# Materials and methods

## Comparing regions’ environmental heterogeneity

In order to compare the EH in both regions, we acquired a broad suite of geospatially-explicit environmental data in the form of raster layers (Table 1). We used satellite-derived environmental data as far as possible in this work to minimise differences in data quality and methodologies between the GCFR and SWAFR. [Is this needed?: “SRTM and MODIS are entirely derived from satellite measurements (refs), whereas CHIRPS is interpolated from weather station data with satellite-derived radiometric measurements (refs). SoilGrids250m is a machine-learning derived product, based on soil measurements as a function of many covariates, including MODIS and STRM sources, using gradient-boosting [see Hengl et al. (2017); other refs]”]. For the soil data considered here, we used the depth-interval weighted average values for each particular soil variable. We summarised climatic and spectral data with their mean annual values. Pronounced seasonality of rainfall is a known feature of mediterranean systems (ref). Consequently, we describe this seasonality by computing computing the precipitation in the driest quarter (PDQ), as in the “biovars” function in the R package “dismo” (ref). These data were re-projected to a common coordinate reference system (WGS84; ref) using the “rgdal” (Bivand et al., 2017) package in R (R Core Team, 2018). All data were then re-sampled to 0.05º resolution using the “resample” function in the R package “raster” (Hijmans, 2016), with the “bilinear” method. The GCFR was treated here as the area occupied by the Succulent Karoo and Fynbos biomes in the current delineation of South Africa’s biome boundaries (Mucina & Rutherford, 2006). The SWAFR was treated as the areas occupied by the Southwest Australia savanna, Swan Coastal Plain Scrub and Woodlands, Jarrah-Karri forest and shrublands, Southwest Australia woodlands, Esperance mallee, and Coolgardie woodlands in the World Wildlife Fund Terrestrial Ecoregions dataset (Olson et al., 2001) in order to closely match the currently delineated SWAFR (Gioia & Hopper, 2017, Hopper & Gioia (2004)).

We developed an index of EH that would account for the spatial configuration of different environmental conditions. Our index, as applied to raster data, employs a moving-window function of nine pixels, representing eight neighbour pixels about a focal pixel (*x*\*). Our index represents the average magnitude of the difference between pixels’ conditions and that of their neighbours, as follows:

We only applied this index when pixels had at least one neighbour pixel. We implemented this using the “focal” function in the R package “raster” (Hijmans, 2016).

Accounting for the spatial scale of measurements is an important consideration in ecological studies (refs). As such, we compared EH in both regions across five principal spatial scales: 0.05º x 0.05º squares (the finest common resolution among the environmental data sources used), eighth degree squares (EDS), quarter degree squares (QDS), half degree squares (HDS) and three-quarter degree squares (3QDS) (sensu Larsen et al., 2009). To do this, the absolute environmental raster data were aggregated to broader spatial scales from 0.05º x 0.05º, using the “aggregate” function in the R package “raster” (Hijmans, 2016), with the “mean” method. We then converted these to rasters of EH using Equation 1. Note, the QDS- and HDS-grids exhibited in these data are not exactly the same grids as those described by Larsen et al. (2009), but are of equal spatial grain. This is because the “aggregate” function was applied to the original 0.05º x 0.05º-scale rasters, averaging pixels in groups of 4, and repeating this multiple times to reach 3QDS scale.

Finally, we also summarised the overall EH in both regions using principle component analysis (PCA). At each of the five scales, rasters of EH were ordinated following log(x + 1)-transformations to ensure normality. The first principle component (PC1) from each of the five PCAs represents the major axis of heterogeneity among the environmental axes considered here. We then also compared these PC1-values between both regions.

## Comparing and decomposing regions’ species richness

In order to compare the species richness and turnover of both regions’ floras, we downloaded geospatially-explicit records of vascular plant occurrences from the Global Biodiversity Information Facility (GBIF, see Table 1). Queries were made for tracheophyte records from within the borders of the GCFR and SWAFR as treated here. Only records with defined species and intra-specific ranks were kept. Intra-specific occurrences were treated as simply being representative of their species. This resulted in XXX and XXX unique species names in the GCFR and SWAFR respectively. We cleaned these data using the R package “taxize” (Scott Chamberlain & Eduard Szocs, 2013; Chamberlain et al., 2018) to check that these species names had accepted-status among taxonomic databases. We queried two major taxonomic databases: the Global Name Resolver (GNR) and the Taxonomic Name Resolution Service (TNRS). Should either one of these services return at least one match for a given name, then that name was accepted. Those names for which no full binomial matches were found in either database were excluded. The number of species names excluded totalled at XXX and XXX for the GCFR and SWAFR respectively. Especially for the SWAFR, these numbers may be deemed appreciably high. But, the occurrence records that would be dropped, as a consequence of these names’ removals, appeared randomly distributed in geographic space in both regions. As such, any effect of the loss of these records in this analysis is likely uniform within the two regions. After the unaccepted names were removed, it was important to ensure that a species was not listed under multiple synonyms in different areas. Such cases would skew estimates of species richness and turnover in this study. In light of this, the remaining names were queried in the Tropicos and Integrated Taxonomic Information System (ITIS) databases for their known synonyms, again using “taxize”. These were used to amend species’ names in the occurrence dataset. We also removed all records of invasive alien or non-indigenous species from both regions, based on existing lists of invasive plants for South Africa and Australia from the IUCN’s Global Invasive Species Database (<http://www.iucngisd.org/gisd/>). The final total plant species richness in each region was XXX and XXX for the GCFR and SWAFR respectively.

