How many sites? Methods to assist design decisions when collecting multivariate data in ecology

Ben Maslen[1], Gordana Popovic[1], Adriana Verges[1], Ezequiel Marzinelli[2], and David Warton[1]

[1]University of New South Wales
[2]The University of Sydney Faculty of Science

April 16, 2024

Abstract

Sample size estimation through power analysis is a fundamental tool in planning an ecological study, yet there are currently no well-established procedures for when multivariate abundances are to be collected. A power analysis procedure would need to address three challenges: designing a parsimonious simulation model that captures key community data properties; measuring effect size in a realistic yet interpretable fashion; and ensuring computational feasibility when simulation is used both for power estimation and significance testing. Here we propose a power analysis procedure that meets these challenges with accompanying R software (ecopower). Our simulation model uses a Gaussian copula model, and expert opinion is leveraged to simplify effect size specification into "increasers", "decreasers" or "no effect" taxa. Computational issues are addressed by using a critical value approach, reducing computation time from days to minutes. The procedure is demonstrated by estimating the sample size required to detect changes for fish abundances.

Hosted file

`Main_Document.tex` available at https://authorea.com/users/738620/articles/712780-how-many-sites-methods-to-assist-design-decisions-when-collecting-multivariate-data-in-ecology

How many sites? Methods to assist design decisions when collecting multivariate data in ecology

Ben Maslen[1], **Gordana Popovic**[1,2], **Adriana Vergés**[2,3,4], **Ezequiel Marzinelli**[4,5,6], **David Warton**[1,2]

[1]*School of Mathematics and Statistics, University of New South Wales, NSW 2052, Australia*

[2]*Evolution and Ecology Research Centre, University of New South Wales, Sydney, NSW 2052, Australia*

[3]*School of Biological, Earth, and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia*

[4]*Sydney Institute of Marine Science, 19 Chowder Bay Rd, Mosman NSW 2088, Australia*

[5]*The University of Sydney, School of Life and Environmental Sciences, Coastal and Marine Ecosystems, Sydney NSW 2006, Australia*

[6]*Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore*

Contact details (respectively):

b.maslen@unsw.edu.au, g.popovic@unsw.edu.au, a.verges@unsw.edu.au, e.marzinelli@sydney.edu.au, david.warton@unsw.edu.au

Short Running Title: Tools for planning multivariate studies

Keywords: sample size, power analysis, restoration, software, statistics, copula, simulation, crayweed, study design

Article type: Method

Statement of Authorship: BM designed the power analysis procedure and built the R package ecopower. DW and GP provided expertise and direction in copula modelling, multivariate analysis and power simulation. AV and EMM provided the fish abundance data, the motivating example and used their domain expertise to help define the effect size of interest. BM wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

Data accessibility statement: Should the manuscript be accepted, the data will be archived in the Dryad public repository and the data DOI will be included at the end of the article.

Number of:

Abstract words: 149

Main text words: 4075

Text box words (respectively): 33, 15

References: 35

Figures: 4 (+2 in Appendix)

Tables: 0

Text boxes: 2

Corresponding author: Ben Maslen, +61438790990, b.maslen@unsw.edu.au

Abstract

Sample size estimation through power analysis is a fundamental tool in planning an ecological study, yet there are currently no well-established procedures for when multivariate abundances are to be collected. A power analysis procedure would need to address three challenges: designing a parsimonious simulation model that captures key community data properties; measuring effect size in a realistic yet interpretable fashion; and ensuring computational feasibility when simulation is used both for power estimation and significance testing. Here we propose a power analysis procedure that meets these challenges with accompanying `R` software (`ecopower`). Our simulation model uses a Gaussian copula model, and expert opinion is leveraged to simplify effect size specification into "increasers", "decreasers" or "no effect" taxa. Computational issues are addressed by using a critical value approach, reducing computation time from days to minutes. The procedure is demonstrated by estimating the sample size required to detect changes for fish abundances.

Introduction

In planning any study it is important to consider how large an effect ("effect size") can be detected by the intended study design (Rosenthal et al., 1994; Fritz and MacKinnon, 2007; Kelley and Preacher, 2012). Such information is often obtained via a power analysis of preliminary data, which can help design a study that has a good chance of detecting effects of ecological interest (Cohen, 1992, 2013). While texts for ecologists frequently discuss the importance of power analysis, and related techniques, in sample size determination (Gerrodette, 1987; Johnson et al., 2015; Green and MacLeod, 2016), there is little guidance for the ecologist on how to undertake such an analysis for multivariate data, i.e. multiple response variables measured per sample. This is particularly the case if analysing abundance or presence-absence data simultaneously collected for many different taxa; hereafter (*multivariate abundances*).

There are three key challenges that would need to be addressed in developing a power analysis procedure for multivariate abundance data. *Challenge 1* is to design a simulation model to randomly generate realistic multivariate abundance data, reflecting key properties of the data to be collected. It should be possible to use pilot data, when available, to tune the settings of the simulation model, to generate data that "looks like" the pilot data. Some simulation approaches are available in the literature (e.g. Xu et al., 2010), however with limited ability to tune using pilot data. Other methods have also been developed using a Bray-Curtis distance based approach (Irvine et al., 2011) or by considering the abundances as continuous random variables and calculating Euclidean distances (Collins et al., 2000; Angeler et al., 2009). These approaches however ignore important mean-variance assumptions and therefore subsequent analyses can be misleading (Warton et al., 2012). Multivariate modelling approaches have been developed recently that could be used to address this, including Gaussian copulas (Popovic et al., 2018; Anderson et al., 2019) and hierarchical models (Warton, Foster, Death, Stoklosa and Dunstan, 2015; Ovaskainen et al., 2017). *Challenge 2* is measuring effect size. When many taxa are to be sampled, many parameters need to be specified *a priori* that will capture the size and nature of the effect the study has been designed to detect. Decisions about these parameters are to be made based on relatively little information, and need to be captured by a relatively simple, interpretable effect size measure in order for results to be useful for study design (Kelley and Preacher, 2012). *Challenge 3* is making the power analysis procedure computationally efficient. Hypothesis testing procedures for multivariate abundance data typically use resampling (Anderson, 2001; Wang et al., 2012), conventionally recalculating a test statistic at least 1000 times across resampled datasets. A power analysis would require this to be done for each of say 1000 simulated datasets, such that estimating power for a typical multivariate abundance dataset would take hours or days.

