Analyses v2

Cape vs SWA

Ruan van Mazijk

2019-06-21

Preamble/outline

Here I layout the "new", second incarnation of the analyses as discussed over the course of May/June 2019, following the first draft of the manuscript.

To reiterate that manuscript, we hypothesise that the greater vascular plant species richness of the GCFR compared to that of the SWAFR is explained by the regions' difference in environmental heterogeneity.

The proposed "story" of questions for the analyses is as follows:

1. Is the GCFR more heterogeneous environmentally than the SWAFR, and does the scale of that heterogeneity differ to that of the SWAFR?
2. Do the regions differ w.r.t. the species richness of both HDS and QDS cells, and, for HDS cells' richness (S_{HDS}), does the explanatory power of mean QDS richness (S_{QDS}) and turnover (T_{QDS}) differ between the regions?
3. Does heterogeneity explain differences in richness and turnover between the regions?

1. Environmental heterogeneity & scale

Is the GCFR more heterogeneous environmentally than the SWAFR, and does the scale of that heterogeneity differ to that of the SWAFR?

In order to determine which region is more environmentally heterogeneous, and what scales heterogeneity is most pronounced, we calculated a measure of environmental heterogeneity at various spatial scales (namely: the base data resolution (0.05º x 0.05º), eighth- (EDS), quarter- (QDS), half- (HDS) and three-quarter-degree-squares (3QDS)).

Environmental "roughness" in both regions was calculated, in moving 3 x 3 cell windows, as the average absolute difference between cells and their (usually) 8 neighbours. Alternatively, for a focal cell x^*, the roughness is based on $x_1, x_2, \ldots, x_i, \ldots, x_8$ neighbour cells as:

$$Roughness(x^*) = f\begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x^* & x_5 \\ x_6 & x_7 & x_8 \end{pmatrix} = \frac{1}{8}\sum_i |x^* - x_i|$$

In R, this is implemented this as follows:

```r
roughness <- function(x) {
  raster::focal(x, matrix(1, nrow = 3, ncol = 3), function(x) {
    focal_cell <- x[5]
    focal_exists <- (!is.na(focal_cell)) & (!is.nan(focal_cell))
    if (focal_exists) {
      neighbour_exists <- (!is.na(x)) & (!is.nan(x)) & (x != focal_cell)
      neighbour_cells <- x[neighbour_exists]
      return(mean(abs(focal_cell - neighbour_cells)))
    } else {
      return(NA)
    }
  })
}
```

Following this, the various forms environmental heterogeneity were ordinated using principal component analysis (PCA), to summarise a major axis of heterogeneity in each region (Figure 1). Portions of the data matrices for each scale for these PCAs are shown in Table 1.

Both the actual environmental heterogeneity values and the principal component of heterogeneity were then compared between the GCFR and SWAFR using common language effect sizes (*CLES*). The *CLES* of GCFR vs SWAFR heterogeneity values was regressed against the spatial scale at which it was calculated using simple linear regression (Figure 2, Table 2).

We can see that PDQ, NDVI, pH and, arguably, elevation are all consistently more heterogeneous in the GCFR than in the SWAFR, regardless of spatial scale (Figure 2). The GCFR is more heterogeneous at finer scales in terms of MAP, surface temperature, CEC and soil carbon (Figure 2). Notably, the GCFR is more pronouncedly heterogeneous at broad scales in terms of clay (Figure 2). In general (i.e. regarding PC1; Figure 2), the GCFR is more environmentally heterogeneous than the SWAFR, and particularly so at fine spatial scales.

Table 1: Portions of the data matrices used in the PCA for this section of the analysis, where roughness values were $log(x + 1)$-transformed to ensure normality.

region	Elevation	MAP	PDQ	Surface.T	NDVI	CEC	Clay	Soil.C	pH
GCFR	5.19	2.52	0.72	1.32	15.13	1.14	1.2	2.46	1.36
GCFR	5	2.7	0.61	1.16	15.01	1.11	1.11	1.74	1.83
GCFR	4.86	2.55	0.72	1.17	15.08	1.18	1.4	1.79	1.65
…	…	…	…	…	…	…	…	…	…
SWAFR	3.27	2.77	1.1	0.71	14.91	0.31	1.19	1.59	0.48
SWAFR	2.36	2.41	1.15	0.7	14.28	0.67	1.29	2.03	1.3
SWAFR	2.86	1.98	1.17	1.09	13.58	0.73	2.27	2.4	2.58

Table 2: Slopes and associated *P*-values from simple linear regressions of *CLES* against scale for each form of environmental roughness (Figure 2).

