

Disentangling the drivers of local species richness using probabilistic species pools

Running title: Probabilistic species pools

Karger, Dirk Nikolaus¹; Wüest, Rafael O.¹, König, Christian²; Sarmento Cabral, Juliano³; Weigelt, Patrick²; Zimmermann, Niklaus E.¹; Linder, H. Peter⁴,

1. Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

2. Biodiversity, Macroecology & Biogeography, University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany

3. Ecosystem Modeling, Center for Computational and Theoretical Biology (CCTB), University of Würzburg, Emil-Fischer-Str. 32, 97074 Würzburg, Germany

4. Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland

Corresponding author: Dr. Dirk N. Karger (dirk.karger@wsl.ch) Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

Tweet: Local species richness is determined by regional species pool size, and additionally constrained by dispersal filters, and environmental suitability. This local pool is then further filtered by local site variables such as wet habitats or shallow soils, to create the observed diversity we observe at a very local scale.

ABSTRACT

Aim: Local richness is a result of both regional richness and local site-specific factors. We quantify the effects of different regional (dispersal, environmental filtering) and local (habitat selection, biotic interactions) community assembly processes in the Cape reeds (*Restionaceae*) of the Cape flora and test if community assembly processes vary spatially.

Location: Southern Africa, Cape Floristic Region

Taxon: Angiosperms, African *Restionaceae* (restios)

Methods: We calculate local probabilistic species pools, based on local species richness, dispersal rates and environmental filtering at a 4 km² resolution. We then compare the relative contribution of the processes shaping the species pool spatially. Finally, we correlate observed species richness (based on 916 circular, 10 m in diameter relevés) with the local species pools to test the explanatory power of additional local edaphic variables that could act as filters at very small scales.

Results: There is substantial spatial variation in the local probabilistic species pools. Within the central part of the Cape, the local pools are mostly influenced by environmental filtering, while along the periphery dispersal limitation dominates. Observed species richness is lower than the predicted species pool size, which forms an upper boundary of the potential species richness for a local site. Local, site-specific edaphic factors help to explain the remaining differences between local species pool size and α -richness.

Main conclusion: The analytical framework for inferring local probabilistic species pools provides a powerful tool for including the influence of regional species richness on variations in relevé α -

richness. This allows us to disentangle the various processes that operate at different spatial scales, and ultimately determine local restio species richness in the Cape Floristic Region.

Keywords: α -diversity, Cape flora, environmental limits, community assembly, regional species pools, Restionaceae

INTRODUCTION

Community richness and composition are shaped by both regional richness and local site-specific factors. In the past two decades it has increasingly become accepted that the regional species pool provides a source of potential local species, and that different processes filter these regional pools into local communities (Cornell & Harrison, 2014). This simple model can be expanded by including species interactions, resulting in what (Lortie, Brooker, Choler *et al.*, 2004) refer to as an “integrated community concept”. The relative contributions of filters related to dispersal, environment, or biotic interactions have attracted much attention and it has become increasingly obvious that all these filters need to be accounted for in an hierarchal fashion to obtain an assessment of their relative importance for building local communities.

Regional richness is the result of three mechanisms: immigration, speciation and extinction (Cornell & Harrison 2014; Ricklefs, 1987). These mechanisms are, in turn, modulated by many different drivers and constraints, including available area, environmental heterogeneity, time, and degree of isolation (Ibanez, Keppel, Baider *et al.*, 2018). The potential composition of any community (i.e. the local species pool, *sensu* Zobel, 1997) is that subset of the regional biota which can (a) reach that focal point, (b) survive in that environment, and (c) is not excluded due to biotic interactions with co-occurring species (Cornell *et al.*, 2014; Lessard, Belmaker, Myers *et al.*, 2012; Zobel, 1997). While the regional richness and the associated local species pools are spatially defined for islands and archipelagos (Ibanez *et al.*, 2018), it is much more difficult to define the local species pools of continental areas. Early approaches used predefined biogeographical regions or fixed environmental limits (e.g. Wüest, Litsios, Forest *et al.*, 2016) to delimit local species pools. Such approaches are, however, categorical, and use arbitrary region boundaries (Karger, 2016). Recently, two approaches to define probabilistic species pools of any focal site have been developed. Both determine the probability that a species can establish in the focal site from the dispersal ability and current distribution of the species and the suitability of the local habitat but differ in the way they take dispersal into account (Karger, 2016; Karger, Cord, Kessler *et al.*, 2016; Lessard, Weinstein, Borregaard *et al.*, 2016).

When calculated probabilistically, the local species pool considers the size of the regional species pool and the factors determining it, and so can be used to test a wide array of hypotheses linked to community assembly (Cornell & Harrison 2014; Harrison *et al.*, 2008, Pärtel, Szava-Kovats, & Zobel, 2013). These include, for example, whether habitats influence species richness, whether there are phylogenetic biases in community assembly (phylogenetic community ecology, testing for phylogenetic clustering or over-dispersion) (Lessard *et al.*, 2016), and whether there are geographical or trait biases (Lessard *et al.*, 2016).

Despite extensive research on the Cape flora of southern Africa (summarized in Allsopp, Colville & Verboom, 2014; Cowling, 1992) and its exceptional species richness (Linder, 2003; Manning & Goldblatt, 2012), we do not understand the processes that influence regional and local diversity in this region (Cowling, 1990; Cowling, Holmes & Rebelo, 1992; Kreft & Jetz, 2007). There is considerable regional variation in plant species richness (Manning *et al.*, 2012), and the regional diversity is higher in the west than the east. The latter pattern has been postulated to be due to different diversification histories (Cowling, Bradshaw, Colville *et al.*, 2017; Cowling & Lombard, 2002). However, α -richness

(i.e. relevé level) in the Cape flora has been reported to show no significant variation across the region as well (Cowling *et al.*, 1992).

Here we aim at disentangling the impact of dispersal limitation, environmental filtering, and biotic interactions on α -diversity. As a test case, we use the Restionaceae (hereafter restios) in the Cape Floristic Region (CFR). Restios are one of the ecologically dominant clades in the fynbos vegetation (Rebelo, Boucher, Helme *et al.*, 2006), which is the most species rich vegetation in the hyper-diverse Cape flora (Linder, 2003). We first predict the LPSPs, and compare the predicted variation in richness to that documented before for restios and other clades in the Cape flora. We do not explore however what forces regional richness (e.g. the drivers of speciation, extinction and dispersal – see Wüest, Boucher, Bouchenak-Khelladi *et al.* (2019). Specifically, we first disentangle the relative contributions of the environment and dispersal in shaping the spatial variation in local restio richness. Then we explore how the relative importance of environment and dispersal varies spatially across the study region. Finally, we use a regression approach to compare potential explanatory edaphic variables for their filtering effect on the α -diversity, when taking the local species pool into account.

MATERIALS AND METHODS

Distribution and environmental data

The geographical area of the study was delimited by the total range of all restios in South Africa. This is somewhat larger than the CFR (Manning *et al.*, 2012), including also Namaqualand north to the Kamiesberg, the escarpment mountains leading to the Kamiesberg, and the adjacent coastal plain (Figure 1). The restios from this whole region constitute the regional species pool or the regional richness.

