ENVIRONMENTAL GENOMICS LABORATORY

STANDARD OPERATING PROTOCOL · EnGeL-SAMPL-POWER

Planning sample size: simulation-based power analysis for soil fauna and other ecological designs

Protocol ID	EnGeL-SAMPL-POWER
Version	1.1.0
Date	2026-11-20
Status	draft
Area	Sampling and experimental design (SAMPL)
Authors	Luís Cunha (lead; analysis scripts; review and approval; ORCID 0000-0002-5870-2537)
Reviewed and approved by	Luís Cunha
Laboratory	Environmental Genomics Laboratory (EnGeL-UC), Environmental Ecology & Ecotoxicology Group (E3G), Centre for Functional Ecology, Department of Life Sciences, University of Coimbra · Laboratório Associado TERRA
Contact	luis.cunha@uc.pt · ewormuc@gmail.com
Licence	Text: CC BY 4.0 · Code: MIT
Main hazards	None: computational protocol

1 Purpose and scope

This guide helps you decide how many plots, samples or individuals a study needs **before** fieldwork, by simulating the planned design and the planned analysis many times and counting how often the analysis detects an effect of a given size (statistical power). It covers four situations common in the lab: differences in community composition (PERMANOVA), differences in a single diversity index, the choice between sampling methods with different numbers of sub-samples per plot, and repeated measures on the same units (for example tree-ring series or repeated soil cores). After sampling, check whether the effort was enough with EnGeL-SAMPL-EFFORT.

2 Principle

Power is the probability that a test detects an effect that is really there; it depends on the effect size, the variability, the number of independent units and the significance level. For small, structured ecological designs the usual formulas do not apply well, so power is estimated by simulation (Johnson et al., 2015): (1) build the design (factors, levels, replicates); (2) generate data in which the effect has a chosen size and the noise a chosen variance; (3) run the planned analysis; (4) repeat many times and record the proportion of significant results. Repeating this over a range of effect sizes gives the **minimum detectable effect (MDE)**, the smallest effect detected with the target power; repeating it over a range of replicate numbers shows how many units the target power requires. Composition is tested with PERMANOVA on a dissimilarity matrix (Anderson, 2001; vegan), a diversity index with a lin-

ear model, and repeated measures with a linear mixed model whose power is simulated with simr (Green & MacLeod, 2016) on lme4 (Bates et al., 2015).

3 Before you start

Item	Details
Hands-on time	[TO CONFIRM]
Computing time	Minutes for a quick check (100–200 simulations); up to hours for final figures (1000 or more simulations, several effect sizes)
Input	The planned design, the planned analysis, and values for effect sizes and variances
Prepare in advance	R packages installed; the question and the test written down; pilot data or published values for the expected effect and variability
Key safety points	Not applicable

CRITICAL The result is only as good as the effect sizes and variances you put in. Take them from pilot data (fit the planned model to the pilot data and use its estimates) or from comparable published studies, and record the source. If no source exists, say so: report power over a range of effect sizes rather than for one invented value.

4 Materials

4.1 Computer

- Computer with R and RStudio (lab workstation or personal computer)

4.2 Consumables

- None

4.3 Software and input data

- R [TO CONFIRM: version] with the packages vegan, lme4, simr, dplyr and ggplot2 [TO CONFIRM: package versions]
- Study design: factors, levels, replicates per cell, sub-samples per plot and sampling methods; effect sizes and variances from pilot data or published studies

5 Reagent preparation

Write the design down first. Name the factors and their levels, the number of independent replicates per combination of levels (cell), the sampling methods, the number of sub-samples per plot for each method, and the test and term of interest (for example the effect of stand age, adjusted for land use). The unit of analysis is the unit that is replicated independently (usually the plot), not the sub-samples taken inside it.

6 Procedure

The examples use a neutral design: **two land uses × two stand ages (young, mature) × 3 replicate plots = 12 plots**. In each plot, soil fauna are sampled with 3 soil monoliths, 5 pitfall traps and 1 Winkler litter sample; sub-samples of the same method are pooled into one value per plot. Replace the numbers with those of your study.