Variable	Slope	*P*	
Elevation	0.044	0.016	*
MAP	-0.313	0.020	*
PDQ	0.010	0.387	
Surface.T	-0.330	0.026	*
NDVI	0.032	0.459	
CEC	-0.126	0.063	.
Clay	0.243	0.013	*
Soil.C	-0.298	0.003	*
pH	-0.010	0.756	
PC1	-0.172	0.010	*

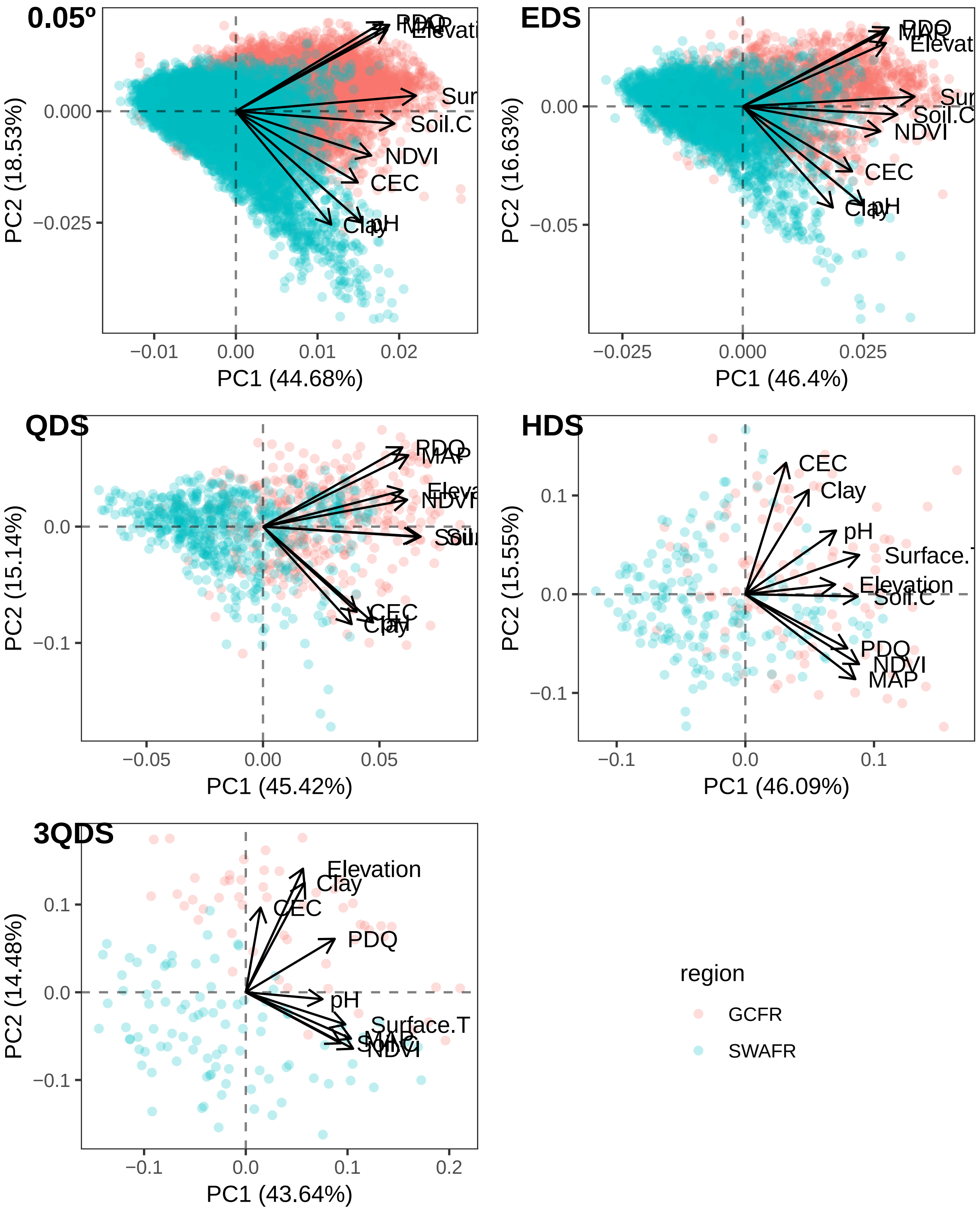


Figure 1: Scatter plots of the first and second principal components (PC1, PC2) of environmental heterogeneity following principal components analyses (PCAs) of the various forms of environmental heterogeneity, repeated at the five spatial scales. The proportion of variation accounted for by each axis is denoted in parentheses. Arrows (labelled) denote the rotational loading of a given form of environmental heterogeneity. Note, the signs of loadings on PC1 have been forced to be positive, while the signs of loadings on PC2 are arbitrary.