We converted these final collections of species occurrence records to raster-layers in R, where pixels’ values represent their species richness. These rasters were produced at the QDS and HDS (sensu Larsen et al., 2009) scales only, as most plant locality data is only recorded to the QDS-scale [and 3QDS? Is this too coarse to be useful or should we do it?]. These richness values (*S*) were compared across both regions. Additionally, we decomposed the richness of pixels (e.g. *S*HDS), following Whittaker’s (ref) original additive formulation of -diversity (). The difference between a pixel’s richness and the average richness of its constituent pixels (e.g. QDS) represents the floristic turnover (*T*QDS), as follows from:

## Environmental heterogeneity as an explanation of species richness

We investigate the explanatory power of various forms of EH with linear models of species richness as a function of EH. To this end, we used the species richness data collated at the QDS- and HDS-scales and measures of EH in those same QDS and HDS. We generated this EH data separately from that described above, in order to align with the Larsen et al. (2009) QDS- and HDS-grids proper. We defined this second EH index as the mean of the mean absolute differences in environmental conditions between each constituent QDS in an HDS (and EDS in a QDS) and the other three squares within that HDS (or QDS for the case of EDS):

Taking QDS within HDS as an example, this index of EH is the result of 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 there are four QDS in an HDS) are averaged. This index allows each of the four values to be similarly independent, and thus more suitable for our averaging and analyses. These values were then log(x + 1)-transformed to ensure residual normality for use in linear models. We also repeated the PCA on these transformed data to produce PC1 values analogous to those described above [reword; avoid ‘above’].

In order to understand the relationships between environmental heterogeneity and species richness differ between the GCFR and SWAFR, we fit both simple and multiple linear regression models of *S*HDS and *S*QDS as functions of EH values. The rationale of the univariate models was to describe empirical patterns of covariance between the each axis of EH and species richness. The multivariate models allow us to account for differences in richness across multiple axes of EH simultaneously.

Table 1: Georeferenced environmental data and vascular plant species occurence data sources used in this study. Data were acquired for the GCFR and SWAFR regions, with the temporal extent of data products used described where applicable. Abbreviations are as follows: MAP, mean annual precipitation; PDQ, precipitation in the driest quarter; CEC, cation exchange capacity.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Source | Temporal extent | Citation(s) |
| Elevation | SRTM v2.0 |  | Farr et al. (2007) |
| NDVI | MODIS (MOD13C2) | Feb. 2000 to Apr. 2017 | NASA (?) |
| Surface temperature | MODIS (MOD11C3) | Feb. 2000 to Apr. 2017 | NASA (?) |
| MAP | CHIRPS v2.0 | Jan. 1981 to Feb. 2017 | Funk et al. (2015) |
| PDQ | CHIRPS v2.0 | Jan. 1981 to Feb. 2017 | Funk et al. (2015) |
| CEC | SoilGrids250m (“CECSOL M 250m”) |  | Hengl et al. (2017) |
| Clay | SoilGrids250m (“CLYPPT M 250m”) |  |  |
| Soil carbon | SoilGrids250m (“OCDENS M 250m”) |  |  |
| pH | SoilGrids250m (“PHIKCL M 250m”) |  |  |
| Plant species occurrences | GBIF |  | GBIF (24 July 2017), GBIF (24 July 2017) |

## References

Bivand, R., Keitt, T., & Rowlingson, B. (2017) rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-7..

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O’Donnell, J., Oksanen, J., Tzovaras, B.G., Marchand, P., & Tran, V. (2018) *Taxize: Taxonomic information from around the web.*

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.

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.

GBIF (24 July 2017) GBIF Occurrence Download..

GBIF (24 July 2017) GBIF Occurrence Download..

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**, 1–15.

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.

Hijmans, R.J. (2016) raster: Geographic Data Analysis and Modeling. R package version 2.5-8..

Hopper, S.D. & Gioia, P. (2004) The Southwest Australian Floristic Region: Evolution and Conservation of a Global Hot Spot of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 623–650.

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**, 382–392.

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

NASA (?) 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*,.

NASA (?) 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*,.

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**, 933–938.

R Core Team (2018) *R: A Language and Environment for Statistical Computing. Version 3.5.0.* R Foundation for Statistical Computing, Vienna, Austria.

Scott Chamberlain & Eduard Szocs (2013) Taxize - taxonomic search and retrieval in r. *F1000Research*,.