Supplementary Information

For “Heterogeneity and species richness”

Ruan van Mazijk, Michael D. Cramer and G. Anthony Verboom

# 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 were 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].

# Tables

**Table S1:** Georeferenced environmental data1 and vascular plant species occurrence data sources used in this study. Data were acquired for the GCFR and SWAFR, with the temporal extent of data products used described where applicable.

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

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

**Table S2:** ANOVAs for the heterogeneity variables (including interactions with region) used in the three multiple regression models of vascular plant species richness, across the GCFR and SWAFR, including species richness hotspots. The variables in each model are arranged in descending order according to their proportion of variance explained. The significance1 of each variable’s contribution to each model is also shown. Abbreviations follow that in Table S1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Response | Term | Variance explained | *P* | | |
| (a) *S*QDS | (Residuals) | 0.75 |  |  |  |
|  | MAP | 0.14 | < 0.001 |  | \*\*\* |
|  | Elevation | 0.06 | < 0.001 |  | \*\*\* |
|  | Region | 0.01 | < 0.001 |  | \*\*\* |
|  | NDVI | 0.01 | < 0.001 |  | \*\*\* |
|  | PDQ | 0.01 | 0.005 |  | \*\* |
|  | pH × Region | 0.01 | 0.011 |  | \* |
|  | PDQ × Region | 0.01 | 0.011 |  | \* |
|  | pH | <0.01 | 0.057 |  | ~ |
|  | NDVI × Region | <0.01 | 0.066 |  | ~ |
|  | MAP × Region | <0.01 | 0.091 |  | ~ |
|  | CEC × Region | <0.01 | 0.325 |  |  |
|  | CEC | <0.01 | 0.601 |  |  |
| (b) *S*HDS | (Residuals) | 0.62 |  |  |  |
|  | MAP | 0.19 | < 0.001 |  | \*\*\* |
|  | NDVI | 0.05 | < 0.001 |  | \*\*\* |
|  | Elevation | 0.04 | < 0.001 |  | \*\*\* |
|  | Clay | 0.02 | 0.008 |  | \*\* |
|  | PDQ × Region | 0.01 | 0.038 |  | \* |
|  | PDQ | 0.01 | 0.050 |  | \* |
|  | pH × Region | 0.01 | 0.067 |  | ~ |
|  | Elevation × Region | 0.01 | 0.081 |  | ~ |
|  | CEC × Region | 0.01 | 0.095 |  | ~ |
|  | MAP × Region | 0.01 | 0.102 |  |  |
|  | pH | 0.01 | 0.126 |  |  |
|  | Soil C × Region | <0.01 | 0.314 |  |  |
|  | Region | <0.01 | 0.458 |  |  |
|  | CEC | <0.01 | 0.518 |  |  |
|  | Soil C | <0.01 | 0.611 |  |  |
| (c) *S*DS | (Residuals) | 0.31 |  |  |  |
|  | MAP | 0.27 | < 0.001 |  | \*\*\* |
|  | Elevation | 0.12 | < 0.001 |  | \*\*\* |
|  | Clay | 0.05 | 0.008 |  | \*\* |
|  | Clay × Region | 0.05 | 0.008 |  | \*\* |
|  | PDQ | 0.04 | 0.015 |  | \* |
|  | Soil C | 0.04 | 0.016 |  | \* |
|  | Surface T × Region | 0.04 | 0.020 |  | \* |
|  | Elevation × Region | 0.03 | 0.034 |  | \* |
|  | pH × Region | 0.03 | 0.035 |  | \* |
|  | pH | 0.02 | 0.123 |  |  |
|  | PDQ × Region | 0.01 | 0.143 |  |  |
|  | Surface T | <0.01 | 0.475 |  |  |
|  | Region | <0.01 | 0.644 |  |  |

