Uganda Malaise trapping 2014–2015 background data

Tapani Hopkins, Heikki Roininen, Ilari E. Sääksjärvi
https://doi.org/10.5281/zenodo.2225643, please consider also citing the associated paper (see https://doi.org/10.5281/zenodo.2225643 for a link).

This dataset contains the background data to the Uganda Malaise trapping 2014 – 2015. We trapped flying insects for a year in tropical forest in Kibale National Park. The sampling itself is described separately, in the associated paper (see https://doi.org/10.5281/zenodo.2225643 for a link). This dataset contains background data such as the weather, vegetation around our traps, maps and descriptions of the trap sites, and data on the 876 insect samples. It also contains the scripts used to process and clean up the data.

Here, we describe the files in this dataset (section Files). We also describe some of the methods used to collect and process the data (section Methods).

The seven files in folder 1 Processed data will usually be of the greatest interest. They contain ready-to-use data on GPS tracks and waypoints, vegetation, weather, trap sites and samples.

Contents

Methods ↑

Malaise traps ↑

We trapped flying insects with 34 Malaise traps Sep 2014 – Sep 2015 in Kibale National Park, Uganda. The traps were in different forest types ranging from farmland outside the national park to undisturbed forest. We describe the study site and the trapping in our associated paper. More detailed information on the trap sites is in Uganda trap data.csv and 5 Trap pictures.

Vegetation data ↑

We collected vegetation data from all 34 trap sites. At each site, we noted what plant species were present in a 5 x 5 metre square centred on the trap (Fig. 1). We also documented the trees in two 50 x 2.5 metre transects (Fig. 1). We included trees of at least 5 cm diameter (at breast height), and also included dead trees and logs. If a tree was in both of the 50 x 2.5 metre transects we made sure it was counted only once.