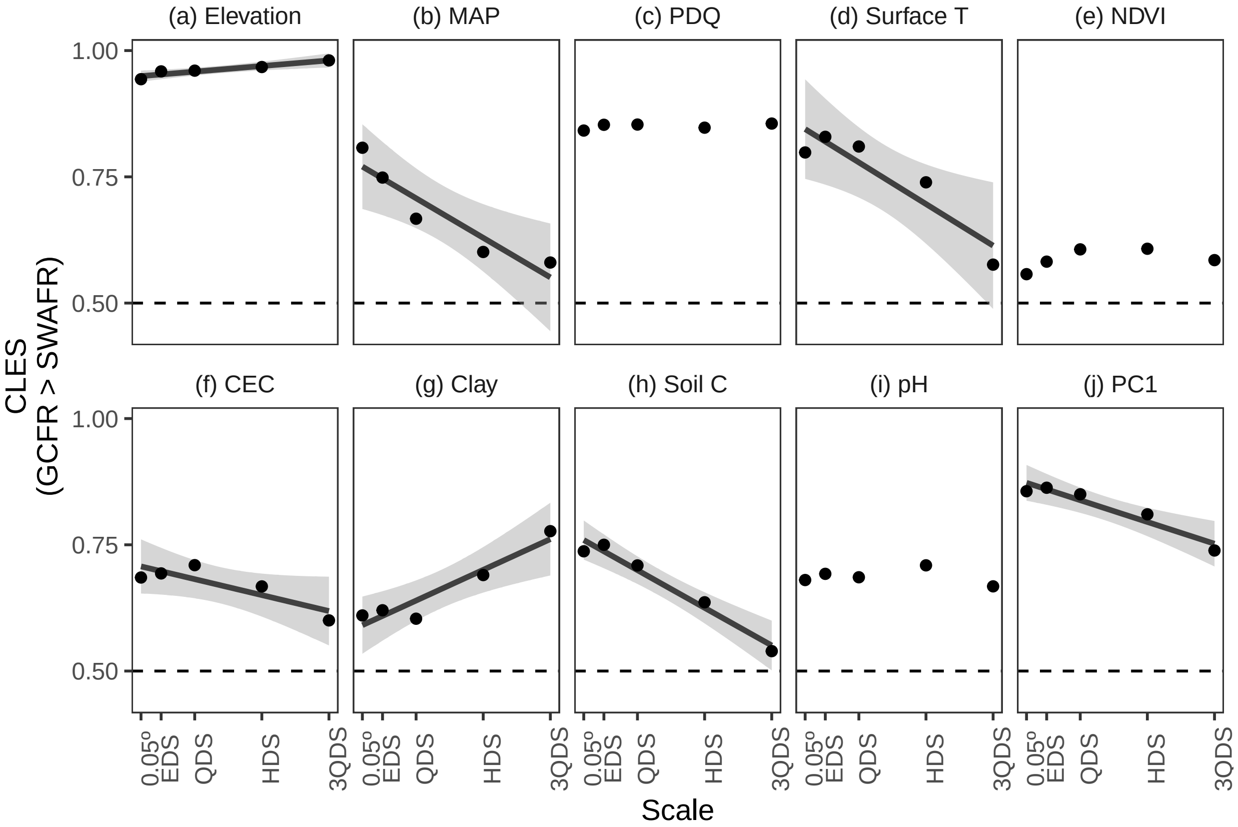


Figure 2: Simple linear regressions of the common language effect size ($CLES$) of various forms of environmental heterogeneity (a–i), and the first principal component of heterogeneity (j; see Figure 1), where the $CLES$ is treated as the effect of GCFR relative to SWAFR values. Only significant or marginally significant fits are plotted (Table 2). Grey bands denote 95% confidence intervals about the fitted lines. Across spatial scales, all $CLES$ values differed significantly from zero following two-sided t-tests ($P < 0.001$).