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

**Table S4:** Comparisons of the standard deviations (*SD*) of residuals from PC1-based and multivariate (MV) models using datasets both including excluding vascular plant species richness hotspots across the GCFR and SWAFR, across the three spatial scales. Hotspots excluded from each model were those with residuals greater than two standard deviations from the mean for that model. *F*-tests of the ratios of GCFR to SWAFR *SD*s in each case were all significant (*P* < 0.01) except at the DS-scale (c) when species richness hotspots were excluded (denoted by † and ‡).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | *SD* of model residuals | | | | |
|  |  |  | Including hotspots | |  | Excluding hotspots | |
| Scale | Region |  | PC1 | MV |  | PC1 | MV |
| (a) QDS | GCFR |  | 335.17 | 315.52 |  | 233.45 | 222.60 |
|  | SWAFR |  | 247.39 | 230.05 |  | 198.36 | 174.13 |
| (b) HDS | GCFR |  | 607.10 | 540.17 |  | 467.83 | 437.31 |
|  | SWAFR |  | 387.00 | 337.31 |  | 343.84 | 299.52 |
| (c) DS | GCFR |  | 965.29 | 638.38 |  | 665.02 † | 383.79 ‡ |
|  | SWAFR |  | 558.41 | 353.93 |  | 554.68 † | 336.43 ‡ |

# Figures



**Figure S1:** Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels) and distributions (diagonal panels) of different forms of (log10-transformed) environmental heterogeneity (QDS-scale) across the GCFR and SWAFR. Abbreviations follow that in Tables S1, S2 and S3.



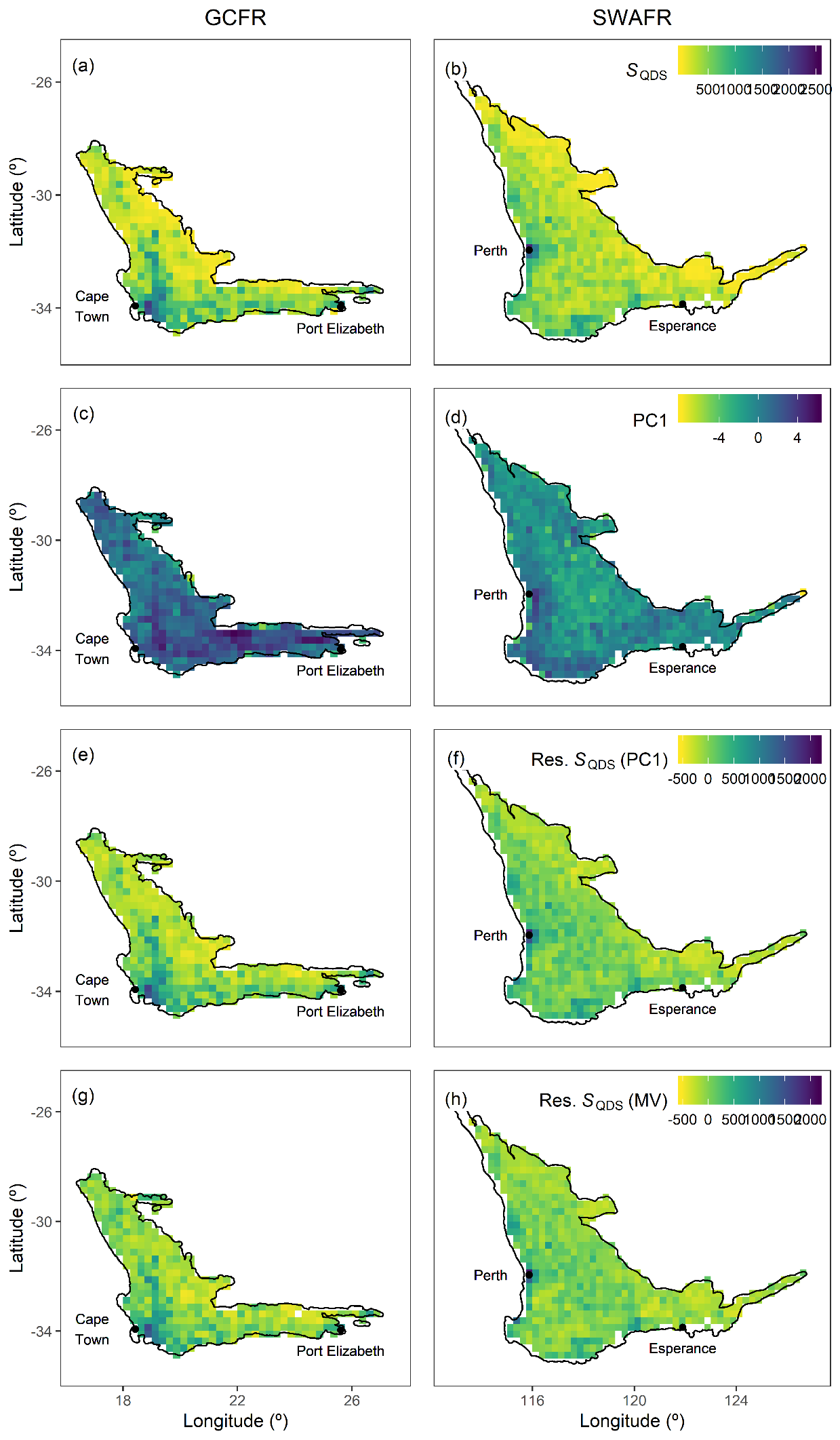
**Figure S2:** Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels) and distributions (diagonal panels) of different forms of (log10-transformed) environmental heterogeneity (HDS-scale) across the GCFR and SWAFR. Abbreviations follow that in Tables S1, S2 and S3 and Figure S1.



