Applying conservation reserve design strategies to define ecosystem monitoring priorities

Irene Martín-Forés[1,2*], Greg R. Guerin[1,2], Samantha E.M. Munroe[1,2], Ben Sparrow[1,2]

[1] School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia

[2] Terrestrial Ecosystem Research Network (TERN), University of Adelaide, Adelaide, South Australia 5005, Australia

*corresponding author: irene.martin@adelaide.edu.au

Authors' Contributions:

All authors conceived the ideas; IMF analysed the data with help from GRG and SEMM; IMF drafted the paper. All authors reviewed the final version of the manuscript and approved its submission.

Abstract

In an era of unprecedented ecological upheaval, accurately monitoring ecosystem change at large spatial scales and over long-time frames is an essential to effective environmental management and conservation. However, economic limitations often preclude revisiting entire monitoring networks at a high enough frequency to accurately detect ecological changes. Thus, a prioritisation strategy is needed to select a subset of sites that meets the principles of complementarity and representativeness of the whole ecological reality. Here, we applied two well-known approaches for conservation design, the ‘minimum set’ and the ‘maximal coverage’ problems, to develop a strategic monitoring prioritisation procedure that compares potential monitoring sites using a suite of alpha and beta biodiversity metrics. To accomplish this, we created a novel function for the R environment that easily performs biodiversity metric comparisons and site prioritisation on a plot-by-plot basis. We tested our procedures using plot data provided by the Terrestrial Ecosystem Research Network (TERN) AusPlots, an Australian long-term monitoring network of 774 vegetation and soil monitoring plots. We selected 250 plots and 80% of the total species recorded for the maximal coverage and minimum set problems, respectively. We compared the results of each approach in terms of ecological complementarity (species accumulation) and the spatial and environmental representativeness of the plots selected by the different biodiversity metrics. We repeated the selection process for clusters of plots to incorporate logistic constraints for field expeditions. We found that prioritisation based on species turnover (i.e. selection of the most dissimilar plots in terms of species composition but ignoring species richness) maximised ecological complementarity and spatial representativeness, while also providing high environmental coverage. Species richness was an unreliable metric for spatial representation, whereas plot selection based on corrected

weighted endemism failed to capture ecological and environmental variation. Range-rarity-richness was a more balanced metric in terms of complementarity and representativeness. Prioritisation based on species turnover is desirable to cover the maximum variability of the whole network.

Synthesis and applications: Our results inform monitoring design and conservation priorities, which should consider changes in the turnover component of the beta diversity instead of being based on univariate metrics.

Introduction

Ecological change and monitoring

Monitoring ecological change is of the utmost importance in the face of increasing anthropogenic encroachment on natural systems and staggering rates of biodiversity loss worldwide (Spellerberg, 2005).

Ecological monitoring programs are a prerequisite for successful environmental policy and decision making, and the development of effective management and conservation programs (Jeffers, 1989; Parr et al., 2003; Sparrow, Edwards et al., 2020). Monitoring programs systematically measure certain ecological variables and processes (e.g. species richness and abundance, net primary production, etc; Vos et al., 2000) over time to inform ecological status and environmental quality (Wolfe et al., 1987; Haase et al., 2018). Updated and thorough long-term records inform about the change between precedent and current ecological status which will guide policy makers, and subsequent restoration and conservation priorities (Lovett et al., 2007; Jones, 2011).

Over the past two decades, monitoring programs have been developed at large scales to incorporate broader ecological processes (Yoccoz et al., 2001; Parr et al., 2003). Field-based sampling of terrestrial ecosystems at multi-scales has therefore been required to build comprehensive ecosystem observation networks (Sparrow, Edwards et al., 2020). As such, there are now several examples of extensive monitoring networks established at the continental scale, including the pan-European Integrated Carbon Observation System (ICOS), the National Ecological Observatory Network (NEON; USA), the Global Ecosystems Monitoring (GEM) network across the tropics (Malhi et al., 2021), and the Terrestrial Ecosystem Research Network (TERN) in Australia (Cleverly et al., 2019).

However, long-term monitoring programs are expensive. Large financial investments are required to provide standardized surveying training, fieldwork organisation, sample preservation and storage, as well as data curation, access and promotion. Therefore, ecosystem monitoring programs are limited by strict financial constraints (Lovett et al., 2007). Consequently, monitoring resources and funding need to be carefully weighted, allocated, and prioritised (James et al., 1999).

Approaches to optimise representation for reserve design and monitoring prioritisation

When resources and funding are limited, ‘conservation prioritisation’ style strategies should be implemented to select an optimal subset of sites for monitoring. This includes ensuring high information content (i.e., the largest possible set of species), and meeting the principles of complementarity and representativeness (Bennett et al., 2014; Guerin, Williams, Sparrow et al., 2020; Guerin, Williams, Leitch et al., 2020) to be effective and useful for decision-making. Thus, the monitored sites should constitute a spatially and environmentally representative subset of locations to obtain an ecologically representative and complementary sample of habitats and communities, which will ensure both ecological and statistical validity (Cullen, 1990; Vos et al., 2000; Sparrow, Foulkes et al., 2020). Hence, to detect ecological changes in the long-term at large spatial scales, a cost-effective resampling strategy needs to be designed to define a subset of priority sites to be revisited, and overcome the existing resource-limited trade-offs between effective spatial and temporal monitoring (Hewitt & Thrush, 2007).

Little agreement has been reached so far in terms of how to best monitor and quantify ecological change (Hill et al., 2016). Normally, fieldwork locations are determined depending on specific objectives (e.g. monitoring biodiversity hotspots; assessing the most anthropogenically disturbed areas, etc.), or ultimately by logistics, feasibility, and administrative constraints. While species richness has been the most commonly employed biodiversity metrics to prioritise areas to monitor and protect, it is not robust or ideal (Gotelli & Colwell, 2001). First, complementarity rather than richness *per se* should be used for location selection through an iterative process (Kirkpatrick, 1983; Justus & Sarkar, 2002). Second, there may be a mismatch between biodiversity hotspots in terms of species richness and centres of endemism (Godoy-Bürki et al., 2014) and it is well known that site selection based on species richness offers a poor optimisation for both monitoring planning and conservation practices (Hillebrand et al., 2018).

The ‘minimum set’ problem and the ‘maximal coverage’ problem are two common approaches to design conservation reserves aiming to maximise conservation benefits at minimum costs (McIntosh et al., 2017). The minimum set problem is based on ecological constraints; it identifies a set of plots that meets certain conservation targets (typically species) within the fewest possible number of sites (Margules & Pressey, 2000). In contrast, the maximal coverage problem is based on economic constrains and consists of maximising the number of species in a given number of sites (Alagador & Cerdeira, 2020).

While heuristic algorithms are effective at optimizing complex reserve design for both minimum set and maximal coverage problems, they can require complex analytical decisions or be computationally expensive (Pressey et al. 1996; Ball et al. 2009). Alternative approaches based on simpler optimisation strategies have been employed, including nature reserve design in California based on rarity-rarity richness (RRR; i.e. the inverse of the number of sites in which it is present; Albuquerque & Beier, 2015), or a trans-frontier conservation area in Southern Africa based on endemism and threatened species (Smith et al., 2008). The species turnover component (i.e. species replacement) of beta diversity (Baselga, 2010) has been proposed as one of the most robust biodiversity metrics to detect ecological changes over time, as it reflects compositional change within plant communities and is a strong indicator of how those communities respond to global change (Hillebrand et al., 2018). Yet, the use of species turnover to define conservation priorities and complementarity is still scarce (but see Socolar et al., 2016). Given the different approaches employed in the literature it is worth investigating how they perform when applying to site prioritisation in a continental ecological monitoring program.

TERN AusPlots as a case-study for the implementation of a practical and efficient prioritisation method

Here, we aim to identify which biodiversity metric is the best optimiser in terms of selecting a subset of plots that maximises the number of species accumulated while ensuring environmental and spatial representativeness. To do this we have developed a free and easy tool for the R environment to perform an optimisation process applying the maximal coverage problem. The function we developed is called *optim_species* (included in the ausplotsR package; Guerin, Saleeba et al., 2020; see Supplementary material S1 for R code details) and it compares a suite of different optimisers related to the biodiversity metrics that are most commonly used in the scientific literature.

We tested this methodology on TERN AusPlots, an Australian long-term monitoring network of ecosystem surveillance and monitoring sites distributed at a continental scale (*sensu* Eyre et al., 2011; Sparrow, Edwards et al., 2020), to select a subset of sites that optimise the complementarity and representativeness of the whole network. We applied both the minimum set and maximal coverage approaches to design a revisiting strategy for the collection of AusPlots. Specifically, we aimed to identify which biodiversity metrics could be most efficiently used to achieve an optimal revisiting strategy that maximises the number of species accumulated when imposing i) an ecological constraint (i.e. minimum set problem –consisting on covering 80% of the total number of species recorded in the dataset–) and ii) an economic constraint (i.e. maximal coverage problem –consisting on selecting a subset of 250 plots–).

For the results of the maximal coverage problem approach (i.e. those obtained with the *optim_species* function), we subsequently compared the environmental and spatial representativeness of the subsets of sites selected by each of those biodiversity metrics as optimisers to determine which is the preferred option to meet our complementarity and representativeness principles.

