Supplement 1: Study area, data and covariates

Peter A. Vesk, William K. Morris, Will C. Neal,

Karel Mokany & Laura J. Pollock

January 2020

Study area

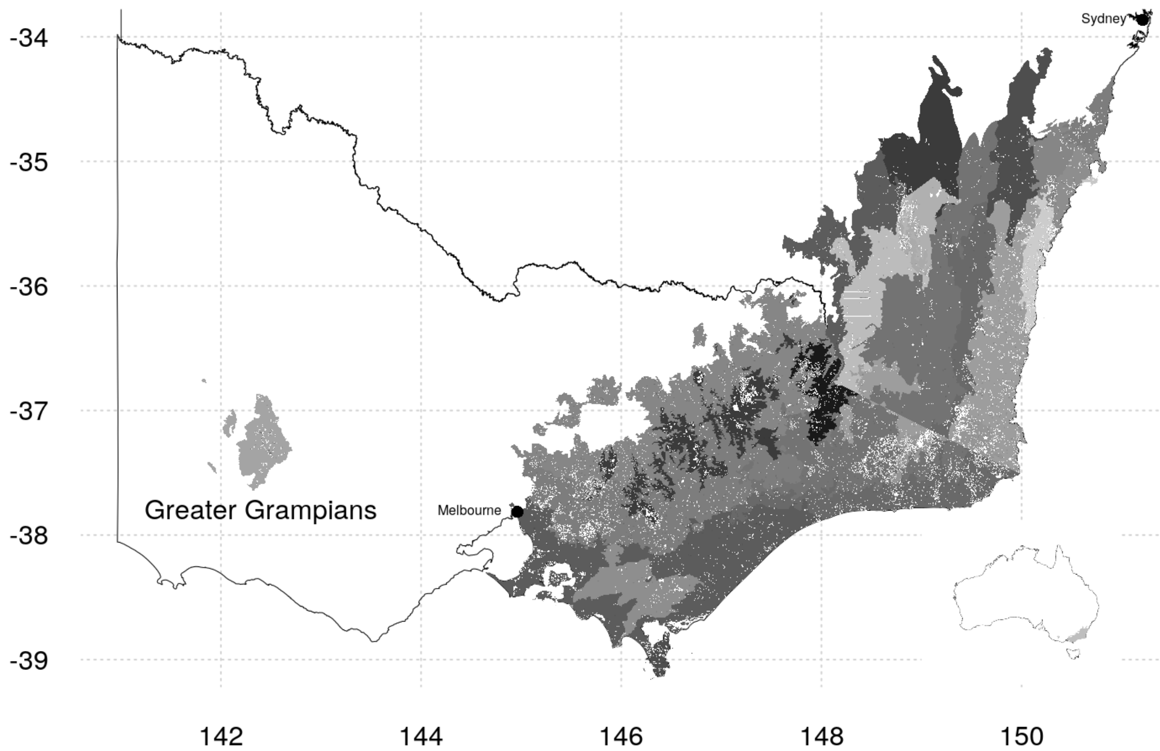

Figure S1.1: Boundaries of IBRA subregions within southeastern Australia and the plot locations within them (white dots).

Table S1.1: IBRA subregions

Region	Area (km^2)	No. of plots	No. of taxa
Gippsland Plain	12500	2442	39
Highlands-Southern Fall	12000	3393	47
Highlands-Northern Fall	14200	1740	34
Strzelecki Ranges	3400	321	21
Victorian Alps	5200	2758	30
Wilsons Promontory	400	265	10
East Gippsland Lowlands	6200	2660	40
South East Coastal Ranges	17300	5267	72
Bondo	5400	664	19
Snowy Mountains	7100	1481	24
Monaro	12700	727	33
Murrumbateman	6300	565	23
Kybeyan-Gourock	4800	1628	42
Bungonia	4300	256	34
Bateman	1700	535	33
Ettrema	1800	384	46
Jervis	1400	439	27
Illawarra	1200	189	28
Greater Grampians	2400	457	20

Selecting covariates

Our objective was to identify a subset of maximally independent covariates that best discriminated between plot species compositions. To do this we used an approach based on identifying discriminatory power of covariates and cluster analysis to construct a dendrogram of covariates (Q-model analysis sensu P. Legendre & L. F. J. Legendre 2012).

Candidate covariates were obtained from the Soil and Landscape Grid of Australia (Grundy *et al.* 2015) and the NSW and ACT Regional Climate Modelling (NARCliM) project (Evans *et al.* 2014).