6.1 1. Set up

1. Load the packages and fix the random seed so the results can be reproduced:

```r
library(vegan)    # PERMANOVA (adonis2), dissimilarities
library(lme4)     # mixed models
library(simr)     # power for mixed models
library(dplyr)
library(ggplot2)
set.seed(42)
```

2. Build the design table:

```r
make_design <- function(n_rep) {
  d <- expand.grid(
    land_use  = factor(c("LU1", "LU2")),
    age_class = factor(c("young", "mature"), levels = c("young", "mature")),
    replicate = seq_len(n_rep)
  )
  d$plot_id <- factor(paste(d$land_use, d$age_class, d$replicate, sep = "_"))
  d
}
design <- make_design(n_rep = 3)   # 12 plots
```

6.2 2. Example 1: community composition (PERMANOVA)

3. Simulate a community matrix (plots × taxa) in which part of the taxa change in abundance with stand age. Each taxon has a baseline log-abundance; its response to age has a **fixed sign** (here 60% of taxa increase and 40% decrease) and a size scaled by `effect_size`; counts are drawn from a Poisson distribution with plot-level noise `noise_sd`:

```r
simulate_community <- function(design, n_taxa = 25, effect_size = 0.3,
                               noise_sd = 0.4, baseline_mean = 2) {
  base  <- rnorm(n_taxa, baseline_mean, 1)
  sign  <- sample(c(1, -1), n_taxa, replace = TRUE, prob = c(0.6, 0.4))
  resp  <- sign * effect_size * rgamma(n_taxa, shape = 2, scale = 0.5)
  old   <- as.numeric(design$age_class == "mature")
  comm  <- sapply(seq_len(n_taxa), function(j)
    rpois(nrow(design), exp(base[j] + old * resp[j] + rnorm(nrow(design), 0, noise_sd))))
  colnames(comm) <- paste0("taxon_", seq_len(n_taxa))
  comm
}
```

CRITICAL Give each taxon's response a consistent sign. If the sign is drawn at random for each taxon, increases and decreases cancel out at the community level, PERMANOVA (which tests a shift of the community centroid) detects nothing at any effect size, and the power curve is wrongly flat at zero. A flat curve even at large effects is a sign of this error.

4. Estimate power: simulate, compute Bray–Curtis dissimilarities, run PERMANOVA with the design factors and count significant results for the term of interest:

```r
permanova_power <- function(design, effect_size, noise_sd = 0.4, nsim = 200,
                            n_taxa = 25, method = "bray", alpha = 0.05) {
  hits <- 0
  for (i in seq_len(nsim)) {
    comm <- simulate_community(design, n_taxa, effect_size, noise_sd)
    if (any(rowSums(comm) == 0)) next          # empty plots break the dissimilarity
    res <- tryCatch(adonis2(vegdist(comm, method) ~ land_use + age_class,
                            data = design, permutations = 199, by = "terms"),
                    error = function(e) NULL)
    if (!is.null(res) && res["age_class", "Pr(>F)"] <= alpha) hits <- hits + 1
  }
  hits / nsim
}
permanova_power(design, effect_size = 0.3)
```

5. Draw the power curve over a range of effect sizes and read off the MDE where the curve crosses the target power:

```
effects <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0)
curve1  <- data.frame(effect = effects,
                      power  = sapply(effects, function(e) permanova_power(design, e, nsim = 150)))
ggplot(curve1, aes(effect, power)) + geom_line() + geom_point() +
  geom_hline(yintercept = 0.8, linetype = "dashed") + ylim(0, 1) +
  labs(x = "Effect size (taxon log-abundance shift)", y = "Power")
```

6.3 3. Example 2: a diversity index

6. For one value per plot (for example Shannon diversity or functional richness, standardised to mean 0 and SD 1 so that effects are in SD units), simulate the 2 × 2 design and test the age effect with a linear model:

```
diversity_power <- function(design, age_eff, land_use_eff = 0, interaction_eff = 0,
                            residual_sd = 1, nsim = 500, alpha = 0.05) {
  old <- as.numeric(design$age_class == "mature")
  lu2 <- as.numeric(design$land_use == "LU2")
  mu  <- land_use_eff * lu2 + age_eff * old + interaction_eff * lu2 * old
  p   <- replicate(nsim, {
    design$y <- mu + rnorm(nrow(design), 0, residual_sd)
    summary(lm(y ~ land_use + age_class, data = design))$coefficients["age_classmature", "Pr(>|t|)"]
  })
  mean(p <= alpha)
}
diversity_power(design, age_eff = 0.5)
```

CRITICAL Test the term that answers your question. With `y ~ land_use * age_class` and the default treatment contrasts, the coefficient `age_classmature` is the age effect **within the first land use only** (half the plots), not the overall age effect, and power is underestimated. For the overall (main) effect use the additive model above; to test whether the age effect differs between land uses, fit `y ~ land_use * age_class`, give `interaction_eff` a value and test the interaction term (`land_useLU2:age_classmature`).