METHODS

Case study dataset

Our study uses the AusPlots database, a network systemically surveyed over ten years by TERN's Ecosystem Surveillance Program, a component of Australia's land ecosystem observatory (Cleverly et al., 2019). We used species presence and cover data from 774 one-hectare plots to compare biodiversity metrics and test our site prioritisation analysis approach. TERN plots are established in homogenous areas of terrestrial vegetation to take quantitative measurements of vegetation and soil characterisation (Guerin et al., 2017). In each plot, vegetation structure and composition are recorded using the point-intercept module (Sparrow, Foulkes et al., 2020). Parallel transects (10x100 m long) are laid out in a 5x5 grid pattern, spaced 20 meters apart. Species identity, cover and growth form are recorded at each 1 m point along each transect, resulting in 1010 survey points per plot. Data for each plot are available and freely accessible within the AusPlots database, and were extracted from the database using *ausplotsR* (v1.2; Guerin, Saleeba et al., 2020; Munroe et al., 2020; TERN 2020).

Some of the sites included in the dataset were revisited (i.e. 99 sites had been revisited, 73 of them twice and 26 three times) and different sets of species were recorded on each revisit. Where repeat visits occurred, each was treated as a sample (hereafter we refer each visit as plots for language simplification).

Biodiversity metrics

Using the function *optim_species* from the ausplotsR package, we compared a selection of often-used biodiversity metrics to be employed as optimisers to define monitoring priorities. The biodiversity metrics included univariate metrics: i) species richness, ii) weighted species richness or range-rarity richness (RRR; Guerin & Lowe, 2017), iii) corrected weighted endemism (CWE; Crisp et al., 2001), iv) Shannon-Wiener diversity index, (Shannon;

Shannon & Weaver, 1949) and v) Simpson diversity index (Simpson; Simpson, 1949). Species turnover-based metrics used included: i) pairwise Simpson dissimilarity (Simpson_Beta; Baselga, 2010) and ii) the most frequent selected plots after simulating pairwise Simpson dissimilarity selection over 1000 iterations (Frequent). See Table 1 and Supplementary material S2 for details of these optimisers.

Data analyses

Multi-site beta diversity

We carried out all statistical analyses in R (R Core Team 2020). To check to what degree biodiversity differences between plots were due to species replacement or species loss we computed multiple-site Sorenson dissimilarities in beta diversity (βsor) accounting for both the spatial turnover (βsim) and the nestedness (βnes) components of beta diversity (βsor = βsim + βnes; Koleff et al., 2003; Baselga et al., 2018).

Conservation reserve design applied to optimise monitoring strategies

We applied the maximum coverage and the minimum set problems to optimise monitoring site selection to prioritise sites to revisit. For both the minimum set problem and the maximal coverage problem we performed the analyses by adding individual plots. In addition, we applied the maximal coverage problem to clusters of plots to consider a more realistic and feasible scenario because it is unlikely a field team would go to a remote area to only sample one plot.

We developed an R function called *optim_species* as part of this study which combines functionalities from the vegan (Oksanen et al., 2019) and betapart (Balsega et al., 2018) packages. The optimisation analysis is captured in this function which can be accessed in ausplotsR (Guerin, Saleeba et al., 2020). The function is thus free and easily accessible and can be run on any similar dataset (see R code as well as another example in the supplementary material S1 and S3 for details). Hence, we performed the analyses employing the *optim_species* function, using as data input the species vs sites matrix in terms of presence/absence, except for Shannon and Simpson, for which we used the matrix including percent cover values.

The maximal coverage problem

To address the maximal coverage problem, we set to 250 the number of plots to be selected for future revisits and monitoring. We decided on 250 plots within the AusPlots monitoring network because it is within our ability to revisit over a three to five years time scale.

The minimum set problem

To address the minimum set problem, we elucidated how many plots we would need to be revisited using each optimiser to account for at least 80% of the overall species richness (2822 species). The minimum set problem was analysed employing the same optimisers described for the maximal coverage problem (Table 1).

Spatial coverage representativeness

To compare spatial coverage representativeness of the plots selected by different optimisers, we computed the Clark and Evans aggregation index (Clark & Evans, 1954) for the spatial point patterns obtained with each of the optimisers using spatstat (Baddeley et al., 2015). We applied the cumulative distribution function *cdf* without edge correction because of corresponding to the mean value of nearest neighbour distance distribution function G(r) from a point pattern within an arbitrary shape. The Clark-Evans test values show whether a spatial point pattern distribution is clustered (R < 1), or ordered or regular (R > 1). We also mapped the location of the 250 selected plots obtained from the maximal coverage problem to visually support the differences in spatial representativeness when applying each of the biodiversity optimisers.

Environmental coverage representativeness

We compared sets of optimised plots for their climatic representativeness across Australia. We extracted data for 25 climatic variables from Harwood et al. (2016) (Supplementary material S4). We assessed plant species composition data from field plots in the order they were selected by the different optimisers, treating successive plots as additions to a cumulative sample of environmental and ecological space. We computed Euclidean distances for environmental variables with the function *vegdist* from the vegan package to assess the environmental representativeness of the subsets of plots selected by different optimisers. We implemented the *betadisper* function to analyse multivariate homogeneity of dispersions (distance to group centroid in principal coordinates space) of the cumulative samples (Anderson et al., 2006) for the different optimisers. We plotted the cumulative mean of environmental variation against the subsets of plots selected and visually compared the representativeness when using each of the biodiversity metrics as optimisers. Finally, we conducted a permutation test for homogeneity of multivariate dispersions with 999 permutations to explore pairwise comparisons between optimisers with regards to environmental coverage.

Monitoring strategy optimisation considering logistics

Spatial clustering

To make the optimisation more realistic in terms of field work feasibility, we clustered the 774 plots by geographic distance using a modified version of the *CalcDists* function (https://gist.github.com/sckott/931445) in which we estimated the distances among plots with the *distCosine* function from the geosphere package (Hijmans, 2019). The final number of clusters was 68, with an average number of eleven plots (nine sites) within each of them. The number of sites within each cluster ranged from three to 24 (Supplementary material S5).

We aggregated the species presence/absence data of species in the plots comprising each cluster. For the cover data, we calculated the Shannon and Simpson indices per plot, and then calculated the average value of the index for all the plots. We set to 20 the number of clusters to be selected via the same optimisation process. We then compared the species accumulation in the top 20 clusters when employing each of the biodiversity metrics.

RESULTS

Multi-site beta diversity across Australia

A total of 3528 species were recorded across all of the sampled plots (n = 774 plots). The multi-site Sorenson dissimilarity index was 0.998, the species turnover component (i.e. Simpson dissimilarity) corresponded to 0.997, while the nestedness component was only 0.001, indicating a very high rate of species replacement across the distributed plot network.

Conservation reserve design applied to optimise monitoring strategies

When comparing species accumulated with each of the optimisers, we observed that the species turnover-based metrics (i.e. the pairwise Simpson dissimilarity with its three implementation variants: Simpson_Beta, Simpson_Random, and Frequent) were the indices that maximised the cumulative number of species (Fig. 1). In particular, the Frequent variant outperformed the other two, with 3,051 species accumulated (86.5% of the species recorded; Supplementary material S6).

For univariate indices, the subsets of plots selected by RRR and species richness when applying the maximal coverage problem, accumulated a greater number of species (2,866 and 2,864, respectively, which accounted for 81.2% of all the species recorded in AusPlots sampling) than the rest of the optimisers. When incorporating species frequencies, the Shannon index outperformed the Simpson index for both the maximal coverage problem and the minimum set problem. CWE was a poor optimiser, with 2,024 species accumulated which accounted for 57.4% of the total number of species recorded (Fig. 1; Supplementary material S6).

Spatial coverage representativeness

All the subsets of plots selected were spatially clustered, but they differed among each other regarding their spatial representativeness. To visually complement the results from the Clark-Evan test, we mapped the subsets of plots selected with different optimisers (Fig. 2). Species richness was the optimiser that presented the most clustered spatial distribution (R = 0.366), followed by Shannon and Simpson optimisers (both displaying R = 0.408). Plots selected with RRR and CWE displayed Clark-Evans values of R = 0.414 and R = 0.428, respectively. From the species turnover-based metrics, pairwise Simpson dissimilarity

(Simpson_Beta) showed better spatial coverage (R = 0.450), whereas the best optimiser in terms of spatial representativeness was Frequent (R = 0.545).

Plot selection based on species richness and RRR was geographically biased towards coastal regions, failing to cover remote areas within the Australian outback. This was also the case for the Shannon and Simpson indices. Contrarily, the opposite trend (i.e. optimised plots located towards central and remote areas) was found when selecting plots based on CWE. This suggests than when selecting plots using univariate diversity metrics the results are geographically biased towards sites located either in biodiversity hotspots and areas with milder environmental conditions (e.g. richness) or in remote centres of endemism (e.g. CWE). Plot selection with pairwise Simpson dissimilarity accounts for the species ID and the turnover component; therefore, the spatial distributions displayed with these indices were more balanced in terms of representation across the whole Australia, than those obtained by univariate biodiversity metrics. This trend was enhanced when selecting the most frequent plots after 1,000 simulations of the pairwise Simpson dissimilarity plot selection.

Environmental coverage representativeness

The permutation test for homogeneity of multivariate dispersions showed significant differences in environmental representativeness among optimisers (F = 6.49; p-value ≤ 0.001; Table 2). We found that optimisation with CWE was the least representative in terms of environmental coverage (CWE: average distance to median = 3.41), showing significant differences with the environmental coverage of the subsets selected by all the other optimisers. Richness and RRR were the most representative with regards to environmental coverage (Richness and RRR: average distance to median = 4.75 and 4.64, respectively), followed by Simpson_Beta and Frequent (Simpson_Beta and Frequent: average distance to median = 4.33 and 4.46, respectively; Fig. 3; Table 2), with only marginally significant differences between Richness and Simson_Beta (Table 2). Shannon and Simpson (Shannon and Simpson: average distance to median = 4.15 and 4.07, respectively) were both significantly less environmentally representative than Richness and RRR, while no significant differences were found between the former two and the results obtained by Simpson_Beta and Frequent. Hence, richness was the biodiversity metric that best covered environmental differences when used as the optimiser. Results of the monitoring strategy optimisation for plot clusters are detailed in Supplementary material S7.

