. https://doi.org/oso/9780199679584.003.0002

Byrne, M. (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, *27*(27–28), 2576–2585. https://doi.org/10.1016/j.quascirev.2008.08.032

Byrne, M., & Hines, B. (2004). Phylogeographical analysis of cpDNA variation in Eucalyptus loxophleba (Myrtaceae). *Australian Journal of Botany*, *52*(4), 459–470.

Carnicer, J., Brotons, L., Herrando, S., & Sol, D. (2013). Improved empirical tests of area-heterogeneity tradeoffs. *Proceedings of the National Academy of Sciences*, *110*(31), E2858--E2860.

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

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547.

Cowling, R. M., & Lombard, A. T. (2002). Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*, *8*(3), 163–179.

Cowling, R. M., Witkowski, E. T. F., Milewski, A. V., & Newbey, K. R. (1994). Taxonomic, Edaphic and Biological Aspects of Narrow Plant Endemism on Matched Sites in Mediterranean South Africa and Australia. *Journal of Biogeography*, *21*(6), 651. https://doi.org/10.2307/2846038

Cowling, Richard M., Potts, A. J., Bradshaw, P. L., Colville, J., Arianoutsou, M., Ferrier, S., Forest, F., Fyllas, N. M., Hopper, S. D., Ojeda, F., Procheş, Ş., Smith, R. J., Rundel, P. W., Vassilakis, E., & Zutta, B. R. (2015). Variation in plant diversity in mediterranean-climate ecosystems: The role of climatic and topographical stability. *Journal of Biogeography*, *42*(3), 552–564. https://doi.org/10.1111/jbi.12429

Cowling, Richard M., Rundel, P. W., Lamont, B. B., Arroyo, M. K., & Arianoutsou, M. (1996). Plant diversity in mediterranean-climate regions. *Trends in Ecology and Evolution*, *11*(9), 362–366. https://doi.org/10.1016/0169-5347(96)10044-6

Cowling, Richard M, Bradshaw, P. L., Colville, J. F., & Forest, F. (2017). Levyns’ Law: explaining the evolution of a remarkable longitudinal gradient in Cape plant diversity. *Transactions of the Royal Society of South Africa*, *72*(2), 184–201.

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

Cramer, M. D., West, A. G., Power, S. C., Skelton, R., & Stock, W. D. (2014). Plant ecophysiological diversity. In *Fynbos: Ecology, Evolution and Conservation of a Megadiverse Region* (pp. 248–272). Oxford University Press. https://doi.org/oso/9780199679584.003.0011

Cramer, M. D., Wootton, L. M., van Mazijk, R., & Verboom, G. A. (2019). New regionally modelled soil layers improve prediction of vegetation type relative to that based on global soil models. *Diversity and Distributions*, *25*(11), 1736–1750.

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754–756. https://doi.org/10.1038/nature07764

Currie, D. J. (1991). Energy and large-scale patterns of animal-and plant-species richness. *The American Naturalist*, *137*(1), 27–49.

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11549–11555. https://doi.org/10.1073/pnas.0801962105

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., Peet, R. K., Violle, C., & Svenning, J.-C. (2015). Limited sampling hampers “big data” estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, *5*(3), 807–820.

Figueiredo, F. O. G., Zuquim, G., Tuomisto, H., Moulatlet, G. M., Balslev, H., & Costa, F. R. C. (2018). Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species. *Journal of Biogeography*, *45*(1), 190–200. https://doi.org/10.1111/jbi.13104

Froend, R., & Sommer, B. (2010). Phreatophytic vegetation response to climatic and abstraction-induced groundwater drawdown: examples of long-term spatial and temporal variability in community response. *Ecological Engineering*, *36*(9), 1191–1200.

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

Goldblatt, P. (1978). An analysis of the flora of Southern Africa: its characteristics, relationships abd origins. *Annals of the Missouri Botanical Garden*, *65*, 369–436.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Hart, S. P., Usinowicz, J., & Levine, J. M. (2017). The spatial scales of species coexistence. *Nature Ecology & Evolution*, *1*(8), 1066–1073. https://doi.org/10.1038/s41559-017-0230-7

Hawkins, B. A., Field, R., Cornell, H. V, Currie, D. J., Guégan, J.-F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O’Brien, E. M., & others. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, *84*(12), 3105–3117.

