SEOSAW dataset manual
v3.1

John L. Godlee

12th January 2024

Contents

1 Introduction

This manual provides information on the structure of the SEOSAW dataset, and how to contribute data to the dataset.

For further guidance on collecting SEOSAW data in the field, you should consult the other SEOSAW field manuals.

2 Description of the SEOSAW dataset

This section gives a brief overview of the contents of the latest iteration of the SEOSAW dataset, which is available for research purposes by approval from the SEOSAW network. At the time of writing the latest version is `v3.1`.

The SEOSAW dataset is built around two key tables:

- `plots.csv` - Plot level data
- `stems.csv` - Stem level data

Each row in the plots table represents a single census in a single plot. Each row in the stems table represents a single stem measurement. Note that a single tree may contain many stems and each stem may be measured multiple times over multiple censuses. See section 4 for a breakdown of columns found in the dataset tables. All data objects are linked by the `plot_id` column, which serves as the primary key for the dataset.

There are also some auxiliary data objects in the SEOSAW data repository:

- `polys.shp` - Polygons for each plot
- `small_stems.csv` - Count data for small stems (below minimum diameter threshold)
- `cwd.csv` - Coarse Woody Debris (CWD) data
- `ground_layer` - Ground layer biomass data

There are also some derived datasets:

- `stems_latest.csv` - Stem data subset for the most recent census of each plot
- `stems_multi.csv` - Stem data subset for plots with multiple censuses
- `plots_latest.csv` - Plot data subset for the most recent census of each plot
- `plots_multi.csv` - Plot data subset for plots with multiple censuses
- `plots_subdiv.csv` - Plot data for 1 ha subdivisions of plots >2 ha
- `polys_subdiv.shp` - Polygons for 1 ha subdivisions of plots >2 ha
- `plot_clusters.csv` - Floristic clusters for a subset of the plot data
- `cluster_dom.csv` - Dominant species per cluster

This section describes important aspects of the plot data in `plots.csv`. See section 4 for a full breakdown of columns in the plots table.

2.1 Nesting and subplots

Researchers occassionally stratify their sampling effort across a plot, often according to stem diameter classes, to reduce the work required to survey a plot. 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.

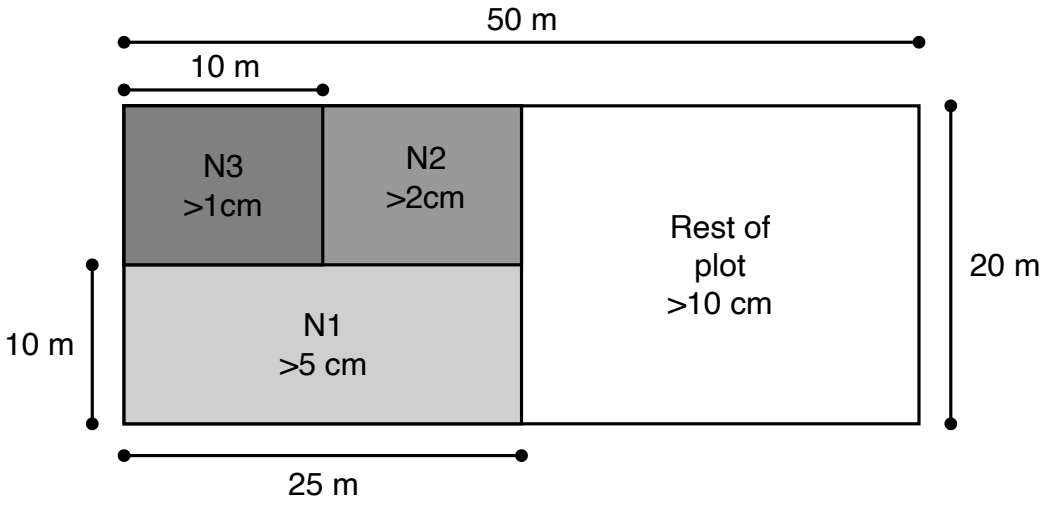

Figure 1: Schematic diagram of a nested plot with three levels of nesting.

There are a number of columns in `plots.csv` which record nesting, prefixed with either `nested_`, `n1`, `n2` or `n3`, for different levels of nesting.

If plots are nested, with variation in sampling effort for trees of different sizes, the Fractional Probability of Inclusion in the `fpc` column of the stem table describes the sampling effort for each stem. In the example, stems >10 cm DBH are measured across the whole plot, so their `fpc` is 1.0. 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 dataset only allows up to three levels of nesting, but this may change in the future.

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.

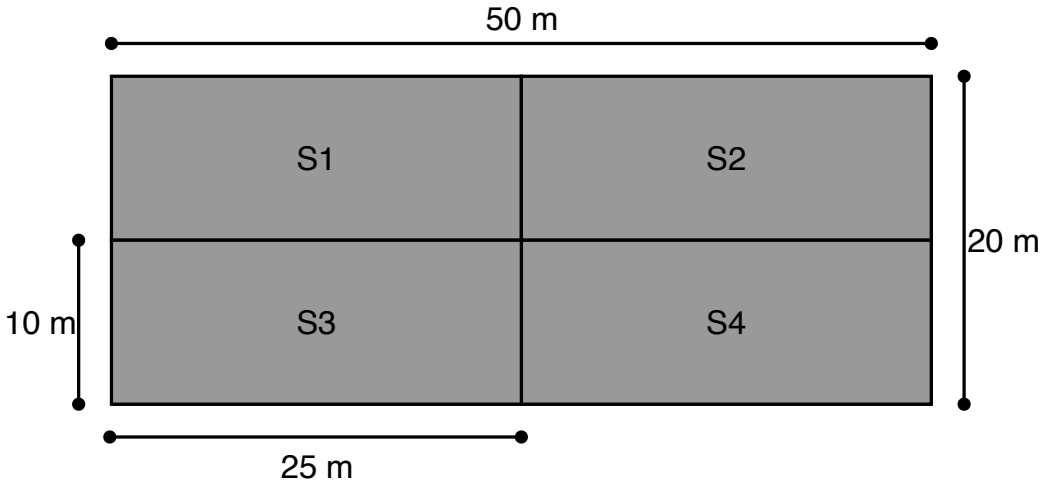

Figure 2: Schematic diagram of a plot with subplots, each occupying 25% of the total plot space.

It is assumed that all subplots have the same dimensions and that subplots are not nested. Columns prefixed with `subplot_` describe subplots.

2.2 Stems or trees

The ideal data for the SEOSAW dataset is organised so that there is one measurement for each stem during each census, with stems grouped into trees and a permanent tag to identify each stem. Understandably however, not all data matches this ideal, and is dependent on the original data collection protocol. There are a few columns in the plots table which help describe how stem data is organised:

- `tags` - Do stems have permanent tags which allow for remeasurement?
- `tag_tree_stem` - Are tags applied to each tree, or each stem?
- `meas_tree_stem` - Are measurements recorded at the tree or stem level?
- `tree_diff` - If `meas_tree_stem == 'stem'`, were stem measurements grouped into separate trees?

2.3 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.

Related, in many permanent plots it is possible to track the growth and mortality of individual stems between multiple censuses within a plot. Normally this is achieved through the use of numbered tags, but can alternatively be achieved with a highly accurate grid coordinate system, a differential GPS, or other tagging methods. The `stem_timeline_cont` column records whether the current census can be successfully stem-matched with the previous census. The `abandoned` column records if a plot has been abandoned, i.e. will not receive any more censuses in the future.

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

Both permanent and one-off plots are accepted by SEOSAW. The `permanent` column in the plots table notes whether a plot is permanent or not.