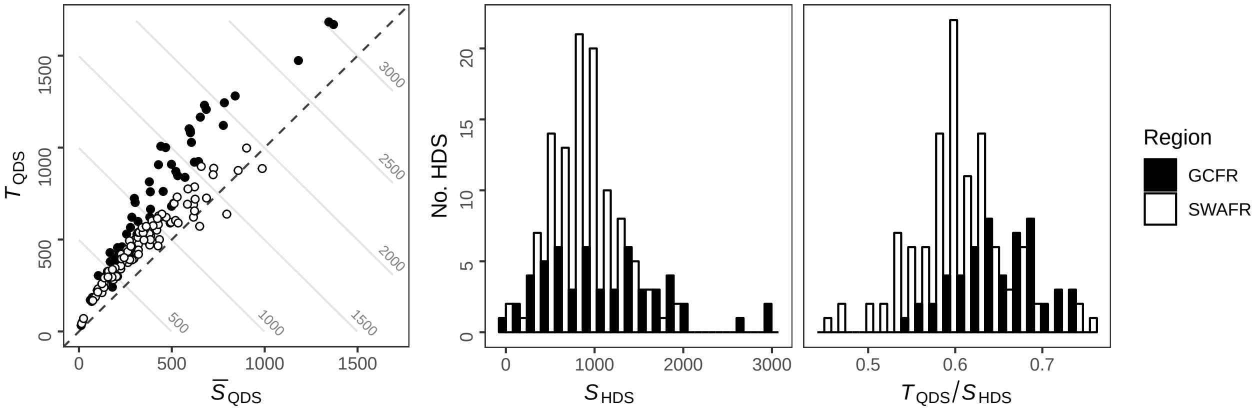


Figure 3: (a) Scatter plot of mean QDS-scale richness ($\overline{S}_{\text{QDS}}$) and turnover ($T_{\text{QDS}}$) with contour lines denoting the S_{HDS} that would arise as their sum (i.e. increasing from lower-left to upper-right). Distributions of (a) HDS-scale species richness (S_{HDS}) and (b) the turnover partition of that richness ($T_{\text{QDS}}/S_{\text{HDS}}$).

2. Species richness & turnover

Do the regions differ w.r.t. the species richness of both HDS and QDS cells, and, for HDS cells' richness (S_{HDS}), does the explanatory power of mean QDS richness (S_{QDS}) and turnover (T_{QDS}) differ between the regions?

To tackle this question, I compare measures of species richness and turnover between the regions. Species richness at the HDS-scale (S_{HDS}) can be partitioned into the average richness of the constituent QDS in HDS ($\overline{S}_{\text{QDS}}$) and species turnover ($T_{\text{QDS}}$) defined[1] as:

$$T_{\text{QDS}} = S_{\text{HDS}} - \overline{S}_{\text{QDS}}$$

The distributions of these data are presented in Figure 3. To test for significant differences between GCFR and SWAFR values, I use Mann-Whitney U-tests and $CLES$ (Table 3), as most of the variables deviate significantly from normality (Shapiro-Wilk normality test; $P < 0.05$).

Additionally, a visualisation of how S_{HDS} is partitioned into $\overline{S}_{\text{QDS}}$ and T_{QDS} is presented in Figure 4.

We can conclude that broad scale species richness (i.e. that at the HDS scale) is more strongly driven by turnover between areas (i.e. QDS) than so in the SWAFR.

Table 3: Results of Mann-Whitney U-tests and the $CLES$ of GCFR vs SWAFR for various species richness and turnover metrics.

Metric	$CLES$	P_U
S_{HDS}	0.612	0.020
S_{QDS}	0.595	< 0.001
$T_{\text{QDS}}/S_{\text{HDS}}$	0.784	< 0.001

3. Relating heterogeneity to species richness & turnover

Does heterogeneity explain differences in richness and turnover between the regions?

[1] following Whittaker's original additive definition: $\gamma = \alpha + \beta$

Here I fit various linear regressions of richness and turnover as functions of environmental heterogeneity across the two regions. The richness and turnover measures used are the same as in the previous section, while the environmental heterogeneity was recalculated in the same grid-wise fashion as the richness and turnover measures. These analyses were carried out at both the HDS- and QDS-scales, insofar as species occurrence data from GBIF is only accurate to the QDS-scale. These analyses were only carried out on HDS-scale data for HDS-cells that contained four QDS-cells, and similarly for QDS-scale data for QDS-cells that contained four EDS-cells.

Environmental "roughness" here was calculated for each HDS- and QDS-cell in both regions as the mean of each consituent QDS- and EDS-cell's mean absolute difference in environmental conditions from the other three cells within that HDS- or QDS-cell.