Distribution data of all restio species were compiled from the herbarium records in the Bolus Herbarium of the University of Cape Town and the Compton Herbarium of the South African National Biodiversity Institute. In addition, 916 circular relevés were sampled for restios. This dataset was carefully checked for both identification and locality errors. The final dataset constituted 12'903 occurrence records for the 350 restio species including two infraspecific taxa.

Relevé data were collected across the whole study area (Figure 1) to document species co-occurrences at a local scale. With some exceptions the relevés were circular, 10 m in diameter, and each placed subjectively in a patch of homogenous vegetation. For each relevé all restio species were recorded (presence / absence data), in addition to a limited set of habitat variables, and the relevés were georeferenced using a GPS. The occurrence dataset, including the relevé data, is available on the DRYAD digital repository (<https://doi.org/10.5061/dryad.1cs77qn>).

Climate data were taken from CHELSA 1.2 (Karger, Conrad, Böhner *et al.*, 2017a; Karger, Conrad, Böhner *et al.*, 2017b), available at www.chelsa-climate.org. We selected a set of seven bioclimatic variables with seven additional climatic variables aggregated from monthly mean-, maximum-, and minimum temperatures, as well as monthly precipitation, plus a further 12 topographic variables and one geological layer. From the set of 27 available environmental predictor layers we selected a subset of predictors to avoid collinearity between predictors. A full set of the variables used and the methodology is described in Supplementary Information Appendix S1.

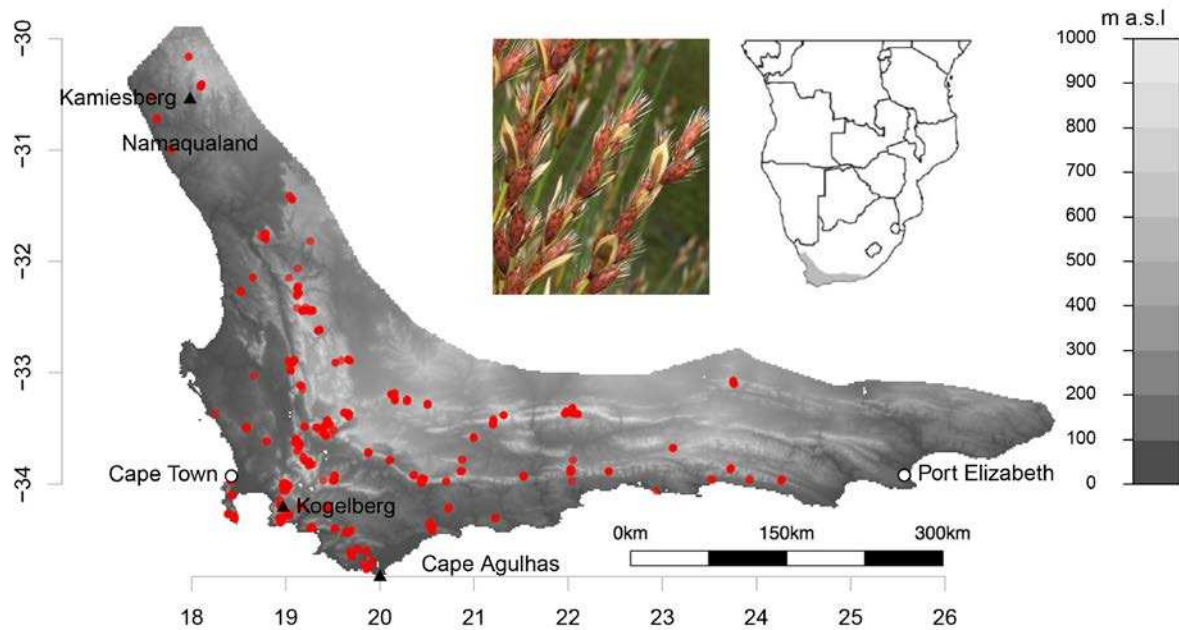


Figure 1: Study area at the southern tip of Africa, terrain elevation, and the locations of the relevés in red. The right inlet shows the location of the Cape Floristic Region with respect to the southern part of Africa. The left inlet shows *Hypodiscus aristatus* (© H. Peter Linder).

Geological data were obtained from the South African Council for Geoscience (downloaded in January 2018 from <http://www.geoscience.org.za>), and describe the geology of South Africa at a resolution of 1:1'000'000 using 496 classes. We simplified these to six classes - calcrete, granite, recent sand, silcrete, sandstones, shale, others (including water-bodies and classes outside the CFR; details of conversion see (Wüest *et al.*, 2019) – which we assume to reflect important soil characteristics for plants.

Building local probabilistic species pools

We used correlative species distribution models (SDMs) to estimate the potential distribution of restio species within the study area. We set the minimal number of occurrences per species to 30, consequently 130 out of the 350 restio species could be modeled (Table S1). In an ensemble approach, we used the four statistical models GLM (Nelder & Wedderburn, 1972), GAM (Hastie & Tibshirani, 1990), Random Forest (Breiman, 2001) and MaxEnt (Phillips, Anderson & Schapire, 2006) to model restio occurrence based on climatic and geological information. We outline in Supplementary Information Appendix S2 how we fitted the four models in R (R Development Core Team, 2017). All models were cross-validated using a split-sample approach, where the data were randomly split into a training part (70% of the data) and a testing part (30% of the data). We repeated the split sampling 100 times and evaluated predictive performance using the True Skills Statistic (TSS, Allouche, Tsoar & Kadmon, 2006). The probability of presence for each species was obtained as an ensemble by unweighted averaging the predicted habitat suitability values from the four models (Supplementary Information Appendix S2). We further generated binary maps, where probabilities of occurrence from each model and each species were transformed into presence-absence by applying a threshold that optimized TSS. A consensus map for each species was generated by summing up the binary maps, where consensus presence was assumed only if all models agreed on presence. The subsequent richness values were then aggregated from these individual models, as stacked SDMs (S-SDM - Guisan & Rahbek, 2011).

The probabilistic species pool size index in a grid cell(${}^i\psi$, *sensu* Karger *et al.* 2016, hereafter LPSP):

$${}^i\psi_{ED} = \sum_{s=1}^S \prod_{x=1}^n p_{xs}$$

was calculated by multiplying the probabilities (P_x) for n filters (E =Environment, D =Dispersal), with x being probabilities based on environmental or dispersal, for a species s , for each 2 x 2 km grid, i.e.: Where s is a species from the regional pool S , x is a probabilistic factor (here: environmental conditions and dispersal probability), and p_{xs} is the occurrence probability of species s in grid cell i given the probabilistic factors $x_{1...n}$. Environmental probabilities come from ensemble SDMs (see Supplement S2).

The environmental pool is derived by summing the probabilities of the stacked SDMs of all included species. As these are based on estimated species environmental niches, this reflects the regions in which species can occur based on their environmental preferences. By using only environmental variables (and neglecting the dispersal capacity and local population dynamics of species), SDMs generally predict potential ranges of the species that are larger than the realized ranges (see Cabral & Schurr, 2010; Schurr, Pagel, Cabral *et al.*, 2012).

To estimate the dispersal pool we first estimated the dispersal rate for each species using the R package “dispeRsal” (Tamme, Gotzenberger, Zobel *et al.*, 2014), which considers the dispersal syndrome (ant, wind, or ballistic), the plant height, and the growth form. This was turned into a dispersal distance by assuming that the average generation time of restios is 13 years, as they germinate only after fire events, assuming an average fire return time of 10 years (Kraaij & Van Wilgen, 2014), and flower after three years, and by allowing dispersal to run for 10k, 20k, 30k, 40k, 50k and 60k generations. As starting points for the dispersal we used the edges of the range maps generated as the intersection of the consensus SDMs and a polygon created from the occurrence points with a 5 km buffer.