2.4 Subdivided large plots

`plots_subdiv.csv` and `polys_subdiv.shp` hold data on 1 ha square subdivisions of large plots. Large plots are subdivided to make them more readily comparable with other 1 ha plots in the dataset, and to make them easier to work with when conducting spatial analyses. Plots are only subdivided if a number of conditions are met:

- Plot is >=2 ha
- Plot is rectangular
- All plot edges are divisible by 100 m

Additionally, stems are only assigned to a subdivision if they have grid coordinates in `x_grid` and `y_grid`. If your stems are located instead by latitude/longitude coordinates, consider using the `gridCoordGen()` function to

approximate grid coordinates originating from the southwest corner of the plot. Subdividing plots is performed by `largePlotSplit()` in the `{seosawr}` R package.

The `subdiv` column in the plots table shows whether a plot has been subdivided. The `subdiv_id` column in the stems table shows which subdivision a stem has been assigned to. `subdiv_id` matches in `plots_subdiv.csv` where subdivision-level calculated values (e.g. `agb_ha`) are stored. The calculated values are calculated for each census per plot, like `plots.csv`. If you want only the most recent census for each plot, use `subsetLatest()`. Subdivision IDs are assigned by appending `_S[1-n]` to the plot ID.

2.5 Floristic clusters

A portion of the SEOSAW dataset has been clustered based on Ward's hierarchical clustering method. Data is clustered based on stem abundances per genus in the most recent census per plot. Plots with fewer than 14 stems were excluded from the analysis. Plots with exotic species were also excluded. Data on which cluster each plot is assigned to is found in `plot_clusters.csv` with the dominant species per cluster found in `cluster_dom.csv`.

2.6 Taxonomy

The taxonomic 'backbone' of the SEOSAW dataset is the CJB African plants database with some notable exceptions, which are discussed below. The dataset also uses the `taxize::gnr_resolve()` utility to query multiple databases to provide higher taxonomic rank information. The queried databases are:

- Open Tree of Life Reference Taxonomy
- The International Plant Names Index
- Tropicos - Missouri Botanical Garden
- GBIF Backbone Taxonomy
- NCBI

The breakup of the *Acacia* genus has been well documented by (Kyalangalilwa et al., 2013). SEOSAW uses their list of synonyms to correct ex-*Acacia* species. The CJB database does not currently acknowledge the *Acacia* split, so queries to CJB retain the *Acacia* name.

All *Rhus* species in SEOSAW's study region belong to *Searsia* and are transferred in this way. All recent phylogenetic results point to this, but the formal taxonomic work to transfer African *Rhus* species is incomplete.

Previously, *Marquesia* species were placed in the *Trillesanthus* genus, however this has since been reverted. SEOSAW corrects *Trillesanthus* to *Marquesia*.

The latest taxonomy states that *Maytenus* is a strictly Neotropical genus, while acknowledging that African *Maytenus* are a "problem". The latest taxonomic literature doesn't tell us what to do with African *Maytenus*. SEOSAW defaults to the CJB taxonomy on *Gymnosporia*/*Maytenus*, but this may change in the future.

There are also some miscellaneous synonym corrections which are maintained throughout the canonical SEOSAW dataset. The vast majority of these corrections reflect previously debunked names which still appear in older datasets.

If a species cannot be identified, it is given the "indet" epithet. If only the genus is known, it is recorded as "*Genus* indet" (e.g. *Brachystegia* indet). If only the family is known, it is recorded as "Family indet" (e.g. Fabaceae indet). As a last resort, if only the local name of the species is known, it is recorded in the `species_orig_local` column.

2.7 Biomass

Estimates of biomass, specifically above-ground woody biomass (AGB) are stored in the `agb` column in the stems table. Additionally, plot level above-ground woody biomass is stored in the `agb_ha` (AGB per hectare) column in the plots table. Biomass is estimated in the SEOSAW dataset by default using the `chave14()` function, which uses Equation 7 from (Chave et al., 2014). This allometric equation takes into account species wood density, stem diameter and an environmental stress index (E), derived from the plot location.

2.8 Plot polygons

Plot outline polygons are provided in `polys.shp`, for each plot where appropriate location and plot dimensions data is available. For rectangular plots, these polygons are normally created from GPS coordinates taken at each corner and used as vertices when transforming to a closed spatial polygon.

The SEOSAW region is provided as a spatial data object in the `{seosawr}` R package (`data(seosaw_region)`), along with a simplified bounding box of the SEOSAW region (`data(seosaw_bbox)`). There is also a useful outline of all countries in Africa (`data(africa)`).

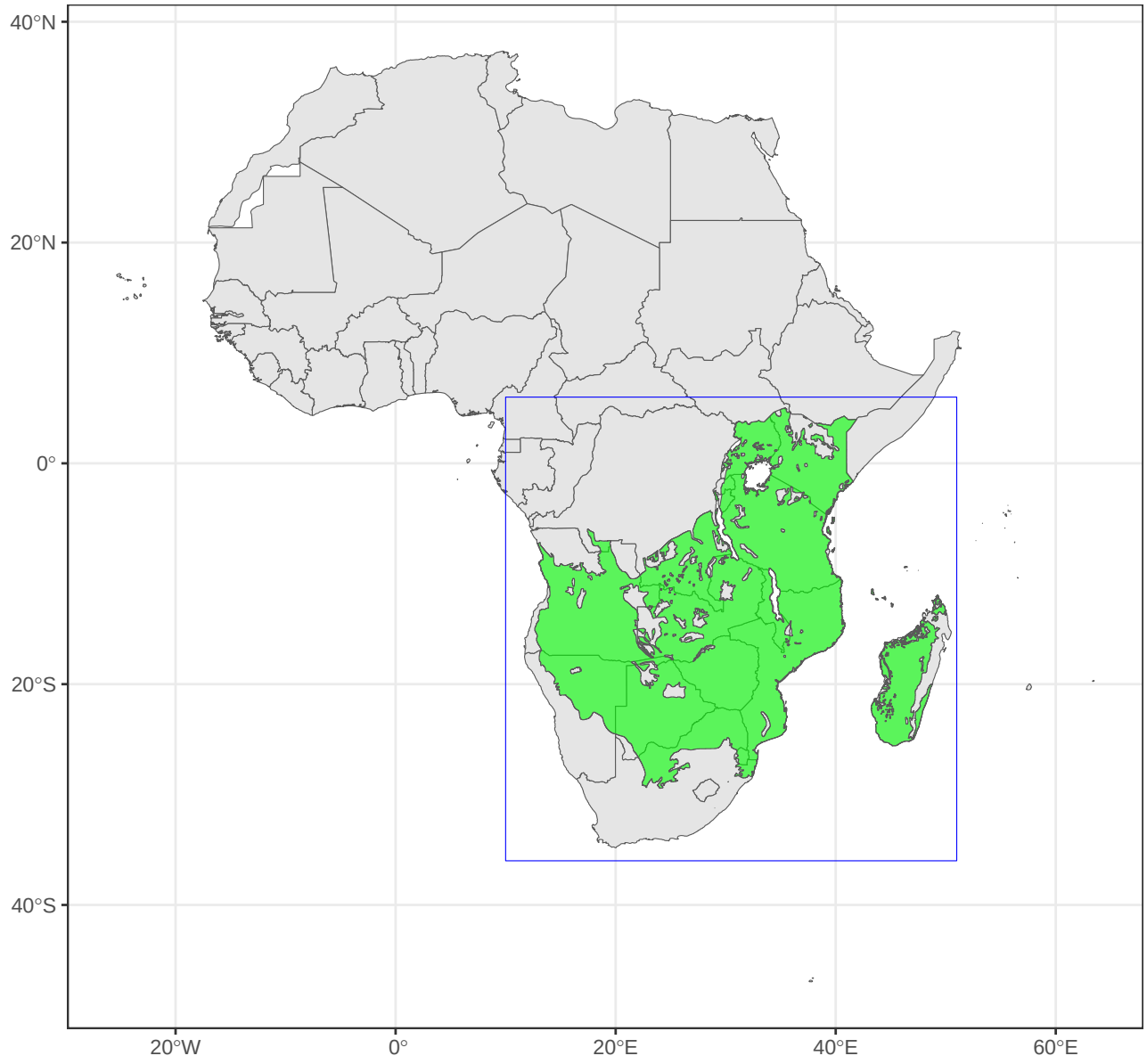

Figure 3: The spatial extent of the SEOSAW region shown in green and the SEOSAW bounding box as a blue rectangle, within Africa.

