SEOSAW data contribution guidelines

John L. Godlee

2020-07-10

Introduction

SEOSAW is building a database of plot data from savannas and woodlands across southern Africa. This guide provides an introduction to preparing data for submission to the SEOSAW database.

The SEOSAW dataset is built around two key tables: plot data and stem data. Each row in the plot table represents a single plot. Each row in the stem table represents a single stem. Note that a single tree may contain many stems. It is recommended that you format your own data in two tables, one for plots and the other for stems. The guidance below details the contents of these two tables, and other metadata.

We have provided blank template data sheets which you can take into the field when setting up a new plot. Additionally, we have provided template data entry files (`plot_data_template.csv`, `stem_data_template.csv`) and a table which will help translate values in the template data sheets to columns in the template data entry files.

Submission guidelines

Please submit tabular data as `.csv` (comma-separated values) files if possible. If working in Microsoft Excel or a similar spreadsheet program, there is usually an option to export as `.csv`. `.csv` files are preferred as they are simple and store data exactly as entered. `.xls` or `.xlsx` files are also acceptable.

Spatial data such as plot outlines are preferred as shapefiles (`.shp`), though plot corner or plot centre locations as latitude-longitude coordinates, stored in a `.csv` file are also acceptable.

If you are not using the SEOSAW template data entry files, please provide a description of the contents of each file in a separate document. For tabular data, try to provide a description of the contents of each column. For example, to describe the contents of a column in the stem table containing stem diameter measurements:

> "`stem_diameter` - Stem diameters, measured in centimetres to one decimal place (i.e. to the nearest mm)."

Provide a description of the methodology used to collect the data, especially any plot nestedness or stratified sampling, as this will determine whether any corrections should be made when calculating plot-level statistics like aboveground biomass.

Vital data

There are 142 columns which can be filled by data contributors in the stem and plot tables, and you should try to provide data for as many as possible. We understand however, that some data simply don't exist depending on the original goal of your data collection, Our stance is that metadata should be left blank if it is not known, rather than guessed at.

When compiling your own data, please use the template data tables where possible. These template tables contain fictitious data to give you an idea of how to format your own data. These links provide

tables of column descriptions for both the plot table and stem table, online, for more details on column contents. Note that the column description tables online are searchable, so just type in the name of a column you would like guidance for and the table will filter to display that column.

Below is a list of the columns which are absolutely necessary for minimum data quality. Data submitted without all of these columns cannot be added to the SEOSAW dataset. The column names as they appear in the template data tables are shown in brackets:

Plot table:

- Plot name, for matching with plot names in stem table (`plot_name`)
- Plot longitude and latitude coordinates (`longitude_of_centre`, `latitude_of_centre`)
- Plot shape and dimensions (`plot_shape`, `plot_length`, `plot_width`, `plot_diameter`)
- The Primary Investigator of the data (`prinv`)

Stem table:

- Plot name, for matching with plot names in stem table (`plot_name`)
- Tree species (`species_orig_binom` or `species_orig_local`)
- Tree diameter (`diam`)
- Height of diameter measurement (`pom`)
- Fractional Probability of Inclusion. If plots are nested with varying sampling effort within plots for stems of different size, the probability of a stem this size occurring in the whole plot (`fpc`)
- Is the stem alive or dead? (`alive`)

If only a few entries from each of these columns is missing due to mistakes during fieldwork, don't worry, you data is probably still usable, but try to fill as many empty values as possible.

Stems or trees

The ideal data for the SEOSAW dataset is organised so that there is one measurement for each stem, and stems are grouped into trees, with a permanent tag to identify each stem. We understand however, that your data may not always match this ideal, and will be dependent on your original data collection protocol. There are a few columns in the table to help describe how your data is organised:

- `tags` - TRUE or FALSE, do stems have permanent tags which allow for remeasurement?
- `tree_stem` - Are measurements in the stem table recorded at the tree level or stem level? Possible values: "tree", "stem".
- `tree_diff` - TRUE or FALSE, if `tree_stem` == 'stem', were stem measurements grouped into separate trees?

