Submit to: Methods in Ecology and Evolution

# APPLICATION

simr : an R Package for Power Analysis of Linear Mixed Models by Simulation

Peter Green1\*, Catriona J. MacLeod1

1 Landcare Research, Private Bag 1930, Dunedin 9054, New Zealand

\* greenp@landcareresearch.co.nz

# Summary

1. The R package simr provides tools that make it simple to test and run simulation experiments to determine whether a given sampling design, or range of designs, have sufficient power to detect a specific change of interest.
2. It includes tools for (a) running a very simple posthoc power analysis for a specified sampling design; (b) using power curves to assess trade-offs in sampling design; (c) testing the effects of varying the sample and effect sizes; and (d) exploring adjusting Type 1 error thresholds.
3. This paper presents a series of tutorials, using a nested sampling design example, to provide a shallow learning curve for the user, guiding them through increasingly complex analyses, but adding only a few commands or options at a time.

Key-words: effect size, monitoring, nested sampling, power curves, random effects, sampling design, Type 1 errors.

# Informing monitoring designs: why use a power analysis?

Environmental monitoring is increasingly used to assess spatial and temporal trends in variables of interest (e.g. population decline, pollution increase) as well as the effectiveness of management policies (Yoccoz *et al.* 2001; Butchart et al. XXX). Often a fundamental aim of environmental monitoring is to ensure that real change is detected and acted upon as promptly as possible. Detecting changes in ecological systems, however, is often technically and logistically challenging, especially when resources are limited. Careful planning and executing of sophisticated analyses of monitoring data are recommended for informing cost-effective and robust monitoring (Field *et al.* 2007). Such analyses can be used to identify monitoring efforts that have no realistic chance of detecting relevant changes and options for improving them (Legg & Nagy 2006; Field *et al.* 2007).

The extent and strength of interferences that can be drawn from monitoring programmes depends on their scale, design and intensity (Yoccoz *et al.* 2001). A robust monitoring design is one with sufficient statistical power to detect a specified change of interest if it actually occurs. This requires: (a) having a sufficient sample size in relation to the variability inherent in the system; (b) setting an ecologically appropriate level of power as a target; and (c) implementing a flexible design that allows for learning and improvement in the future (Field *et al.* 2007).

Although there are a range of tools available for carrying out power analyses (e.g. add refs?), there is currently no single tool available that can automate the power analysis process for a range of arbitrary models. This paper describes a package simr, now available in R (ref), which provides tools that make it simple to set up and run simulation experiments to help better inform, assess and improve monitoring designs. Key features of the package are as follows:

1. *Applicable to simple and complex survey designs*: With simr the user can perform a power analysis based on any linear model (fitted with the lm command from base R) or linear mixed model (fitted with the lmer function in the lme4 package). When the design is complicated (e.g. spatial and/or temporal nested sampling protocols are implemented), linear mixed models allows the user to model the random variation among units as well as the effect of explanatory variables (Bolker 2008). By building on existing functions within lme4, simr can handle both crossed and nested survey designs.
2. *A flexible and robust process for assessing trade-offs in survey design*: Simulation is a general procedure for determining the properties of a statistical method or design (simSummary [ref] and simFrame [ref]?). With simulation the user does not need to find an analytical approach specialised to their particular analysis (e.g. pamm ref?) or rely on an approximation that may not be robust to departures from its assumptions (e.g. longpower ref?). This can be especially important with mixed models where the sampling distributions of parameter estimates are difficult to work with (ref Bates?).
3. *Suitable and efficient for a range of analytical capabilities*: Simulation studies can be difficult or time consuming to set up and run. They would normally involve some degree of programming by the investigator (e.g. pamm ref?), which might be beyond their current ability or might simply take longer than they might like. The ready-made package makes power analyses not only accessible to a wide range of scientists, but also efficient to run. Investigators who might otherwise be limited to fitting a model in R [ref] could supplement their study with a power analysis without much additional effort. For scientists comfortable with R coding, the turnkey package could save them time and let them focus their efforts on a more advanced analysis.