The SEOSAW region is based largely on White's vegetation map, which defines the locations of savannas and woodlands (White, 1983).

A key issue was defining the north-eastern boundary of the SEOSAW region. It is defined here as the borders of Kenya and Uganda, excluding Ethiopia, Somalia, and South Sudan. While the decision to use country boundaries was largely pragmatic, it is backed up by environmental data. The mountainous Ethiopian highlands on the border between Kenya, Ethiopia and Somalia cause quite a clear change in precipitation, probably due them blocking dry Arabian air moving further southwest.

The north-western boundary largely follows the rift valley lakes, with everything to the west of them being excluded. Some areas to the east of the lakes are also excluded as they are classed by White's vegetation map as

Congo Forest Transitions and Mosaics.

The south-east of the region is further reduced from White's vegetation map according to the Terrestrial Ecoregions of the World (Dinerstein et al., 2017). White classes some areas as woodland or forest in this area, but for our purposes they are considered too temperate, and are thus excluded from the SEOSAW region.

2.9 Dataset caveats

The SEOSAW dataset is aggregated from many individual datasets contributed by researchers around the world. While effort has been made to ensure that datasets are compatible with each other, particularly with respect to stem data like stem `diam` and `height`, some pieces of metadata should not be compared among datasets. For example, `canopy_cover` is collected with hemispherical photos, different types of densitometer, aerial photos and many other methods. Due to the subjectivity of all these methods, it is not recommended to use this data as a variable in statistical analysis when considering multiple datasets; it should be used only as metadata.

SEOSAW takes the stance that metadata should be left blank if it is not known, rather than guessed at. This means that many metadata columns consist mainly of `NA` values. For example, `charcoal_harvesting` is left mostly blank because most data originators did not record whether wood collection for charcoal harvesting had occurred in the plot. This also applies to stem data. If for example, the data originator did not record whether stems were located on termite mounds, `on_termites` is left as `NA`, even if it is more likely that most stems were not located on termite mounds.

3 Contributing data

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. Make sure to describe the Coordinate Reference System (CRS) used for coordinates, WGS84 Latitude-Longitude (EPSG: 4326) is preferred.

Please provide a description of the contents of each file in a separate document. For tabular data, if not using the SEOSAW column definitions, provide a description of the contents of each column. For example, to describe the contents of a column in the stems table containing stem diameter measurements:

> `S diam` - 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 in a separate document, especially any plot nested-ness or stratified sampling, as this will determine whether any corrections should be made when calculating plot-level statistics like aboveground biomass.

3.1 Vital data

There are many 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.

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:

Plots table:

- `plot_name` - Plot name, for matching with plot names in stems table
- `longitude_of_centre`, `latitude_of_centre` - Plot longitude and latitude coordinates
- `plot_shape`, `plot_length`, `plot_width`, or `plot_radius` - Plot shape and dimensions
- `prinv` - The Primary Investigator of the data
- `nested_1`, `n1_*`, etc. - Information on plot nested-ness, including nested plot dimensions and stem size thresholds.

Stems table:

- `plot_name` - Plot name, for matching with plot names in plots table
- `species_orig_binom` or `species_orig_local` - Tree species name
- `diam` - Tree diameter
- `pom` - Height of diameter measurement
- `fpc` - 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, based on the proportional area of the nested plot. If plots are not nested, `fpc` = 1.

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.

3.2 Additional 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.

4 Dataset table columns

This section provides a comprehensive description of every column found in the SEOSAW dataset, across the plots table, stems table, small stems table, coarse woody debris table, and ground layer table.

Table 1: Column names for plots table.

Column	Class	Description	Units	Example
plot_id	chr	Standardised method of identifying plots. A three letter code, followed by a number referring to the particular plot. Three letter codes are normally, but not always, created like this: Angola (A), in Bicuar National Park (B), by John Godlee (G). The three letter code should be conserved within a dataset, i.e. a group of plots owned by a particular researcher within a particular locale.	NA	SSM_1
census_date	chr	The date of the census. This can be formatted as either YYYY, YYYY-MM, or YYYY-MM-DD	date	2008
plot_cluster	chr	If plots occur in a spatial cluster, a single 'plot_id' used to identify the cluster.	NA	ABG_1
plot_name	chr	Plot name as designated by those who originally collected the data. 1, Wedza_01, MHR-04b are all valid. No spaces please.	NA	Wedza_01
country_iso3	fct	ISO 3166-1 alpha-3 code to refer to the country in which the plot is situated. Only countries within Africa. See 'data(africa_iso3)' for available country codes. use countrycode::countrycode() to convert a country name to an ISO3 code	NA	AGO
prinv	chr	Name of the Principal Investigator. This can be an organisation, a person's name, a surname, initials, etc.	NA	Godlee J. L.
other_authors	chr	Names of other authors who contributed to the collection of the dataset, separated by semi-colons	NA	Godlee J. L.; Maiato F.; Dexter K. G.
plot_shape	fct	Shape of the plot. Possible values are: "rectangle", "circle" or "irregular". Note that a square is a special case of "rectangle".	NA	rectangle
plot_width	int	If plot_shape == "rectangle", the shorter side length, in metres to the nearest metre	m	100
plot_length	int	If plot_shape == "rectangle", the longer side length, in metres to the nearest metre	m	50
plot_radius	int	If plot_shape == "circle", the radius of the plot, measured in metres to the nearest metre	m	25
subplot_n	int	Number of subplots within the larger plot. Subplots (as opposed to nested plots) have the same sampling effort as the entire plot, generally used for ease of relocating trees.	n	10
subplot_shape	fct	Shape of the subplots. Possible values are the same as plot_shape	NA	circle
subplot_length	int	If subplot_shape == "rectangle", the longest side length, in metres to the nearest metre	m	50
subplot_width	int	If subplot_shape == "rectangle", the shorter side length, in metres to the nearest metre	m	10

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
subplot_radius	int	If subplot_shape == "circle", the radius of the plot, measured in metres to the nearest metre	m	10
min_diam_thresh	int	Minimum stem diameter value considered within this plot, in centimetres to the nearest centimetre. If sampling is nested this should be the minimum value outside of all nested subplots.	cm	5
min_height_thresh	num	Minimum stem height value considered within this plot, in metres to the nearest 10 centimetres. If sampling is nested this should be the maximum value outside of all nested subplots.	m	1.5
nested_1	log	Does the plot have at least one level of nested sampling?	NA	TRUE
n1_min_diam_thresh	int	If nested_1 == TRUE, the minimum stem diameter value considered within the first nesting level, in centimetres to the nearest centimetre.	cm	10
n1_min_height_thresh	num	If nested_1 == TRUE, the minimum stem height value considered within the first nesting level, in metres to the nearest metre.	m	20
n1_plot_shape	fct	If nested_1 == TRUE, the shape of the subplot, same values as plot_shape	NA	circle
n1_plot_width	int	If nested_1 == TRUE & n1_plot_shape == "rectangle" the shorter side length, in metres to the nearest metre	m	20
n1_plot_length	int	If nested_1 == TRUE & n1_plot_shape == "rectangle" the longer side length, in metres to the nearest metre	m	10
n1_plot_radius	int	If nested_1 == TRUE & n1_plot_shape == "circle" the radius of the subplot, in metres to the nearest metre	m	14
nested_2	log	Does the plot have at least two levels of nested sampling?	NA	TRUE
n2_min_diam_thresh	int	If nested_2 == TRUE, the minimum stem diameter value considered within the second nesting level, in centimetres to the nearest centimetre.	cm	5
n2_min_height_thresh	num	If nested_2 == TRUE, the minimum stem height value considered within the second nesting level, in metres to the nearest metre.	m	2.5
n2_plot_shape	fct	If nested_2 == TRUE, the shape of the subplot, same allowed values as plot_shape	NA	rectangle
n2_plot_width	int	If nested_2 == TRUE & n2_plot_shape == "rectangle" the shorter side length, in metres to the nearest metre	m	30
n2_plot_length	int	If nested_2 == TRUE & n2_plot_shape == "rectangle" the longer side length, in metres to the nearest metre	m	20
n2_plot_radius	int	If nested_2 == TRUE & n2_plot_shape == "circle" the radius of the subplot, in metres to the nearest metre	m	5
nested_3	log	Does the plot have at least three levels of nested sampling?	NA	TRUE