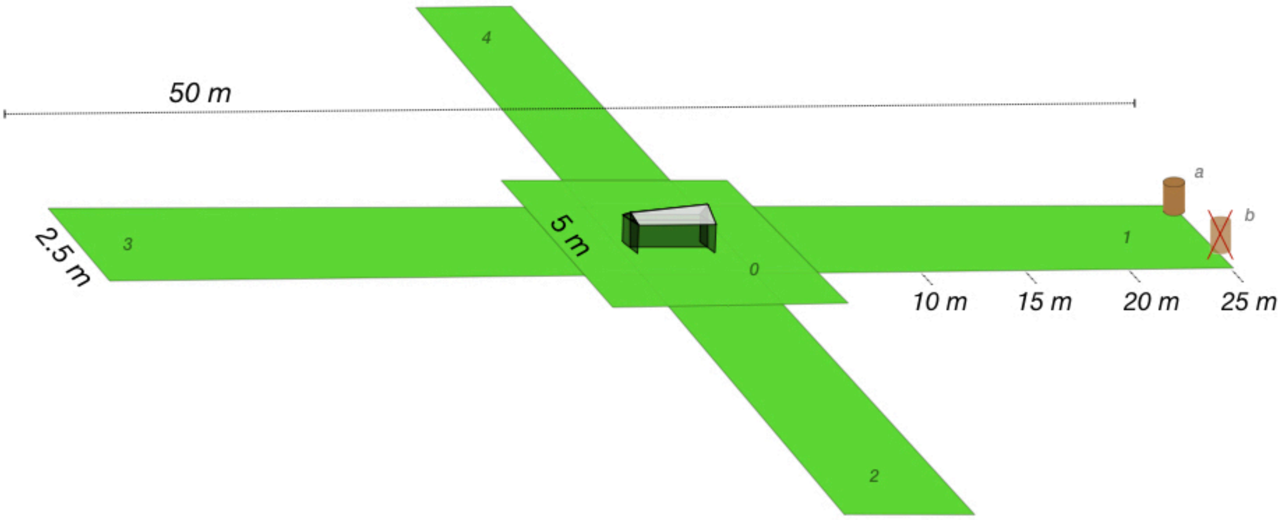

Figure 1. Vegetation transects at the 34 Malaise traps. We collected vegetation data in a 5 x 5 metre square (**0**), and two 50 x 2.5 metre transects (**1-4**). In the square we noted what plant species were present. In the two transects we documented the species, distance from the trap along the transect, and diameter at breast height of trees (dbh ≥ 5 cm). We also noted the direction from the trap (**1**, **2**, **3** or **4**). We included live trees whose centre line was inside the transect (e.g. tree **a,** but not tree **b**). We included all dead, fallen trees that overlapped the transect and measured their diameter near the middle.

At some of the trap sites, we stretched the tree transects to 100 x 2.5 metres. This data is available in the raw vegetation data, but any trees outside 50 x 2.5 metres have been pruned from the processed data. We felt that this was the easiest way to make the trap sites comparable.

Comparing our data on dead trees to that of other sampling programs may be tricky. This is because we included all dead, fallen trees that overlapped a transect. Other sampling programs will either have to use the same kind of transects as we did, or figure out how to estimate the density of dead wood (per surface area) in our sampling. With hindsight, collecting the vegetation data from a circle around the trap instead of transects would have been a good idea.

Number of insects from live and dead trees

↑

From the viewpoint of a Malaise trap, trees are not all the same. A large tree will support more insects in its canopy than a small one, and insects from a nearby tree are more likely to reach the trap. We would thus expect most of the (tree-living) insects caught in a trap to come from large nearby trees.

We estimated the contribution of each tree to the overall catch based on its size and inverse distance to the trap (Fig. 2). For each tree, we calculated dbh^2 / r or dbh^3 / r, and added these to the processed data (with both *dbh* and *r* in metres, see Uganda vegetation.csv). This is based on the assumption that insects live in the tree canopy, and are trapped when occasionally travelling along the ground surface. Insects that do not follow this assumption (e.g. leaf litter inhabitants) should be unaffected by tree sizes and distances.

We also calculated the total contribution of trees to the catch by summing dbh^2 / r and dbh^3 / r at each trap site. We did this separately for live and dead trees. This can be useful for detecting which species live in the tree canopy or in dead wood: we use it in our associated paper, for example, to show that traps with large amounts of dead wood nearby catch more rhyssine wasps. Note that these values cannot be directly compared to those of other sampling programs unless the transects were of the same shape and size.