To illustrate the range of functions available within simr, this paper runs through several tutorials. These are intended to provide a shallow learning curve, guiding the user through increasingly complex analyses, but adding only a few commands or options at a time, specifically: (1) running a very simple posthoc power analysis for a specified sampling design; (2) using power curves to assess trade-offs in sampling design; (3) testing the effects of varying the sample and effect sizes; and (4) exploring adjusting Type 1 error thresholds.

Although all functions within simr can be readily applied to simple survey designs meeting the requirements of a linear modelling approach (using the lm command), here a nested survey design example is applied here to illustrate their utility. The example dataset has response variable y (e.g. soil quality, bird abundance) measured at ten levels of the explanatory variable x (e.g. year) for three groups g (e.g. study sites) (Figure 1).

# Post-hoc power analysis for a specified sampling design

We start by fitting a simple mixed effects model in lme4. In this case we have a random intercept model, where each group (g) has its own intercept but the groups share a common trend.

fit <- **lmer**(y ~ x + (1 | g), data = example)

**summary**(fit)

## Linear mixed model fit by REML ['merModLmerTest']

## Formula: y ~ x + (1 | g)

## Data: example

##

## REML criterion at convergence: 97.07

##

## Random effects:

## Groups Name Variance Std.Dev.

## g (Intercept) 11.136 3.337

## Residual 0.972 0.986

## Number of obs: 30, groups: g, 3

##

## Fixed effects:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 10.6734 1.9655 5.43 0.02811 \*

## x -0.2398 0.0627 -3.83 0.00073 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Correlation of Fixed Effects:

## (Intr)

## x -0.175

The effect size for x is -0.2398 and is significant even at the 0.001 level. If this was the true effect size, and we were to repeat the experiment, what would be our power to detect this trend? We can do this kind of post-hoc power analysis very easily in simr:

**power**(fit)

## [1] "98.00% (87.12, 99.72)"

This power function assumes a number of default settings to make it simple for the user to implement, but it can also be readily modified to meet specific needs. For example, the user can derive a more precise power estimate (i.e. a smaller confidence interval for the power) by increasing the number of simulations, by modifying the nsim argument from its default setting of 1000 to 2000:

>power(fit, nsim = 2000)

## 96.85% (95.99, 97.53)

The user can also calculate the power for a different effect size if they have a specific scientifically relevant value in mind [see tutorial #x].

Add example?

# Assessing trade-offs in sampling design and power

## identifying Minimum sampling size required

Assuming that collecting data on many levels of x is costly, the user might want to collect only as much data as needed to reach a certain level of statistical power. The powerCurve function in simr can be used to explore trade-offs between sample size and power. Here, we explore the effect of varying in sample size (x) in the example dataset:

pc2 <- powerCurve(fit)

plot(pc2)

In this case, XX simulated datasets were analysed (as specified using the default setting for the nsim argument), and for each of those simulated datasets fitting the model to 8 different subsets (varying the sample size from 3 to 10 levels of x). The results can be visualised in Figure 2, which shows the power to detect the specified effect size increases with sampling size. A minimum sample size of 8 would be required to have 80% power to detect the specified effect size.

This function uses a XX step process to………. The time taken to estimate the power curve will increase as the number of simulations and the range of sample sizes considered increases.

## specifying effect size of interest

The first two analyses made heavy use of the default settings in simr to keep things as simple as possible. For example, the simulated trend defaults to the trend estimated in the fitted model. Often the user will have a specific value for an ecologically significant effect, and be interested in determing the power to detect an effect of that size. We can access the fixed effects in an lmm with fixef(fit). If we want a specific fixed effect, say the effect for x we use: fixef(fit)[‘x’]. Suppose that our ecologically significant effect size for x is -0.1. simr uses the (hopefully) obvious idiom to change the size of fixed effects:

fixef(fit)[‘x’] <- -0.1

We can now calculate a power curve for our modified fitted model the same way as above:

pc3a <- powerCurve(fit)

plot(pc3a)

Again, this will take some time. However you can see the results now in Figure 3a. This analysis shows us that we have insufficient power at this effect size for any of the sample sizes considered.

## trade-offs in sampling designs

To get a better picture of the trade-off between power and sample size we need to increase the number of levels of x, which we can do using the extend command:

fit <- extend(fit, along=’x’, n=20)

pc3b <- powerCurve(fit)

plot(pc3b)

Figure 3b shows the power curve for a larger range of levels of x (3—20).