Discussion

Large-scale monitoring networks aim for high spatial coverage but resource constraints enforce trade-offs between spatial and temporal sampling. Our results, as applied to the TERN Ausplots dataset, clearly demonstrate that to design monitoring strategies that track the greatest number of species, it is better to focus on the turnover component (i.e. species replacement) through diversity partitioning than on univariate diversity indices.

Species turnover best optimised the selection of plots from a monitoring network to be revisited more often when applying both the minimum set and the maximal coverage approaches. Species turnover maximised species complementarity and spatial representativeness, without being significantly worse than the other optimisers regarding environmental representativeness. We obtained more robust results when we ran 1,000 random-seed iterations and extracted the most frequently selected plots (‘Frequent’) compared to using a predetermined, fixed seed (‘Simpson_Beta’). These results make sense considering that turnover was the most relevant source of change (accounting for 99% of total multi-site beta diversity) among the Australian vegetation communities sampled.

When comparing different univariate biodiversity metrics, our results indicated species richness was the worst performing biodiversity metric optimiser in terms of spatial representativeness. This is consistent with previous findings that have also demonstrated richness-based decisions do not meet the complementarity principle (Gotelli & Colwell, 2001; Godoy-Bürki et al., 2014), and are biased towards spatial clustering around more tropical climatic conditions (Veach et al., 2017). Based on these results, we provide further evidence that species richness is not an efficient measure of biodiversity and its change over time (Hillebrand et al., 2018).

Optimisation based on corrected weighted endemism (i.e. CWE) failed to be environmentally, spatially and ecologically representative in terms of biodiversity, with the lowest number of species accumulated across the whole network (worse than random) and the worst environmental representativeness. The poor performance of optimisation based on CWE has implications for monitoring strategies and conservation planning. While conservation reserves could aim to protect endemic species and therefore should consider this metric (Pelletier et al., 2018), monitoring priorities should not be based on CWE, as it will neither meet the principle of complementarity nor representativeness of the whole network and will fail to inform on the ecological reality.

Among the univariate biodiversity metrics, RRR was the most balanced, capturing a great number of species and being spatially and environmentally representative. Its estimation is straightforward from incidence datasets; therefore, when seeking a simple but relatively reliable way to select sites for a monitoring program, from the univariate metrics we recommend using RRR as an alternative to species turnover-based prioritisation. Our findings are supported by previous work demonstrating the great ecological representation of this index, as well as its effectiveness as a surrogate for biodiversity when fitted to environmental models to predict biodiversity in the absence of available data (Albuquerque & Beier, 2015, 2016).

The Shannon and Simpson optimisers performed comparatively poorly in the three dimensions studied here (i.e. ecological, spatial and environmental representativeness). Hence, plot selection prioritisation processes should preferably not be based on these metrics.

Both reserve design approaches (minimum set problem and maximal coverage problem) displayed similar results in terms of species accumulation, but with important consideration of the threshold *a priori* selected regardless of the approach. For example, as observed in Figure 1, species accumulations curves for some metrics crossed-over when reaching approximately either the 70% of the species (minimum set problem) or 150 plots (maximal coverage problem). This suggests the target matters and robust results may be jeopardised if thresholds are set too low.

When implementing optimisation for clusters of plots, differences in ecological representativeness were diluted relative to plot-by-plot selection (except in the case of CWE, which still performed significantly worse than the rest of the optimisers). Nevertheless, selection based on species turnover (most specifically employing the Frequent optimiser) performed best, with Frequent and Simpson_Beta approaches the most, and second most, environmentally and spatially representative, respectively. This has implications for hands-on applications of the current findings, since the prioritisation of clusters of plots will need to be carefully supervised to ensure complementarity and representativeness. We therefore suggest that practitioners perform plot-by-plot optimisation to get the ideal subset and then apply logistic principles to determine clusters of plots in a given geographic area.

Our results have potential application to conservation reserve design, whereby species turnover metrics could be considered to optimise complementarity and representativeness. Various criteria have been followed to design conservation reserves historically, including maximising species richness or genetic diversity, protecting rare or endemic species or restoring impacted or degraded areas (Margules et al., 1982; Kingsland, 2002). In this sense, Simpson pairwise dissimilarity is potentially useful as it selects a representative subset of the habitats and flora within a region.

The application of the findings reported here has some limitations. For example, optimisation and therefore reserve design based on species turnover relies on already available ground data and sampled communities and in some cases this information is incomplete or even non-existent. The optimisation process employed in this study (and the tools developed for the analysis) are able to be implemented in a variety of studies, and that can potentially be extended to similar approaches such as site selection based on phylogenetic or functional alpha and beta diversity. Similarly, it could be used to detect change in ecosystem composition over time in the context of a spatial framework; or within a temporal framework to identify sites with the most dissimilar samples among revisits, i.e. sites where vegetation is shifting more rapidly over time. These techniques will enable large-scale monitoring programs to maximise the value of information at a given resourcing level.

In summary, monitoring ecological state, function and change over time has become essential across national scales. The selection of sites for regular monitoring based on univariate biodiversity metrics (e.g. richness, CWE) often fails to meet the principles of complementarity and representativeness. We have therefore developed a practical, free and easy-to-use tool that can be used in any species versus sites dataset. The tool uses a set of alpha and beta diversity metrics to optimise species representation in a subset of monitoring sites to maximise species complementarity and spatial and environmental representation. Our results demonstrate that a representative subset of monitoring sites can be selected by finding the most ecologically dissimilar communities. This approach targets differences in composition instead of focusing on univariate metrics such as species richness, while also capturing spatial and environmental diversity. Long-term monitoring sampling strategies need to be carefully planned and designed. Applying reserve design approaches based on spatial vegetation compositional differences to maximise coverage constitutes a cost-effective and easily updated strategy to define monitoring priorities that leverages ground data already collected. This will in turn help policy, decision-making and conservation practices ensuring them to be based on accurate information that meets the complementarity and representativeness principles.

Acknowledgements

We acknowledge the TERN Ecosystem Surveillance field team (especially Emrys Leitch), and volunteers as well as support by the Australian Government through the National Collaborative Research Infrastructure Strategy. We have no conflicts of interest to declare.

Data availability

The dataset will be published in a public repository upon the acceptance of the manuscript

References

Alagador, D., & Cerdeira, J. O. (2020). Revisiting the minimum set cover, the maximal coverage problems and a maximum benefit area selection problem to make climate-change-concerned conservation plans effective. *Methods in Ecology and Evolution, 11*, 1325–1337.

Albuquerque, F., & Beier, P. (2015). Rarity-weighted richness: a simple and reliable alternative to integer programming and heuristic algorithms for minimum set and maximum coverage problems in conservation planning. *PloS One, 10,* e0119905.

Albuquerque, F., & Beier, P. (2016). Predicted rarity-weighted richness, a new tool to prioritize sites for species representation. *Ecology and Evolution, 6*, 8107–8114.

Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters, 9*, 683–693.

Baddeley, A., Rubak, E., & Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. London:Chapman and Hall/CRC Press, 2015. URL http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/

Ball, I. R., Possingham, H. P., & Watts, M. (2009). Marxan and relatives: software for spatial conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods and computational tools*, 185-195.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography, 19*, 134–143.

Baselga, A., Orme, D., Villeger, S., De Bortoli, J. & Leprieur, F. (2018). betapart: Partitioning Beta Diversity into Turnover and Nestedness Components. R package version 1.5.1. https://CRAN.R-project.org/package=betapart

Bennett, J. R., Sisson, D. R., Smol, J. P., Cumming, B. F., Possingham, H. P., & Buckley, Y. M. (2014). Optimizing taxonomic resolution and sampling effort to design cost-effective ecological models for environmental assessment. *Journal of Applied Ecology*, 51, 1722–1732.

Clark, P J. & Evans, F.C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology, 35,* 445–453.

Cleverly, J., Eamus, D., Edwards, W., Grant, M., Grundy, M. J., Held, A., Karan, M., Lowe, A. J., Prober, S. M., Sparrow, B. & Morris, B. (2019). TERN, Australia's land observatory: addressing the global challenge of forecasting ecosystem responses to climate variability and change. *Environmental Research Letters, 14,* 095004

Csuti, B., Polasky, S., Williams, P., Pressey, R. L., Camm, J., Kershaw, M., Kierster, A. R., Downs, B., Hamilton, R., Huso, M., & Sahr, K. (1997). A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biological Conservation, 80*, 83–97.

Cullen, P. (1990). Biomonitoring and environmental management. *Environmental Monitoring Assessment, 14,* 107–114.

Godoy-Bürki, A. C., Ortega-Baes, P., Sajama, J. M., & Aagesen, L. (2014). Conservation priorities in the Southern Central Andes: mismatch between endemism and diversity hotspots in the regional flora. *Biodiversity and Conservation, 23,* 81–107.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters, 4*, 379–391.

Guerin, G. R., Sparrow, B., Tokmakoff, A., Smyth, A., Leitch, E., Baruch, Z., Lowe, A.J. (2017) Opportunities for integrated ecological analysis across inland Australia with standardised data from AusPlots Rangelands. *PLoS ONE, 12,* e0170137

Guerin, G., Saleeba, T., Munroe, S., Blanco-Martin, B., Martín-Forés, I., Tokmakoff, A. (2020). ausplotsR: TERN AusPlots analysis package. R package version 1.2. https://CRAN.R-project.org/package=ausplotsR

Guerin, G. R., Williams, K. J., Sparrow, B., & Lowe, A. J. (2020). Stocktaking the environmental coverage of a continental ecosystem observation network. *Ecosphere, 11*, e03307.