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

Hopper, S. D. (1979). Biogeographical Aspects of Speciation in the Southwest Australian Flora. *Annual Review of Ecology and Systematics*, *10*, 399–422. http://www.jstor.org/stable/2096798

Hopper, S. D. (2009). OCBIL theory: Towards an integrated understanding of the evolution, ecology and conservation of biodiversity on old, climatically buffered, infertile landscapes. *Plant and Soil*, *322*(1), 49–86. https://doi.org/10.1007/s11104-009-0068-0

Hopper, S. D., & Gioia, P. (2004a). 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. https://doi.org/10.1146/annurev.ecolsys.35.112202.130201

Hopper, S. D., & Gioia, P. (2004b). 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

Jiménez, I., & Ricklefs, R. E. (2014). Diversity anomalies and spatial climate heterogeneity. *Global Ecology and Biogeography*. https://doi.org/10.1111/geb.12181

Keppel, G., Robinson, T. P., Wardell-Johnson, G. W., Yates, C. J., Van Niel, K. P., Byrne, M., & Schut, A. G. T. (2017). A low-altitude mountain range as an important refugium for two narrow endemics in the Southwest Australian Floristic Region biodiversity hotspot. *Annals of Botany*, *119*(2), 289–300. https://doi.org/10.1093/aob/mcw182

Kreft, H., & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences*, *104*(14), 5925–5930. https://doi.org/10.1073/pnas.0608361104

Laliberte, E., Zemunik, G., & Turner, B. L. (2014). Environmental filtering explains variation in plant diversity along resource gradients. *Science*, *345*(6204), 1602–1605. https://doi.org/Doi 10.1126/Science.1256330

Lambers, H., Brundrett, M. C., Raven, J. A., & Hopper, S. D. (2010). Plant mineral nutrition in ancient landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. *Plant and Soil*, *334*(1–2), 11–31. https://doi.org/10.1007/s11104-010-0444-9

Lambers, H., Shane, M. W., Cramer, M. D., Pearse, S. J., & Veneklaas, E. J. (2006). Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. *Annals of Botany*, *98*(4), 693–713. https://doi.org/10.1093/aob/mcl114

Lamont, B. B., & He, T. (2017). When did a Mediterranean-type climate originate in southwestern Australia? *Global and Planetary Change*, *156*, 46–58. https://doi.org/10.1016/j.gloplacha.2017.08.004

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

Levin, L. A., Sibuet, M., Gooday, A. J., Smith, C. R., & Vanreusel, A. (2010). The roles of habitat heterogeneity in generating and maintaining biodiversity on continental margins: an introduction. *Marine Ecology*, *31*, 1–5. https://doi.org/10.1111/j.1439-0485.2009.00358.x

Linder, H. P. (2019). Rare species, Restionaceae, and the Cape flora. *Journal of Biogeography*, 1–14. https://doi.org/10.1111/jbi.13709

McLaughlin, B. C., Ackerly, D. D., Klos, P. Z., Natali, J., Dawson, T. E., & Thompson, S. E. (2017). Hydrologic refugia, plants, and climate change. *Global Change Biology*, *23*(8), 2941–2961.

Meadows, M E, Chase, B. M., & Seliane, M. (2010). Holocene palaeoenvironments of the Cederberg and Swartruggens mountains, Western Cape, South Africa: pollen and stable isotope evidence from hyrax dung middens. *Journal of Arid Environments*, *74*(7), 786–793.

Meadows, M E, & Sugden, J. M. (1991). A vegetation history of the last 14,500 years on the Cederberg, SW Cape, South Africa. *South African Journal of Science*, *87*(3).

Meadows, Michael E, & Sugden, J. M. (1993). The late Quaternary palaeoecology of a floristic kingdom: the southwestern Cape South Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *101*(3–4), 271–281.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species’ distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069.