Alternatively, the user may choose to increase the number of groups (g) that were sampled along:

fit1 <- extend(fit, along = "g", n = 8)

pc3c <- powerCurve(fit1)

plot(pc3c)

Figure 3c shows that doubling the number of groups sampled (from 3 to 6) will increase the power of the study design from <20% to 80%.

## altering p-value thresholds

Include Type I error calculations – effects of changing the p-value thresholds in Figure 4 – in relation to varying levels of x and g. Higher p-value thresholds (pval = 0.01) reduces the power to detect change. For example, increasing the pval threshold from 0.05 to 0.1, would require an increased sampling effort (x = 14 vs. 16 levels; g = 6 to 7 groups).

pc4 <- pc3b

plot(pc4, pval = c(0.01, 0.05, 0.1))

plot(pc5, pval = c(0.01, 0.05, 0.1))

# Further Work

As currently implemented, simr uses data from a pilot study to inform the power analysis, with the structure of the pilot data providing defaults for the simulation settings. Future versions will also include the ability to create data sets from scratch; this will not require a pilot study, but will require some domain expertise to select sensible parameters.

Version 1.0 of simr is designed for linear mixed models using lmer in lme4. The next version will add support for generalised linear models (glm in base) and generalised linear mixed models (glmm in lme4). At some point tools will be added to make it simple to create interfaces to arbitrary R packages.

## Notes: Which Dataset to Use?

Simulated dataset: conceptually simple, allows the package to be the focus of the paper.

Real dataset: allows practical scientists to get a handle on what’s actually happening. But we have to deal with introducing the data and with any peculiarities; this all takes away from the main focus of the paper.

Simulated “Real” Dataset: Combines advantages of both approaches. We generate a clean simulated dataset; this means we spend no time explaining data cleaning assumptions. But the variables are given meaningful names so that people can develop an intuitive grasp of what’s going on.

### Acknowledgements

### References

Bolker 2008

Butchart et al.

Field et al. 2007

Legg & Nagy 2006;

Yoccoz et al.

longpower ref

pamm ref

simSummary [ref]

simFrame [ref]

## 

Figure 1: A scatterplot of the example data with fitted lines

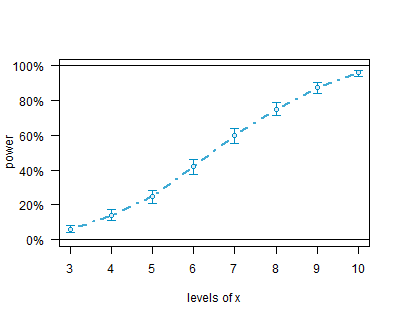
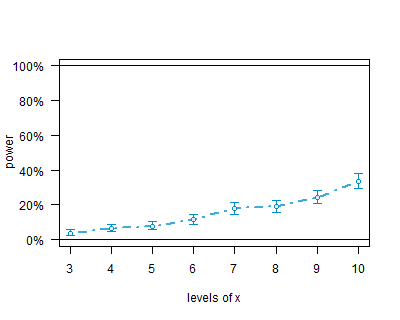
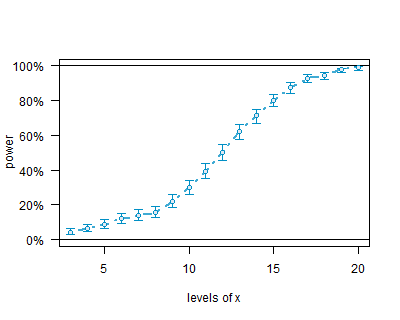


Figure 2: Power (mean ± 95%CI) to detect a specified trend in relation to sample size, in this case increasing the number of levels of x, calculated using the powerCurve function.

(a)

(b)



(c)

Add figure showing effect of varying g

Figure 3: Power (mean ± 95%CI) to detect a modified trend (x = -0.1) in relation to sample size, calculated using the powerCurve function, increasing the number of: (a) levels of x from 3 to 10; (b) levels of x from 3 to 20; and (c) groups (g) from 3 to 10.



(b)



Figure 4: Changes in power for multiple p-value thresholds in relation to varying the number of levels of (a) x; and (b) g.