Dispersal probabilities P_D were calculated assuming that each species has a probability of reaching a cell (n) based on its presence in a total of N occupied cells located distance d_n from each occupied cell:

$$P_{D,n} = 1 - \prod_{n=1}^N (1 - e^{-kd_n})$$

(Bischoff 2005) with N being the total number of cells occupied by the species, and k being a rate constant representing the dispersal ability of species over a given distance d and time.

To address the geographical patterns in the relative contributions of the three pools, we divided the local (cell-wise) species richness resulting from each pool by each other pool.

Table 1. Diversity at different spatial grains, their data sources, and major governing processes.

Name	Grain	Source	Processes and filters
Regional species pool (RSP)	Cape Floristic Region	All restios occurring in the Cape Floristic Region	Speciation, extinction and migration/dispersal, see Wüest <i>et al.</i> 2019
Local probabilistic species pool (LPSP)	2 x 2 km	Calculated “Probpool”	Environment and dispersal
α -diversity = observed relevé	78.5 m ²	Observed	local habitat variables at the relevé

Environmental variables at the scale of the relevé

As measure of observed relevé α -diversity, we used the species richness in each individual relevé. As α diversity is bounded by the LPSP, we needed to account for the variation in the LSP before analysing the unique contribution of the local habitat variables. We did so by using the difference between the LPSP and α -diversity (or “dark diversity”, *sensu* Pärtel 2011) as response in a generalized linear model with Gaussian error distribution.

To investigate the effect of local environmental factors, we used the richness predicted by the LPSP to account for the bounded diversity. Climatic variation cannot be seen as one of the local environmental factors, as many of the relevés are spatially clustered, less than 1 km apart, and climatic variation is at a much larger spatial scale. Furthermore, climate variables were already included in LPSPs. This leaves soil and bedrock variables that have been suggested to be important in the CFR (Linder, 2005). Bedrock type has long been seen as critical for determining variation in the vegetation, with fynbos vegetation (and restios) most commonly found on sandy soils derived from sandstones, quartzites, granites and calcretes, and shrubby renosterveld vegetation more commonly inhabiting heavier soils derived from shales (Rebello *et al.*, 2006). Ground-water availability is an important variable for Cape restios in particular (Araya, Silvertown, Gowing *et al.*, 2011; Huber & Linder, 2012) and for the Cape fynbos in general (Sieben, Boucher & Mucina, 2004). The rockiness, or degree of rock covering, has long been included in fynbos studies (e.g. McDonald, 1993a; McDonald, 1993b; McDonald, 1993c). In each of these edaphic variables, the most common state was taken as the norm (for bedrock type “sandstone”, for ground water “well-drained”, and for rockiness of soils “pebbly”). Additionally, we included the elevation of each relevé as a predictor variable.

We used a linear regression model with rockiness, bedrock type and ground-water availability predictors and the difference between LSP and α diversity as the response variable. We generated 95% confidence intervals (CIs) of the effects by sampling the estimated distribution of the effect (given by the point estimate and its standard error) using the ‘arm’ R-package (Gelman & Su, 2018). We assessed significance of the effects by inspecting whether the CIs included zero.

All analysis have been done using R (R Development Core Team, 2017) and the R package probpool is available at <https://github.com/ChrKoenig/probpool>.

RESULTS

Species richness

The stacked binary occurrences show that the diversity of restios is unevenly distributed across the Cape with a richness centre in the SW Cape and a reduction in richness to the N, E and NE (Figure 2a). Species richness is mostly concentrated in the mountains, with a much lower species diversity on coastal plains in inter-montane valleys.

Local probabilistic species pool

The three probabilistic pools are summarized in Figure 2. The environmental pool (Figure 2b) is very similar to the stacked binary occurrences (Figure 2a), but as it uses the summed probabilities it shows a more gradual transition in richness. The dispersal pool (Figure 2c) is strongly influenced by the number of generations (see SI Figure S1). With dispersal calculated over 10 000 generations (Fig. 3) the major valleys in the SW Cape are not crossed, and with 60 000 generations all of the SW Cape is within reach of the species located on the SW Cape mountains. The resultant spatial pattern of the probabilistic pools (Figure 2d) is little influenced by the number of generations used for the dispersal pool (Supplementary Information Appendix S3). The strongest influence is visible on the coastal plains between the Cape

Peninsula, the Kogelberg, and along the coast towards Cape Agulhas. The absolute number of species however increases with the number of generations considered.

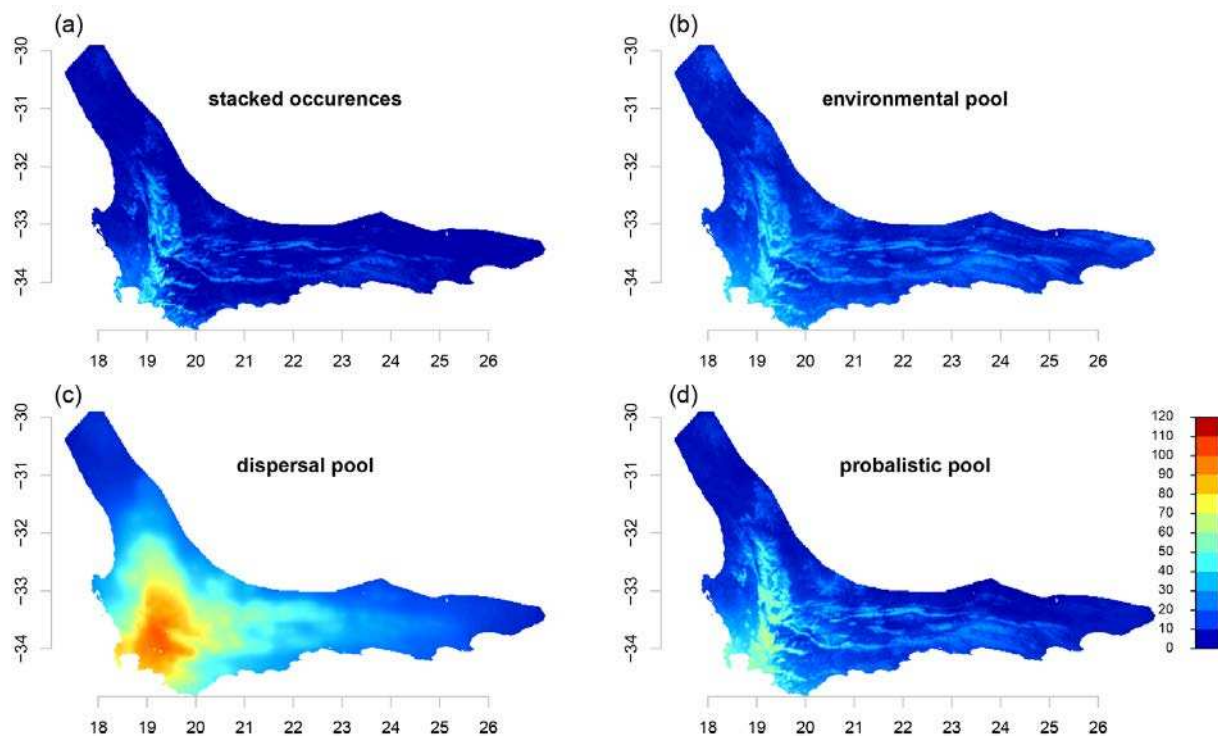


Figure 2. Spatial variation in the sizes of the probabilistic species pools for Restionaceae in the Cape Floristic Region. a) Inferred species richness obtained by summing the binary stacked species distribution models (SDM). b) Environmental pool, obtained by the summed individual probabilities of the SDM. c) Dispersal pool, summed from the binary SDM with dispersal kernel around each species occurrence, calculated over 30 000 generations. d) LPSP, which is the environmental pool x dispersal pool.