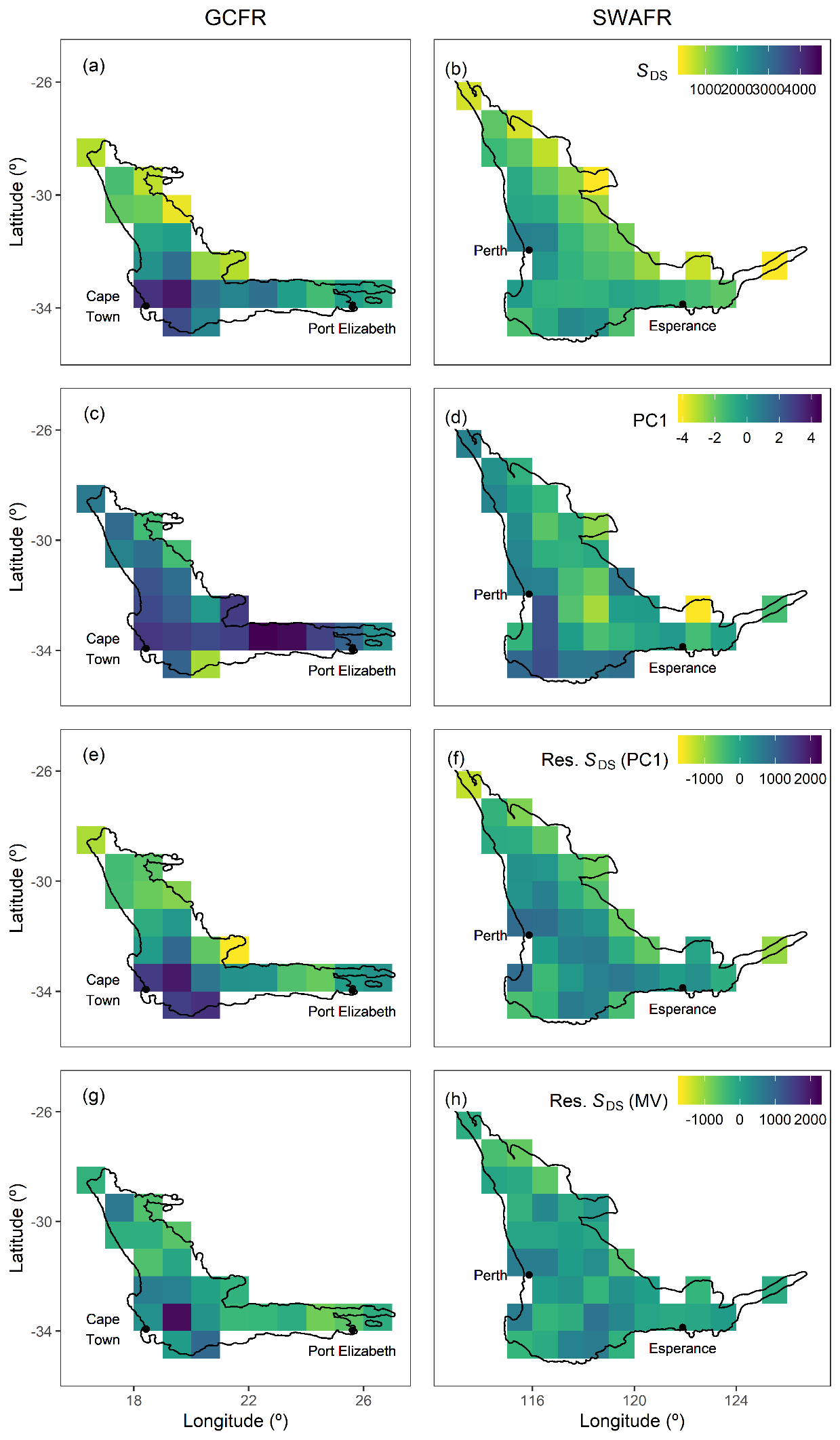
**Figure S3:** Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels) and distributions (diagonal panels) of different forms of (log10-transformed) environmental heterogeneity (DS-scale) across the GCFR and SWAFR. Abbreviations follow that in Tables S1, S2 and S3 and Figure S1 and S2.