First, we performed a hierarchical cluster analysis using a Raup distance on the the Grampians plots, choosing 20 clusters to reflect the 20 taxa in the dataset. Thus, we obtained clusters that tended to be dominated by each of the 20 species, respectively. Attached to these clusters were all the environmental covariates we had. We then calculate F-ratios of between- to within-cluster variance from a generalised linear mixed model using the R (R Core Team 2019) package lme4 (Bates *et al.* 2015).

We next performed a hierarchical cluster analysis (Chavent *et al.* 2012) of the environmental covariates based on their correlations across the plots. We then plotted the dendrogram from that cluster analysis, attaching the F-ratios to the labels of the covariates (Fig. S1.2). This plot revealed the relationships among the covariates. We used this dendrogram to sequentially identify clusters of covariates that were most different. Then we could choose among the covariates within a cluster based on higher F-ratios, representing better discriminatory power. We bypassed the radiation-related

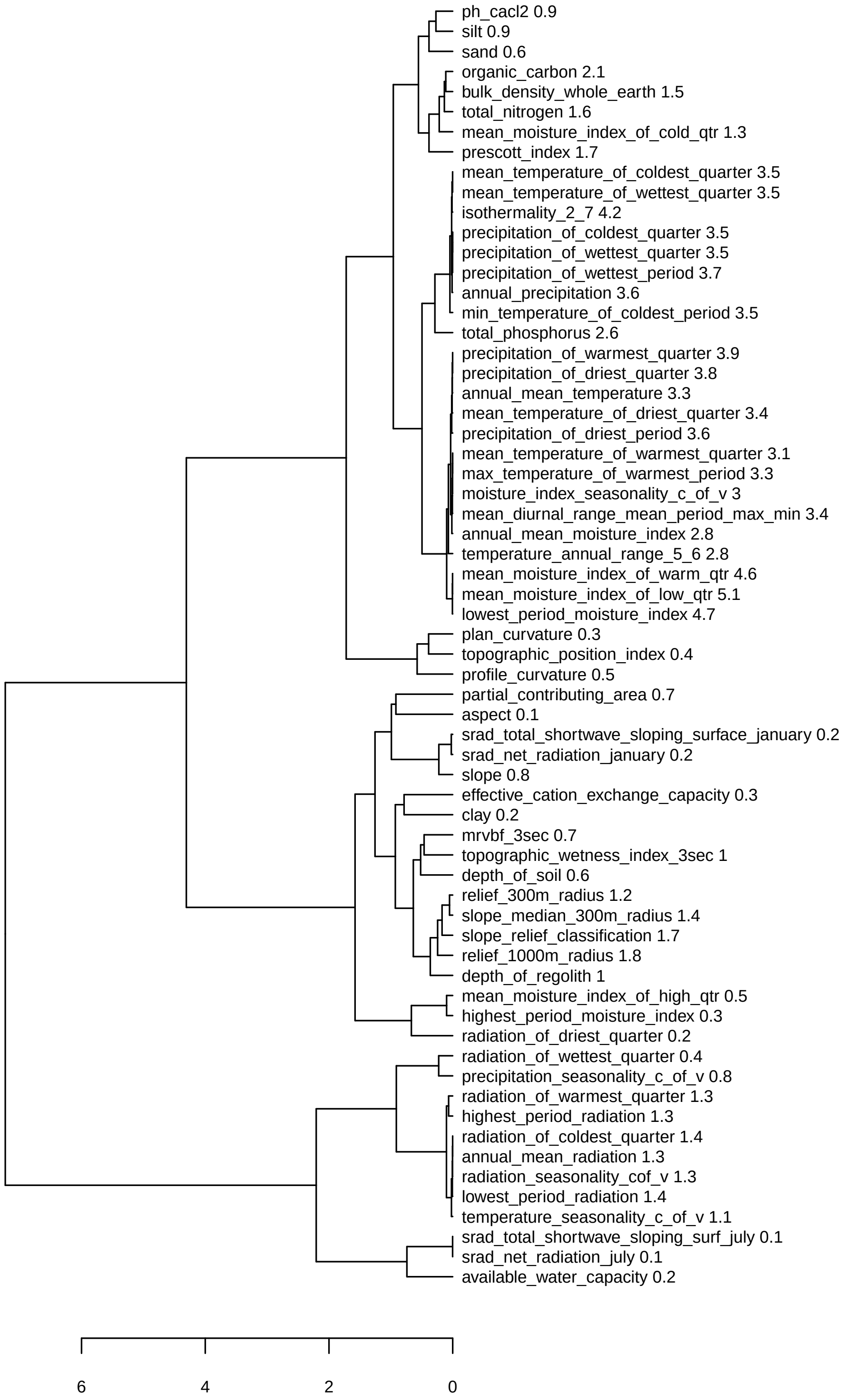

Figure S1.2: Hierarchical cluster analysis of covariates data for the Grampians eucalypt occurrence plots. Labels show the F-ratio of each variable denoting that variables ability to discriminate among eucalypt communities in the Grampians.