This paper proposes a power analysis procedure for multivariate abundance data that addresses each of the above three challenges. *Challenge 1* is addressed using a Gaussian copula-factor analysis model to simulate data in a parsimonious fashion, an approach for which methods to tune the model using training data were only recently proposed (Popovic et al., 2018). *Challenge 2* is addressed using a simple parameterisation for effect size that requires *a priori* information only concerning taxa that are likely to be affected, and the direction of the effect. *Challenge 3* is addressed using a novel "critical value" approach to power analysis of resampled statistics, which can reduce computation time from days to minutes. The procedure will be illustrated on a marine habitat restoration project which involves regular monitoring of ecological communities, where we are interested in the number of samples required for a future monitoring period in order to likely detect community differences in abundance across treatments, as well as the size of effects that are able to be detected under different sampling designs.

Materials and Methods

Operation Crayweed Restoration Project

Researchers within the Operation Crayweed Restoration Project in Sydney are restoring the locally extinct macro-algae *Phyllospora comosa* (“crayweed”: see Coleman et al., 2008; Campbell et al., 2014; Vergés et al., 2020) and are interested in the effect of this restoration on associated ecological communities (Marzinelli et al., 2014, 2016; Wood et al., 2019). Pilot data have already been collected, where the abundance of fish species in nine open ocean sites have been recorded. We are interested in observing if there is a change in mean fish abundance between three treatments: control (sites in Sydney without crayweed), restored (similar to control sites where crayweed has recently been planted) and reference (sites north and south of Sydney with extant crayweed forests), as in Figure 1. There are plans to collect more data in the future, however there is an upper bound of approximately 24 possible spatially independent restored or control coastal bays/sites within the Sydney region and surroundings.

We are interested in answering the following experimental design questions from the pilot data, which will act as the motivating examples for this paper:

1. How many sites are required in order to likely detect 20% differences between treatments?
2. Under the maximum balanced sampling design of 12 sites for each treatment, what are the size of effects that are likely to be detected?

Power analysis

Consider a situation where the objective of the study is to test a specific null hypothesis (H_0) by looking for evidence of some alternative (H_1). The effectiveness of any given study design and testing procedure can be evaluated using power, i.e. the probability of rejecting the null hypothesis H_0 given that a particular alternative hypothesis H_1 is true. Power generally increases as sample size and effect size increases, and variability decreases (Cohen, 2013). Thus, given an understanding of the variability to be expected in the data, power can be used to work out how large a sample size is needed to detect an effect of a given size, or what sizes of effects can be detected in study of a given sample size (Cohen, 1992, 2013).

In order to undertake a power analysis, it is necessary to work through the following steps:

1. Specify a model for data, which captures key properties of the data that are expected to be collected. For multivariate abundances, as in Operation Crayweed, this is a non-trivial task (*Challenge 1*).
2. Decide on a measure of effect size that is ecologically meaningful. When there are many taxa, as for fish assemblages sampled in Operation Crayweed, there are many different effect size parameters that need to be considered (*Challenge 2*).
3. Decide on the testing procedure. Hypothesis tests of multivariate abundance data typically use resampling to ensure valid inference (Anderson, 2001; Wang et al., 2012); in this paper testing will make use of a generalised estimating equations approach (Wang et al., 2012, using the `mvabund` package) that is increasingly common in ecology. For the crayweed fish abundance data, we assumed abundances have a negative binomial distribution, with diagnostic plots suggesting this adequately captured the mean-variance trend in the data (Figures 5 and 6 in Appendix).

4. Estimate power. In some simpler settings this can be done analytically (Cohen, 2013), however for multivariate abundances this needs to be done by simulation, generating data under the assumed model, then applying the testing procedure for each simulated dataset, and recording the proportion of times the null hypothesis was rejected. When using a resampling-based testing procedure, this involves two levels of simulation and will be very computationally intensive (*Challenge 3*).

The following sections detail the proposed solutions to the three challenges identified above.

Challenge 1 – Data generating model

A data generating model is needed that can capture key properties of multivariate abundance data that will be collected. Multivariate abundances are discrete, with many zeros, and high dimensional, with a large number of responses p relative to the sample size n. The ability to specify a parametric statistical model that can be fitted to multivariate abundances is a relatively recent development (Popovic et al., 2018; Warton, Blanchet, OHara, Ovaskainen, Taskinen, Walker and Hui, 2015). This is an important advance because it allows the simulation model to be tuned to pilot data in order to generate data statistically similar to what will actually be observed when the study is undertaken. In particular, power can be strongly affected by mean abundance, variability, and correlation across taxa (Warton, 2011), all of which can vary considerably from one study to another. These data properties need to be tuned to the study in question for a power analysis to be informative.

This paper adopts a Gaussian copula approach. To date, copulas have rarely been used in ecology (Popovic et al., 2018; Anderson et al., 2019; Popovic et al., 2019). A specific advantage of a copula approach is that it specifies a marginal (unconditional) model, making parameters more interpretable. For example, if we set an effect size parameter that ensures a two-fold change in mean abundance between treatment and control groups, we can be sure that there will in fact be a two-fold change. In hierarchical models this does not always happen, because they operate as conditional models, which can induce some surprising behaviour when interpreted marginally (Breslow and Lin, 1995; Lin and Breslow, 1996; Gurka et al., 2011). A nice feature of the copula model is that it assumes the same marginal model as the testing procedure to be used here (Wang et al., 2012), and so is a suitable simulation model when using a Generalised Estimating Equation (GEE) procedure. Copulas are also used in Anderson et al. (2019), although to compare different test procedures, rather than for sample size determination.