Guerin, G. R., Williams, K. J., Leitch, E., Lowe, A. J., & Sparrow, B. (2020). Using generalised dissimilarity modelling and targeted field surveys to gap-fill an ecosystem surveillance network. *Journal of Applied Ecology*.

Haase, P., Tonkin, J. D., Stoll, S., Burkhard, B., Frenzel, M., Geijzendorffer, I. R., Häuser, C., Klotz, S., Kühn, I., McDowell, W. H., Mirtl, M., Müller, F., Musche, M., Penner, J., Zacharias, S., & Schmeller, D. S. (2018). The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. *Science of the Total Environment, 613*, 1376–1384.

Harwood, T., Donohue, R., Harman, I., McVicar, T., Ota, N., Perry, J., & Williams, K. (2016) 9s climatology for continental Australia 1976-2005: Summary variables with elevation and radiative adjustment. v1. In: CSIRO (ed.) Data Collection. CSIRO, https://doi.org/10.4225/08/5afa9f7d1a552.

Hewitt, J. E., & Thrush, S. F. (2007). Effective long-term ecological monitoring using spatially and temporally nested sampling. *Environmental Monitoring and Assessment, 133*, 295–307.

Hijmans, R. J. (2019). geosphere: Spherical Trigonometry. R package version 1.5-10. https://CRAN.R-project.org/package=geosphere

Hill, S. L. L., Harfoot, M., Purvis, A., Purves, D. W., Collen, B., Newbold, T., Burgess, N. D. & Mace, G. M. (2016). Reconciling biodiversity indicators to guide understanding and action. *Conservation Letters, 9*, 405–412.

Hillebrand, H., Blasius, B., Borer, E. T., Chase, J. M., Downing, J. A., Eriksson, B. K., Filstrup, C. T., Harpole, W.S., Hodapp, D., Larsen, S., Lewandowska, A. M., Seabloom, E. W., Van de Waal, D. B., & Ryabov, A. B. (2018). Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology, 55*, 169–184.

James, A.N., Green, M.J.B., Paine, J.R. (1999). A Global Review of Protected Area Budgets and Staff, WCMC–World Conservation: Cambridge.

Jeffers, J. N. R. (1989) Environmental monitoring. *Biologist, 36*, 171

Jones, J.P.G. (2011) Monitoring species abundance and distribution at the landscape scale. *Journal of Applied Ecology, 48*, 9–13.

Justus, J., & Sarkar, S. (2002). The principle of complementarity in the design of reserve networks to conserve biodiversity: a preliminary history. *Journal of Biosciences, 27*, 421–435.

Kingsland, S. E. (2002). Creating a science of nature reserve design: perspectives from history. *Environmental Modeling and Assessment, 7*, 61–69.

Kirkpatrick, J. B. (1983). An iterative method for establishing priorities for the selection of nature reserves: an example from Tasmania. *Biological Conservation, 25*, 127–134.

Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology, 72*, 367–382.

Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., Shanley, J. B., Likens, G. E., & Haeuber, R. (2007). Who needs environmental monitoring?. *Frontiers in Ecology and the Environment, 5*, 253–260.

Malhi, Y., Girardin, C., Metcalfe, D. B., Doughty, C. E., Aragão, L. E., Rifai, S. W., ... & Phillips, O. L. (2021). The Global Ecosystems Monitoring network: Monitoring ecosystem productivity and carbon cycling across the tropics. *Biological Conservation, 253*, 108889.

Margules, C., Higgs, A. J., & Rafe, R. W. (1982). Modern biogeographic theory: are there any lessons for nature reserve design?. *Biological Conservation, 24*, 115–128.

Margules, C.R. & Pressey, R.L. (2000). Systematic conservation planning. *Nature, 405*, 243–253

McIntosh, E. J., Pressey, R. L., Lloyd, S., Smith, R., & Grenyer, R. (2017). The impact of systematic conservation planning. *Annual Review of Environment and Resources, 42*, 677–697.

Moilanen, A., Wilson, K. & Possingham, H.P. (Eds.) (2009) Spatial conservation prioritization. Oxford: Oxford Univ. Press.

Munroe, S., Guerin, G., Saleeba, T., Martín-Forés, I., Blanco-Martin, B., Sparrow B. & Tokmakoff, A. (2020) ausplotsR: An R package for rapid extraction and analysis of vegetation and soil data collected by Australia’s Terrestrial Ecosystem Research Network. *Ecoevorxiv*. Doi: 1032942/osf.io/25phx

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Parr, T. W., Sier, A. R. J., Battarbee, R. W., Mackay, A., & Burgess, J. (2003). Detecting environmental change, science and society – perspectives on long-term research and monitoring in the 21st century. *Science of the Total Environment*, 310, 1–8.

Pelletier, T. A., Carstens, B. C., Tank, D. C., Sullivan, J., & Espíndola, A. (2018). Predicting plant conservation priorities on a global scale. *Proceedings of the National Academy of Sciences, 115*, 13027–13032.

Pressey, R. L., Possingham, H. P., & Margules, C. R. (1996). Optimality in reserve selection algorithms: when does it matter and how much?. *Biological Conservation*, *76*, 259–267.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shannon, C.E. & Weaver, W. (1949) The mathematical theory of communication. The University of Illinois Press, Urbana, 117pp.

Simpson, E. H. (1949) Measurement of species diversity. *Nature, 163*, 688.

Smith, R. J., Easton, J., Nhancale, B. A., Armstrong, A. J., Culverwell, J., Dlamini, S. D., Goodman, P. S., Loffler, L., Matthews, W. S., Monadjem, A., Mulqueeny, C. M., Ngwenya, P., Ntumi, C. P., Soto, B. & Leader-Williams, N. (2008). Designing a transfrontier conservation landscape for the Maputaland centre of endemism using biodiversity, economic and threat data. *Biological Conservation, 141*, 2127–2138.

Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation?. *Trends in Ecology and Evolution, 31,* 67–80.

Sparrow, B. D., Edwards, W., Munroe, S. E., Wardle, G. M., Guerin, G. R., Bastin, J. F., Morris, B., Christensen, R., Phinn, S. & Lowe, A. J. (2020). Effective ecosystem monitoring requires a multi-scaled approach. *Biological Reviews, 95*, 1706–1719.

Sparrow, B., Foulkes, J., Wardle, G., Leitch, E., Caddy-Retalic, S., van Leeuwen, S., Tokmakoff, A., Thurgate, N. Y., Guerin, G. R. & Lowe, A. J. (2020). A vegetation and soil survey method for surveillance monitoring of rangeland environments. *Frontiers in Ecology and Evolution, 8*, 157.

Spellerberg, I. F. (2005). Monitoring ecological change. Cambridge University Press.

TERN (2020) AusPlots ecosystem surveillance monitoring dataset (http://aekos.org.au/). Obtained via the ausplotsR R package, accessed June 2020.

Veach, V., Di Minin, E., Pouzols, F. M., & Moilanen, A. (2017). Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions, 23*, 715–726.

Vos, P., Meelis, E., Ter Keurs, W.J. (2000). A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environmental Monitoring and Assessment 61, 317–344

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution, 16*, 446–453.

Table 1. Optimisers description

Optimiser ID	Optimiser name	Description	Special utilisation / Best used
Richness	Species richness	Count of the number of species present in a given site	Identify biodiversity hotspots
RRR	Range-rarity richness	Inverse of the number of sites in which a species occurs. RRR = $\sum_1^n 1/c_i$, where c_i is the number of map grid cells occupied by species *i* and *n* is the number of species	When the goal is to identify areas of high biodiversity and biological uniqueness
CWE	Corrected weighted endemism	Range rarity richness (RRR) divided by species richness. CWE = $(\sum_1^n 1/c_i)/n$, where c_i is the number of map grid cells occupied by species *i* and *n* is the number of species	When the goal is to identify centres of endemism highlighting range-restricted species
Shannon	Shannon-Wiener diversity index	Combines species richness and the evenness or equitability by computing the species' relative abundances. H': $-\sum_{i=1}^{S} p_i log_n p_i$, where S is the species richness and p_i is the relative abundance of the species	It assumes that all species are represented in a sample and that they are randomly sampled
Simpson	Simpson diversity index	Combines species richness and the evenness or equitability by computing the species' relative abundances $D = 1 - \sum p_i^2$, where p_i is the proportional abundance of species *i*	It is the complement of Simpson's original dominance index, and represents the probability that two randomly chosen individuals belong to different species
Simpson_Beta	Pairwise Simpson dissimilarity index	It is based on diversity partitioning, which separates species replacement (i.e. turnover) from species loss (i.e. nestedness). The Simpson dissimilarity corresponds to the turnover component of the Sorensen dissimilarity. Considering two sites, $\beta sim = min(b,c) / (a + min(b,c))$, where *a* is the number of	It is used to maximise species turnover

	species present in both sites, *b* is the number of species present in the first site, but not in the second, and *c* is the number of species present in the second site, but not in the first.
Frequent	The most frequent plots selected over 1,000 iterations with a randomised starting seed using the pairwise Simpson dissimilarity index
Simpson_Random	The pairwise Simpson dissimilarity index with a randomised starting seed iterated 1,000 times

Table 2. Pairwise comparisons between optimisers with regards to environmental representativeness when applying maximal coverage problem at plot level. The observed p-value are located in the below diagonal, while the permuted p-value are in the above diagonal. Only significant differences are highlighted in bold. Notice that marginally significant values (p-value ≤0.1) are shown although not highlighted.