variables at the bottom of the dendrogram as they tended to be diffuse N-S gradients, that we could see no direct physiological mechanism for. Temperature and precipitation variables clustered together and with moisture index, leading use to chose mean moisture index of the low quarter first (F=5.1). Relief within 1000m radius (F=1.8) was then selected from a cluster of topographic variables reflecting ruggedness. At the top of the dendrogram, Total Nitrogen (F=1.6) was selected from a cluster of soil-related variables that reflected areas with high bulk density, supporting organic soils. Finally, Topographic Wetness Index (F=1.0) was selected from a cluster of variables reflecting landscapes shedding or collecting water, including multi-resolution valley bottom flatness and soil depth.

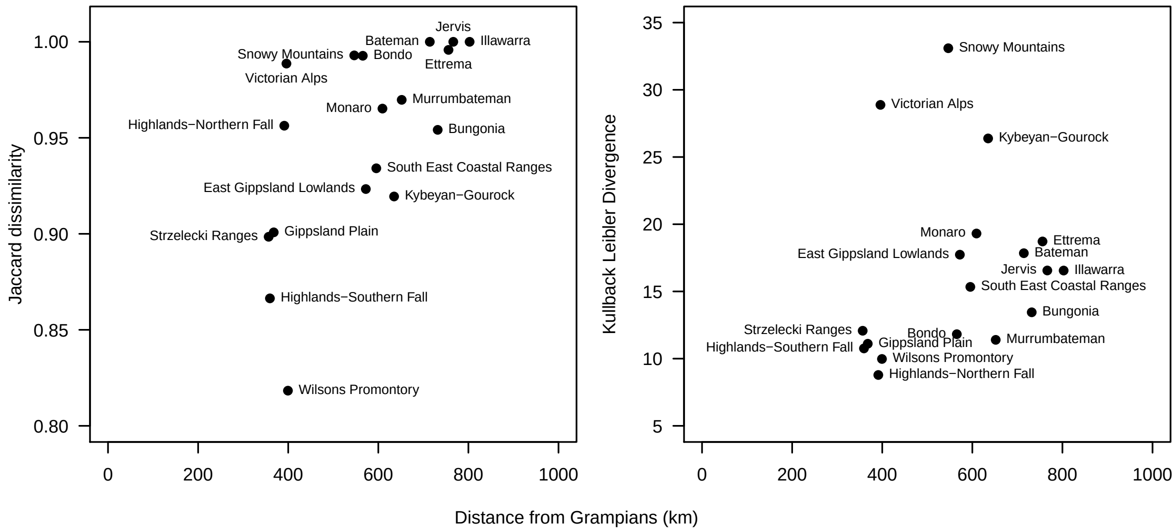

Figure S1.3: Distance of each region from the Grampians, measured as: (x-axis) geographical distance (km) between the centroid of the polygon delimiting the Grampians and the target region, and (left panel, y-axis) the Jaccard dissimilarity between the eucalypt community of the Grampians and the target region, and (right panel, y-axis) the Kullback-Leibler divergence across the four environmental covariates.

The distribution of our chosen environmental covariates across the Grampians region and the test bioregions of southeastern Australia can be seen in Fig. S1.4. Scatterplots (Fig. S1.5) and a principal components analysis (Fig. S1.6) illustrate the coverage of the environmental covariates in the Grampians and southeastern Australia. The Grampians tends to be drier, with low soil nutrients and greater sand content. Environments not well covered in the Grampians include high moisture in general but also in combination with low topographic relief or low nitrogen (Fig. S1.5).