A discrete Gaussion copula (Popovic et al., 2018, 2019) can model correlated discrete data y_{ij} using latent Gaussian variables z_{ij},

$$Y_{ij} = F_j^{-1}(\Phi(Z_{ij})) \tag{1}$$

Where F_j is the marginal distribution for taxon j (e.g. Poisson with mean $\lambda = \exp(XB)$). Φ is a multivariate Gaussian distribution with zero mean and covariance structure Σ,

$$z_{ij} \sim \mathcal{N}_p(0, \Sigma).$$

This model is estimated using maximum likelihood to obtain $\hat{F}_j$ and $\hat{\Sigma}$. To simulate new multivariate abundances we simply simulate new latent variables $z_{ij} \sim \mathcal{N}_p(0, \hat{\Sigma})$ and then obtain abundances by transformation $y_{ij} = {F^*}_j^{-1}(\Phi(z_{ij}))$, where $\hat{F}_j^*$ has been parameterised via an effect of interest.

One difficulty with this approach however is estimating a covariance structure across responses. With p taxa, there are $p(p-1)/2$ pairwise correlations to estimate, and these will typically be estimated from a small amount of pilot data. Thus a parsimonious method of modelling correlations is needed if the simulation model is to be trained using pilot data. This issue is addressed here by assuming the p variables are driven by a shared response to a few ($q \ll p$) unobserved latent variables through the use of factor analysis. These latent variables can be interpreted as unobserved environmental covariates (Warton, Blanchet, OHara, Ovaskainen, Taskinen, Walker and Hui, 2015). This approach requires $p(q+1) - q(q-1)/2$ elements to be estimated to formulate a covariance matrix Σ_{FA}, which can be much smaller than $p(p-1)/2$. For our fish abundance data set with $p = 34$ species, if we take $1, 2$ or 3 factors we have 68, 101 or 133 covariance parameters to estimate, which is much less than the 561 parameters that would otherwise have been needed.

This process has been implemented within the `ecopower` package, using an internal function called `extend`. This function takes a `cord` object (obtained by fitting a Gaussian copula to a `manyglm` object using the `cord` function from the `ecocopula` package; Popovic et al., 2021) and simulates `N` multivariate abundances using the above procedure (to date, it can handle Poisson, negative binomial and binomial distributions). The function then refits the simulated responses to a `manyglm` object with a data frame that is 'extended' in a manner that preserves the original or pre-specified design (Box 2).

Challenge 2 – Specifying an interpretable measure of effect size

With a large number of taxa, there are a large number of ways that these taxa can respond to a treatment. In order to undergo a power analysis, an effect needs to be specified in a way that is interpretable, such that the researcher can understand how large the effect is in the context of their study, thus helping to understand whether or not it is necessary to increase the sample size. This interpretable measure of effect size also needs to be specified with relatively little *a priori* information, which is generally lacking in an ecological setting (at least in comparison to the possible complex relations that the taxa can respond to a treatment).

A simple approach proposed here is to decide on:

1. Which species/response variables are expected to be (i) positively related to a given treatment (e.g. species that increase in abundance;“increasers”), (ii) those expected to be negatively related (species that decrease in abundance; “decreasers”), or (iii) those not related to the treatment at all.
2. The size of effect (ρ) that is negatively or positively related to the mean abundance μ of taxa, on the proportional scale. That is, $\rho = 2$ means that the mean abundance doubles for increasers, and halves for decreasers.

This is a relatively simplistic scenario, however it enables the effect size to be captured in a single coefficient ρ, and for expert opinion to inform the way in which different taxa are likely affected. Being on the proportional scale it also allows regression coefficients for simulated models to be easily specified, for example with log link; $\log \rho = \log \frac{\hat{\mu}_{Treat=1}}{\hat{\mu}_{Cont=0}} = \beta_{1j}$ for species that increase in abundance and $\log 1/\rho = -\log \frac{\hat{\mu}_{Treat=1}}{\hat{\mu}_{Cont=0}} = -\beta_{1j}$ for species that decrease in abundance in the treatment group, relative to the control.

For the Crayweed Restoration Project, we believe that restored sites will lie somewhere between reference and control sites (e.g. because the extent of crayweed restored is still small relative to extant, natural populations; Layton et al., 2020), and by using existing data of fish surveys along the NSW coastline, as well as results from Curley et al. (2002), assumptions were made as to which fish species will be expected to change in sites with crayweed (reference and restored), compared to those without crayweed (control). This allows the desired parameterisation of three types of fish species, whose mean abundances increase, decrease or do not change in reference and restored sites relative to control sites (Figure 2, with the magnitude of these effects being specified with ρ for changes between restored and control sites, and ρ^2 for changes between reference and control sites).

By using pilot data to estimate mean abundance of each species, then taking for example $\rho = 1.2$ to specify effects across different treatments, the mean abundances of data to be simulated from marginal distributions F_j can be specified as in Figure 2. $\rho - 1$ can also be interpreted as the $\%$ change in mean abundance across treatments, so to answer our first experimental design question, we simply specify $\rho = 1.2$.

This approach has been implemented within the `ecopower` package, with the `effect_alt` function. Users input a `manyglm` object, the name of the predictor of interest (`term`), an effect size of interest (`effect_size`) and a list of taxon that are "`increasers`" or "`decreasers`". The function then returns a parameterised coefficient matrix that can be used in ensuing power simulations, where taxa not specified as "`increasers`" or "`decreasers`" are assumed to not be affected. There are also options to specify more complicated effect sizes for effects that change over multiple levels of a categorical predictor. To produce the coefficient matrix in Figure 2 we would use the code in Box 1.

```
> library(ecopower)
> fit <- manyglm(fish˜ Site.Type, family="negative.binomial",data = X)
> increasers <- c("Aplodactylus.lophodon","Atypichthys.strigatus",
+                 "Cheilodactylus.fuscus","Olisthops.cyanomelas",
+                 "Pictilabrius.laticlavius" )
> decreasers <- c("Abudefduf.sp","Acanthurus.nigrofuscus","Chromis.hypsilepis",
+                 "Naso.unicornis","Parma.microlepis","Parupeneus.signatus",
+                 "Pempheris.compressa","Scorpis.lineolatus","Trachinops.taeniatus")
> coeff.alt <- effect_alt(fit,effect_size=1.2,increasers,decreasers,
+                 term="Site.Type")
```

Box 1: Code used to generate the effect size specification in Figure 2.