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
n3_min_diam_thresh	int	If nested_3 == TRUE, the minimum stem diameter value considered within the third nesting level, in centimetres to the nearest centimetre.	cm	5
n3_min_height_thresh	num	If nested_3 == TRUE, the minimum stem height value considered within the third nesting level, in metres to the nearest metre.	m	2.5
n3_plot_shape	fct	If nested_3 == TRUE, the shape of the subplot, same allowed values as plot_shape	NA	rectangle
n3_plot_width	int	If nested_3 == TRUE & n3_plot_shape == "rectangle" the shorter side length, in metres to the nearest metre	m	30
n3_plot_length	int	If nested_3 == TRUE & n3_plot_shape == "rectangle" the longer side length, in metres to the nearest metre	m	20
n3_plot_radius	int	If nested_3 == TRUE & n3_plot_shape == "circle" the radius of the subplot, in metres to the nearest metre	m	5
lianas_sampled	log	Were lianas sampled? I.e. woody stems that are not self-supporting and use other trees to grow towards the canopy	NA	TRUE
soil_sampled	log	Were soil samples collected?	NA	FALSE
ground_layer_sampled	log	Was the ground layer sampled for grasses, forbs, shrubs, or small trees that did not fit the size thresholds for the plot?	NA	TRUE
dead_stems_sampled	log	Were dead stems sampled in this plot?	NA	FALSE
stumps_sampled	log	Were tree stumps sampled?	NA	TRUE
coarse_woody_debris_sampled	log	Was coarse woody debris sampled?	NA	FALSE
small_stems_sampled	log	Were woody stems smaller than the minimum size thresholds counted?	NA	TRUE
tags	log	Are stems or trees within the plot tagged or otherwise identifiable between censuses?	NA	FALSE
permanent	log	Is the plot a permanent plot (TRUE) or a one-off census (FALSE)? Could the plot be re-measured in the future?	NA	TRUE
meas_tree_stem	fct	Are measurements taken for each "tree", or each "stem"?	NA	tree
tag_tree_stem	fct	Are tag IDs at the "tree" or "stem" level?	NA	tree
tree_diff	log	Are stems grouped into trees, where trees have multiple stems?	NA	TRUE
abandoned	log	Is this the last census and the plot is now abandoned? Common reasons for abandonment: Clear-felling or nearly all trees removed. Often combined with tag loss and irreconcilable damage to the plot which would make it hard to match the stems with a previous census.	NA	TRUE
stem_timeline_cont	log	If tags == TRUE and permanent == TRUE, is it impossible to link stems between this census and any previous censuses?	NA	FALSE

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
sampling_notes	chr	Extra notes relating to the sampling process	NA	Only stumps greater than 2 m tall were sampled
sampling_design_code	chr	File name for document containing the description of your methods/protocol for data collection, provided to SEOSAW, or a reference to such a document. If not documented then NA	NA	SSAW_V1.gdoc
plot_notes	chr	Include any long form comments about the plot here, if they don't fit into other columns.	NA	It looks like this plot is a rubbish dump for household waste.
longitude_of_centre	num	Longitude of the centre of gravity of the plot. If you have GPS points of the plot corners, send those and we will calculate this. Measured in decimal degrees, WGS84 EPSG:4326	deg	-50.425622
latitude_of_centre	num	Latitude of the centre of gravity of the plot. If you have GPS points of the plot corners, send those and we will calculate this. Measured in decimal degrees, WGS84 EPSG:4326	deg	30.446453
elevation	num	Elevation of the plot. Measured in metres above International Mean sea level to 1 decimal place (nearest 10 cm). In order of preference: 1) Average of GPS readings across the plot, 2) a single GPS reading, 3) Manual lookup on topographic map, 4) NA. We can query SRTM for missing values.	m asl	1234
slope_method	fct	Method used to determine plot slope. Allowed values: "rangefinder" if a laser or sonic rangefinder was used, "trig" if slope was estimated with tape measures and-or a clinometer, "eye" if estimated by eye, "other" if another method.	NA	trig
slope_method_notes	chr	Extra notes relating to the slope method	NA	20 points taken with 20 m minimum distance, randomly located
slope	num	Plot slope, measured in degrees from 0 (flat) to 90 (vertical)	deg	4.4
aspect	num	Aspect of the plot, i.e. which direction is the slope facing, in degrees (0-360)	deg	54
polygons	log	Polygon data of the plot shape provided by the data originator?	NA	TRUE
local_allometry	fct	Most appropriate aboveground woody biomass allometry for the plot. In order of preference: (1) DOI, (2) reference, (3) other URL, (4) shorthand allometry name	NA	ryan11

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
allometry_notes	chr	Extra notes relating to the appropriate local allometry	NA	ryan11 should be used for everything except the eucalyptus trees present in plot ABC_5
catenal_position	fct	Where on the catena is plot? Possible values are:"ridge_crest", "concave_slope", "midslope", "convex_slope", "valley_floor", "flat"	NA	midslope
termites	int	If termites were counted, the number of termite mounds found on the plot	n	34
canopy_cover	prop	The proportion of the ground (periscope, aerial photo, canopy dimensions), or sky hemisphere (hemi-photo, mirror densitometer), covered by a tree canopy. Measured as a proportion between 0 and 1, e.g. 60% = 0.6	proportion	0.28
canopy_cover_method	fct	If canopy cover was estimated, what method was used? Possible values are: "periscope_densitometer", "mirror_densitometer", "hemispherical _photo", "aerial_photo", "canopy_dimensions", "other"	NA	hemispherical_photo
canopy_cover_method_notes	chr	Extra notes relating to the canopy cover method	NA	40 hemispherical photos in a regular grid with 20 m spacing. Analysed with Hemiphot.
xy_method	fct	Method used to estimate stem location within the plot, either grid or GPS coordinates. Possible values: "tape", "gps", "dgps", "angle" (angular direction and distance from plot centre in circular plot), "other".	NA	tape
manipulation_experiment	log	Has any aspect of the plot's land use or environment been manipulated by researchers?	NA	FALSE
experiment_ref	chr	If manipulation_experiment == TRUE, a link, citation, or contact details relating to the experimental manipulation of this plot. In order of preference: (1) DOI, (2) reference, (3) other URL, (4) contact details	NA	10.1890/06-1664
manipulation_experiment_notes	chr	If manipulation_experiment == TRUE, notes relating to the experimental treatment applied to the plot	NA	Tilling of soil 15-20 cm throughout
pa	log	Is the plot in a protected area?	NA	TRUE