	Richness	RRR	CWE	Shannon	Simpson	Simpson_Beta	Frequent
Richness		0.68	**≤0.001**	**≤0.01**	**≤0.05**	≤0.1	0.25
RRR	0.68		**≤0.001**	≤0.1	**≤0.05**	0.20	0.49
CWE	**≤0.001**	**≤0.001**		**≤0.01**	**≤0.01**	**≤0.001**	**≤0.001**
Shannon	**≤0.05**	≤0.1	**≤0.01**		0.78	0.48	0.21
Simpson	**≤0.01**	**≤0.05**	**≤0.01**	0.75		0.29	0.12
Simpson_Beta	≤0.1	0.22	**≤0.001**	0.48	0.30		0.58
Frequent	0.25	0.48	**≤0.001**	0.21	0.12	0.58	

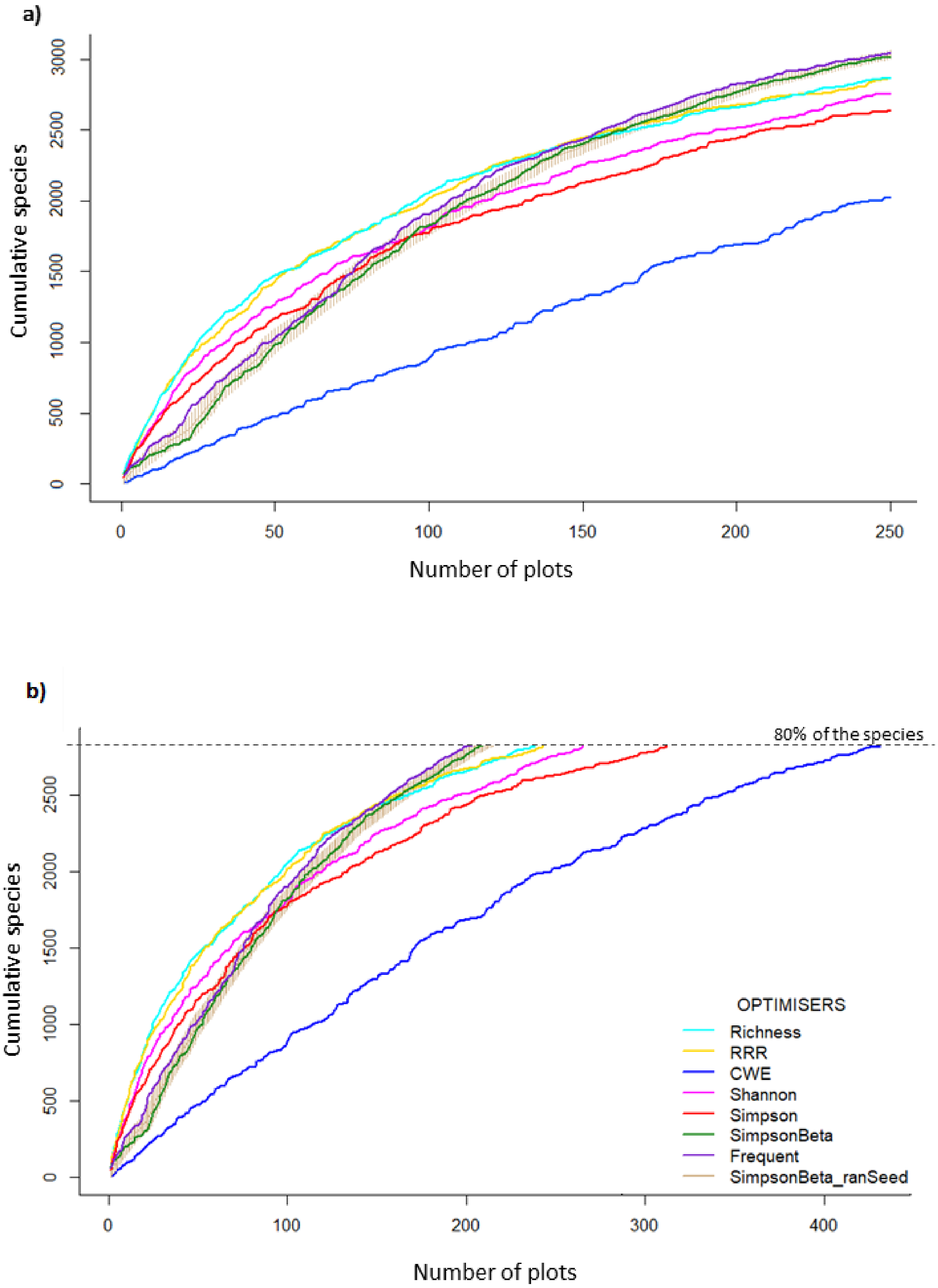


Fig. 1. Site optimisation process applying conservation reserve design strategies based on: a) Maximum coverage problem (selection of 250 sites), and b) Minimum set problem (selection of the minimum number of plots that allow including 80% of the species,

represented by dashed line). Optimisation has been performed in both cases employing different optimisers, including species richness, range rarity richness (RRR), corrected weighted endemism (CWE), Shannon-Wienner diversity index (Shannon), Simpson diversity index (Simpson), the turnover component of beta diversity or pairwise Simpson dissimilarity index (Simpson beta), the most frequent plots selected in 1000 iterations with a randomised starting seed using the pairwise Simpson dissimilarity index (Frequent), and the plots selected with a randomised seed using the pairwise Simpson dissimilarity index (SimpsonBeta_random seed).

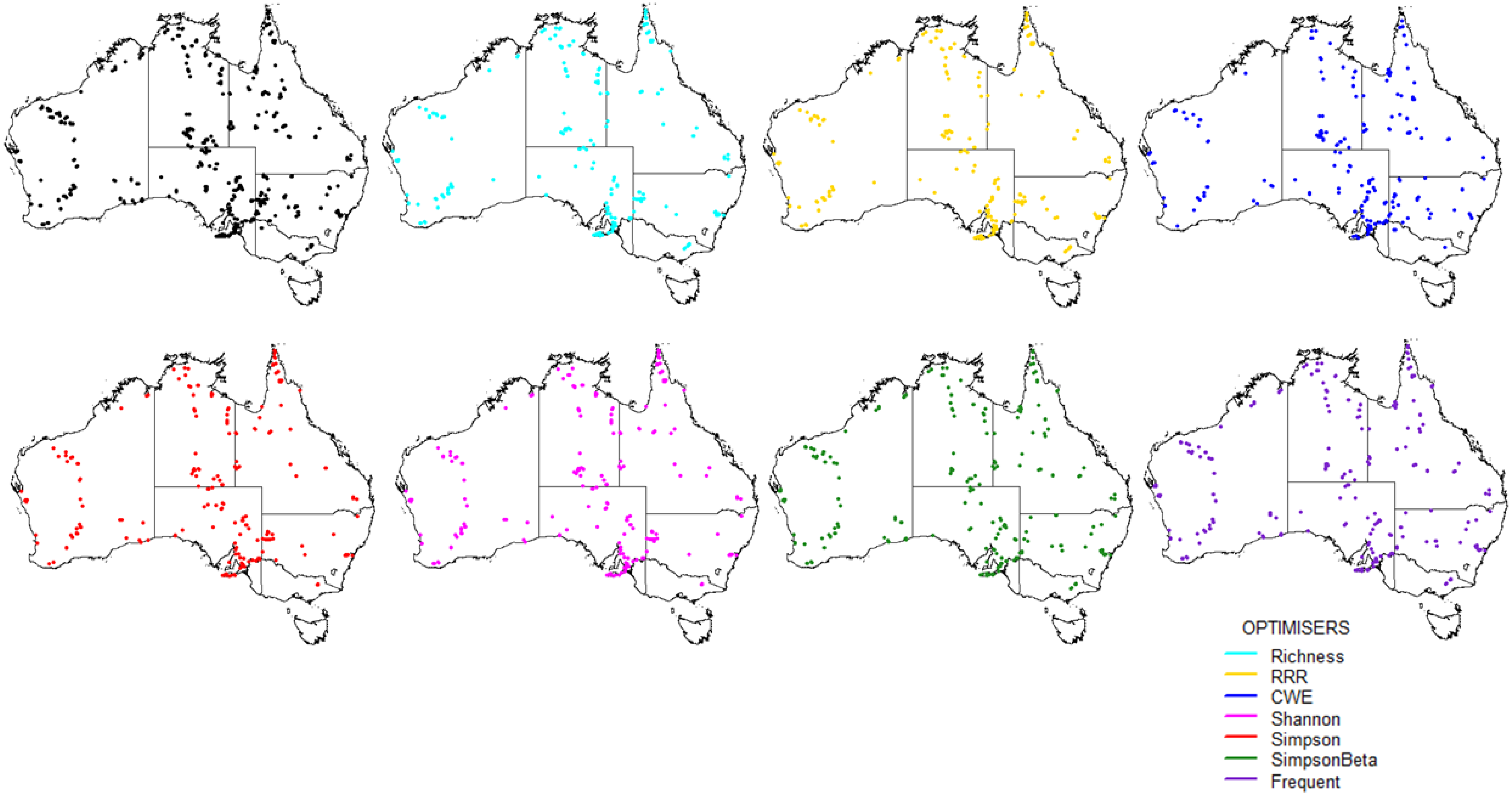


Fig. 2. Geographic location of the selected plots (N = 250) applying the maximum coverage problem. Black dots correspond to all the plots established. Colour dots refer to each of the selection employing different optimisers.