Challenge 3 – Managing computation time

Power at significance level α can be estimated from a set of n_{power} simulated datasets, where the ith dataset returned P-value p_i, as follows:

$$\widehat{\text{Power}} = \frac{1}{n_{power}} \sum_{i=1}^{n_{power}} I_{\{p_i \leq \alpha\}},$$

where $I_{\{\cdot\}}$ is the indicator function. That is, we estimate the power at significance level α as the sample proportion of P-values p_i less than or equal to α. In order to get a reliable Monte Carlo estimate of power we would conventionally use approximately $n_{power} = 1000$ datasets. In a multivariate abundance setting, the significance of multivariate Wald or score GEE test statistics is calculated via resampling techniques (Wang et al., 2012). The standard asymptotic chi-squared techniques we would otherwise use are not suitable here because they assume large sample size (n) and a fixed number of responses (p), however for multivariate abundances, p is rarely small compared to n. Nevertheless, if resampling is used for testing, then for each n_{power} simulated data set we would have to resample an additional $n_{resamp} \approx 1000$ times to estimate a simulated p-value p_i as $\hat{p}_i$. Each of these test statistics involve fitting a model to p correlated taxa, which will involve fitting $p \times (n_{power} + n_{power} \times n_{resamp})$ GLMs overall and takes approximately 1.3 hours for a small data set of size $n = 21$ and $p = 34$ taxa, using parallel computing on a standard processing computer with 12 logical processes.

A key innovation proposed in this paper is to use a critical value approach to estimate power, globally testing the significance of n_{power} simulated test statistics T_{1_i} using a critical value $\hat{c}_\alpha$. The critical value can be estimated as the upper $1 - \alpha$ qauntile of simulated test statistics under the null hypothesis $T_{0_j} \in H_0$ satisfying

$$\frac{1}{n_{resamp}} \sum_{j=1}^{n_{resamp}} \mathbb{P}(T_{0_j} > \hat{c}_\alpha) = \alpha,$$

and power can then be estimated as

$$\widehat{\text{Power}}_{crit} = \frac{1}{n_{power}} \sum_{i=1}^{n_{power}} I_{\{T_{1_i} > \hat{c}_\alpha\}}.$$

In this approach power is estimated using test statistics T_{1i} from simulated data, rather than P-values. This means that there is no longer a need to resample the simulated datasets, leading to a huge computational saving. The above approach involves only $n_{power} + n_{resamp} \approx 2000$ simulated or resampled datasets, a reduction by a factor of 500, and reducing computation time for the Crayweed restoration dataset from over an hour to just 42 seconds. Note this is an approximation to power, essentially, because we are assuming that $\hat{c}_\alpha$ is constant across all simulated datasets, when it may vary slightly. As sample size increases the accuracy of this approximation will improve, with the simulations below investigating how well this approximation works in our context.

The critical value approach to estimating power is implemented in the `ecopower` package, through the `powersim` function (which calls the internal `extend` function). It takes a `cord` object, a coefficient matrix (`coeffs`) that can be specified using the `effect_alt` function, a total sample size `N`, the name of the predictor of interest (`term`), number of simulations (`nsim`) and returns a power estimate. Computation time is also reduced by running simulations in parallel over a series of clusters (`ncores`), which defaults to one less then the number of cores available on your machine. To estimate power, for a sample size of $N =$

100, using the pre-specified effect of $\rho = 1.2$ from `coeff.alt`, we would use the code in Box 2 (computation time is in seconds).

```
> fit_factors.cord = cord(fit)
> powersim(fit_factors.cord,N=100,coeff.alt,term="Site.Type",nsim=1000)
    Power Comp time
  0.43700  42.02651
```

Box 2: `ecopower` code to obtain a power estimate over `nsim=1000` simulations for a sample size of `N=100` using the `powersim` function by first fitting a copula model from our `manyglm` object using the `cord` function and then using the effect size (`coeff.alt`) generated in Box 1.

Results

Validating method

The effectiveness of estimating $\widehat{\text{Power}}$ as $\widehat{\text{Power}}_{crit}$ can be observed through direct application with simulations. Both methods are applied to answer our first experimental design question. That is, to find the number of samples required in order to likely detect an effect size of 20% change in mean abundance of fish species between the three treatments (where some species have been specified to increase, decrease or not change across the treatments, Figure 2).

As can be seen in Figure 3, the use of critical test statistics approximates power quite well, even for small study designs. Both approaches recommended at least $n^* = 57$ sites per treatment group in order to likely detect 20% changes in mean abundances. The critical value approach also reduced computation time from over 3 days to just 11 minutes. It tended to give slightly more variable power estimates, which could be overcome by increasing the number of resamples (n_{resamp}) or simulations (n_{power}) at the cost of increased computation time. The critical value approach would still however be appreciably faster (unless n_{resamp} and n_{power} were increased to $500,000$!).

Example

While Figure 3 suggests $n^* = 57$ sites per treatment group, we note this is not feasible under a balanced experimental design, because only 24 control or restored sites are available within the region where this restoration project was undertaken. Hence, it is unlikely that we will detect 20% changes in the mean abundances of fish species between the treatment levels. This is most likely due to the variability of the observed fish abundances creating a lot of noise in the data and the large (66%) proportion of zero counts observed in the data.

The second experimental design question we are interested in is: what effect sizes are likely to be observed at 12 sites per treatment? In order to answer this question, we can simply plot power curves over a range of effect size specifications. We actually plotted power curves as a function of sample size n for a broader perspective (Figure 4), for effect size specifications $\rho_{10\%} = 1.1, \rho_{20\%} = 1.2, \rho_{30\%} = 1.3, \rho_{40\%} = 1.4, \rho_{50\%} = 1.5, \rho_{60\%} = 1.6, \rho_{70\%} = 1.7, \rho_{80\%} = 1.8$. Computing each power curve took approximately 5 minutes of computation time on a computer with 12 logical processors, less time than for Figure 3, because the sample sizes under consideration here were smaller. We observe as expected that as we increase the effect size specifications, we increase the power of each experimental design (Figure 4), since larger effect sizes are easier to detect. Importantly, we observe that under the maximum balanced sample size of $n^* = 12$ sites per treatment, the smallest effect size likely to be observable is $\rho_{50\%} = 1.5$, or equivalently 50% changes in the mean abundances of fish species. The pilot survey included $n = 6$ sites per treatment, which would only be able to detect quite extreme effects.