Strength of regional filters

We found a strong spatial pattern in the relative importance of the dispersal and environmental filtering in assembling the LPSPs (Fig.3). Dispersal has a larger impact than environmental filtering towards the N and E ends of the region at low generation times, whereas environmental filtering seems more efficacious than dispersal inside the CFR, mostly in the lowlands, where both soils and climate are not suitable for restios (Figure 3). Additionally, when we look at dispersal over a large number of generations, environmental filtering becomes more important relative to dispersal, indicating that there are large environmental barriers for restios that largely prevent them from spreading from their area of origin. In the environmentally unsuitable intermontane valleys surrounded by environmentally suitable mountains, there is no dispersal limitation, and hence the major impact is environmental. The eastern and northern extremes might be partly environmentally suitable for restios, but few species can reach these regions. The number of generations over which the dispersal pool is calculated has a major impact on the extent of the region predominantly influenced by dispersal, but not on the basic pattern of environmental filtering. After only 10 000 generations, the dispersal pool is relatively more important than the environmental filtering over about half of the study region. This includes, in addition to the northern and eastern extremes, the northern and central parts of the western coastal plain, parts of the southern coastal plain, and the lowlands inland of the fold mountains. With an increase in the number of generations for calculating the dispersal pool, the relative filtering effect of dispersal is reduced and the absolute number of species in the local probabilistic pools increases (SI Appendix S3).

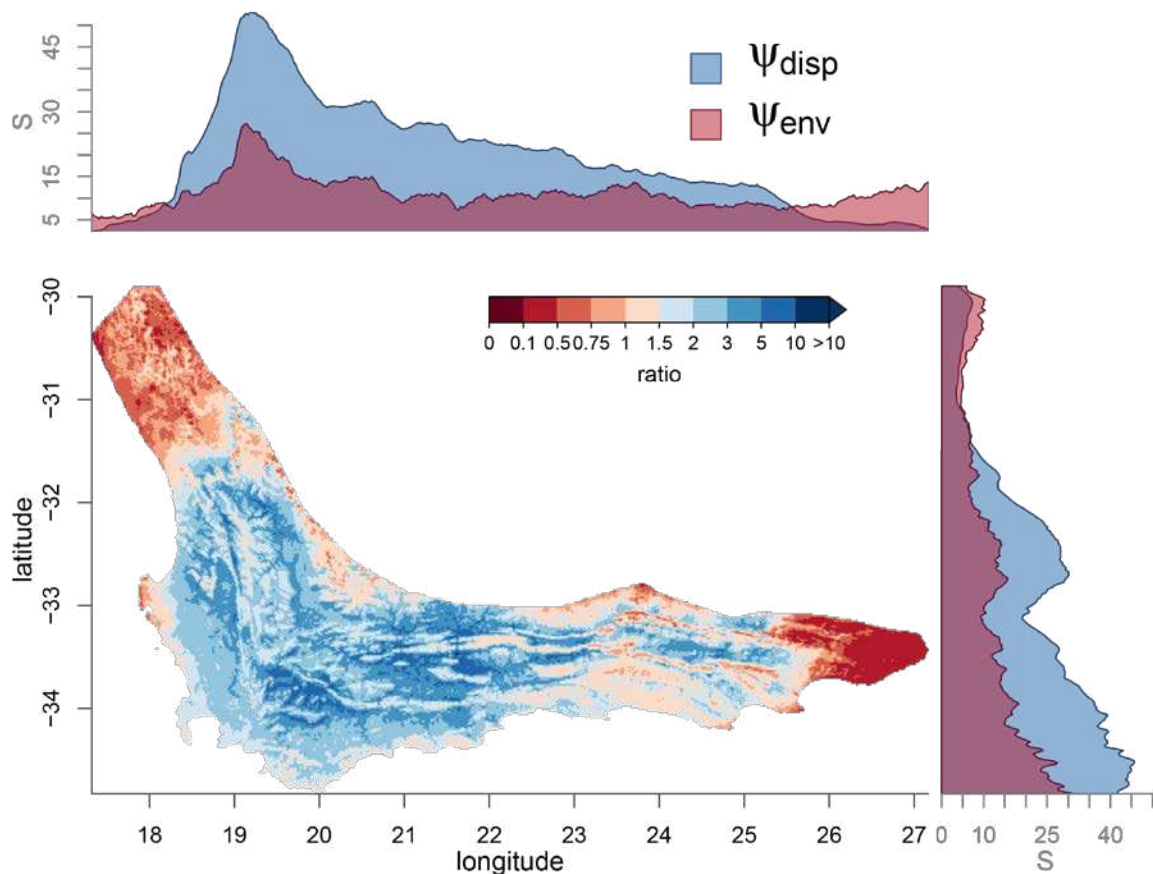


Figure 3. Spatial variation in the relative importance of environment versus dispersal filters (ratio: dispersal pool (Ψ_{disp})/environmental pool (Ψ_{env})) for restio species richness in southern Africa. The lower the ratio, the stronger the pools are dispersal filtered. The marginal plots are created by calculating the mean along latitude and longitudes for Ψ_{disp} , and Ψ_{env} . Ψ_{disp} is based on 30'000 generations. For examples of other dispersal pools see Supplement S1.

Drivers of species richness at the scale of the relevé

The difference between the predicted LPSP and the observed α -diversity at the relevé scale (Figure 4A) likely has two sources. The difference in absolute values is probably a result of the much larger spatial scale of the predicted pools (4 km²) compared to the relevés (78.5 m²), and the variation in this expected relationship (Figure 4B), could be due to habitat filtering at the scale of the relevé.