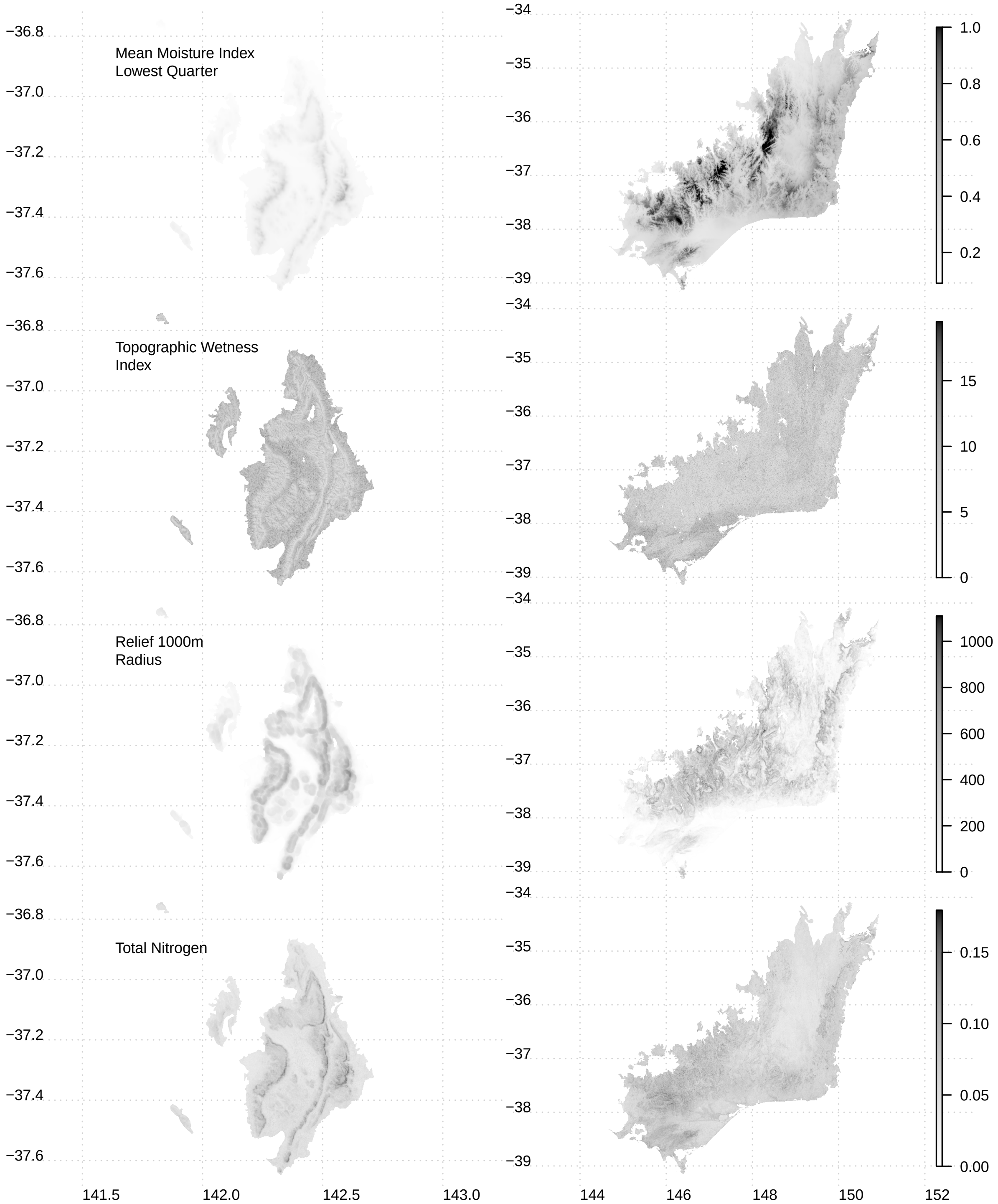

Figure S1.4: Chosen model covariate values across the Greater Grampians and the broader southeast region.

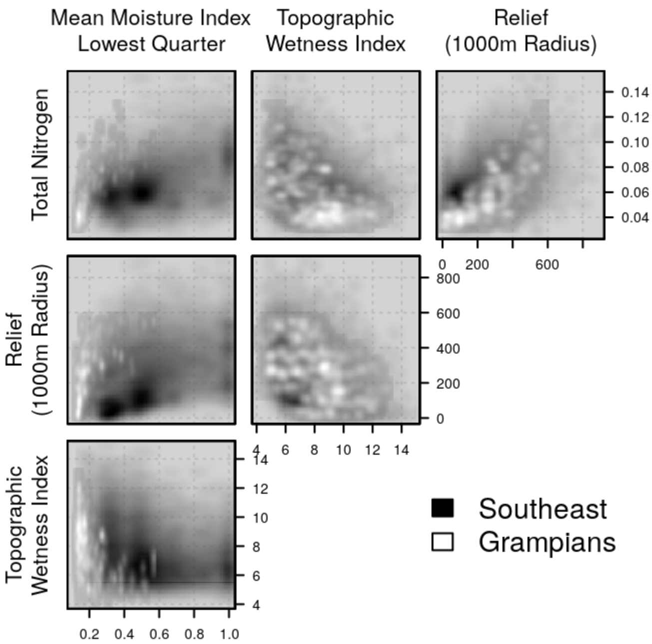

Figure S1.5: Bivariate relationship among the environmental covariates used in the trait-environment model for the Grampians (green) and the South-east regions as a whole (purple). Point clouds are derived from a kernel density estimation.