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
pa_name	chr	If pa == TRUE, the name of the protected area. If the protected area is listed by the WDPA, please use the name provided by the WDPA	NA	Bicuar National Park
pa_type	chr	If pa == TRUE, the type of protected area. If the protected area is listed by the WDPA, use their type classification.	NA	Forest reserve
wdpa_id	chr	If pa == TRUE and the protected area is listed by the WDPA, the WDPA ID value for that protected area. See https://www.protectedplanet.net	NA	370
local_community_rights	chr	Explain the rights of the local community in using the plot for resource extraction, grazing, farming etc.	NA	Dead wood collection by local village
other_user_rights	chr	Explain the rights of any others in using the plot for resource extraction, grazing, farming etc.	NA	Yearly game drive in early May
timber_harvesting	log	Is there evidence of tree harvesting for timber products, e.g. planks?	NA	TRUE
high_graded_100_years	log	Is it likely that high quality timber trees were removed from the plot the last 100 years?	NA	FALSE
clear_cut	log	Has any part of the plot ever been clear cut (all trees removed)?	NA	TRUE
mining	log	Has any part of the plot ever been used for mining?	NA	FALSE
fuel_wood_harvesting	log	Is there evidence of fuel wood collection, excluding charcoal production?	NA	TRUE
charcoal_harvesting	log	Is there evidence of trees being harvested for charcoal production?	NA	FALSE
other_woody_product_harvesting	log	Is there evidence of people harvesting woody material for any other use, e.g. string, beehives, canoes?	NA	TRUE
ntfp_harvesting	log	Is there evidence of people harvesting non-wood tree products, e.g. mushrooms honey, fruit etc.?	NA	FALSE
farmed_30_years	log	Is it likely that the plot was used for farming crops in the last 30 years?	NA	TRUE
livestock_30_years	log	Is it likely that the plot was used for grazing livestock in the last 30 years?	NA	FALSE
land_use_notes	chr	Extra notes relating to land use in the plot	NA	FALSE
fire_exclusion	log	Is there a fire exclusion experiment in this plot? If TRUE, manipulation_experiment should also = TRUE.	NA	TRUE
fire_exclusion_success	fct	If fire_exclusion == TRUE, to what extent were fires successfully excluded from the plot? Possible values: "all", "most", "few"	NA	all
fire_treatment	log	Is or was there a fire treatment experiment in this plot, e.g. controlled fires every two years? If TRUE, manipulation_experiment should also = TRUE.	NA	FALSE

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
fire_treatment_season	fct	If fire_treatment == TRUE, the season in which fires are set. Possible values: "early_dry_season", "late_dry_season", "early_wet_season", "late_wet_season"	NA	early_dry_season
fire_return_interval	int	Number of years between fires. If there has been some sort of fire control experiment, the treatment fire return interval measured in years, e.g. 1 for annual burning. If no treatment, the estimated mean interval between fires.	year	5
fire_last_date	chr	Date of the last fire.	date	2022
fire_cause	chr	What commonly causes fires in the plot? E.g. Management, accidental, lightning, agriculture.	NA	Adjacent field fires
fire_fuel	chr	What do fires on this plot typically burn? E.g. Canopy, understorey, litter	NA	Tree canopies
fire_notes	chr	Extra notes relating to fire regime or fire treatment	NA	Fires are set with drip torches in the late evening, after a heavy rain
elephants	log	Do elephants frequent the plot	NA	TRUE
elephant_density	num	If elephants are present, at what approximate density do they occur within the area? Measured as elephants per km^2	n km^-2	5
elephant_data_source	chr	A link, citation, or contact details relating to elephant density estimates. In order of preference: (1) DOI, (2) reference, (3) other URL, (4) contact details	NA	10.6088/ijes.002
exclosure	fct	Are herbivores excluded from the plot by a fence? If so, what sizes? Possible values are: "extra_large" (>1000 kg), "large" (100-1000 kg), "medium" (30-100 kg), "small" (<30 kg). Multiple values possible separated by semi-colon.	NA	extra_large
small_nonsocial_browsers	log	Do small non-social browsers frequent this plot? e.g. dik-diks, duikers and grysbok, steenbok (these categories are based on Hempson et al 2015). TRUE or FALSE	NA	TRUE
large_browsers	log	Do large browsers frequent this plot? e.g. giraffe, okapi, kudu	NA	FALSE
medium_sized_social_mixed_diet	log	Do medium sized social mixed diet herbivores frequent this plot? e.g. gazelle, oryx	NA	TRUE
nonruminants_excluding_suids	log	Do non-ruminant herbivores frequent this plot? e.g. zebra, hippo, rhino	NA	FALSE
water_dependent_grazers	log	Do water-dependent grazers frequent this plot? e.g. roan, sable, bluebuck, waterbuck, lechwe	NA	TRUE
goat_grazing	log	Do goats feed in this plot?	NA	FALSE

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
sheep_grazing	log	Do sheep feed in this plot?	NA	TRUE
cattle_grazing	log	Do domestic cattle feed in the plot?	NA	FALSE
herbivore_notes	chr	Extra notes relating to herbivores in the plot.	NA	Dik Diks, palankas, Olongo, elephants like to knock over the B. massaiensis
cyclone	log	Do cyclones (category 3+) ever occur in this plot?	NA	TRUE
cyclone_last_date	chr	When was the last cyclone?	NA	2022
drought	log	Do major droughts ever occur in this plot?	NA	FALSE
drought_last_date	chr	When was the last major drought?	NA	2021
wind	log	Do wind storms or tornadoes ever occur in this plot?	NA	TRUE
wind_last_date	chr	When was the last wind storm or tornado?	NA	2020
earthquake	log	Do earthquakes ever occur in this plot?	NA	FALSE
earthquake_last_date	chr	When was the last earthquake?	NA	2019
landslide	log	Do landslides ever occur in this plot?	NA	TRUE
landslide_last_date	chr	When was the last landslide?	NA	2018
flood	log	Do major floods ever occur in this plot?	NA	FALSE
flood_last_date	chr	When was the last major flood?	NA	2017
frost	log	Do major frost events (<-2C for three consecutive nights) ever occur in this plot?	NA	TRUE
frost_last_date	chr	When was the last major frost event?	NA	2016
data_ref	chr	A link, citation, or contact details relating to the publication of this plot's data. In order of preference: (1) DOI, (2) reference, (3) other URL, (4) contact details	NA	10.3390/d12040140
photos	chr	Filenames of photos of the plot, separated by semi-colons	NA	DSC_134.jpg;DSC_-135.jpg
date_metadata	num	Year that dataset metadata values, i.e. herbivory, land use, plot location, fire regime etc. were last updated	NA	2020
file_plots	chr	Filepath of the raw plot metadata file provided to SEOSAW. Include multiple files in a semi-colon separated list.	NA	archibald_-kruger.xlsx;Kruger_-2024.csv
file_stems	chr	Filepath of the raw stem data file provided to SEOSAW. Include multiple files in a semi-colon separated list.	NA	Archibald_kruger.csv
subdiv	log	Is the plot subdivided into smaller 1 ha square parcels? If TRUE, plot-level calculated values can be found in plots_subdiv.csv	NA	TRUE
dataset_id	chr	The dataset ID, normally defined as the three letter code in plot_id	NA	SSM
plot_area	num	Flat area of the plot, in hectares	ha	0.25

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
subplot_area	num	Flat area of the subplots, in hectares	ha	0.005
plot_perimeter	num	Perimeter of the plot, in metres	m	400
diam_max	num	Maximum stem diameter found in the plot	cm	50.4
diam_min	num	Minimum stem diameter found in the plot	cm	5
diam_mean	num	Mean stem diameter found in the plot	cm	5.6
diam_median	num	Median stem diameter found in the plot	cm	10.3
diam_sd	num	Standard deviation of stem diameter found in the plot	cm	1.5
diam_ge5_mean	num	Mean stem diameter found in the plot, greater than 5 cm diameter	cm	5.4
diam_ge5_median	num	Median stem diameter found in the plot, greater than 5 cm diameter	cm	10.6
diam_ge5_sd	num	Standard deviation of stem diameter found in the plot, greater than 5 cm diameter	cm	1.5
diam_ge10_mean	num	Mean stem diameter found in the plot, greater than 10 cm diameter	cm	15.7
diam_ge10_median	num	Median stem diameter found in the plot, greater than 10 cm diameter	cm	15.9
diam_ge10_sd	num	Standard deviation of stem diameter found in the plot, greater than 10 cm diameter	cm	2.5
diam_ge20_mean	num	Mean stem diameter found in the plot, greater than 20 cm diameter	cm	15.7
diam_ge20_median	num	Median stem diameter found in the plot, greater than 20 cm diameter	cm	15.9
diam_ge20_sd	num	Standard deviation of stem diameter found in the plot, greater than 20 cm diameter	cm	2.5
n_diam	num	Number of stems for which diameter was measured	n	524
n_height	num	Number of stems for which height was measured	n	532
height_max	num	Maximum stem height found in the plot	m	10.5
height_min	num	Minimum stem height found in the plot	m	1.6
height_mean	num	Mean stem height found in the plot	m	1.6
height_median	num	Median stem height found in the plot	m	5.4
height_sd	num	Standard deviation stem height found in the plot	m	4.5
height_ge5_mean	num	Mean stem height found in the plot, greater than 5 cm diameter	m	3.5
height_ge5_median	num	Median stem height found in the plot, greater than 5 cm diameter	m	5.4
height_ge5_sd	num	Standard deviation of stem height found in the plot, greater than 5 cm diameter	m	1.5
height_ge10_mean	num	Mean stem height found in the plot, greater than 10 cm diameter	m	3.5
height_ge10_median	num	Median stem height found in the plot, greater than 10 cm diameter	m	6.6
height_ge10_sd	num	Standard deviation of stem height found in the plot, greater than 10 cm diameter	m	1.4
height_ge20_mean	num	Mean stem height found in the plot, greater than 20 cm diameter	m	3.5