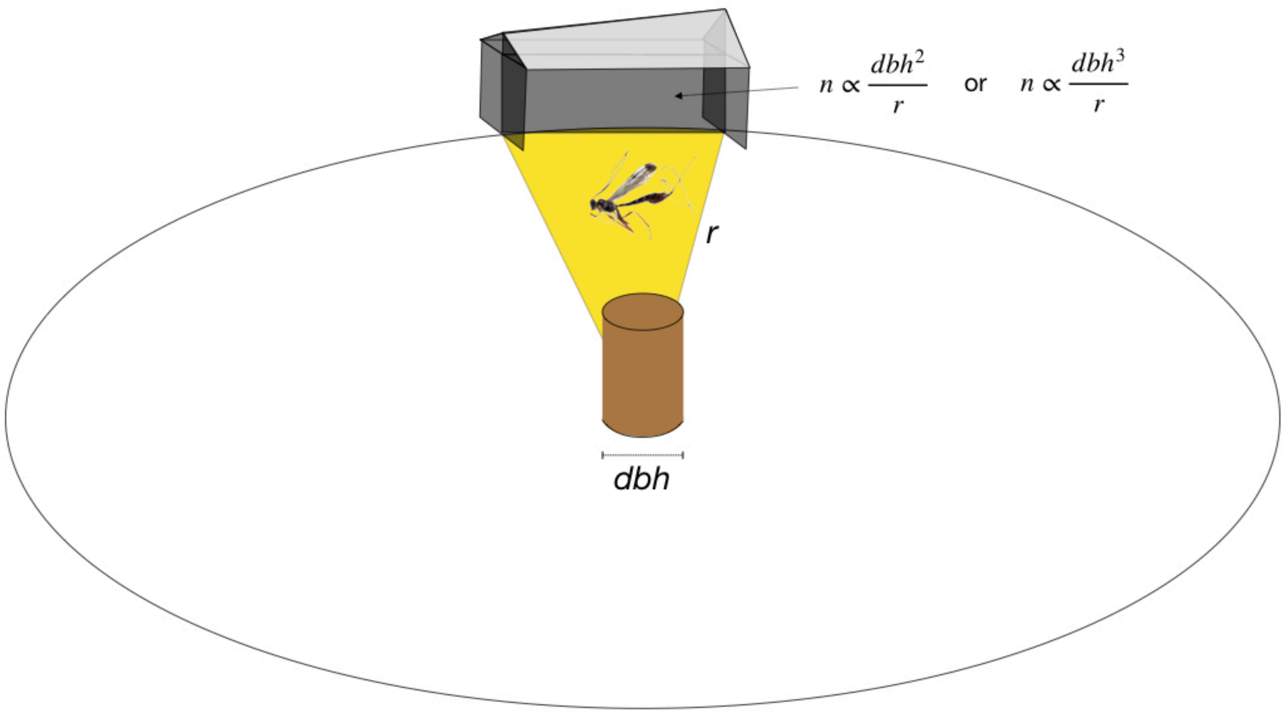

Figure 2. How we estimated the contribution of different trees to the insect catch in our Malaise traps. We assumed that the number of insects leaving a tree corresponds to its surface area (**dbh**^{2}) or volume (**dbh**^{3}). We also assumed that insects spread out evenly in all directions, so that the numbers arriving in a trap correspond to 1/**r**. The number of insects coming from a tree thus corresponds to **dbh**^{2} / **r** or **dbh**^{3} / **r**.

Forest types

↑

To classify the forest around the 34 Malaise traps into different forest types, we used a mix of known successional history and vegetation data. Some classifications were based solely on successional history (e.g. column *forest.2a* in Uganda trap data.csv), and are described in our associated paper. Here we describe classifications which also took into account the vegetation near the traps, such as column *forest.3* in Uganda trap data.csv.

To visualise how the vegetation of the trap sites differed, we created ordination plots of the sites. These were based on the vegetation in a 5 x 5 metre square (Fig. 3) and in two 50 x 2.5 metre transects (Fig. 4) centred on each trap. The plots placed trap sites that had similar vegetation (low Bray-Curtis distance) close to each other using non-metric multidimensional scaling. We used the R package vegan to calculate and draw the plots (see 3 Process raw data.R).

Based on the ordination plots, we modified a forest type classification based on successional history (Fig. 3a, Fig. 4a) by separating waterlogged primary forest from other primary forest (Fig. 3b, Fig. 4b). We also moved two traps into a different forest type: R93T1 into disturbed forest and R93T2 into primary forest. These two traps were in a former plantation area, clearcut in 1993, which is very close to disturbed and primary forest.

We also saved the positions of the trap sites in the two plots. We interpreted the two axes of Figure 3 as representing gradients from dark forest to grassy vegetation, and from high elevations to swampy ground (columns *grassy* and *swampy* in Uganda trap data.csv). We interpreted the axes of Figure 4 as representing gradients from large trees to shrubby vegetation, and from drier to swampy ground (columns *shrubby* and *swampy2* in Uganda trap data.csv). These variables are intended as an aid to visualising the vegetation; they may not be very useful in analyses.

