Online SSP - Estimation of sampling effort in community ecology

Edlin Guerra-Castro1, Juan Carlos Cajas1, Juan José Cruz-Motta2, Nuno Simoes1 and Maite Mascaro1.

1Universidad Nacional Autónoma de México, 2Universidad de Puerto Rico-Mayagüez

Rationale of **SSP**

**SSP** is an **R** package designed to estimate sample effort in studies of ecological communities using intensive sampling over several sets of simulated data. Our procedure is based on the previous definition of MultSE but eludes the double resampling over a unique pilot data set (Anderson and Santana-Garcon 2015). The protocol in **SSP** consists of simulating several extensive data matrices that mimic some of the relevant ecological features of the community of interest using a pilot data set. For each simulated data, several sampling efforts are repeatedly executed and *MultSE* is calculated to each one. The mean value, and the 0.025 and 0.975 quantiles of *MultSE* for each sampling effort across all simulated data; are then estimated and plotted. The mean values are standardized in relation to the lowest sampling effort (consequently, the worst precision), and an optimal sampling effort can be identified as that in which the increase in sample size do not improve the precision beyond a threshold value (*e.g.* 3 %).

**SSP** includes seven functions: assempar for extrapolation of assemblage parameters using pilot data; simdata for simulation of several data sets based on extrapolated parameters; datquality for evaluation of plausibility of simulated data; sampsd for repeated estimations of *MultSE* for different sampling designs in simulated data sets; summary\_sd for summarizing the behavior of *MultSE* for each sampling design across all simulated data sets, ioptimum for identification of the optimal sampling effort, and plot\_ssp to plot sampling effort vs *MultSE*.

**SSP** is free and open source, distributed under GNU Public License ver. 2 (GPL-2). This package is available on the comprehensive R archive network (CRAN) <https://cran.r-project.org/web/packages/SSP/index.html> and is also hosted and developed in GitHub <https://github.com/edlinguerra/SSP>.

The online version of **SSP**

The sequence of using **SSP** online is the same as in R.

1. First, you must start with 1. assempar. For this, upload the pilot data as a .csv file (species names in columns and samples as rows, the first column should indicate the site to which the sample belongs, regardless of whether a single site has been sampled). Identify the type of data (presence / absence, counts, coverage) as well as the extrapolation method of species richness. Then press GO assempar. Wait for the progress bar to finish to move to 2. simdata.
2. Once in 2. simdata, define the number of data sets to be simulated (cases), the number of sites (this applies only if you have more than 1 site in the pilot data, keep at 1 if you do not have multiple sites), and the number of samples to be simulated in each site. Press GO simdata and wait for the progress bar to finish.
3. In 3. sampsd, define the maximum sampling effort (m for sites, n for samples) as well as the number of repetitions (k) for each combination. Remember, if you are simulating a single site keep m at 1. You should also identify the appropriate pretreatment for abundances and choose the similarity measure. Press GO sampsd and wait. Don’t skip to the next panel until the progress bar ends, this process can take a while, especially if you requested a lot of datasets, sites, samples, and repetitions. We suggest using few simulations (cases) and few repetitions to obtain exploratory results. Then you can repeat the process by increasing these arguments.
4. In the fourth panel 4. summary\_ssp & datquality you must identify whether you simulated one site (by default) or several. At this point you can assess the quality of the simulated data with GO datquality. Then, press Go summary to calculate the *MultSE* behavior across all simulated data sets for each sampling effort. The results will be displayed in the *Summary MultSE* tab. Both outputs can be downloaded as .csv files.
5. Once the summary is displayed, you can go to 5. ioptimum & plot. You can choose the optimization breakpoints regarding the lowest sampling (by default 10%, 5% and 3%) as well as the format of the file to download (png or pdf). Press Go ioptimum, the plot will be displayed in the *MultSE plot* tab.

This online version was built with [Shiny](https://shiny.rstudio.com/), and is hosted at [Shinyapps.io](https://https/www.shinyapps.io). This application has a limit of 25 active hours per month on this server, so depending on the use of the app during this month, it may not be able to run until the beginning of next month.