In other words, roughness was calculated by first calculating the average absolute-difference in environmental values between each QDS and it's three neighbours in a given HDS. Then, these four values (assuming four QDS in an HDS) are averaged. This roughness index is presented mathematically below. This index allows each of the four values to be similarly independent, and thus more sutiable for our averaging and analyses, as opposed to if it were simly the direct average of pairwise differences [expand?].

$$Roughness_{cellular}(\{x_1, x_2, x_3, x_4\}) = \frac{1}{4}\sum_i f(x_i) = \frac{1}{4}\sum_i \left(\frac{1}{3}\sum_{j \neq i} |x_i - x_j|\right)$$

In R, this is implemented this as follows:

```
roughness_cells <- function(x) {
  out <- vector(mode = "numeric", length = length(x))
  for (i in seq_along(x)) {
    out[[i]] <- mean(abs(x[i] - x[-i]))
  }
  mean(out)
}
```

3.1. Separate-regions models with combinations of variables

```
## # A tibble: 4 x 5
##   term             estimate std.error statistic  p.value
##   <chr>               <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          813.      97.3      8.36 4.69e-11
## 2 Clay_roughness       185.      76.7      2.42 1.93e- 2
## 3 MAP_roughness        738.     115.       6.41 5.04e- 8
## 4 pH_roughness        -323.     113.      -2.85 6.33e- 3
```

```
## $GCFR_HDS_richness
## [1] "Clay_roughness + MAP_roughness + pH_roughness"
##
## $GCFR_QDS_richness
## [1] "MAP_roughness + NDVI_roughness + PDQ_roughness + pH_roughness + Soil.C_roughness"
##
## $GCFR_QDS_turnover
## [1] "Clay_roughness + MAP_roughness + Soil.C_roughness"
##
## $SWAFR_HDS_richness
## [1] "CEC_roughness + Clay_roughness + Elevation_roughness + MAP_roughness + PDQ_roughness + Surface.T
##
## $SWAFR_QDS_richness
## [1] "CEC_roughness + Clay_roughness + Elevation_roughness + MAP_roughness + PDQ_roughness + Surface.
##
## $SWAFR_QDS_turnover
## [1] "CEC_roughness + Clay_roughness + Elevation_roughness + MAP_roughness + PDQ_roughness + pH_rough
```

Table 4: Results of bi-directional stepwise multiple linear regressions of three richness and turnover responses in the against additive combinations of environmental heterogeneity variables. The stepwise regression procedure started with all variables included. (See Figure 5 for a graphical representation.)

Region	Response	Predictor	Slope	P_{slope}	
GCFR	S_{HDS}	Clay	185.456	0.019	*
GCFR	S_{HDS}	MAP	738.358	0.000	*
GCFR	S_{HDS}	pH	-322.625	0.006	*
GCFR	S_{QDS}	MAP	136.688	0.003	*
GCFR	S_{QDS}	NDVI	139.568	0.000	*
GCFR	S_{QDS}	PDQ	-45.541	0.147	
GCFR	S_{QDS}	pH	-164.670	0.000	*
GCFR	S_{QDS}	Soil.C	97.764	0.009	*
GCFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	Clay	-0.017	0.016	*
GCFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	MAP	-0.026	0.010	*
GCFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	Soil.C	0.024	0.015	*
SWAFR	S_{HDS}	CEC	-111.775	0.000	*
SWAFR	S_{HDS}	Clay	56.676	0.036	*
SWAFR	S_{HDS}	Elevation	200.297	0.000	*
SWAFR	S_{HDS}	MAP	108.435	0.001	*
SWAFR	S_{HDS}	PDQ	180.511	0.001	*
SWAFR	S_{HDS}	Surface.T	99.867	0.027	*
SWAFR	S_{QDS}	CEC	-28.862	0.012	*
SWAFR	S_{QDS}	Clay	18.683	0.094	.
SWAFR	S_{QDS}	Elevation	42.177	0.014	*
SWAFR	S_{QDS}	MAP	97.709	0.000	*
SWAFR	S_{QDS}	PDQ	116.652	0.000	*
SWAFR	S_{QDS}	Surface.T	47.573	0.002	*
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	CEC	0.014	0.008	*
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	Clay	-0.011	0.022	*
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	Elevation	-0.035	0.000	*
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	MAP	-0.009	0.066	.
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	PDQ	-0.015	0.113	
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	pH	0.011	0.020	*
SWAFR	$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	Soil.C	-0.012	0.046	*

Table 5: Adjusted R^2-values of the models in Table 5.