DISCUSSION

Here we have described a sample size estimation procedure for multivariate abundance data. Techniques and software for this are much needed in ecology, however the problem is rather difficult technically, given the three aforementioned challenges. The main innovations of our procedure are to: specify a joint data generating model through discrete margin copulas that is able to be tuned to the properties of the pilot data; implement a simple and interpretable approach to specifying an effect size for complex multivariate effects; reduce computation time by a factor of 500, which was achieved using a "critical value" approach to power estimation. The procedure has been coded in `R` with general purpose functions in the `ecopower` package, which can be downloaded from Cran (Maslen and Lim, 2021). This procedure can be applied to a vast range of ecological studies with multivariate abundance data to answer experimental design questions and give sample size recommendations, an important yet technically difficult task not previously addressed in the literature.

Being a general-purpose procedure, power can also be estimated for more complex designs. For instance, the methods proposed here (and the software tools in the `ecopower` package) can estimate power for effects on both categorical and quantitative predictors, handle the most common distributions for handling abundance or presence-absence data (to date; Poisson, negative binomial and binomial distributions) and handle designs with multiple covariates. It can also investigate unbalanced designs. In our fish abundance example, we could study the change in power when fixing the number of control and restored sites at $n_1 = n_2 = 12$, but varying the number of reference sites, as more of these are available regionally. It is also possible to use `ecopower` to specify different and potentially more complicated effect size structures, if desired. Note however that all effect size parameters need to be set *a priori*, and the more intricate the scenario, the more *a priori* decisions would be needed.

It would be interesting to apply this procedure to other study designs and consider whether the patterns seen here apply elsewhere too. Specifically, this study's intended sampling design ($n = 6$) was underpowered for the size of effects that we would like to detect. Is this more generally true – do ecological monitoring studies tend not to sample enough sites to detect effects of practical interest?

Acknowledgements

David Warton and Gordana Popovic are supported by the Australian Research Council Discovery Grant DP210101923. Ezequiel Marzinelli and Adriana Verges were both supported by the Australian Research Council Linkage Grant LP160100836, the NSW DPI Recreational Fishing Trust, and Environmental Trust. Ezequiel Marzinelli and Adriana Verges were also supported by Australian Research Council Discovery Grants DP180104041 and DP190102030 respectively. We would also like to thank Alexandra Campbell for field sampling.

References

Anderson, M. J. (2001), 'A new method for non-parametric multivariate analysis of variance', *Austral Ecology* **26**(1), 32–46.

Anderson, M. J., de Valpine, P., Punnett, A. and Miller, A. E. (2019), 'A pathway for multivariate analysis of ecological communities using copulas', *Ecology and Evolution* **9**(6), 3276–3294.

Angeler, D. G., Viedma, O. and Moreno, J. M. (2009), 'Statistical performance and information content of time lag analysis and redundancy analysis in time series modeling', *Ecology* **90**(11), 3245–3257.

Breslow, N. E. and Lin, X. (1995), 'Bias correction in generalised linear mixed models with a single component of dispersion', *Biometrika* **82**(1), 81–91.

Campbell, A. H., Marzinelli, E. M., Vergés, A., Coleman, M. A. and Steinberg, P. D. (2014), 'Towards restoration of missing underwater forests', *PloS One* **9**(1), e84106.

Cohen, J. (1992), 'Statistical power analysis', *Current Directions in Psychological Science* **1**(3), 98–101.

Cohen, J. (2013), *Statistical power analysis for the behavioral sciences*, Routledge.

Coleman, M. A., Kelaher, B. P., Steinberg, P. D. and Millar, A. J. (2008), 'Absence of a large brown macroalga on urbanized rocky reefs around Sydney, Australia, and evidence for historical decline', *Journal of Phycology* **44**(4), 897–901.

Collins, S. L., Micheli, F. and Hartt, L. (2000), 'A method to determine rates and patterns of variability in ecological communities', *Oikos* **91**(2), 285–293.

Curley, B. G., Kingsford, M. J. and Gillanders, B. M. (2002), 'Spatial and habitat-related patterns of temperate reef fish assemblages: implications for the design of Marine Protected Areas', *Marine and Freshwater Research* **53**(8), 1197–1210.

Fritz, M. S. and MacKinnon, D. P. (2007), 'Required sample size to detect the mediated effect', *Psychological Science* **18**(3), 233–239.

Gerrodette, T. (1987), 'A power analysis for detecting trends', *Ecology* **68**(5), 1364–1372.

Green, P. and MacLeod, C. J. (2016), '`simr`: an `r` package for power analysis of generalized linear mixed models by simulation', *Methods in Ecology and Evolution* **7**(4), 493–498.

Gurka, M. J., Edwards, L. J. and Muller, K. E. (2011), 'Avoiding bias in mixed model inference for fixed effects', *Statistics in medicine* **30**(22), 2696–2707.

Irvine, K. M., Dinger, E. C. and Sarr, D. (2011), ‘A power analysis for multivariate tests of temporal trend in species composition’, *Ecology* **92**(10), 1879–1886.

Johnson, P. C., Barry, S. J., Ferguson, H. M. and Müller, P. (2015), ‘Power analysis for generalized linear mixed models in Ecology and Evolution’, *Methods in Ecology and Evolution* **6**(2), 133–142.

Kelley, K. and Preacher, K. J. (2012), ‘On effect size.’, *Psychological Methods* **17**(2), 137.

Layton, C., Coleman, M. A., Marzinelli, E. M., Steinberg, P. D., Swearer, S. E., Vergés, A., Wernberg, T. and Johnson, C. R. (2020), ‘Kelp forest restoration in australia’, *Frontiers in Marine Science* **7**, 74.

Lin, X. and Breslow, N. E. (1996), ‘Bias correction in generalized linear mixed models with multiple components of dispersion’, *Journal of the American Statistical Association* **91**(435), 1007–1016.

Marzinelli, E., Campbell, A., Vergés, A., Coleman, M., Kelaher, B. P. and Steinberg, P. (2014), ‘Restoring seaweeds: does the declining fucoid phyllospora comosa support different biodiversity than other habitats?’, *Journal of Applied Phycology* **26**(2), 1089–1096.