Milewski, A. V. (1981). A comparison of vegetation height in relation to the effectiveness of rainfall in the mediterranean and adjacent arid parts of Australia and South Africa. *Journal of Biogeography*, *8*(2), 107. https://doi.org/10.2307/2844553

Mouchet, M., Levers, C., Zupan, L., Kuemmerle, T., Plutzar, C., Erb, K., Lavorel, S., Thuiller, W., & Haberl, H. (2015). Testing the effectiveness of environmental variables to explain European terrestrial vertebrate species richness across biogeographical scales. *PLoS ONE*, *10*(7), 1–16. https://doi.org/10.1371/journal.pone.0131924

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

Oliver, E. G. H., Linder, H. P., & Rourke, J. P. (1983). Geographical distribution of present-day Cape taxa and their phytogeographical significance. *Bothalia*, *14*(3/4), 427–440.

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*(11), 933–938.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Rebelo, T. G. (2001). *A field guide to the proteas of South Africa*. Fernwood.

Ricklefs, R. E. (1987). Community diversity: relative roles of local and regional processes. *Science, New Series*, *235*(4785), 167–171.

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, *7*(1), 1–15.

Shane, M. W., Cramer, M. D., & Lambers, H. (2008). Root of edaphically controlled Proteaceae turnover on the Agulhas Plain, South Africa: phosphate uptake regulation and growth. *Plant, Cell & Environment*, *31*(12), 1825–1833.

Slingsby, J. A., February, E. C., & Rebelo, T. G. (2018). Water: at what cost to our unique flora? *Veld & Flora*, *June*, 72–75.

Snijman, D. A. (2013). *Plants of the Greater Cape Floristic Region. 2: The Extra Cape flora*. South African National Biodiversity Institute.

Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*(2), 295–315. https://doi.org/10.1111/j.1558-5646.2009.00877.x

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, *17*(7), 866–880. https://doi.org/10.1111/ele.12277

Stock, W. D., & Verboom, G. A. (2012). Phylogenetic ecology of foliar N and P concentrations and N: P ratios across mediterranean-type ecosystems. *Global Ecology and Biogeography*, *21*(12), 1147–1156.

Thuiller, W., Midgley, G. F., Rouget, M., Cowling, R. M., F. Midgley, G., Rougeti, M., & M. Cowling, R. (2006). Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, *29*(5), 733–744. https://doi.org/10.1111/j.0906-7590.2006.04674.x

van Rensburg, B. J., Chown, S. L., & Gaston, K. J. (2002). Species richness, environmental correlates, and spatial scale: a test esing South African birds. *The American Naturalist*, *159*(5), 566–577. https://doi.org/10.1086/339464

Veech, J. A., Summerville, K. S., Crist, T. O., & Gering, J. C. (2002). The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, *99*(1), 3–9.

Verboom, G. A., Linder, H. P., Forest, F., Hoffmann, V., Bergh, N. G., & Cowling, R. M. (2014). Cenozoic assembly of the Greater Cape flora. In N. Allsopp, J. F. Colville, & G. A. Verboom (Eds.), *Fynbos: Ecology, Evolution and Conservation of a Megadiverse Region*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199679584.001.0001/acprof-9780199679584

Whittaker, R. H. (1960). Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the {tidyverse}. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiens, J. J. (2004a). Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution*, *58*(1), 193–197.

Wiens, J. J. (2004b). What is speciation and how should we study it? *The American Naturalist*, *163*(6), 914–923.

Wüest, R. O., Boucher, F. C., Bouchenak-Khelladi, Y., Karger, D. N., & Linder, H. P. (2019). Dissecting biodiversity in a global hotspot: Uneven dynamics of immigration and diversification within the Cape Floristic Region of South Africa. *Journal of Biogeography*, *46*(9), 1936–1947. https://doi.org/10.1111/jbi.13625

# Biosketches

**Ruan van Mazijk** is a graduate student interested in macroevolution, macroecology and phylogenetic comparative biology, primarily focusing on plants.

**Michael D. Cramer** is an ecophysiologist interested in physiological specialization in the hyper‐diverse Cape flora and the link between nutrient‐impoverished soils and plant 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 analyses, which were then carried out by RvM, who then wrote the first draft of the manuscript. All authors contributed equally to the writing thereafter.

# 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>