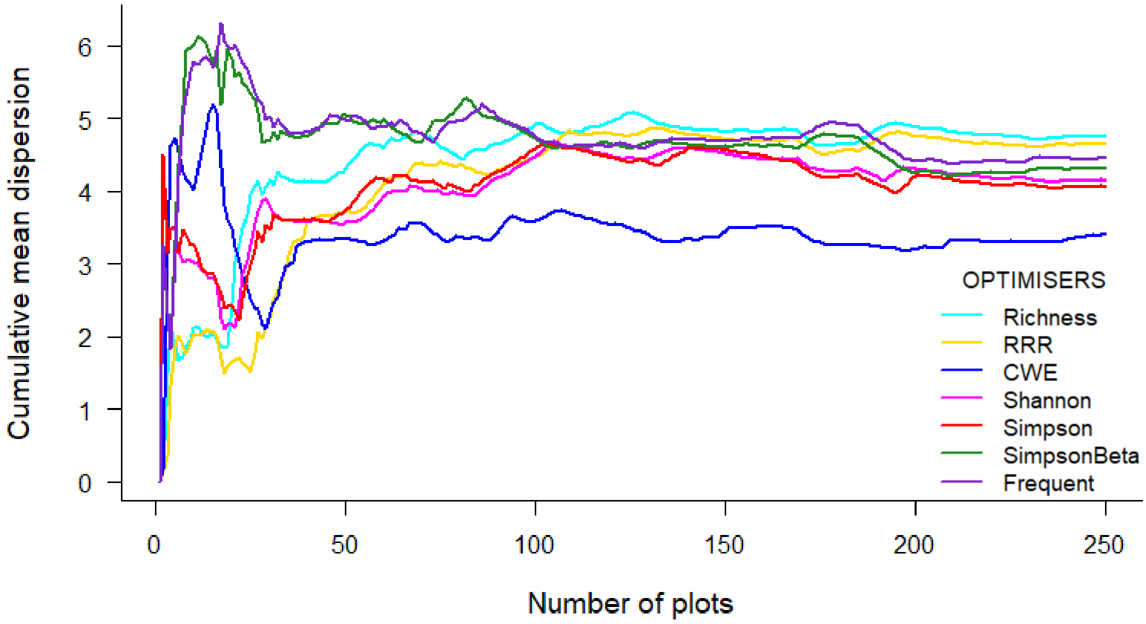


Fig. 3. Environmental representativeness of the 250 selected plots using different optimisers reflected by the cumulative mean dispersion. All environmental variables employed in the analyses are described in the Supplementary material S4.

Supplementary material S1.

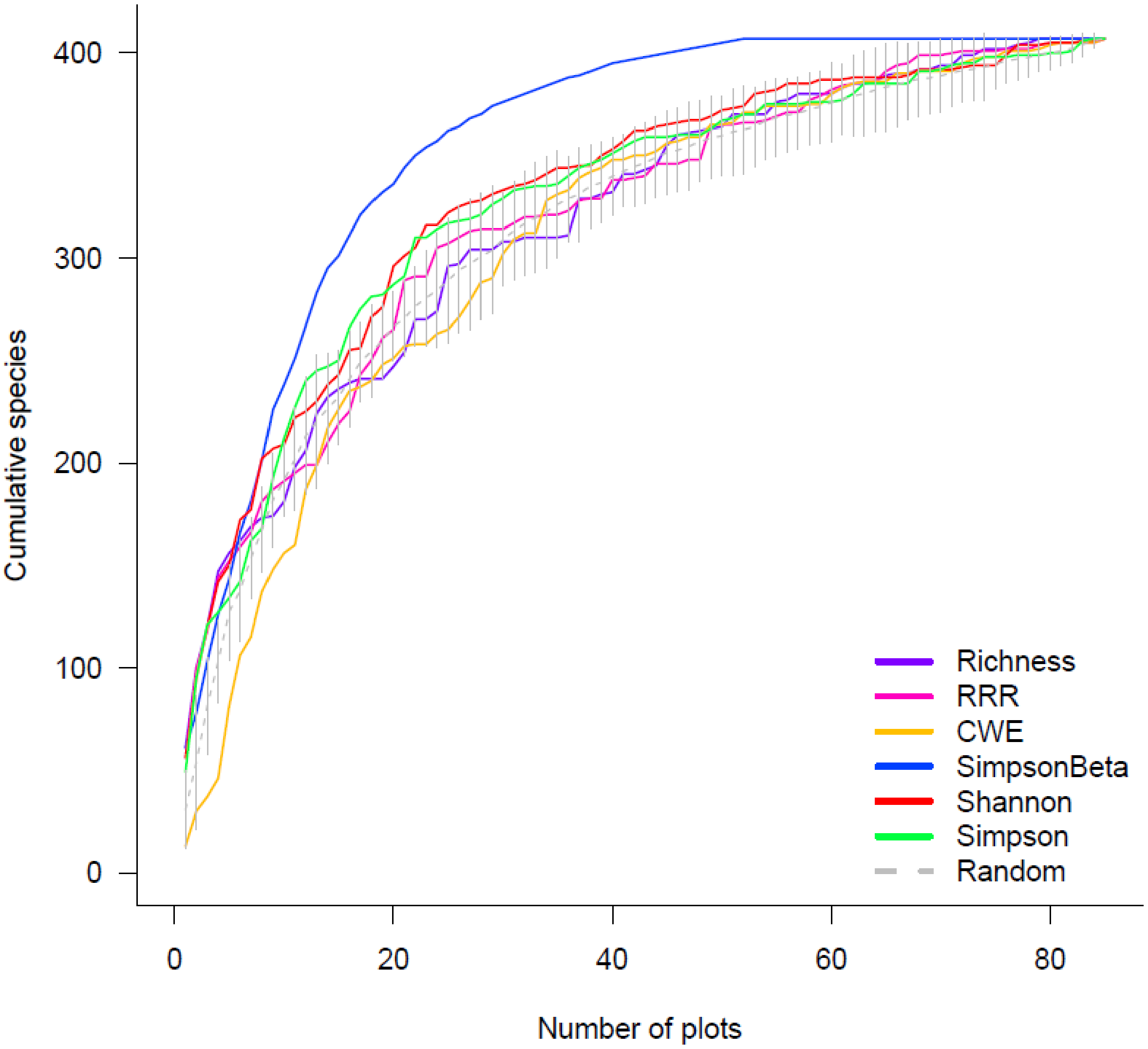
Rcode

Not included in the preprint.

Supplementary material S2. Additional explanation of biodiversity metrics

Species richness is simply the count of the number of species present in a given site and it is best used as an optimiser when the goal is to identify areas containing the greatest number of unique species. RRR is a rarity-weighted richness calculated as the sum of the inverse of the number of sites in which a species occurs and it is used to identify areas of high biodiversity and biological uniqueness. CWE is calculated as RRR corrected by the species richness and is used to identify centres of endemism highlighting range-restricted species. Shannon combines species richness and evenness or equitability by computing the species' relative abundances; it assumes that all species are represented in a sample and that they are randomly sampled. Simpson is the complement of Simpson's original dominance index, and represents the probability that two randomly chosen individuals belong to different species. The species turnover-based metric is based on diversity partitioning, which separates species replacement (i.e. turnover) from species loss (i.e. nestedness); it corresponds to the turnover component of the Sorensen dissimilarity and describes spatial turnover without the influence of richness gradients (because unshared species in the larger species sample are disregarded; Lennon et al., 2001; Baselga, 2010). For the species turnover-based metric, as pairwise Simpson dissimilarity differs depending on the plot that is chosen as seed, we computed three variants: fixed seed (Simpson_Beta), in which we selected the plot with the greatest richness as the starting point; random seed (Simpson_Random), in which we randomly selected the plot at the starting point and we iterated this process 1,000 times; and the most frequent plots (Frequent), by computing the most frequently selected pots in the former 1,000 simulations. When employing the pairwise Simpson dissimilarity, the most dissimilar plot to the one used as seed in terms of pairwise Simpson dissimilarity was added by the function. Subsequently the species present in both plots are pooled together and the most dissimilar one to the combination of both is added. This process was repeated until reaching the top 250 plots that maximised species turnover.

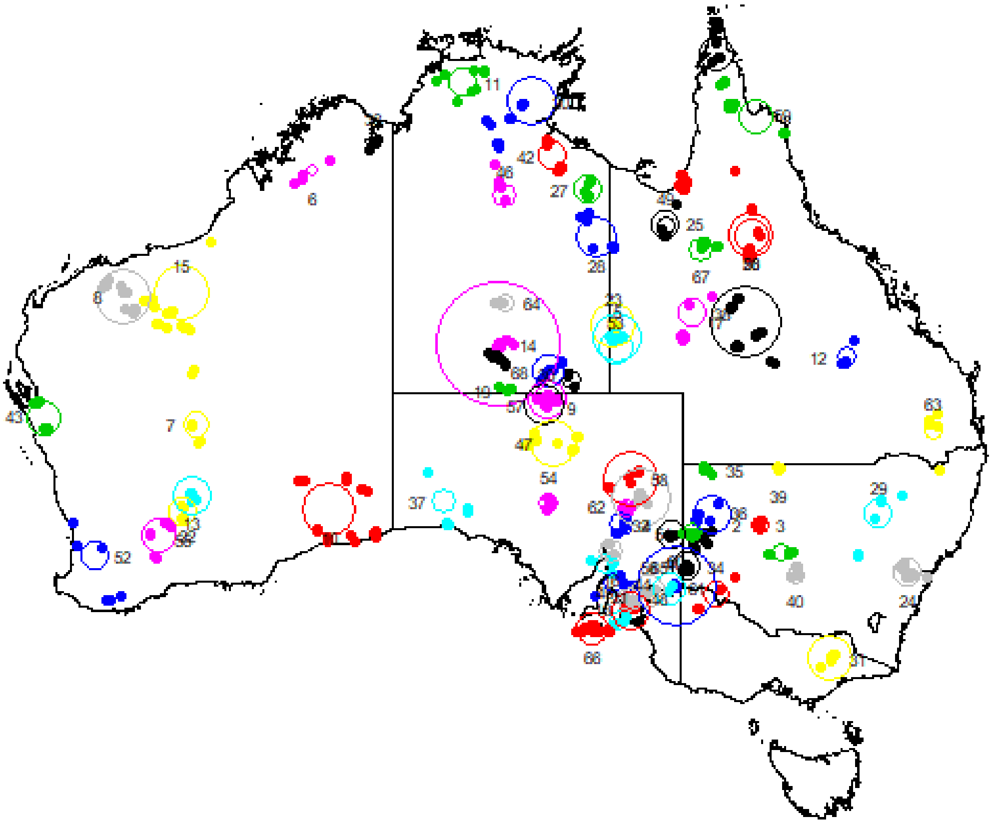
Supplementary material S3. Example of the comparison of different optimisers after applying the *optim_species* function to an alternative dataset. In this case the dataset corresponds to plots located in South Australia within the frame of the project Transect for Environmental Monitoring and Decision Making (TREND). We applied the maximal coverage problem selecting 85 plots to optimise.