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
height_ge20_median	num	Median stem height found in the plot, greater than 20 cm diameter	m	6.6
height_ge20_sd	num	Standard deviation of stem height found in the plot, greater than 20 cm diameter	m	1.4
height_per95	num	95th percentile of stem height, i.e. the height below which 95% of stems can be found	m	4.3
n_stems	num	Number of stems in the plot	n	5025
n_stems_ge5	num	Number of stems greater than 5 cm diameter	n	5010
n_stems_ge10	num	Number of stems greater than 10 cm diameter	n	2040
n_stems_ge20	num	Number of stems greater than 20 cm diameter	n	505
n_stems_ge40	num	Number of stems greater than 40 cm diameter	n	20
n_stems_ge60	num	Number of stems greater than 60 cm diameter	n	1
lorey_height	num	Lorey height of the plot. Calculated as the sum of basal areas multiplied by height, divided by the total basal area. This variant is adjusted for nested sampling with fpc	NA	3.5
n_determ_genus	num	Number of stems where genera successfully identified	n	20
n_determ_species	num	Number of stems where species epithets successfully identified	n	49
dom_1	chr	Most dominant species in the plot, by stem abundance	NA	Brachystegia spiciformis
dom_2	chr	Second most dominant species in the plot, by stem abundance	NA	Ochna pulchra
dom_3	chr	Third most dominant species in the plot, by stem abundance	NA	Burkea africana
dom_4	chr	Fourth most dominant species in the plot, by stem abundance	NA	Indet indet
dom_5	chr	Fifth most dominant species in the plot, by stem abundance	NA	Pseudolachnostylis maprouneifolia
prop_1	num	Proportional abundance of the most dominant species in the plot	proportion	20.4
prop_2	num	Proportional abundance of the second most dominant species in the plot	proportion	14.5
prop_3	num	Proportional abundance of the third most dominant species in the plot	proportion	12.7
prop_4	num	Proportional abundance of the fourth most dominant species in the plot	proportion	8.4
prop_5	num	Proportional abundance of the fifth most dominant species in the plot	proportion	3.6
richness	num	Species richness of plot	n	53
rarefy	num	Rarefied species richness at 25 individuals	NA	42.65
shannon	num	Shannon diversity index of plot	NA	4.3
simpson	num	Simpson diversity index of plot	NA	2.5
invsimpson	num	Inverse Simpson diversity index of plot	NA	2.6

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
evenness	num	Abundance evenness of plot	NA	2.6
stocking_ha	num	Number of stems per hectare, adjusted for nested sampling with fpc	n ha^-1	50.2
stocking_ge5_ha	num	Number of stems greater than 5 cm diameter per hectare, adjusted for nested sampling with fpc	n ha^-1	45.5
stocking_ge10_ha	num	Number of stems greater than 10 cm diameter per hectare, adjusted for nested sampling with fpc	n ha^-1	32.5
stocking_ge20_ha	num	Number of stems greater than 20 cm diameter per hectare, adjusted for nested sampling with fpc	n ha^-1	32.5
ba_ha	num	Basal area per hectare, adjusted for nested sampling with fpc	m^2 ha^-1	3.6
agb_ha	num	Aboveground woody biomass per hectare, for living stems only (stem_status == "a" | is.na(stem_status)), adjusted for nested sampling with fpc	t DM ha^-1	42.6
agb_ge5_ha	num	Aboveground woody biomass per hectare for stems greater than 5 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6
agb_ge10_ha	num	Aboveground woody biomass per hectare for stems greater than 10 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6
agb_ge20_ha	num	Aboveground woody biomass per hectare for stems greater than 20 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6
diam_scale	num	Diameter distribution scaling parameter. The slope of a linear model of log-transformed diameter class frequency values in 5 cm diameter bins.	NA	1.55
diam_scale_se	num	Standard error on the slope parameter of the diameter distribution scaling parameter.	NA	0.544
weibull_scale	num	Scale parameter of a two-parameter diameter Weibull probability distribution.	NA	1.47
weibull_shape	num	Shape parameter of a two-parameter diameter Weibull probability distribution.	NA	0.63
weibull_scale_se	num	Standard error on the scale parameter of the diameter class distribution	NA	0.444
weibull_shape_se	num	Standard error on the shape parameter of the diameter class distribution	NA	0.135
baobab	log	Do members of the Adansonia genus comprise $> 10\%$ of the aboveground biomass? This column is included because Adansonia spp. peculiar growth form can drastically inflate the aboveground biomass of a plot	NA	TRUE
stocking	num	Number of stems per plot adjusted for nested sampling with fpc	n	54.23

Table 1: Column names for plots table. *(continued)*

Column	Class	Description	Units	Example
stocking_ge5	num	Number of stems greater than 5 cm diameter per plot, adjusted for nested sampling with fpc	n	43.2
stocking_ge10	num	Number of stems greater than 10 cm diameter per plot, adjusted for nested sampling with fpc	n	34.5
stocking_ge20	num	Number of stems greater than 20 cm diameter per plot, adjusted for nested sampling with fpc	n	3.24
ba	num	Basal area per plot, adjusted for nested sampling with fpc	m^2 ha^-1	3.2
agb	num	Aboveground woody biomass per plot, adjusted for nested sampling with fpc	t DM ha^-1	5.4
agb_ge5	num	Aboveground woody biomass per plot for stems greater than 5 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6
agb_ge10	num	Aboveground woody biomass per plot for stems greater than 10 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6
agb_ge20	num	Aboveground woody biomass per plot for stems greater than 20 cm diameter, adjusted for nested sampling with fpc	t DM ha^-1	42.6

Table 2: Column names for stems table.

