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].

# Supplementary 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.

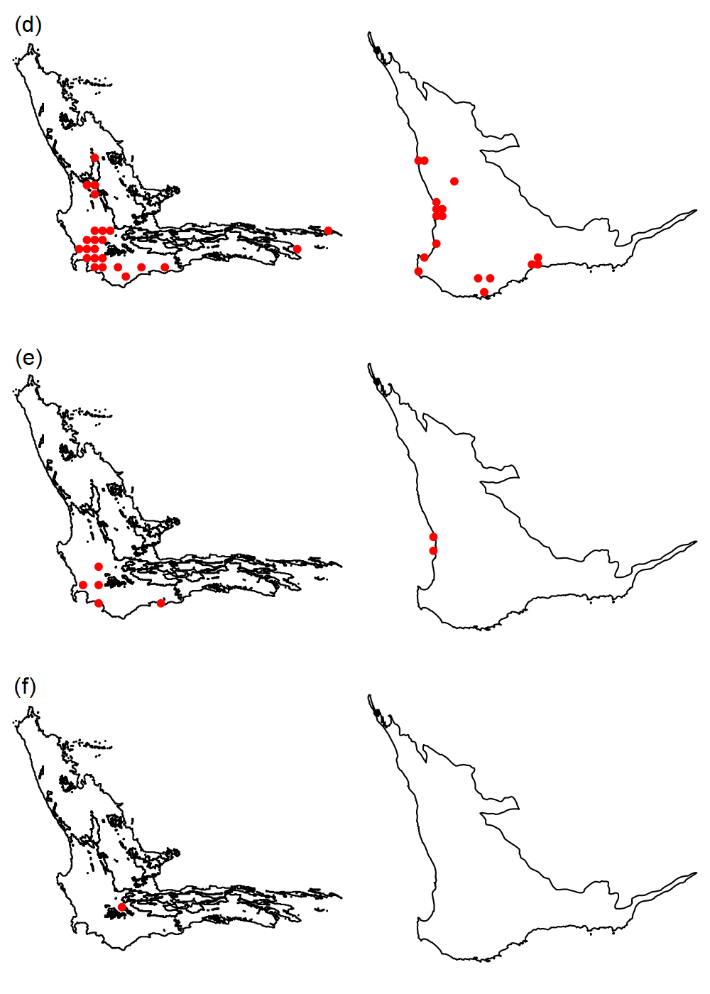
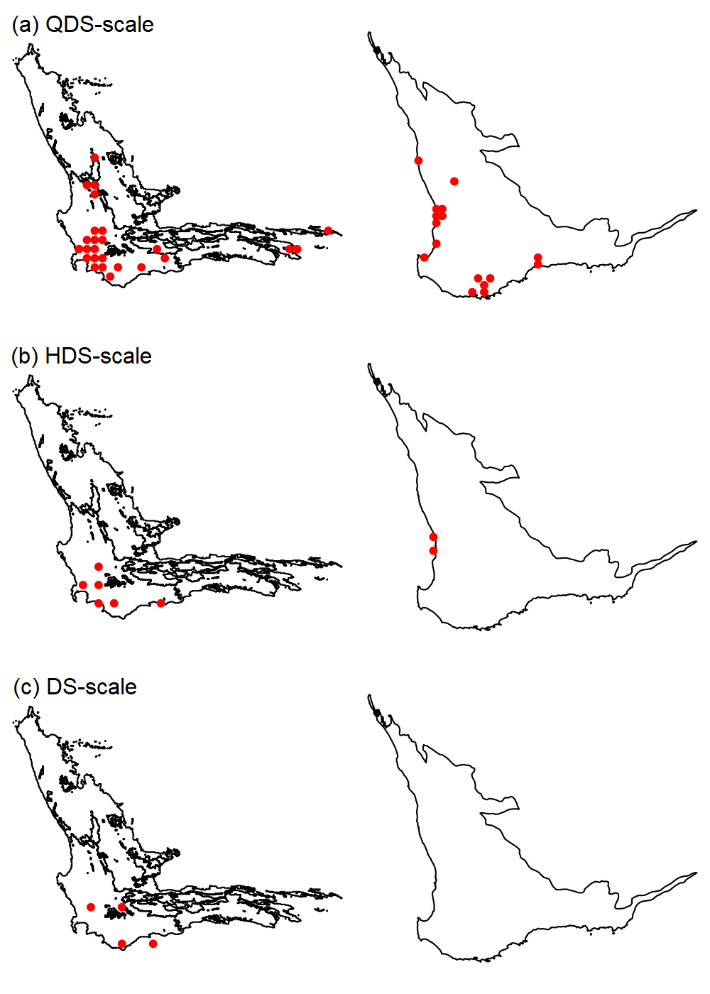
**Table S2:** […] ANOVA for mv models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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.

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.

# Supplementary figures



**Figure SXX:** […] (a) to (c) PC1 outliers; (d) to (f) mv outliers



**Figure SX:** […] QDS-scale



**Figure SY:** […] HDS-scale



**Figure SZ:** […] DS-scale



**Figure SZZ:** […]

# Supplementary references

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

Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45, 1–33. DOI: […]

Funk, C.C., Peterson, P.J., Landsfeld, M., Pedreros, D.H., Verdin, J., Shukla, S., Husak, G., Rowland, J.D., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066. DOI: […]

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

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

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti?, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12, e0169748. DOI: […]

NASA (2017a). Vegetation indices monthly l3 global 0.05Deg cmg (MOD13C2) v[Version]. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, U.S.A. DOI: […]

NASA (2017b). Land surface temperature/emissivity monthly l3 global 0.05Deg cmg (MOD11C3) v[Version]. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, U.S.A. DOI: […]