Supplementary material S4. Climate variables employed to calculate the environmental representativeness of the selected plots. The variables were obtained from Harwood et al. (2016)

Code	Brief description
Aridity index	
ADI	Minimum monthly aridity index (proportion)
ADM	Mean annual aridity index (annual precipitation/annual potential evaporation) (proportion)
ADX	Maximum monthly aridity index (proportion)
Evaporation	
EPA	Annual potential evaporation (mm)
EPI	Minimum monthly potential evaporation (mm)
EPX	Maximum monthly potential evaporation (mm)
EAA	Annual total actual evapotranspiration terrain scaled using MODIS (mm)
EAAS	Annual total actual evapotranspiration modelled using terrain-scaled water holding capacity (mm)
Maximum temperature	
TXM	Maximum temperature – Annual mean °C
TXI	Maximum temperature - monthly minimum °C
TXX	Maximum temperature - monthly maximum °C
Minimum temperature	
TNM	Minimum temperature – Annual mean °C
TNI	Minimum temperature - monthly minimum °C
TNX	Minimum temperature - monthly maximum °C
Temperature range	
TRI	Minimum monthly mean diurnal temperature range °C
TRX	Maximum monthly mean diurnal temperature range °C
TRA	Annual temperature range (TXX – TNI) °C
Precipitation	
PTA	Annual precipitation mm
PTI	Minimum monthly precipitation mm
PTX	Maximum monthly precipitation mm
PTS1	Precipitation seasonality 1- solstice seasonality composite factor ratio
PTS2	Precipitation seasonality 2- equinox seasonality composite factor ratio
Water deficit	
WDA	Annual atmospheric water deficit (annual precipitation – annual potential evaporation) mm
WDI	Minimum monthly atmospheric water deficit (precipitation - potential evaporation) mm
WDX	Maximum monthly atmospheric water deficit (precipitation - potential evaporation) mm

Supplementary material S5.



S5-Fig. 1. Map representing geographic clusters of plots to translate the findings into a realistic and feasible approach. Colours indicate different clusters, number indicate the cluster ID (see Supplementary material S3) and the size of the circle indicates the number of plots within each cluster.

S5-Table 1. Clusters of plots, coordinates of their centroids and number of plots within them

Cluster ID	Number of plots	Number of sites	Latitude	Longitude
clu1	12	12	-11.9637	142.3816
clu2	6	6	-31.4776	144.2586
clu3	6	6	-32.6306	145.1296
clu4	8	5	-31.4964	138.544
clu5	17	6	-23.7562	138.3792
clu6	4	4	-16.8637	125.6894
clu7	9	8	-27.353	120.8142
clu8	19	19	-22.0562	117.8034
clu9	7	7	-25.5573	136.4409
clu10	18	18	-30.8827	126.355
clu11	10	10	-13.1988	131.9348
clu12	7	7	-24.568	147.8653
clu13	13	13	-30.3101	120.6808
clu14	42	18	-24.0693	133.4427
clu15	19	19	-21.8701	120.2695
clu16	8	4	-32.6554	138.0497
clu17	24	24	-23.1361	143.6343
clu18	13	7	-35.0169	138.8766
clu19	4	4	-25.9328	133.86
clu20	17	17	-13.9922	134.7901
clu21	7	4	-35.5107	138.4406
clu22	12	12	-31.9722	119.3232
clu23	15	5	-23.2451	138.2123
clu24	10	9	-33.4671	150.4113
clu25	6	6	-19.079	140.3453
clu26	12	12	-19.5532	143.9305
clu27	10	10	-17.6146	137.0923
clu28	14	14	-19.6014	137.4888
clu29	10	10	-31.0635	149.1966
clu30	13	13	-26.3325	135.3958
clu31	15	15	-37.0249	147.163
clu32	20	10	-30.4005	139.3001
clu33	4	4	-15.7952	128.1739
clu34	9	9	-34.3382	142.4097
clu35	4	4	-29.2578	142.0513
clu36	13	11	-31.0322	142.2075
clu37	8	8	-30.5226	131.1769
clu38	10	10	-22.7827	141.4962
clu39	4	4	-29.2031	145.0439
clu40	6	6	-33.5424	145.6879
clu41	10	6	-31.9441	140.5602

clu42	10	10	-16.27	135.6802
clu43	12	9	-27.0428	114.4813
clu44	7	4	-33.951	138.207
clu45	7	6	-33.1889	137.8196
clu46	8	8	-17.9138	133.6234
clu47	16	16	-28.1194	135.6589
clu48	8	4	-34.5859	138.8827
clu49	10	10	-19.079	140.3453
clu50	15	15	-19.5532	143.9305
clu51	7	7	-31.8375	141.4666
clu52	10	10	-32.7933	116.6572
clu53	12	4	-24.1515	138.2374
clu54	7	7	-30.5942	135.4801
clu55	10	10	-30.9752	120.2899
clu56	11	8	-34.2875	139.6342
clu57	14	14	-26.5941	135.2836
clu58	19	11	-29.5789	138.9052
clu59	12	12	-14.5616	144.045
clu60	27	16	-34.0285	140.7325
clu61	11	7	-34.1447	140.4778
clu62	6	3	-30.7738	138.67
clu63	6	6	-27.5596	151.4464
clu64	6	6	-22.29	133.654
clu65	10	10	-33.2362	141.2395
clu66	11	11	-35.8077	137.3207
clu67	8	8	-20.1054	141.8365
clu68	11	11	-25.1547	135.4541

Supplementary material S6. Species accumulated when applying the maximum coverage problem to select a subset of 250 plots to revisit. For details of selected plots see associated datasets in Dryad (Martin-Fores et al. 2021).

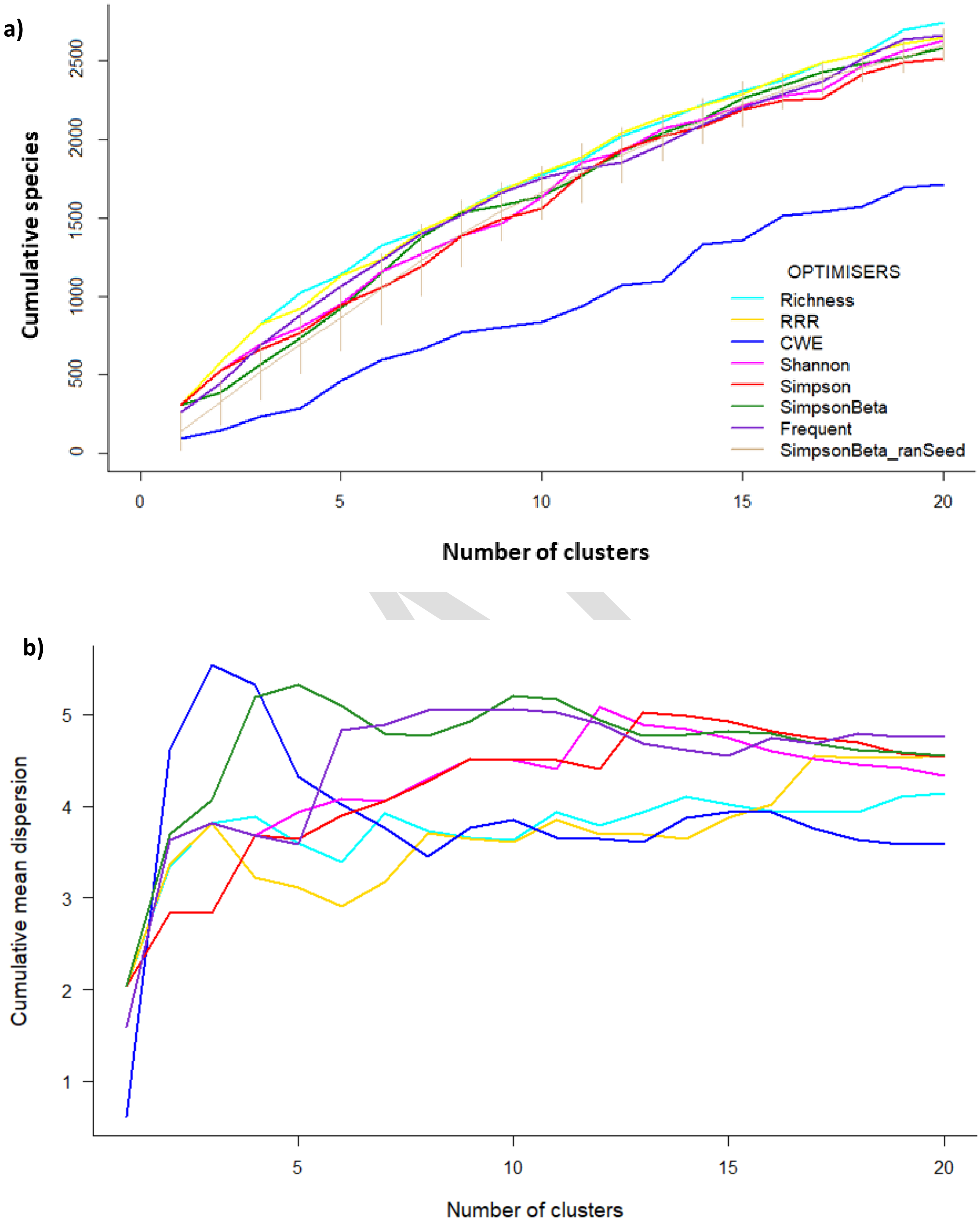
Optimiser	**Number of plots**	**Number of sites**	**Species accumulated**	**% of total species**
Richness	250	229	2864	81.2%
RRR	250	233	2866	81.2%
CWE	250	235	2024	57.4%
Shannon	250	231	2756	78.1%
Simpson	250	234	2633	74.6%
Simpson_Beta	250	245	3021	85.6%
Frequent	250	245	3051	86.5%
Simpson_Random	250		3030.9 ± 14.7	85.9%