**Figure S4:** Scatter plots of the first (PC1) and second (PC2) axes following PCAs of the nine forms of environmental heterogeneity (log10-transformed and re-scaled) across the GCFR and SWAFR, calculated at the (a) 0.10°×0.10°-, (b) QDS-, (c) HDS- and (d) DS-scales. The percentage of variance in environmental heterogeneity explained by each axis is noted in parentheses in each panel. Arrows for each heterogeneity variable show each variable’s associations with PC1 and PC2.



**Figure S5 (previous page):** QDS-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.



**Figure S6 (previous page):** DS-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.



**Figure S7 (previous page):** Scatter plots and regressions of vascular plant species richness (*S*) at the QDS-, HDS- and DS-scales against (a–i) different axes of environmental heterogeneity (log10-transformed) and (j) overall environmental heterogeneity (PC1) across the GCFR (red, where applicable) and SWAFR (blue, where applicable). The fitted lines are from the best-fitting univariate regressions of vascular plant species richness against each from of heterogeneity (Table 2).



**Figure S8 (previous page):** Grid-cells identified as having outstanding vascular plant species richness (i.e. hotspots) across the GCFR and SWAFR, following univariate regressions (Table 2, Figure 3) against the major axis of environmental heterogeneity (PC1) from the PCA (Figure S4). Hotspots were identified as those cells with residual richness greater than two standard deviations from the mean for that model. No hotspots were found for the SWAFR at the DS-scale. Map projection used: WGS84.

****

**Figure S9 (previous page):** Grid-cells identified as having outstanding vascular plant species richness (i.e. hotspots) across the GCFR and SWAFR, following multivariate regressions (Table S2, Figure 4) against the the nine axes of environmental heterogeneity. Hotspots were identified as those cells with residual richness greater than two standard deviations from the mean for that model. No hotspots were found for the SWAFR at the DS-scale. Map projection used: WGS84.



**Figure S10:** […] [As for Figure 4 but without outliers]



**Figure S11:** […] [As for Figure 3 but without outliers] [note QDS and HDS have additive models as best fitting]

# 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.org (24 July 2017a) GBIF Occurrence Download <https://doi.org/10.15468/dl.n6u6n0>

GBIF.org (24 July 2017b) GBIF Occurrence Download <https://doi.org/10.15468/dl.46okua>

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: […]