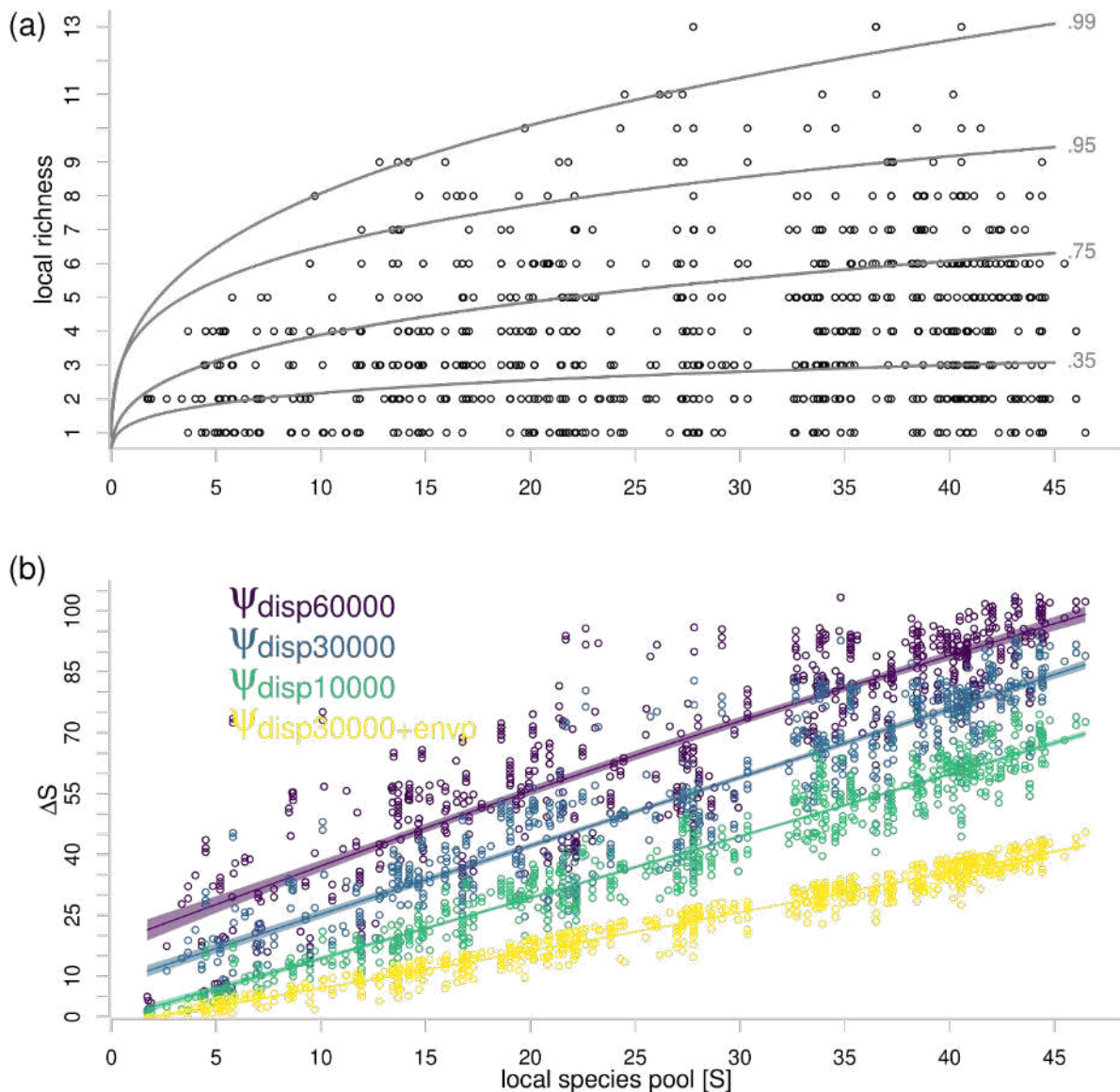


Figure 4. Relevé α -diversity per relevé of Cape restios versus the probabilistic species pool size at 2 x 2 km resolution. a) Scatterplot of observed local species richness per relevé (78.5 m², y-axis) against the probabilistic local species pool (4 km², x-axis), showing that the maximum observed richness in relevés is bounded by the local species pool. Grey lines were fitted using a quantile regression with the respective quantile indicated at the right (.99,.95,.75,.35). b). Scatterplot of the difference between LPSP sizes and relevé α -diversity (ΔS) plotted against the LPSP. Three different dispersal only pools, calculated for three different numbers of generations, are shown. Observed species richness is a more or less constant proportion of the probabilistic species pool. The variance between dispersal pools and observed richness can be explained by macro environmental filtering plus local forcing variables. The remaining deviation from a linear relationship is to be explained by local forcing variables in case of the $\Psi_{disp30000+envp}$.

The degree to which the different predictors used in the generalized linear model contributed to filtering species pools into local assemblages varies considerably (Figure 5). Among bedrock types the greatest reduction in species richness is found on limestone, deep sand and shale, with a tendency to lower richness on granite, and no significant impact from laterites. No filtering is assumed for the widespread sandstones. For soil drainage, wetland habitats (marshes, seepages, streambanks and areas

with impeded drainage) have reduced species richness, and the well-drained habitats and valley-bottoms without a filtering effect. Rockiness of the soil also has an impact, with relevés with bedrock with significant lower species richness than predicted.

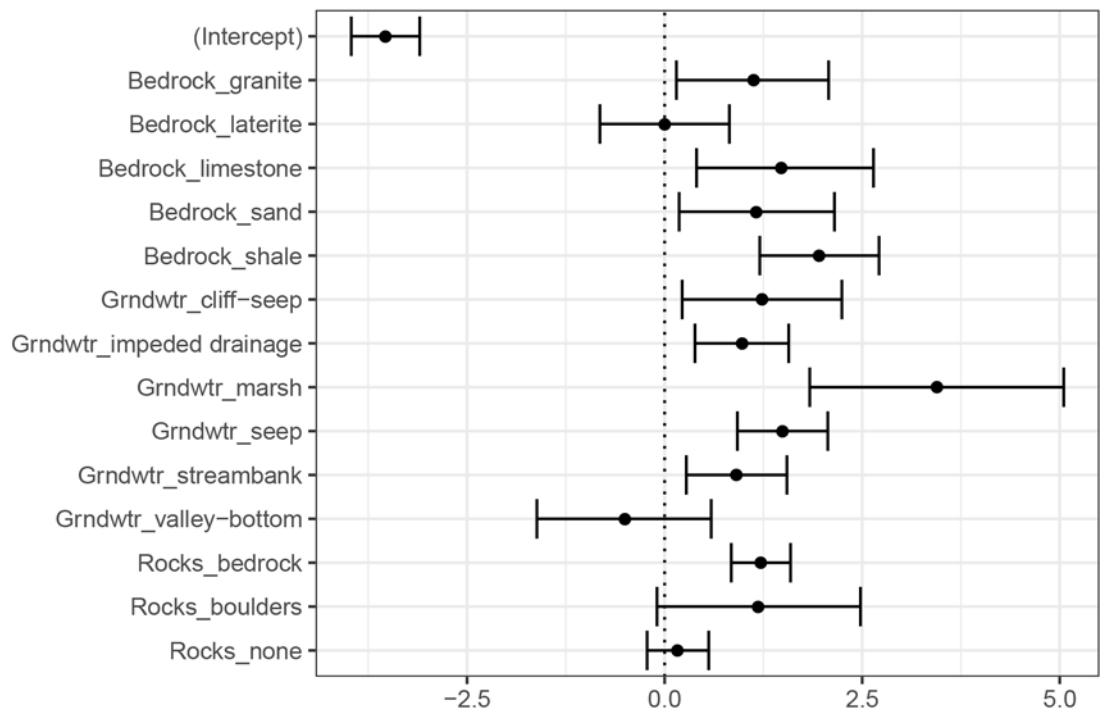


Figure 5. Confidence intervals for the parameter estimates of the generalized linear model of habitat predictors of differences between LPSPs and relevé α diversity in the restios of the Cape Floristic Region. Positive values result from a negative interaction between the variable and the residuals of observed richness, and indicate that this variable acts as a filter. Significant variables have confidence intervals that do not encompass zero. Abbreviations: Grndwtr = Groundwater.