Marzinelli, E. M., Leong, M. R., Campbell, A. H., Steinberg, P. D. and Vergés, A. (2016), ‘Does restoration of a habitat-forming seaweed restore associated faunal diversity?’, *Restoration Ecology* **24**(1), 81–90.

Maslen, B. and Lim, M. (2021), ‘Package ecopower’. R package version 0.1.0.
URL: *https://CRAN.R-project.org/package=ecopower*

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. and Abrego, N. (2017), ‘How to make more out of community data? a conceptual framework and its implementation as models and software’, *Ecology letters* **20**(5), 561–576.

Popovic, G. C., Hui, F. K. and Warton, D. I. (2018), ‘A general algorithm for covariance modeling of discrete data’, *Journal of Multivariate Analysis* **165**, 86–100.

Popovic, G. C., Hui, F. K. and Warton, D. I. (2021), ‘Fast model-based ordination with copulas’, *bioRxiv* .

Popovic, G. C., Warton, D. I., Thomson, F. J., Hui, F. K. and Moles, A. T. (2019), ‘Untangling direct species associations from indirect mediator species effects with graphical models’, *Methods in Ecology and Evolution* **10**(9), 1571–1583.

Rosenthal, R., Cooper, H. and Hedges, L. (1994), ‘Parametric measures of effect size’, *The handbook of research synthesis* **621**, 231–244.

Vergés, A., Campbell, A. H., Wood, G., Kajlich, L., Eger, A. M., Cruz, D., Langley, M., Bolton, D., Coleman, M. A., Turpin, J. et al. (2020), ‘Operation crayweed: Ecological and sociocultural aspects of restoring sydneys underwater forests’, *Ecological Management & Restoration* **21**(2), 74–85.

Wang, Y., Naumann, U., Wright, S. T. and Warton, D. I. (2012), ‘mvabund–an `r` package for model-based analysis of multivariate abundance data’, *Methods in Ecology and Evolution* **3**(3), 471–474.

Warton, D. I. (2011), ‘Regularized Sandwich Estimators for Analysis of High-Dimensional Data Using Generalized Estimating Equations’, *Biometrics* **67**(1), 116–123.

Warton, D. I., Blanchet, F. G., OHara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C. and Hui, F. K. (2015), ‘So many variables: joint modeling in community ecology’, *Trends in Ecology & Evolution* **30**(12), 766–779.

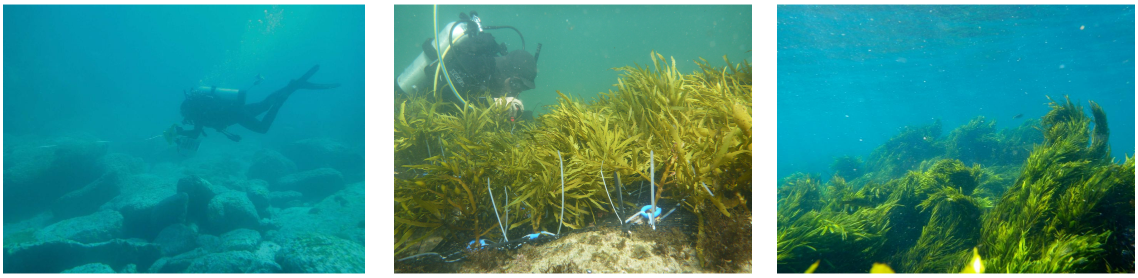
Warton, D. I., Foster, S. D., Death, G., Stoklosa, J. and Dunstan, P. K. (2015), ‘Model-based thinking for community ecology’, *Plant Ecology* **216**(5), 669–682.

Warton, D. I., Wright, S. T. and Wang, Y. (2012), ‘Distance-based multivariate analyses confound location and dispersion effects’, *Methods in Ecology and Evolution* **3**(1), 89–101.

Wood, G., Marzinelli, E., Coleman, M., Campbell, A. H., Santini, N., Kajlich, L., Verdura, J., Wodak, J., Steinberg, P. and Vergés, A. (2019), ‘Restoring subtidal marine macrophytes in the anthropocene: trajectories and future-proofing’, *Marine and Freshwater Research* **70**(7), 936–951.

Xu, J., Nelson, B. L. and Hong, J. (2010), ‘Industrial strength compass: A comprehensive algorithm and software for optimization via simulation’, *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **20**(1), 3.

Figures



(a) Control

(b) Restored

(c) Reference

Figure 1: Images of characterised environments from each treatment level. Control sites are characterised as urchin barren habitats in Sydney shallow rocky reefs where crayweed has been lost. Restored sites are similar to control sites where crayweed has recently been transplanted. Reference sites are shallow rocky reefs outside of the Sydney region with extant crayweed populations.