Column	Class	Description	Units	Example
plot_id	chr	Unique identifier for plots. A three letter code, followed by a number referring to the particular plot. Three letter codes are normally, but not always, created like this: Angola (A), in Bicuar National Park (B), by John Godlee (G). The three letter code should be conserved within a dataset, i.e. a group of plots owned by a particular researcher in a particular locale.	NA	ABG_1
subplot_id	chr	If the plot is divided into multiple parcels, but not "nested", i.e. sampling effort is equal across all parcels, the parcel within which the stem is found.	NA	A
tree_id	chr	Unique identifier for trees within a plot, normally the first stem ID within the tree.	NA	45
stem_id	chr	Unique identifier for stems within a plot. Normally the tag ID if the plot has tags.	NA	6
tag_id	chr	The code of the tag attached to the tree or stem in the field. If tags were not used, leave this column blank.	NA	BP094
census_date	chr	The date of the census. This can be either YYYY, YYYY-MM, or YYYY-MM-DD.	NA	26/06/2022
species_orig_binom	chr	The binomial name of the tree species, as provided by the data originator	NA	Burkea africana
species_orig_local	chr	The local name of the tree species/morphospecies, as provided by the data originator	NA	mumwei
notes_stem	chr	Miscellaneous notes related to the tree or stem	NA	Leaning around 45 degrees and with a beehive in the fluted trunk
diam	num	Stem diameter measurement in cm	cm	6.4
pom	num	The height at which diam was taken, in metres to one decimal place (nearest 10 cm)	m	1.3
x_grid	num	The x coordinate of the stem within the plot if a grid method was used. In metres to nearest 10 cm	m	54.2
y_grid	num	The y coordinate of the stem within the plot if a grid method was used. In metres to nearest 10 cm	m	35.6
latitude	num	The latitude coordinate of the stem within the plot if a GPS method was used. WGS84 EPSG4326	decimal degrees	-50.425622
longitude	num	The longitude coordinate of the stem within the plot if a GPS method was used. WGS84 EPSG4326	decimal degrees	30.446453

Table 2: Column names for stems table. *(continued)*

Column	Class	Description	Units	Example
angle	num	If 'plot_shape' == "circle", and angular stem coordinates were used, the angle (0 = North) direction of the stem from the plot centre, in degrees (0-360)	degrees	43.2
distance	num	If 'plot_shape' == "circle", and angular stem coordinates were used, the distance of the stem from the plot centre. In metres to the nearest 10 cm	m	3.5
height	num	The height of the tree, in metres to one decimal place (nearest 10 cm)	m	3.5
bole_height	num	The height from the ground to the base of the lowest branch also called 'clear length'	m	6.3
crown_x	num	If two crown diameters are taken perpendicular to one another, the first of these measurements. If one crown diameter was taken, that crown diameter, in metres to one decimal place (nearest 10 cm).	m	5.2
crown_y	num	If two crown diameters are taken perpendicular to one another, the second of these measurements. If one crown diameter was taken, this remains blank, in metres to one decimal place (nearest 10 cm).	m	3.5
stump_diam_top	num	If the stem is a stump, the diameter at the top of the stump, in cm.	cm	10.4
stump_diam_base	num	If the stem is a stump, the diameter at the base of the stump (30 cm), in cm.	cm	12.5
stem_status	fct	What is the growth status of the stem. Possible values are: "r" if the stem is topkilled and resprouting, "d" if the stem and the rest of the tree is completely dead (i.e. base appears dead and all stems are dead), "a" if the stem is alive with no abnormal growth status.	NA	r
broken_per_remain	num	If the stem is broken, approximately how much is remaining? As a decimal, e.g. 60% of stem remaining = 0.6, to one decimal place (nearest 10%)	proportion	0.2
broken_height	num	If the stem is broken, at what height is it broken. In metres to one decimal place (nearest 10 cm)	m	4.5
wound_area	num	If the stem is wounded, the area of the wound.	cm^2	6

Table 2: Column names for stems table. *(continued)*

Column	Class	Description	Units	Example
stem_mode	fct	Various flags denoting stem mode. Possible values are: "b" - broken (material above break is detached), "c" - has climbing or strangling or liana plants in canopy, "f" - fallen, "g" - hollow, "j" - bark peeling or lost, "k" - stem has bracket fungi, "l" - leaning (>15deg), "m" - on a termite mound, "n" - a new recruit, "p" - snapped (material above break is attached), "q" - the stem location is unknown and was not found (contrast with "v"), "s" - standing, "t" - stump i.e. broken below POM, "u" - uprooted, "v" - vanished (location of tree found but stem missing), "w" - stem is a liana, "y" - wounded, "z" - sick (i.e. lost most leaves, discoloured foliage, drooping branches) or rotten wood	NA	P
damage_cause	fct	If the stem is damaged, what caused it? Possible values are: "a" if animals (not elephants, not insects), "e" if elephants, "f" if fire, "h" if humans (.e.g. cut, ringbarked, broken or beehive), "i" if frost, "l" if lightning, "m" if termites, "n" if neighbouring tree, "o" if other (describe in damage_cause_notes), "q" if unknown, "w" if wind	NA	N
damage_cause_notes	chr	If damage_cause = "o", describe the specific method of damage.	NA	chainsaw marks and machete on small branches
cambium	num	The proportion of trunk surface area with exposed vascular cambium, between the ground and 2 m.	proportion	0.22
decay	num	What level of decay is the stem? Scale runs from 1 (no decay) to 5 (very decayed). Follow the "Decay class" table in the SEOSAW protocol. To 0.5 of a decay level (e.g. 1.5, 1, 3.5)	NA	4.5
diam_method	fct	What method was used to derive the stem diameter measurement? Possible values are: "dbh_tape" if a tape measure with a diameter conversion scale was used, "tape" if a normal tape measure was used and the circumference converted to diameter, "caliper_1" if a single caliper measurement was used, "caliper_2" if two caliper measurements at 90 degrees with the mean taken, "other".	NA	dbh_tape

Table 2: Column names for stems table. *(continued)*

Column	Class	Description	Units	Example
height_method	fct	What method was used to derive the height measurement? Possible values are: "eye" if estimated by eye, "rangefinder" if estimated using laser or ultrasonic apparatus with an electronic tilt sensor e.g. Nikon Forestry Pro, "clinometer" if estimated using a mechanical Clinometer, "hyposometer" if estimated using a laser vertex hypsometer from below the tree using last return, "direct" if estimated by climbing the tree or adjacent tower and measuring the height with tape or stick, "other"	NA	eye
voucher_id	chr	If a herbarium voucher has been collected, what is the code to identify it? 1, Wedza_01, MHR-04b are all valid. No spaces please.	NA	KEW_FK_42
next_census_check	chr	If any measured values seem wrong or, note checks to be performed during the next census.	NA	Appears to have resurrected check mortality
fpc	num	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.	proportion	0.35
crown_area	num	Area of the tree crown, measured as an ellipse using either one or two crown diameter measurements, if present. In m^2	m^2	4.5
agb	num	Aboveground biomass of stem, in tonnes of dry matter	t DM	0.093
wood_density_mean	num	Mean estimated wood density of stem, in g cm^-3	g cm^-3	0.042
wood_density_sd	num	Standard deviation of estimated wood density of stem, in g cm^-3	g cm^-3	0.0004
wood_density_lev	chr	Level at which wood density was estimated	NA	species
wood_density_n	num	Number of individuals used to estimate wood density	NA	613
ba	num	Basal area of stem, in m^2	m^2	0.46
diam_adj	num	Adjusted stem diameter at 1.3 m height accounting for tree taper	cm	5.3
diam_inc	num	Diameter growth increment between current census and previous census.	cm	5.4
census_int	num	Census interval, in years, between current census and previous census	years	3
diam_inc_annual	num	Annual diameter growth increment between current census and previous census.	cm yr^-1	5.2
species_name_sanit	chr	Species name after automated taxonomy cleaning by sanitizeTaxa()	NA	Protea cf. petiolaris
species_name_clean	chr	Species name after taxize checks and synonymy corrections	NA	Burkea africana
genus_clean	chr	Cleaned genus name, taken from ‘species_name_clean‘	NA	Burkea
species_clean	chr	Cleaned species epithet, taken from ‘species_name_clean‘	NA	africana
subspecies_clean	chr	Cleaned subspecies epithet, taken from ‘species_name_clean‘	NA	africana

Table 2: Column names for stems table. *(continued)*

Column	Class	Description	Units	Example
confer_clean	chr	Cleaned cf. part of species epithet, taken from 'species_name_clean'	NA	polyantha
variety_clean	chr	Cleaned var. part of species epithet, taken from 'species_name_clean'	NA	micrantha
family_gnr	chr	Matched family name using taxize::gnr_resolve()	NA	Fabaceae
species_name_gnr	chr	Name matched during the GNR resolve	NA	Burkea africana
taxon_rank_gnr	chr	All rank names of classification from GNR resolve, separated by "|"	NA	|superkingdom|kingdom|phylum|subphylu
taxon_gnr	chr	All ranks of classification from GNR resolve, separated by "|"	NA	|Eukaryota|Viridiplantae|Streptophyta|Str makondensis
base_rank_gnr	chr	The lowest rank identified in GNR resolve	NA	species
species_name_cjb	chr	Species name matched with the CJB African Plants database	NA	Burkea africana
subdiv_id	chr	If the plot has been divided into smaller 1 ha subdivisions, the names of 1 ha plot subdivisions to which the stem belongs.	NA	ABG_1_S4

Table 3: Column names for small stems table.