DISCUSSION

Our analytical framework showed that the relative importance of environmental filtering and dispersal limitation in the restios of the Cape Floristic Region of southern Africa displays substantial spatial variation. Dispersal is more limiting at the northern and eastern extremes, whereas environmental filtering is more important in the central parts of the region. We calculated the LPSPs at a resolution of 2 x 2 km by combining modelled distributions with dispersal and environmental filters and used these LPSP predictions as the expected species richness, and showed the degree at which observed richness at the relevé scale is significantly modulated by bedrock type, soil drainage and rockiness of the habitats.

Local probabilistic species pools

This is the first high-resolution estimate of the variation in local restio richness across the Cape Floristic Region, consequently its reliability is important. The LPSP size is highest in the western mountains, from Cape Agulhas northward to the Cederberg, and including the Cape Peninsula, and is low in the lowlands and the Cape east of 20 degrees E (the southern and eastern Cape). This pattern closely matches with what is known of the variation in the floristic species richness in the Cape (Cowling *et al.*, 2002; Levyns, 1954; Oliver, Linder & Rourke, 1983) and is also evident from relevé data. Cowling *et al.* (2017) referred to this east-west diversity gradient as “Levyns’ Law”, and discussed the possible reasons for this richness gradient. This pattern has also been repeatedly shown for the restios, first by Oliver *et al.* (1983), subsequently by Linder and Mann (1998) for *Thamnochortus* and Moline and

Linder (2006) for *Elegia*. Consequently, our restio probabilistic pool is consistent with existing knowledge about restio species richness patterns.

However, how accurate are our results in detail? There are several potential refinements which might improve accuracy of our predictions. Here, the most important is probably a more realistic dispersal rate for individual restio species. Although we used an estimate of the dispersal rate of each species as baseline for the kernel, this is based on rather crude assumptions. In reality, we have no empirical estimates of this rate. Furthermore, setting probabilistic species pools requires a time frame for which these pools are calculated. Dispersal is a rate over time, consequently the pool size is time dependent. Here, we incorporated this by using six timeframes for comparison. The general pattern does not change, but for shorter time frames the absolute values of the species pool size are lower and the influence of the dispersal filter relative to the environmental filter is higher. By giving a time frame, we can however set a reference frame for which the statistical inference holds true, for example the duration during which conditions in current barriers are suitable for the species.

A second issue might arise from the interaction matrix. We assume that co-occurrence patterns are indicative of biotic interactions. The available dataset might be too small for a robust assessment of the co-occurrence rates, as these have shown to be very data-intense (Gallien, Zurell & Zimmermann, 2018). Furthermore, calculating co-occurrence assumes that there is no microhabitat variation within a relevé. This assumption may hold for deep soils, but is probably not true for rocky soils or complex landscapes, such as those found in the mountains. Finally, we assume that circular relevés of 10 meter diameter are small enough to capture co-occurrences well, but we are aware that fixed relevés have edge effects.

The final issue concerns the species for which SDMs could not be constructed. Species distribution models could be constructed for only 130 of the 350 restio species, for which enough location points were available. Many of these species for which SDMs could not be constructed are particularly range restricted, some 40 species are known only from single populations, and most are found in the western part of the Cape (Linder, in prep). Consequently, this could lead to an underestimation of the LPSPs in the western part of the Cape, relative to the eastern part. To avoid such effects, we removed all species for which SDMs could not be built from all subsequent analysis.

The processes leading to these regional pool differences (immigration, speciation and extinction) are discussed in Wüest *et al.* (2019). Although our probabilistic species pools contain information on the composition of the restio flora, we focus only on the species richness in this study.

Dispersal and environment filters

Dispersal is relatively more important in the northern and eastern extremes of the Cape region, whereas environment is more important in the central region. The environmental gradients are much steeper in the south-western mountains than in the more subdued landscapes in the eastern and northern extremes, these steeper gradients most likely increase the intensity of environmental filtering in the SW mountains. The extent of the region where dispersal is more limiting than environmental filtering depends on the time-scales over which dispersal is calculated, and on the absolute dispersal rates used. This is intuitively appealing, as dispersal, and consequently probabilistic species pools, are time related. This time component is usually not considered in species pool delineations, but the question “*which species can occur where*” naturally is followed by the question: “*when*”? Given that enough time has passed, species should be able to achieve a global or at least continental distribution. Interestingly, however, the strength of the filters at the relevé scale (mainly soil) are not influenced by the time frame chosen.

The implications are that, given the current climate, ranges in the centre of the Cape Floristic Region are environmentally limited, whereas in the northern and eastern extremes they are dispersal limited.

However, it seems more likely that during the glacial climate change at the extremes may have had a relatively greater impact on the restios, than the centre (summarized in Wüest *et al.*, 2019). The temporal context is that the Last Glacial Maximum was less than 10.000 generations ago. The implications of this spatial variation in the mechanisms of regeneration regional richness has not yet been adequately explored.

Local probabilistic species pool and relevé richness

Our results show that the size of the LPSP strongly impacts species richness at the relevés, and that this relationship is non-linear (e.g. Ibanez *et al.*, 2018; Ricklefs & He, 2016), which Rosenzweig and Ziv (1999) referred to as the “echo pattern”. This could imply presence of community saturation. In one of the first studies to explore this relationship Terborgh and Faaborg (1980) showed extensive saturation of communities for West Indian birds. Karger, Weigelt, Amoroso *et al.* (2014) found that saturation increases with decreasing spatial grain size for fern communities in Melanesia. There are, however, several problems with interpreting our results as evidence for or against saturation. Saturation is always context dependent (Olivares *et al.* 2018), and in the case of the restios, we could argue that some relevés (the ones reaching the upper asymptote of Figure 4a), are saturated with respect to the LPSP. However, we cannot show that these relevés are also at equilibrium, or that other plant groups may affect the competitive relationships, which would be additionally necessary to confirm saturation (Olivares *et al.* 2018). In an ecosystem with remarkable turnover in species composition at relevé level, as shown for the Cape Point reserve (Slingsby, Merow, Aiello-Lammens *et al.*, 2017), and with high habitat dynamics (e.g. fire on a decadal scale), data on temporal turnover would be needed to confirm saturation. Overall, our results suggest that relevé richness is not saturated (Cornell *et al.*, 2014).

Environmental variables at the scale of the relevé

With four exceptions (lateritic bedrock, valley-bottom groundwater, and habitats with boulders, and habitats without rocks), the rarer habitats have fewer species compared to the more common habitats (sandstone bedrock, well drained, and soils without groundwater, and pebbly soils). Well-drained sandstone slopes with pebbly soils may be regarded as the zonal habitat of fynbos, and generally can be considered the matrix in which pockets of poorly drained soils, other bedrock types, and local patches of soil without pebbles, or with bedrock, are embedded. Presumably this more extensive habitat on the regional scale means that it is more species rich due to species area relationships, ergo more widespread habitats contain more species. Consequently, relevé in these habitats should also have, as we show, a higher α diversity. In addition, the fragmented distribution of these rarer habitats might mean that they undersample the regional habitat-defined species pools, whereas this does not apply to the unfragmented widespread habitat types. This could further reduce their α diversity. The implication is that it is not the habitat per se that limits α diversity, but the spatial extent and structuring of these habitats. To evaluate the effect of only the habitat variables, we would need to factor out the spatial extent of the habitats, information which is currently not yet available.