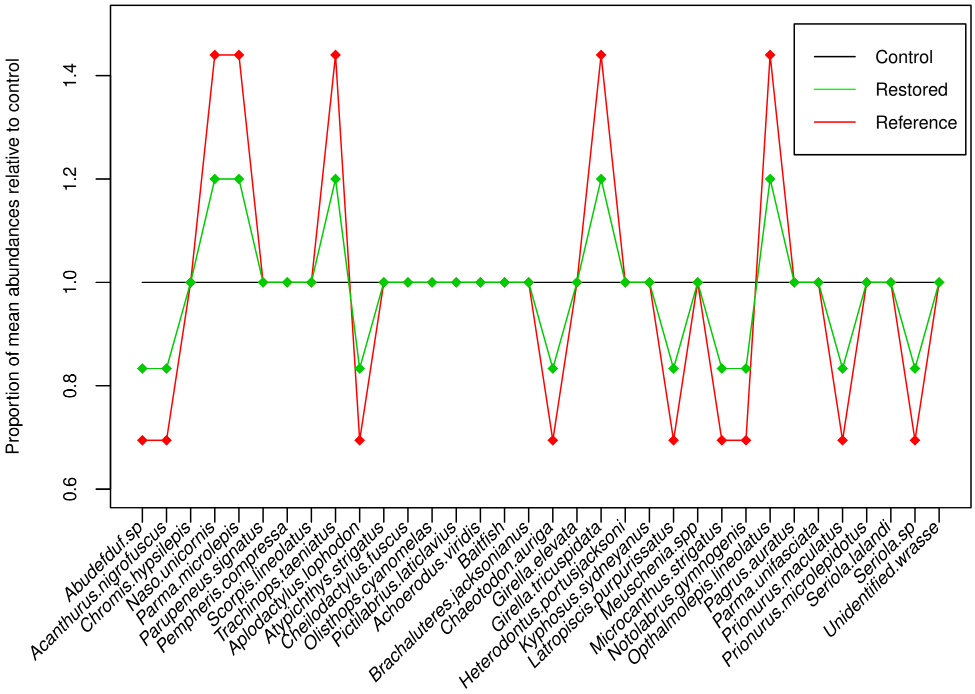


Figure 2: Plot of mean abundance proportions in restored and reference sites relative to the control sites for a specified effect size of $\rho = 1.2$.

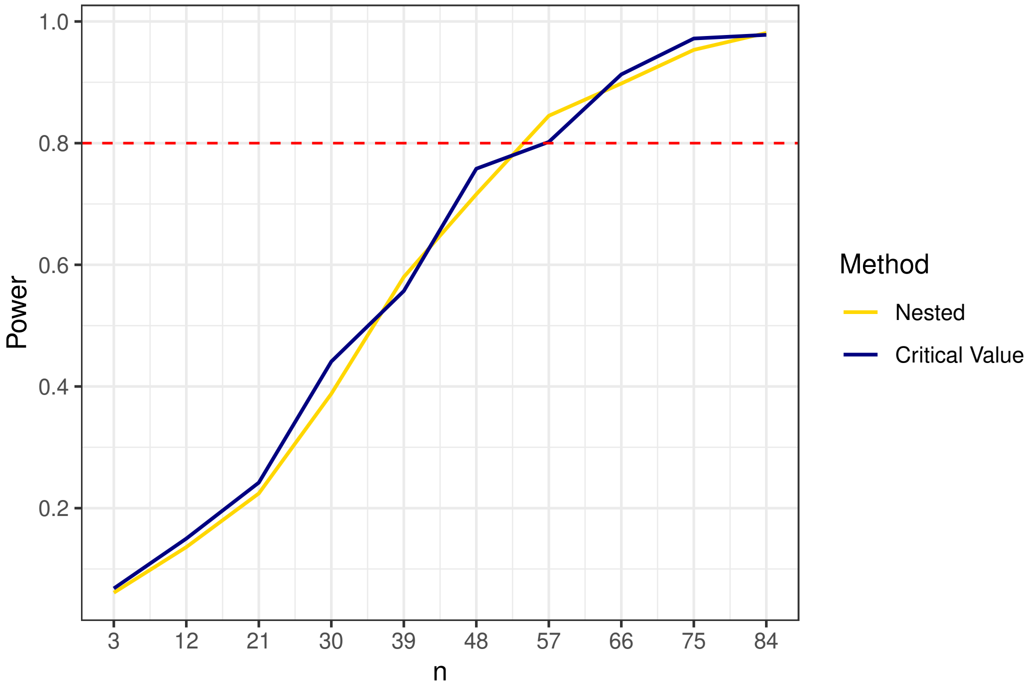


Figure 3: In blue; power curve utilising a critical value approach $\widehat{\text{Power}}_{crit}$ (taking 11 minutes to compute with $n_{resamp}+n_{power} = 2000$ simulated models). In gold; power curve utilising $n_{resamp} = 1000$ and $n_{power} = 1000$ simulated models with a nested approach to estimating power $\widehat{\text{Power}}$ (taking over 3 days to compute with $(n_{resamp}+1)n_{power} = 1,001,000$ simulated models). Power has been estimated for a fish abundance data set ($p = 34$ species) and an effect size specification of 20% changes between three treatments. 'n' is the number of samples per group, not the total sample size.

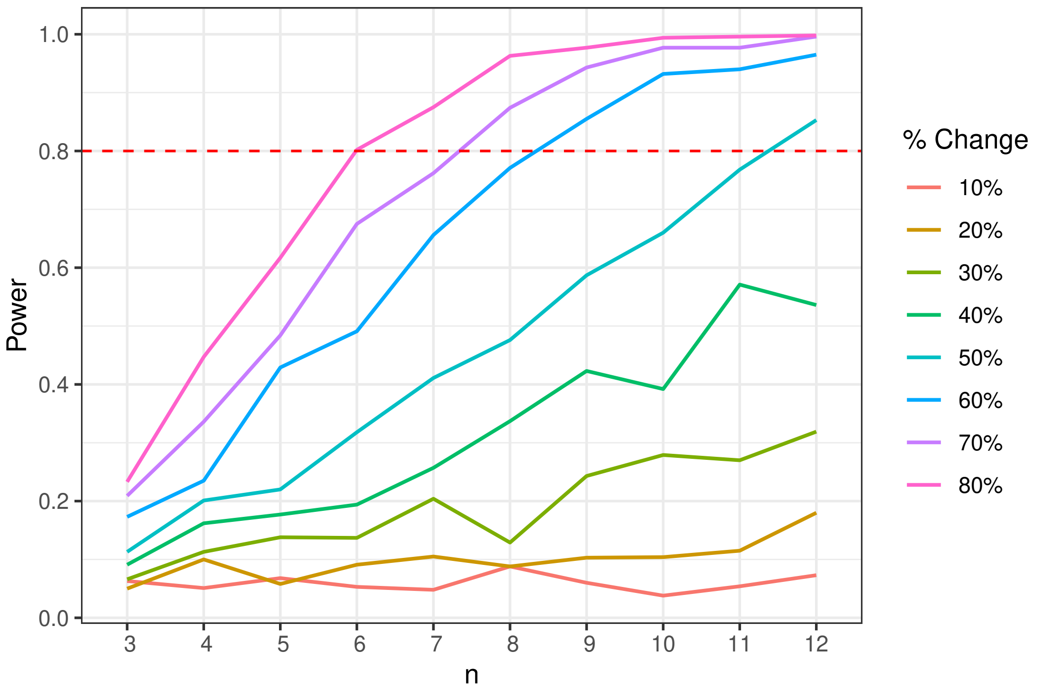


Figure 4: Power curves for a range of effect size specifications $\rho_{10\%} = 1.1, \rho_{20\%} = 1.2, \rho_{30\%} = 1.3, \rho_{40\%} = 1.4, \rho_{50\%} = 1.5, \rho_{60\%} = 1.6, \rho_{70\%} = 1.7, \rho_{80\%} = 1.8$. 'n' is the number of samples per group, not the total sample size.