Response	GCFR ${R^2}_{\mathrm{adj.}}$	SWAFR ${R^2}_{\mathrm{adj.}}$
S_{HDS}	0.429	0.510
S_{QDS}	0.262	0.323
$T_{\mathrm{QDS}}/S_{\mathrm{HDS}}$	0.139	0.424

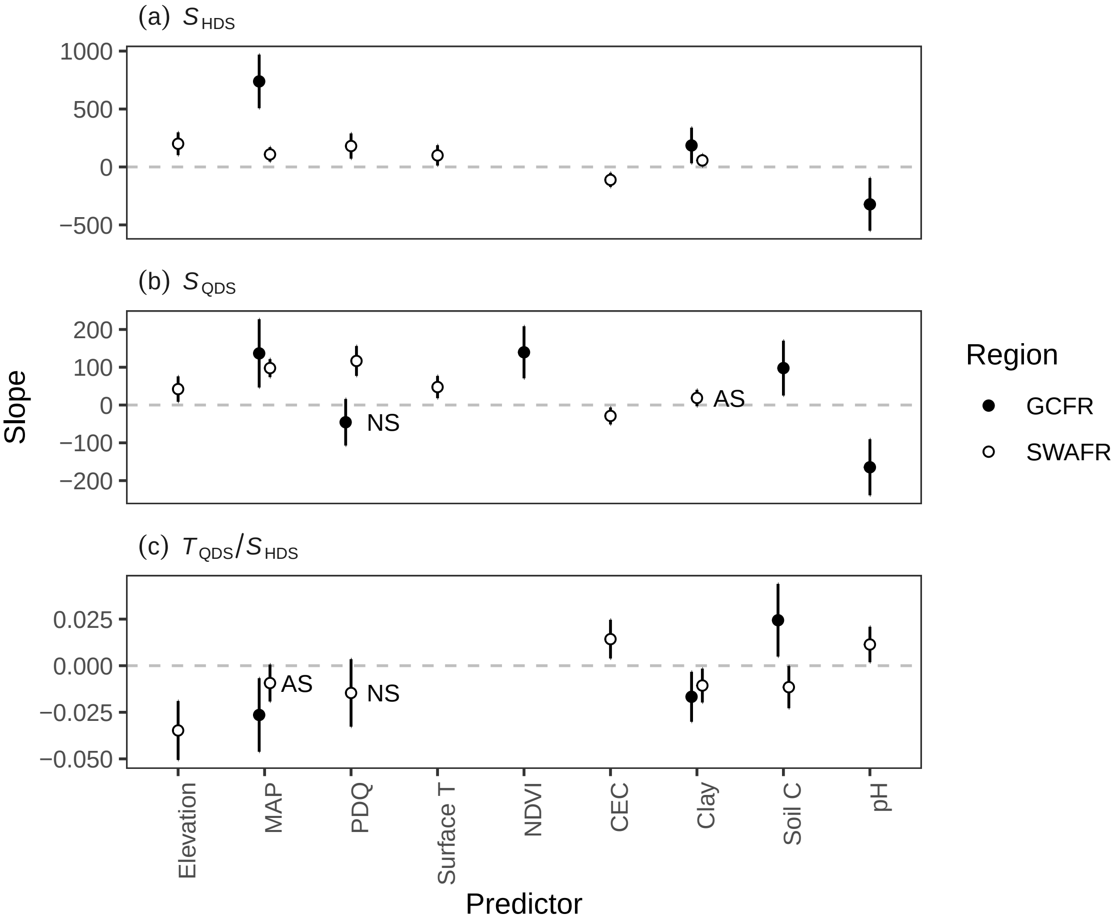


Figure 4: Slopes from Table 5, with error bars denoting 95% confidence intervals about each slope estimate.

3.2. Combined-regions models with individual variables

3.2.1. Environmental heterogeneity variables

[Plots of the models referred to in Tables 6–8 can be discussed in a meeting.]

Table 6: Results of separate simple linear regressions of S_{HDS} against environmental heterogeneity variables with no region-term.

Predictor	R^2	P_{slope}	
CEC_roughness	0.014	0.141	
Clay_roughness	0.038	0.014	*
Elevation_roughness	0.140	0.000	*
MAP_roughness	0.315	0.000	*
NDVI_roughness	0.164	0.000	*
PDQ_roughness	0.171	0.000	*
pH_roughness	0.026	0.042	*
Soil.C_roughness	0.148	0.000	*
Surface.T_roughness	0.125	0.000	*

Table 7: Results of separate simple linear regressions of S_{HDS} against environmental heterogeneity variables with an additive region-term.