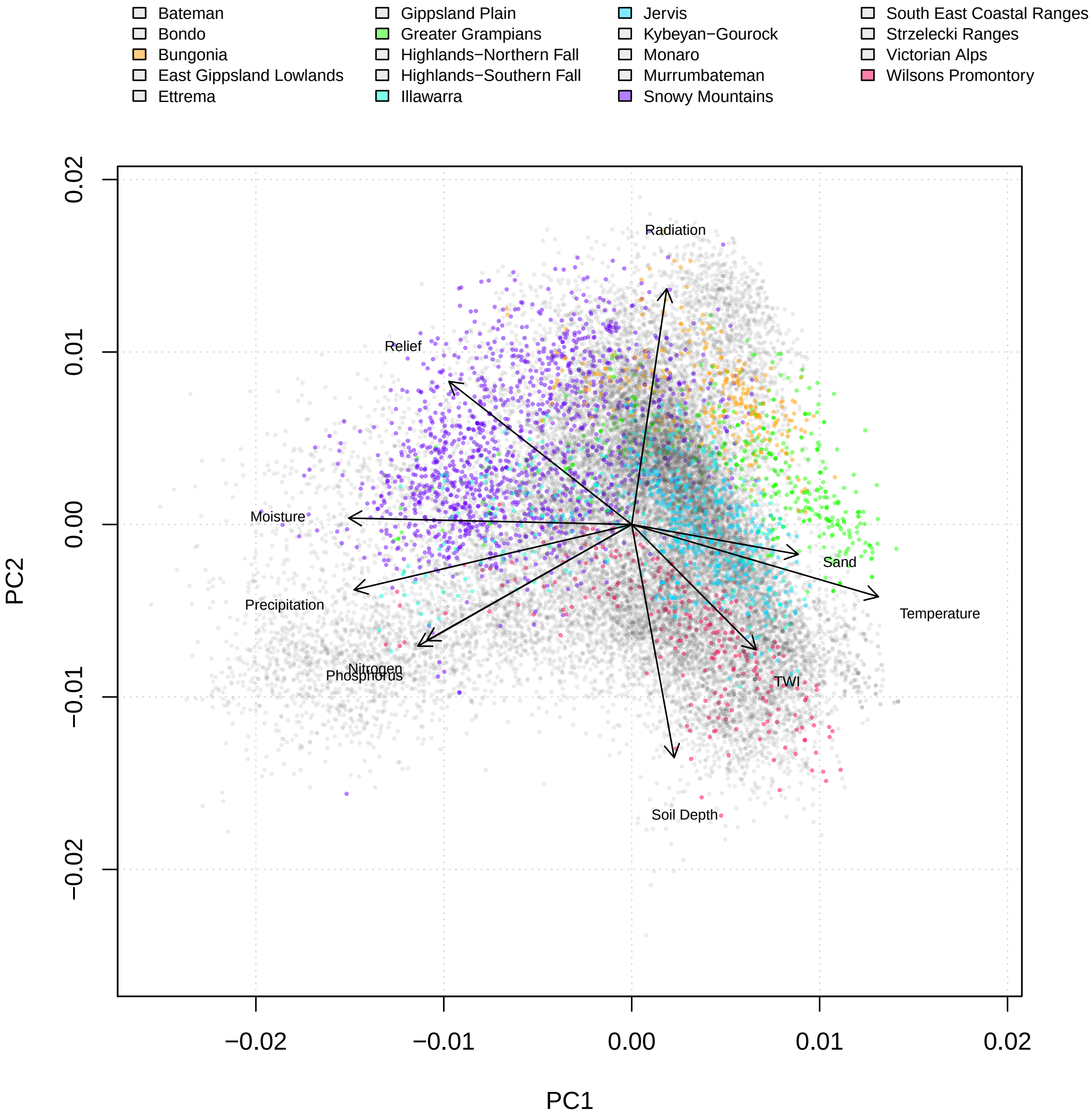

Figure S1.6: Principal components analysis of some important dimensions of environmental space. Each point represents a sample plot where eucalypt species are present. The plots of some regions have been highlighted for comparison. Note the Grampians, where models were fitted, in green.