Appendix

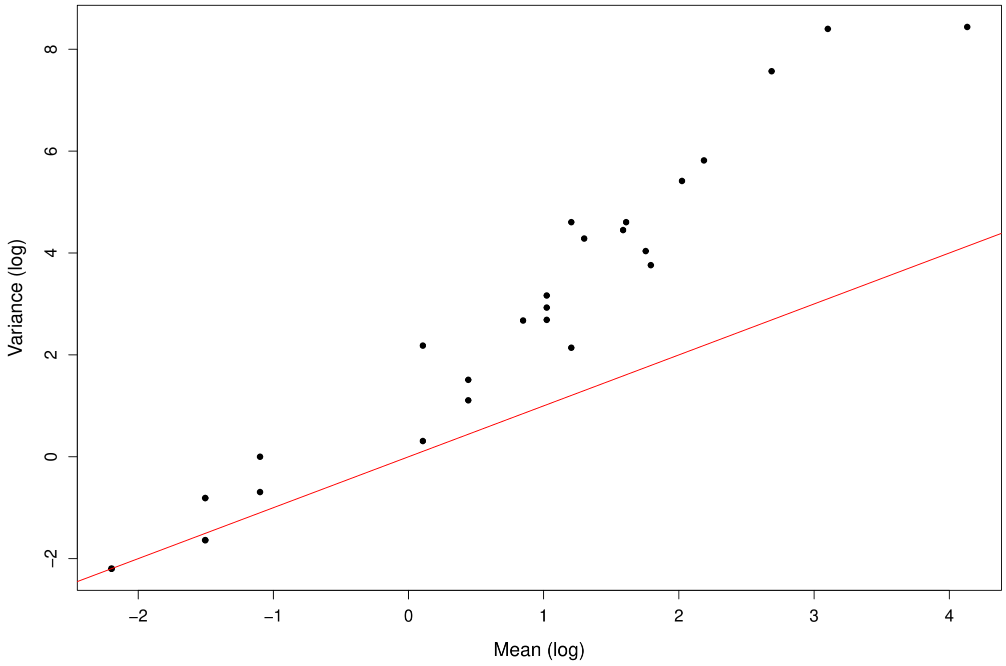


Figure 5: Mean-variance plot of fish abundances for p=34 species. The red line depicts the mean-variance assumption under a Poisson model (mean=variance). The variance appears larger than the mean for species with larger abundances, indicating over-dispersion relative to the Poisson distribution, with the negative binomial distribution being preferred for this dataset.

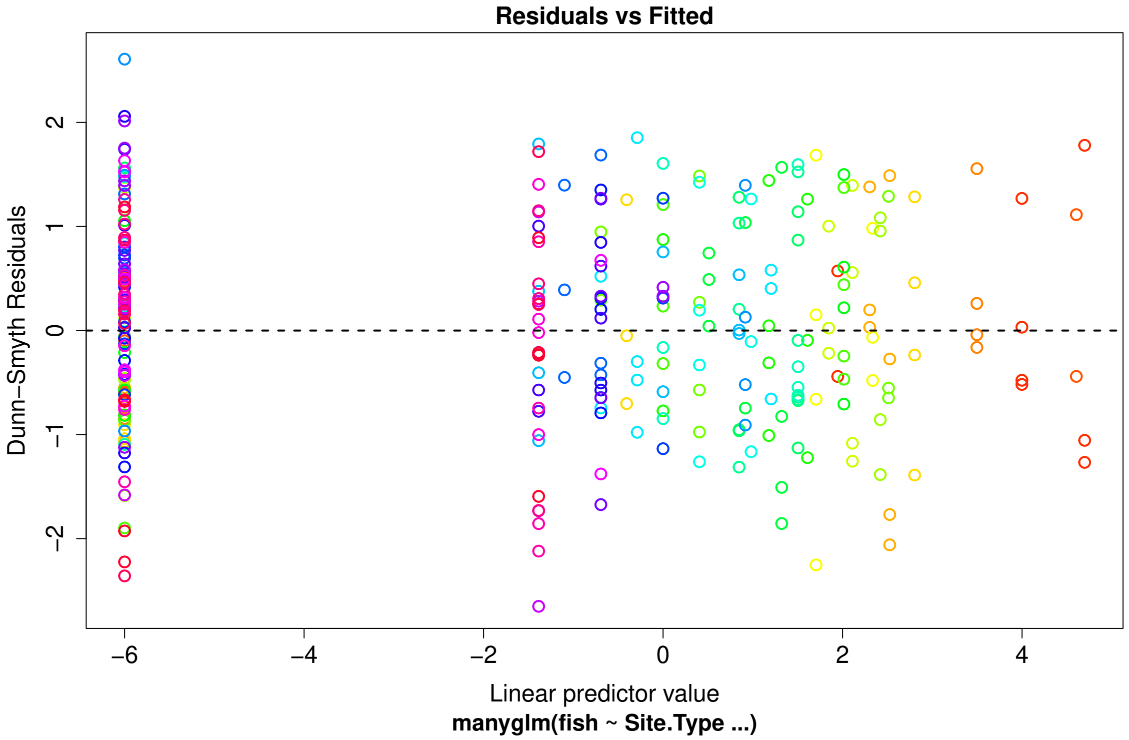


Figure 6: Diagnostic plot of a negative binomial model fitted to the fish abundance data from the Crayweed Restoration Project using `manyglm`. The large cluster of data on the left hand side of this graph refers to observations predicted to have zero abundance in a treatment. The lack of fan shape in the residual vs. fitted plot implies the negative binomial distribution has adequately accounted for the mean-variance relationship in the data shown in Figure 5.