Species names

Try to be precise as possible with species names. If you only know the genus of a species, record it as 'Genus indet' (e.g. 'Brachystegia indet'). Similarly, if you only know the family, record it as 'Family indet' (e.g. 'Fabaceae indet'). As a last resort, if you only know the local name of the species, this is also acceptable and should be recorded in the `species_orig_local` column.

Nesting and subplots

It is common for researchers to stratify their sampling effort across a plot, often according to stem diameter classes. This stratification is often referred to as 'nesting'. See the diagram below for an example of a nested plot. In this example, there are three levels of nesting. Across the whole plot (0.2 ha) stems >10 cm DBH are measured. Within the first level of nesting (N1, 0.1 ha), stems >5 cm DBH are measured, within the second level (N2, 0.05 ha) stems >2 cm DBH, and within the third level (N3, 0.025 ha) stems >1 cm DBH.

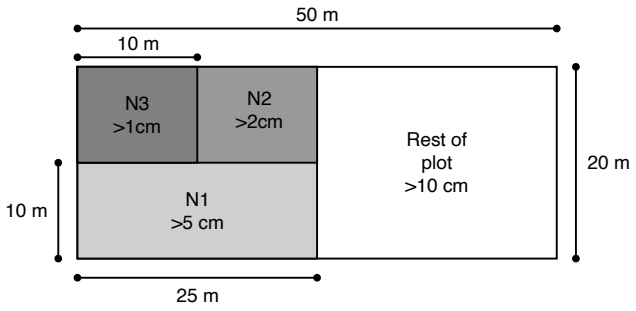

Figure 1: Schematic diagram of a nested plot with three levels of nesting

If plots are nested, with variation in sampling effort for trees of different sizes, you must include the Fractional Probability of Inclusion in the `fpc` column of the stem table. In the example, stems >10 cm DBH are measured across the whole plot, so their `fpc` is 1. Stems >5 cm DBH are measured in an area 20x25 m (0.1 ha), which is half the total plot area, so their `fpc` is 0.5. Stems >2cm are measured over 0.05 ha, which is a quarter of the total plot area, so their `fpc` is 0.25, and so on. At present, the SEOSAW database only allows up to three levels of nesting, but contact us if you have data with more levels.

It is important to distinguish between nesting and subplots. Nested plots, as demonstrated above, are smaller parcels within a plot which denote variation in sampling effort. In contrast, subplots are merely administrative sub-divisions of a plot which can help with relocating stems for future censuses, for example. Sampling effort should be the same across all subplots. Below is an example of a plot with four 20x25 m subplots.

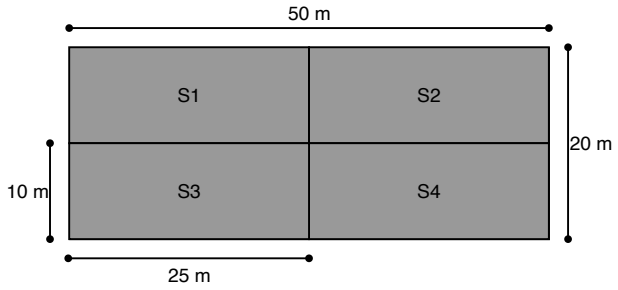

Figure 2: Schematic diagram of a plot with subplots, each occupying 25% of the total plot space.

Permanent or one-off plot

A permanent plot in the SEOSAW dataset is defined as one where trees and new recruits can be measured at a later date. The boundaries of a permanent plot must be clearly identifiable, i.e. with corner posts for a rectangular plot or a central marker for a circular plot. Individual trees must also be identifiable in a permanent plot. Normally this is achieved through the use of numbered tags, but can alternatively be achieved with a highly accurate grid coordinate system, or a differential GPS.

One-off plots are plots where only a single census has been made, with no way of exactly identifying the plot boundary or tree identity in the future.

Both permanent and one-off plots are accepted by SEOSAW. Make sure to fill in the `permanent` column in the plot table with `TRUE` or `FALSE`, accordingly.

Extra data

Please provide any metadata you think might be even slightly useful to gain a more detailed understanding of your plots. For example, if you have meteorological data for a nearby weather station, herbivore dung counts, photos of the plots, land use history timelines, soil core profiles, or fire management notes. As with the plot and stem data, try to provide a detailed description of the data, the methodology used to collect the data, and any caveats.