A plain linear model is enough here: each plot gives one value, so there are no repeated measures and no random effects (simr is not needed and has no function for fixed-effects-only models).

6.4 4. Example 3: sampling methods, sub-samples and replicates

7. Pooling more sub-samples per plot makes the plot value less noisy. As a first approximation, the plot-level noise falls with the square root of the number of pooled sub-samples. Scale the noise of each method relative to a reference method:

```
subsamples <- c(monolith = 3, pitfall = 5, winkler = 1)
noise_for  <- function(base_sd, method, ref = "pitfall")
  base_sd * sqrt(subsamples[[ref]] / subsamples[[method]])
by_method <- data.frame(method = names(subsamples))
by_method$power_permanova <- sapply(by_method$method, function(m)
  permanova_power(design, 0.3, noise_sd = noise_for(0.4, m)))
by_method$power_diversity <- sapply(by_method$method, function(m)
  diversity_power(design, 0.5, residual_sd = noise_for(1, m)))
by_method
```

CRITICAL The square-root rule assumes that sub-samples within a plot are as independent as plots are. Traps or cores close together often share microhabitat, so the real gain from pooling is smaller and the estimated power for methods with many sub-samples is optimistic. The number of plots, not the number of sub-samples, remains the sample size of the test.

8. See how power grows with the number of replicate plots per cell, at a fixed effect size:

```
reps   <- c(3, 5, 8, 10)
curve3 <- data.frame(replicates = reps,
                     permanova  = sapply(reps, function(r) permanova_power(make_design(r), 0.3, nsim =
150)),
```

```
                    diversity  = sapply(reps, function(r) diversity_power(make_design(r), 0.5)))
curve3
```

6.5 5. Example 4: repeated measures on the same units (tree-ring series, simr)

This example uses a dendroecology design: trees of two species in two forest types and two age classes (15 trees per combination, 120 trees), each with an annual series of ring-width index (RWI) values matched to a yearly climate variable (for example a drought index). The question is whether the climate sensitivity of growth differs between young and mature trees (the age × climate interaction). The same approach applies to any design with repeated measurements on the same units.

9. Build the unit-by-time data and a climate series:

```
n_years <- 25
trees <- expand.grid(forest_type = factor(c("FT1", "FT2")),
                     age_class   = factor(c("young", "mature"), levels = c("young", "mature")),
                     species     = factor(c("sp1", "sp2")),
                     tree_rep    = 1:15)
trees$tree_id <- factor(with(trees, paste(forest_type, age_class, species, tree_rep, sep = "_")))
df <- trees[rep(seq_len(nrow(trees)), each = n_years), ]
df$year <- rep(seq_len(n_years), times = nrow(trees))
df <- left_join(df, data.frame(year = seq_len(n_years), climate = rnorm(n_years)), by = "year")
```

10. Create the model with chosen fixed effects, random-effect variances (a random intercept and a random climate slope per tree) and residual SD, then estimate power for the interaction:

```
fixed_eff <- c("(Intercept)" = 0, "age_classmature" = 0,
               "climate" = 0.40, "age_classmature:climate" = -0.20)   # illustrative values
vc  <- matrix(c(0.30, 0, 0, 0.05), 2)   # tree intercept and slope variances: from pilot data
mod <- makeLmer(RWI ~ age_class * climate + (1 + climate | tree_id),
                fixef = fixed_eff, VarCorr = vc, sigma = 0.6, data = df)
powerSim(mod, test = fixed("age_classmature:climate", "z"), nsim = 100)
```

11. Repeat over a range of interaction sizes for the MDE, and over the number of trees per combination with `extend()` and `powerCurve()`:

```
mod_big <- extend(mod, along = "tree_rep", n = 30)
plot(powerCurve(mod_big, test = fixed("age_classmature:climate", "z"),
                along = "tree_rep", breaks = c(5, 10, 15, 20, 25, 30), nsim = 50))
```

The length of the series (`n_years`) also matters: young trees have shorter series; rebuild `df` with the realistic number of years for each age class and compare.

> **PAUSE POINT** Each example can be run on its own. Use low `nsim` (50–200) while setting up, then rerun with `nsim` = 1000 or more for the values you report.