Previous studies (Cowling, 1990; Cowling *et al.*, 1992) have failed to find habitat correlates of relevé α -diversity. However, these included all species (not only restios), did not include soil drainage, and possibly most importantly did not factor out the effects of regional diversity. It is possible that our results were biased by including only species with 10 or more occurrences for which SDMs could be modelled, and that the rare habitats have disproportionally more rare species than widespread habitats. However, as the local species pool is assessed at a 2 x 2 km scale, but the observed richness at a 10 m diameter scale, this should result in an overestimation of the diversity in the rare habitats (the habitat specific rare species were included in our relevé data used as estimate of α diversity, but they were most likely not included in the estimation of the regional pool size). We argue that therefore, our results underestimate the differences between the rare and the common habitats.

This is the first demonstration in the Cape flora fynbos that habitats (or the spatial extent of habitats) impact relevé, α -diversity, albeit only for restios. It is unclear whether this applies to the flora as a whole, and consequently to what extent restios can be used as proxy for the whole fynbos flora.

CONCLUSIONS

We present an analytical framework that is very effective at exploring the factors limiting α -diversity, as it sets up null expectation in the form of probabilistic local species pools which mimics the influence of regional-scale processes. This allows us to disentangle both macro- and local-scale process, while retaining their hierarchical structure, namely the contributions of dispersal, environmental filtering, and habitat parameters at the scale of the relevé. Dispersal and environmental filters are not equally distributed in the Cape Floristic Regions. What is not explained by regional factors, can be explained by variables varying at the scale of the relevés such as soil conditions.

We show for restios in the Cape flora that, at a macro-scale, environmental filtering is relatively important where the environmental gradients are steep, and dispersal limitation is important away from centres of diversity. Having disentangled regional variables, we can show, at a micro-scale, soil conditions are what is limiting the local α -diversity.

BIOSKETCH

Dirk Nikolaus Karger is currently a research scientist at the Swiss Federal Research Institute WSL. His research is focused on eco-climatic modelling and macroecology with a specific focus on the processes that biodiversity – climate interactions across spatial scales.

The concept was developed by DNK, the SDM and regressions were run by DNK, RW. CK, PW, DNK, JSC wrote the R package, NEZ and DNK developed the environmental variables, the writing was led by HPL and DNK, with input from all co-authors.

ACKNOWLEDGEMENTS

We acknowledge financial support from the Swiss National Science foundation, grant 31003A_152982 / 1 to HPL; Cape Nature for collecting permits; the development of the package was funded by sDiv (sREGPOOL2).

DATA AVAILABILITY STATEMENT

Distribution data: DRYAD digital repository (<http://dx.doi.org/XXXX>). Climate data used in the study is also available at www.chelsa-climate.org

REFERENCES

- ALLOUCHE O., TSOAR, A. & KADMON, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**, 1223-1232.
- ALLSOPP N., COLVILLE, J. & VERBOOM, G. A. (2014). Fynbos. Ecology, Evolution, and Conservation of a Megadiverse Region. Oxford, Oxford.
- ARAYA Y. N., SILVERTOWN, J., GOWING, D. J., MCCONWAY, K. J., LINDER, H. P. & MIDGLEY, G. F. (2011). A fundamental, eco-hydrological basis for niche segregation in plant communities. *New Phytologist* **189**, 253-258.
- BLOIS J. L., ZARNETSKY, P. L., FITZPATRICK, M. C. & FINNEGAN, S. (2013). Climate Change and the Past, Present, and Future of Biotic Interactions. *Science* **341**, 499-504.
- BREIMAN L. (2001). Random Forests. *Machine Learning* **45**, 5-32.
- CABRAL J. S. & SCHURR, F. M. (2010). Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography* **19**, 85-97.

- CORNELL H. V. & HARRISON, S. P. (2014). What are species pools and when are they important? *Annual Review of Ecology, Evolution, and Systematics* **45**, 45-67.
- COWLING R. M. (1990). Diversity components in a species-rich area of the Cape Floristic Region. *Journal of Vegetation Science* **1**, 699-710.
- COWLING R. M. (1992). The ecology of fynbos. Nutrients, fire and diversity, pp. 411. Oxford University Press, Cape Town.
- COWLING R. M., BRADSHAW, P. L., COLVILLE, J. F. & FOREST, F. (2017). Levyns' Law: explaining the evolution of a remarkable longitudinal gradient in the Cape plant diversity. *Transactions of the Royal Society of South Africa* **72**, 184 - 201
- COWLING R. M., HOLMES, P. M. & REBELO, A. G. (1992). Plant diversity and endemism. In *The Ecology of Fynbos: Nutrients, Fire and Diversity*, vol. chtr 4 (ed. R. M. Cowling), pp. 62-112. Oxford University Press, Cape Town.
- COWLING R. M. & LOMBARD, A. T. (2002). Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions* **8**, 163-179.
- GALLIEN L., ZURELL, D. & ZIMMERMANN, N. E. (2018). Frequency and intensity of facilitation reveal opposing patterns along a stress gradient. *Ecology and Evolution* **8**, 2171-2181.
- GELMAN A. & SU, Y.-S. (2018). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models.
- GRIFFITH D. M., VEECH, J. A. & MARSH, C. J. (2016). cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software* **69**, 1-17.
- GUISAN A. & RAHBEK, C. (2011). SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* **38**, 1433-1444.
- HARRISON S. & CORNELL, H. (2008). Toward a better understanding of the regional causes of local community richness. *Ecology Letters* **11**, 969-979.
- HASTIE T. J. & TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC, London.
- HUBER M. & LINDER, H. P. (2012). The evolutionary loss of aerenchyma limits both realized and fundamental ecohydrological niches in the Cape reeds (Restionaceae). *Journal of Ecology* **100**, 1338-1348.
- IBANEZ T., KEPPEL, G., BAIDER, C., BIRKINSHAW, C., CULMSEE, H., CORDELL, S., FLORENS, F. B. V., FRANKLIN, J., GIARDINA, C. P., GILLESPIE, T. W., LAIDLAW, M., LITTON, C. M., MARTIN, T. G., OSTERTAG, R., PARTHASARATHY, N., RANDRIANAIVO, R., RANDRIANJANAHARY, M., RAJKUMAR, M., RASINGAM, L., RATOVOSON, F., REZA, L., SACK, L., AIBA, S. I., WEBB, E., WHITFELD, T. J. S., ZANG, R. & BIRNBAUM, P. (2018). Regional forcing explains local species diversity and turnover on tropical islands. *Global Ecology and Biogeography* **27**, 474-486.
- KARGER D. N. (2016). Recent advances in probabilistic species pool delineations. *Frontiers of Biogeography* **8.2**, e30545.
- KARGER D. N., CONRAD, O., BÖHNER, J., KAWOHL, T., KREFT, H., SORIA-AUZA, R., ZIMMERMANN, N. E., LINDER, H. P. & KESSLER, M. (2017a). Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**, 170122 (2017).
- KARGER D. N., CONRAD, O., BÖHNER, J., KAWOHL, T., KREFT, H., SORIA-AUZA, R. W., ZIMMERMANN, N. E., LINDER, H. P. & KESSLER, M. (2017b). Data from: Climatologies at high resolution for the earth's land surface areas. In *Dryad Digital Repository*.
- KARGER D. N., CORD, A. F., KESSLER, M., KREFT, H., KÜHN, I., POMPE, S., SANDEL, B., SARMENTO CABRAL, J., SMITH, A. B., SVENNING, J.-C., TUOMISTO, H., WEIGELT, P. & WESCHE, K. (2016). Delineating probabilistic species pools in ecology and biogeography. *Global Ecology and Biogeography* **25**, 489-501.
- KARGER D. N., WEIGELT, P., AMOROSO, V. B., DARNAEDI, D., HIDAYAT, A., KREFT, H. & KESSLER, M. (2014). Island biogeography from regional to local scales: evidence for a spatially scaled echo pattern of fern diversity in the Southeast Asian archipelago. *Journal of Biogeography* **41**, 250-260.