Predictor	R^2	P_{slope}		P_{region}	
CEC_roughness	0.071	0.827		0.002	*
Clay_roughness	0.095	0.038	*	0.002	*
Elevation_roughness	0.142	0.000	*	0.546	
MAP_roughness	0.315	0.000	*	0.785	
NDVI_roughness	0.186	0.000	*	0.041	*
PDQ_roughness	0.171	0.000	*	0.815	
pH_roughness	0.075	0.383		0.004	*
Soil.C_roughness	0.156	0.000	*	0.227	
Surface.T_roughness	0.127	0.002	*	0.494	

Table 8: Results of separate simple linear regressions of S_{HDS} against environmental heterogeneity variables with an interaction-region-term.

Predictor	R^2	P_{slope}		P_{region}		$P_{slope:region}$	
CEC_roughness	0.073	0.720		0.007	*	0.556	
Clay_roughness	0.097	0.478		0.002	*	0.563	
Elevation_roughness	0.157	0.181		0.899		0.099	
MAP_roughness	0.360	0.000	*	0.334		0.001	*
NDVI_roughness	0.197	0.000	*	0.090		0.143	
PDQ_roughness	0.183	0.003	*	0.915		0.143	
pH_roughness	0.078	0.282		0.015	*	0.455	
Soil.C_roughness	0.156	0.024	*	0.318		0.795	
Surface.T_roughness	0.135	0.147		0.368		0.226	

3.2.2. PC1 models

Here, I present my findings with raw R-code, because I don't have the time to format it neatly.

Table 9: Richness (HDS)

Model	AIC	ΔAIC	w_{Akaike}
PC1	2433.144	0.000	0.451
PC1+Region	2433.558	0.414	0.367
PC1:Region	2434.958	1.814	0.182

Table 10: Richness (QDS)

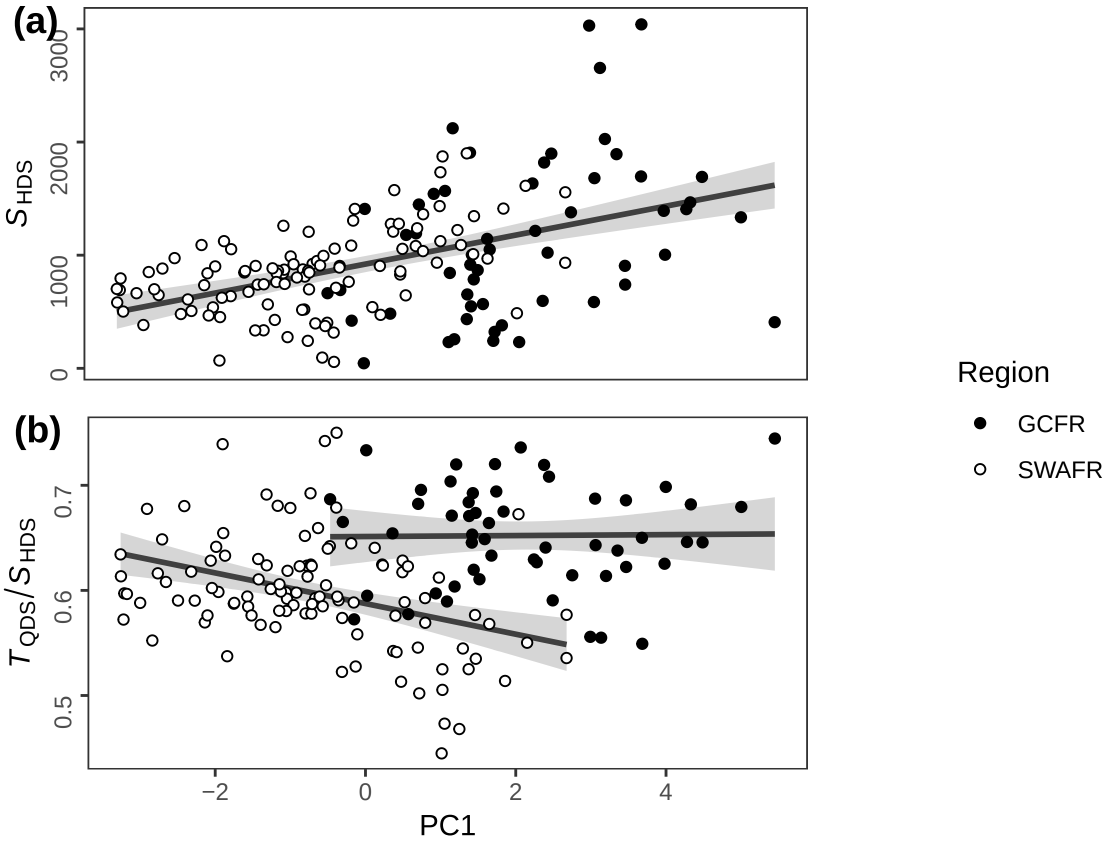
Model	AIC	ΔAIC	w_{Akaike}
PC1	9205.960	0.999	0.262
PC1+Region	9204.961	0.000	0.432
PC1:Region	9205.652	0.691	0.306