Column	Class	Description	Units	Example
plot_id	chr	Standardised method of identifying plots. A three letter code, followed by a number referring to the particular plot. Three letter codes are normally, but not always, created like this: Angola (A), in Bicuar National Park (B), by John Godlee (G). The three letter code should be conserved within a dataset, i.e. a group of plots owned by a particular researcher within a particular locale.	NA	SSM_1
census_date	chr	The date the measurement was taken. This can be formatted as either YYYY, YYYY-MM, or YYYY-MM-DD	NA	05/05/2022
transect_id	chr	ID of transect	NA	T1
transect_length	num	Length of transects, in m to nearest 10 cm. If 'transect_id' is not provided, the total length of all transects in the plot.	m	100
transect_width	num	Width of belt transect, in m to nearest 10 cm	m	2
species_orig_binom	chr	The binomial name of the tree species, as provided by the data originator	NA	Burkea africana
species_orig_local	chr	The local name of the tree species/morphospecies, as provided by the data originator	NA	mumwei
n_alive	int	Number of living stems found for a given species	NA	50
n_dead	int	Number of dead stems found for a given species	NA	4
notes_small_stems	chr	Miscellaneous notes on each species of small stem	NA	Could be Combretum collinum

Table 4: Column names for coarse woody debris table.

Column	Class	Description	Units	Example
plot_id	chr	Standardised method of identifying plots. A three letter code, followed by a number referring to the particular plot. Three letter codes are normally, but not always, created like this: Angola (A), in Bicuar National Park (B), by John Godlee (G). The three letter code should be conserved within a dataset, i.e. a group of plots owned by a particular researcher within a particular locale.	NA	SSM_1
census_date	chr	The date the measurement was taken. This can be formatted as either YYYY, YYYY-MM, or YYYY-MM-DD	NA	05/05/2022
transect_id	chr	ID of transect	NA	NW
transect_length	num	Length of transects, in m to nearest 10 cm. If 'transect_id' is not provided, the total length of all transects in the plot.	m	100
piece_id	chr	ID of piece of coarse woody debris. A piece is a single contiguous bit.	NA	1
section_id	chr	ID of section of coarse woody debris. A section is part of a piece.		
tag_id	chr	The code of any tags attached to the piece in the field.	NA	BP094
species_orig_binom	chr	The binomial name of the tree species, as provided by the data originator	NA	Burkea africana
species_orig_local	chr	The local name of the tree species/morphospecies, as provided by the data originator	NA	mumwei
section_diam	num	Diameter of section of CWD at point of intersection with transect, in cm.	cm	6.8
section_length	num	Length of section of CWD, in m to nearest cm	m	4.5
decay	num	What level of decay is the stump? Scale runs from 1 (no decay) to 5 (very decayed). Follow the "Decay class" table in the SEOSAW protocol. To 0.5 of a decay level (e.g. 1.5, 1, 3.5)	NA	4.5
notes_cwd	chr	Miscellaneous notes related to the piece of CWD	NA	Approximately half of the trunk has broken off

Table 5: Column names for ground layer biomass table.

Column	Class	Description	Units	Example
plot_id	chr	Standardised method of identifying plots. A three letter code, followed by a number referring to the particular plot. Three letter codes are normally, but not always, created like this: Angola (A), in Bicuar National Park (B), by John Godlee (G). The three letter code should be conserved within a dataset, i.e. a group of plots owned by a particular researcher within a particular locale.	NA	SSM_1
census_date	chr	The date the measurement was taken. This can be formatted as either YYYY, YYYY-MM, or YYYY-MM-DD	NA	11/06/2022
quadrat_id	chr	Unique identifier for quadrats within a plot	NA	Q1.1
quadrat_area	num	Area covered by the quadrat.	m^2	0.25
sward_height	num	Height of the grass sward to nearest cm	cm	25
bare_cover	num	Proportional bare ground cover. Bare ground has no litter, just bare soil.	proportion	0.45
rock_cover	num	Proportional bare rock cover.	proportion	0.73
dung	log	Is there animal dung larger than a human fist inside the quadrat?	NA	TRUE
fire	log	Is there evidence of a previous fire inside the quadrat, e.g. charcoal.	NA	FALSE
under_canopy	log	Is the quadrat directly under a tree canopy?	NA	TRUE
canopy_cover	prop	The proportion of the ground (periscope, aerial photo, canopy dimensions), or sky hemisphere (hemi-photo, mirror densitometer), covered by a tree canopy. Measured as a proportion between 0 and 1, e.g. 60% = 0.6	proportion	0.28
canopy_cover_method	fct	Possible values are: "periscope_densitometer", "mirror_densitometer", "hemispherical_photo", "aerial_photo", "canopy_dimensions", "other"	NA	periscope_densitometer
lai_above	num	Leaf Area Index above the grass sward, measured with an AccuPAR Ceptometer.	m^2 m^-2	1.72
lai_below	num	Leaf Area Index below the grass sward, measured with an AccuPAR Ceptometer.	m^2 m^-2	4.55
litter_depth	num	Depth of leaf litter inside the quadrat, to nearest cm.	cm	12
grass_wet_mass	num	The wet mass of grass material collected inside the quadrat.	g	52.5
other_leaf_wet_mass	num	The wet mass of non-grass leaf material collected inside the quadrat.	g	25.6
wood_wet_mass	num	The wet mass of woody litter collected inside the quadrat.	g	42.8
litter_wet_mass	num	The wet mass of leaf litter collected inside the quadrat.	g	24.7
succ_wet_mass	num	The wet mass of succulent material collected inside the quadrat.	g	35.6
grass_dry_mass	num	The dry mass of grass material collected inside the quadrat.	g	5.2
other_leaf_dry_mass	num	The dry mass of non-grass leaf material collected inside the quadrat.	g	2.5
wood_dry_mass	num	The dry mass of woody litter collected inside the quadrat.	g	4.2

Table 5: Column names for ground layer biomass table. *(continued)*

Column	Class	Description	Units	Example
litter_dry_mass	num	The dry mass of leaf litter collected inside the quadrat.	g	2.4
succ_dry_mass	num	The dry mass of succulent material collected inside the quadrat.	g	3.5
notes_quadrat	chr	Miscellaneous notes on each quadrat	NA	Pig diggings inside quadrat
grass_moisture	num	Moisture content of grass	g	3.6
other_leaf_moisture	num	Moisture content of other leaves	g	2.8
wood_moisture	num	Moisture content of woody litter	g	2.6
litter_moisture	num	Moisture content of leaf litter	g	7.4
succ_moisture	num	Moisture content of succulent material	g	7.3

References

Chave, J. et al. (2014). 'Improved allometric models to estimate the aboveground biomass of tropical trees'. In: *Global Change Biology* 20.10, pp. 3177–3190. DOI: 10.1111/gcb.12629.

Dinerstein, Eric et al. (2017). 'An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm'. In: *BioScience* 67.6, pp. 534–545. DOI: 10.1093/biosci/bix014.

Kyalangalilwa, Bruce et al. (2013). 'Phylogenetic position and revised classification ofAcacia s.l.(Fabaceae: Mimosoideae) in Africa, including new combinations inVachelliaandSenegalia'. In: *Botanical Journal of the Linnean Society* 172.4, pp. 500–523. DOI: 10.1111/boj.12047.

White, F. (1983). *The Vegetation of Africa: A descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa.* Paris, France: UNESCO. DOI: http://dx.doi.org/10.2307/2260340.