- KRAAIJ T. & VAN WILGEN, B. W. (2014). Drivers, ecology, and management of fire in fynbos. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region* (ed. N. Allsopp, G. A. Verboom and J. Colville), pp. 47-72. Oxford University press, Oxford.
- KREFT H. & JETZ, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the USA* **104**, 5925-5930.
- LESSARD J.-P., BELMAKER, J., MYERS, J. A., CHASE, J. M. & RAHBEK, C. (2012). Inferring local ecological processes amid species pool influences. *Trends in Ecology & Evolution* **27**, 600-607.
- LESSARD J.-P., WEINSTEIN, B. G., BORREGAARD, M. K., MARSKE, K. A., MARTIN, D. R., MCGUIRE, J. A., PARRA, J. L., RAHBEK, C. & GRAHAM, C. H. (2016). Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *The American Naturalist* **187**, 75-88.
- LEVYNS M. R. (1954). The genus *Muraltia*. *Journal of South African Botany, Supplementary volume* **2**, 1-247.
- LINDER H. P. (2003). The radiation of the Cape flora, southern Africa. *Biological Reviews* **78**, 597-638.
- LINDER H. P. (2005). A protocol for the systematic documentation of the ecology of Cape plants. *South African Journal of Botany* **71**, 201-210.
- LINDER H. P. & MANN, D. M. (1998). The phylogeny and biogeography of *Thamnochortus* (Restionaceae). *Botanical Journal of the Linnean Society* **128**, 319-357.
- LORTIE C. J., BROOKER, R. W., CHOLER, P., KIKVIDZE, Z., MICHALET, R., PUGNAIRE, F. I. & CALLAWAY, R. M. (2004). Rethinking plant community theory. *Oikos* **107**, 433-438.
- MANNING J. & GOLDBLATT, P. (2012). Plants of the Greater Cape Floristic Region. 1. The core Cape Flora. In *Strelitzia*, vol. 29. South African National Biodiversity Institute, Pretoria.
- MCDONALD D. J. (1993a). The vegetation of the southern Langeberg, Cape Province. 2. The plant communities of the Marloth Nature Reserve. *Bothalia* **23**, 153-174.
- MCDONALD D. J. (1993b). The vegetation of the southern Langeberg, Cape Province. 3. The plant communities of the Bergfontein, Rooiwaterspruit and Phesantefontein areas. *Bothalia* **23**, 239-263.
- MCDONALD D. J. (1993c). The vegetation of the southern Langeberg, Cape Province. I. The plant communities of the Boosmansbos Wilderness Area. *Bothalia* **23**, 129-151.
- MOLINE P. M. & LINDER, H. P. (2006). Input data, analytical methods and biogeography of *Elegia* (Restionaceae). *Journal of Biogeography* **33**, 47-62.
- NELDER J. A. & WEDDERBURN, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* **135**, 370-384.
- OLIVER E. G. H., LINDER, H. P. & ROURKE, J. P. (1983). Geographical distribution of present-day Cape taxa and their phytogeographical significance. *Bothalia* **14**, 427-440.
- PHILLIPS S. J., ANDERSON, R. P. & SCHAPIRE, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.
- R DEVELOPMENT CORE TEAM. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- REBELO A. G., BOUCHER, C., HELME, N. A., MUCINA, L. & RUTHERFORD, M. C. (2006). Fynbos biome. In *The Vegetation of South Africa, Lesotho and Swaziland*, vol. 19. *Strelitzia* (ed. L. Mucina and M. C. Rutherford), pp. 53-219. South African National Biodiversity Institute, Pretoria.
- RICKLEFS R. E. (1987). Community diversity - relative roles of local and regional processes. *Science* **235**, 167-171.
- RICKLEFS R. E. & HE, F. (2016). Region effects influence local tree species diversity. *Proceedings of the National Academy of Sciences* **113**, 674-679.
- ROSENZWEIG M. L. & ZIV, Y. (1999). The echo pattern of species diversity: pattern and processes. *Ecography* **22**, 614-628.
- SCHURR F. M., PAGEL, J., CABRAL, J. S., GROENEVELD, J., BYKOVA, O., O'HARA, R. B., HARTIG, F., KISSLING, W. D., LINDER, H. P., MIDGLEY, G. F., SCHRÖDER, B., SINGER, A. & ZIMMERMANN, N. E. (2012). How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography* **39**, 2146-2162.

- SIEBEN E. J. J., BOUCHER, C. & MUCINA, L. (2004). Vegetation of high-altitude fens and restio marshlands of the Hottentots Holland Mountains, Western Cape, South Africa. *Bothalia* **34**, 141-153.
- SLINGSBY J. A., MEROW, C., AIELLO-LAMMENS, M., ALLSOPP, N., HALL, S., KILROY MOLLMANN, H., TURNER, R., WILSON, A. M. & SILANDER, J. A. (2017). Intensifying postfire weather and biological invasion drive species loss in a Mediterranean-type biodiversity hotspot. *Proceedings of the National Academy of Sciences*.
- SVENNING J.-C., GRAVEL, D., HOLT, R. D., SCHURR, F. M., THUILLER, W., MÜNKEMÜLLER, T., SCHIFFERS, K. H., DULLINGER, S., EDWARDS, T. C., HICKLER, T., HIGGINS, S. I., NABEL, J. E. M. S., PAGEL, J. & NORMAND, S. (2014). The influence of interspecific interactions on species range expansion rates. *Ecography* **37**, 1198-1209.
- TAMME R., GOTZENBERGER, L., ZOBEL, M., BULLOCK, J. M., HOOFTMAN, D. A. P., KAASIK, A. & PARTEL, M. (2014). Predicting species' maximum dispersal distances from simple plant traits. *Ecology* **95**, 505-513.
- TERBORGH J. W. & FAABORG, J. (1980). Saturation of Bird Communities in the West Indies. *The American Naturalist* **116**, 178-195.
- VEECH J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography* **22**, 252-260.
- WÜEST R. O., BOUCHER, F. C., BOUCHENAK-KHELLADI, Y., KARGER, D. N. & LINDER, H. P. (2019). Dissecting biodiversity in a global hotspot: uneven dynamics of immigration and diversification within the Cape Floristic Region of South Africa. *Journal of Biogeography* **46**, (Early Online).
- WÜEST R. O., LITSIOS, G., FOREST, F., LEXER, C., LINDER, H. P., SALAMIN, N., ZIMMERMANN, N. E. & PEARMAN, P. B. (2016). Resprouter fraction in Cape Restionaceae assemblages varies with climate and soil type. *Functional Ecology* **30**, 1583-1592.
- ZOBEL M. (1997). The relative of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology & Evolution* **12**, 266-269.
- ZURELL D., POLLOCK, L. J. & THUILLER, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography* **41**, 1812-1819.