Table 11: Turnover

Model	AIC	ΔAIC	w_{Akaike}
PC1	2240.186	0.000	0.536
PC1+Region	2242.178	1.993	0.198
PC1:Region	2241.587	1.401	0.266

Table 12: Turnover (proportional)

Model	AIC	ΔAIC	w_{Akaike}
PC1	-458.545	45.812	0.000
PC1+Region	-499.982	4.374	0.101
PC1:Region	-504.357	0.000	0.899



```
## List of 12
##  $ coefficients : Named num [1:2] 917 129
##   ..- attr(*, "names")= chr [1:2] "(Intercept)" "PC1"
##  $ residuals    : Named num [1:161] -480.7 -244.7 188.2 -99.8 -553.7 ...
##   ..- attr(*, "names")= chr [1:161] "1" "2" "3" "4" ...
##  $ effects      : Named num [1:161] -12042 -3126 226 -55 -508 ...
##   ..- attr(*, "names")= chr [1:161] "(Intercept)" "PC1" "" "" ...
##  $ rank         : int 2
##  $ fitted.values: Named num [1:161] 964 1113 991 1103 1122 ...
##   ..- attr(*, "names")= chr [1:161] "1" "2" "3" "4" ...
##  $ assign       : int [1:2] 0 1
##  $ qr           :List of 5
##   ..$ qr   : num [1:161, 1:2] -12.6886 0.0788 0.0788 0.0788 0.0788 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:161] "1" "2" "3" "4" ...
##   .. .. ..$ : chr [1:2] "(Intercept)" "PC1"
##   .. ..- attr(*, "assign")= int [1:2] 0 1
##   ..$ qraux: num [1:2] 1.08 1.05
##   ..$ pivot: int [1:2] 1 2
##   ..$ tol  : num 1e-07
##   ..$ rank : int 2
##   ..- attr(*, "class")= chr "qr"
##  $ df.residual  : int 159
##  $ xlevels      : Named list()
##  $ call         : language lm(formula = HDS_richness ~ PC1, data = HDS)
##  $ terms        :Classes 'terms', 'formula'  language HDS_richness ~ PC1
##   .. ..- attr(*, "variables")= language list(HDS_richness, PC1)
##   .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:2] "HDS_richness" "PC1"
##   .. .. .. ..$ : chr "PC1"
##   .. ..- attr(*, "term.labels")= chr "PC1"
##   .. ..- attr(*, "order")= int 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(HDS_richness, PC1)
##   .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. ..- attr(*, "names")= chr [1:2] "HDS_richness" "PC1"
##  $ model        :'data.frame':   161 obs. of  2 variables:
##   ..$ HDS_richness: num [1:161] 483 868 1179 1003 568 ...
##   ..$ PC1         : num [1:161] 0.36 1.52 0.57 1.44 1.59 ...
##   ..- attr(*, "terms")=Classes 'terms', 'formula'  language HDS_richness ~ PC1
##   .. .. ..- attr(*, "variables")= language list(HDS_richness, PC1)
##   .. .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. .. ..$ : chr [1:2] "HDS_richness" "PC1"
##   .. .. .. .. ..$ : chr "PC1"
##   .. .. ..- attr(*, "term.labels")= chr "PC1"
##   .. .. ..- attr(*, "order")= int 1
##   .. .. ..- attr(*, "intercept")= int 1
##   .. .. ..- attr(*, "response")= int 1
##   .. .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. .. ..- attr(*, "predvars")= language list(HDS_richness, PC1)
##   .. .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. .. ..- attr(*, "names")= chr [1:2] "HDS_richness" "PC1"
##  - attr(*, "class")= chr "lm"
```

3.3. Combined-regions models with combinations of variables

[To be discussed in meeting.]