7 Calculations

Power = number of simulations with $p \le \alpha$ ÷ number of simulations. Its precision depends on the number of simulations: the standard error is $\sqrt{(P(1-P)/n_{sim})}$; for $P = 0.8$ it is about 0.04 with 100 simulations and about 0.013 with 1000.

Example: 163 significant results in 200 simulations give power = 163 ÷ 200 = 0.82 (standard error $\sqrt{(0.82 \times 0.18 / 200)} \approx 0.027$).

Symbol	Meaning	Unit
P	Power	proportion (0–1)
α	Significance level	proportion

Symbol	Meaning	Unit
n_{sim}	Number of simulations	count

8 Quality control

Check	Acceptance criterion
Target power and significance level	Power ≥ 0.80 at α = 0.05 (lab target)
Number of simulations for reported values	1000 or more (lower values only for setting up)
Sanity check of the simulation	Power ≈ α when the effect size is 0 (in a test run: 0.03–0.06 at α = 0.05), and close to 1 at a large effect; a curve that stays flat at large effects indicates an error in the simulation
Source of effect sizes and variances	Recorded for every value (pilot data, publication, or "illustrative")

9 Data and metadata

- **Outputs:** power for the planned design, power curves (effect size and number of replicates), the MDE, and the script.
- **Where:** project folder in the lab's shared folder; the script in the project's `scripts` folder.

Minimum to record

- Design: factors, levels, replicates per cell, sub-samples per plot and per method
- Test and term of interest; significance level
- Effect sizes, variances and residual SD, each with its source
- Number of simulations and random seed
- R and package versions (`sessionInfo()`)

10 Safety and waste

Not applicable (computational protocol).

11 Troubleshooting

Problem	Likely cause	Solution
PERMANOVA power stays near 0 even at large effects	Taxon responses with random signs cancel out	Give responses a consistent sign (step 2, CRITICAL).
Errors in `vegdist()` or `adonis2()`	Plots with no individuals (all-zero rows)	Skip those simulations (as in the function) or raise the baseline abundance.
Power of a main effect far lower than expected	Main effect read from a model with an interaction (treatment contrasts)	Use the additive model, or test the interaction term itself (Example 2, CRITICAL).
`could not find function "makeLm"`	simr has no fixed-effects-only model	Use the manual simulation of Example 2.
`boundary (singular) fit` messages in simr	Random-effect variance near zero for the simulated data	Check the variances; the warnings are common with small variances and do not stop the simulation.
Power changes between runs	Too few simulations	Increase `nsim`; fix the seed.

Problem	Likely cause	Solution
Non-monotonic power curve	Simulation noise at low `nsim`	Rerun with more simulations before interpreting.

12 Remarks

Report power honestly: if the planned design reaches the target power only for large effects, state it as a limitation and consider more replicate units; a non-significant result from an underpowered design carries little evidence. Effect sizes for community composition have no standard scale: the `effect_size` in Example 1 describes how strongly taxa respond in the simulation model, so calibrate it on pilot abundance data when possible. Where several sampling methods are used, the method with the fewest pooled sub-samples sets the weakest link. The source notebooks were written for a thesis design and are kept with the lab's source documents; this guide uses a neutral example.

13 References

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Green, P., 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. https://doi.org/10.1111/2041-210X.12504

Johnson, P.C.D., Barry, S.J.E., Ferguson, H.M., Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. Methods in Ecology and Evolution, 6(2), 133–142. https://doi.org/10.1111/2041-210X.12306

Oksanen, J., et al. vegan: Community Ecology Package. R package [TO CONFIRM: version].

14 Version history and citation

Version	Date	Change	By
1.0.0	2026-11-20	First EnGeL-UC version, from the lab's power-analysis notebooks (community composition, diversity index, repeated measures), generalised with a neutral example.	Luís Cunha
1.1.0	2026-11-20	Lab target confirmed by Luís Cunha: power 0.80 at α = 0.05; additive model for the overall effect in the diversity-index example confirmed.	Luís Cunha

How to cite

Cunha L (2026). Planning sample size: simulation-based power analysis for soil fauna and other ecological designs. EnGeL-UC protocol EnGeL-SAMPL-POWER, version 1.1.0. Zenodo. DOI [TO CONFIRM].

Licence

Text: CC BY 4.0. Code: MIT. Logos are the property of their institutions and are not covered by the licence.

15 Open questions

1. R and package versions (vegan, lme4, simr); hands-on time.
2. Typical effect sizes and variances for the lab's main study types (soil fauna composition, diversity indices, tree-ring series), from pilot data or literature, to quote as starting points.
3. DOI.