Supplementary material S7. Results of monitoring strategy optimisation for plot clusters

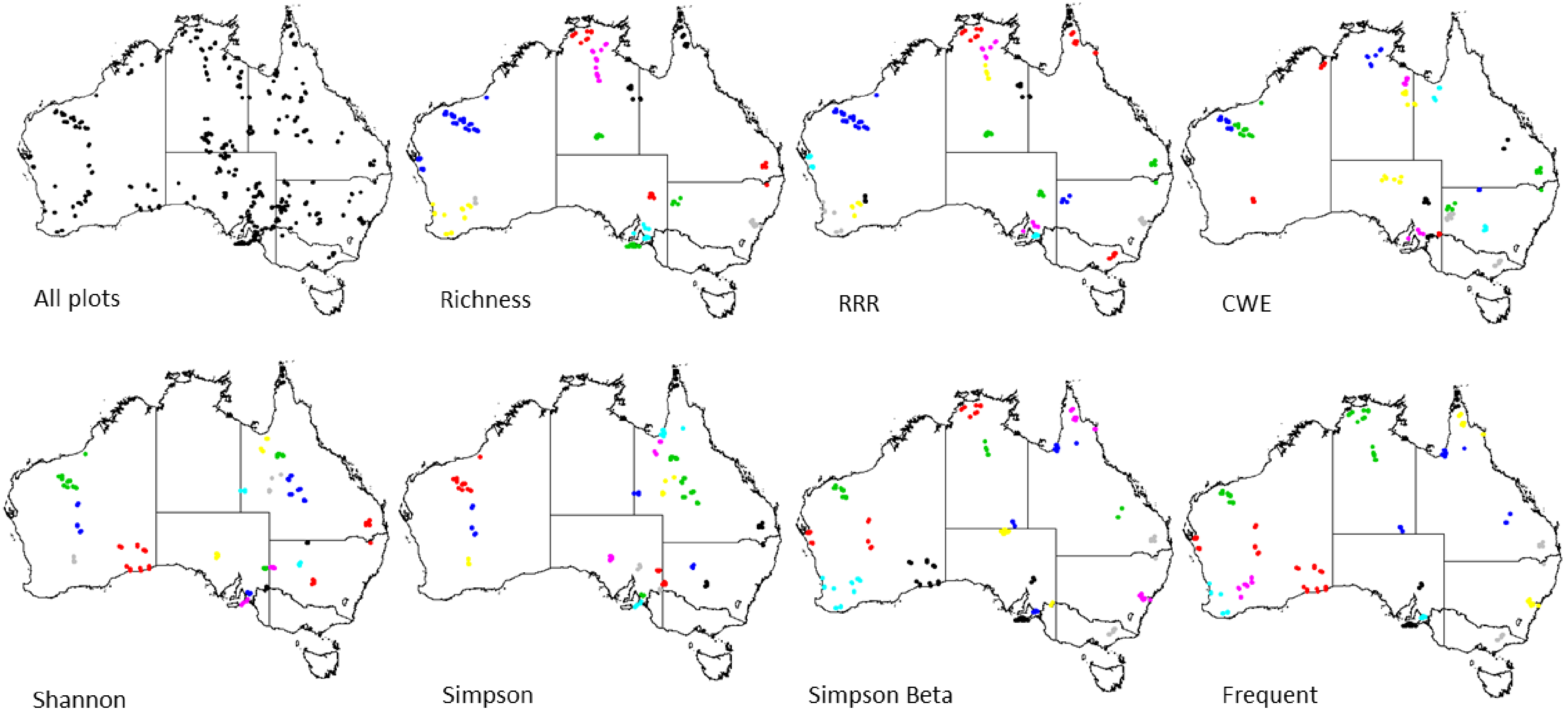
The results obtained when analysing clusters of plots differed from those obtained for single plots. In terms of ecological representativeness, the differences in species accumulation among most of the different optimisers become diluted when selecting clusters of plots instead of single plots. The only biodiversity metric that had a considerably worse performance was CWE (S7-Fig. 1).

Regarding spatial representativeness, when selecting clusters of plots, we obtained opposite trends than when selecting single plots. Species richness was the best optimiser regarding spatial representativeness (R = 0.218), followed by RRR, Frequent and CWE optimisers (R = 0.188, R = 0.184 and R = 0.172). Pairwise Simpson dissimilarity (Simpson_Beta) showed Clark-Evans values of R = 0.134. Clusters of plots selected with Shannon and Simpson metrics displayed both the most clustered spatial coverages (R = 0.011 and R = 0.009, respectively; S7-Fig. 2; Supplementary material S8).

The most environmentally representative subset of selected clusters was obtained using species turnover, with Frequent the best optimiser in terms of environmental representation (Frequent: average distance to median = 4.753, respectively; S7-Fig. 1; S7-Table 1), followed by RRR, Simpson_Beta and Simpson (RRR, Simpson_Beta and Simpson: average distance to median = 4.56, 4.55 and 4.53, respectively), with no significant differences among them. 'Frequent' displayed marginally significantly better environmental representativeness than Shannon (Shannon: 4.34), whereas all the four former biodiversity metrics performed significantly (for Frequent) or marginally significantly (for the later three) better than richness in terms of environmental representativeness (Richness: 4.14). Finally, all the optimisers performed significantly better than CWE in terms of environmental representativeness (CWE: average distance to median = 3.60; S7-Fig. 1; S7-Table 1).



S7-Fig. 1. a) Site optimisation process applying conservation reserve design strategies based on Maximum coverage problem (selection of 20 clusters); b) environmental representativeness of the 20 selected clusters using different optimisers reflected by the cumulative mean dispersion



S7-Fig. 2. Geographic location of the selected clusters (n = 20) applying the maximal coverage problem. Black dots correspond to all the plots established. Colour dots refer to each of the selection employing different optimisers.

S7-Table 1. Pairwise comparisons between optimisers with regards to environmental representativeness when applying maximal coverage problem selecting spatial clusters. The observed p-value are located in the below diagonal, while the permuted p-value are in the above diagonal. Only significant differences are highlighted in bold. Notice that marginally significant values (p-value ≤0.1) are shown although not highlighted.

	Richness	RRR	CWE	Shannon	Simpson	SimpsonBeta	Frequent
Richness		≤0.1	**≤0.05**	**≤0.001**	**≤0.001**	**≤0.05**	**≤0.01**
RRR	≤0.1		**≤0.01**	**≤0.001**	**≤0.001**	0.99	0.42
CWE	**≤0.05**	**≤0.001**		**≤0.001**	**≤0.05**	**≤0.001**	**≤0.001**
Shannon	**≤0.001**	**≤0.001**	**≤0.01**		0.36	**≤0.001**	**≤0.001**
Simpson	**≤0.001**	**≤0.001**	**≤0.05**	0.35		**≤0.001**	**≤0.001**
SimpsonBeta	**≤0.05**	0.99	**≤0.001**	**≤0.001**	**≤0.001**		0.39
Frequent	**≤0.05**	0.42	**≤0.001**	**≤0.001**	**≤0.001**	0.41	

Supplementary material S8. Selected clusters with the different optimisers, number of sites and plots included and species accumulated

Optimiser	**Clusters selected**	**Number of plots**	**Number of sites**	**Species accumulation**
Richness	clu59, clu52, clu8, clu18, clu15, clu43, clu1, clu14, clu55, clu22, clu24, clu36, clu20, clu11, clu28, clu46, clu63, clu32, clu66, clu44	278	230	2645
RRR	clu59, clu8, clu52, clu15, clu43, clu14, clu18, clu1, clu55, clu36, clu20, clu28, clu63, clu22, clu24, clu11, clu31, clu32, clu46, clu44	282	234	2693
CWE	clu56, clu31, clu63, clu39, clu60, clu12, clu40, clu36, clu8, clu28, clu49, clu27, clu25, clu20, clu33, clu44, clu47, clu13, clu32, clu15	245	216	1698
Shannon	clu59, clu46, clu1, clu66, clu63, clu44, clu11, clu24, clu55, clu43, clu36, clu31, clu51, clu37, clu52, clu18, clu47, clu48, clu22, clu49	210	173	2435
Simpson	clu59, clu46, clu1, clu66, clu11, clu63, clu44, clu55, clu24, clu43, clu37, clu47, clu31, clu25, clu48, clu12, clu9, clu51, clu36, clu52	195	184	2351
Simpson_Beta	clu59, clu61, clu66, clu31, clu22, clu30, clu52, clu24, clu8, clu11, clu63, clu43, clu4, clu9, clu50, clu18, clu46, clu10, clu12, clu7	227	212	2628
Frequent	clu8, clu66, clu52, clu50, clu43, clu31, clu24, clu11, clu46, clu18, clu63, clu7, clu12, clu10, clu55, clu1, clu9, clu59, clu22, clu4	225	213	2706
Simpson_Random				2626.5 ± 49.8