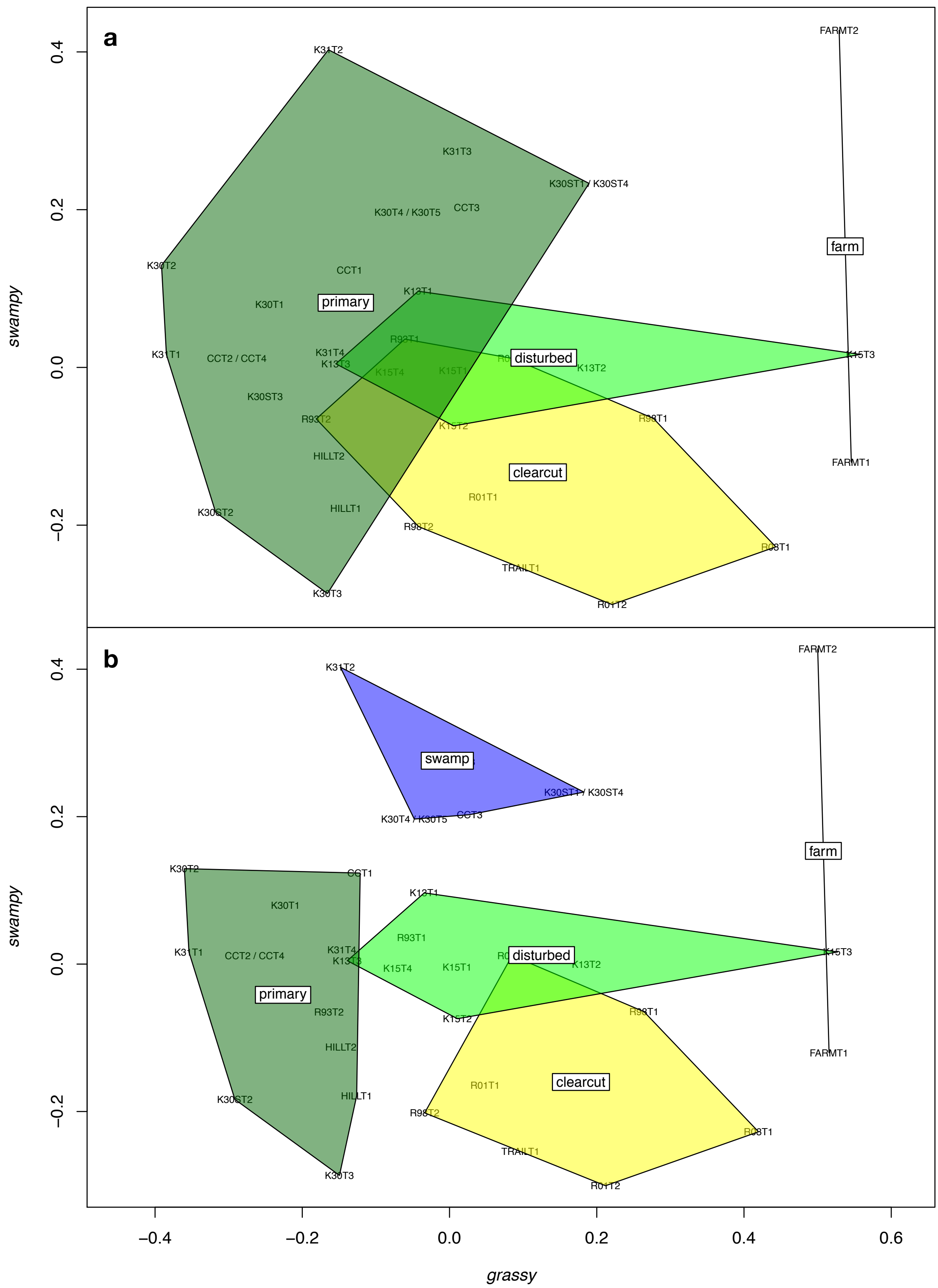

Figure 3. Visualisation of the vegetation surrounding the traps, used to get the forest type. Traps which are close together had similar vegetation in a 5 x 5 metre square centred on the trap (Bray-Curtis distances based on presence/absence data, non-metric multidimensional scaling). We modified the original forest classification which was based on known successional history (**a**) by separating swampy primary forest and moving trap R93T1 to disturbed forest and R93T2 to primary forest (**b**).

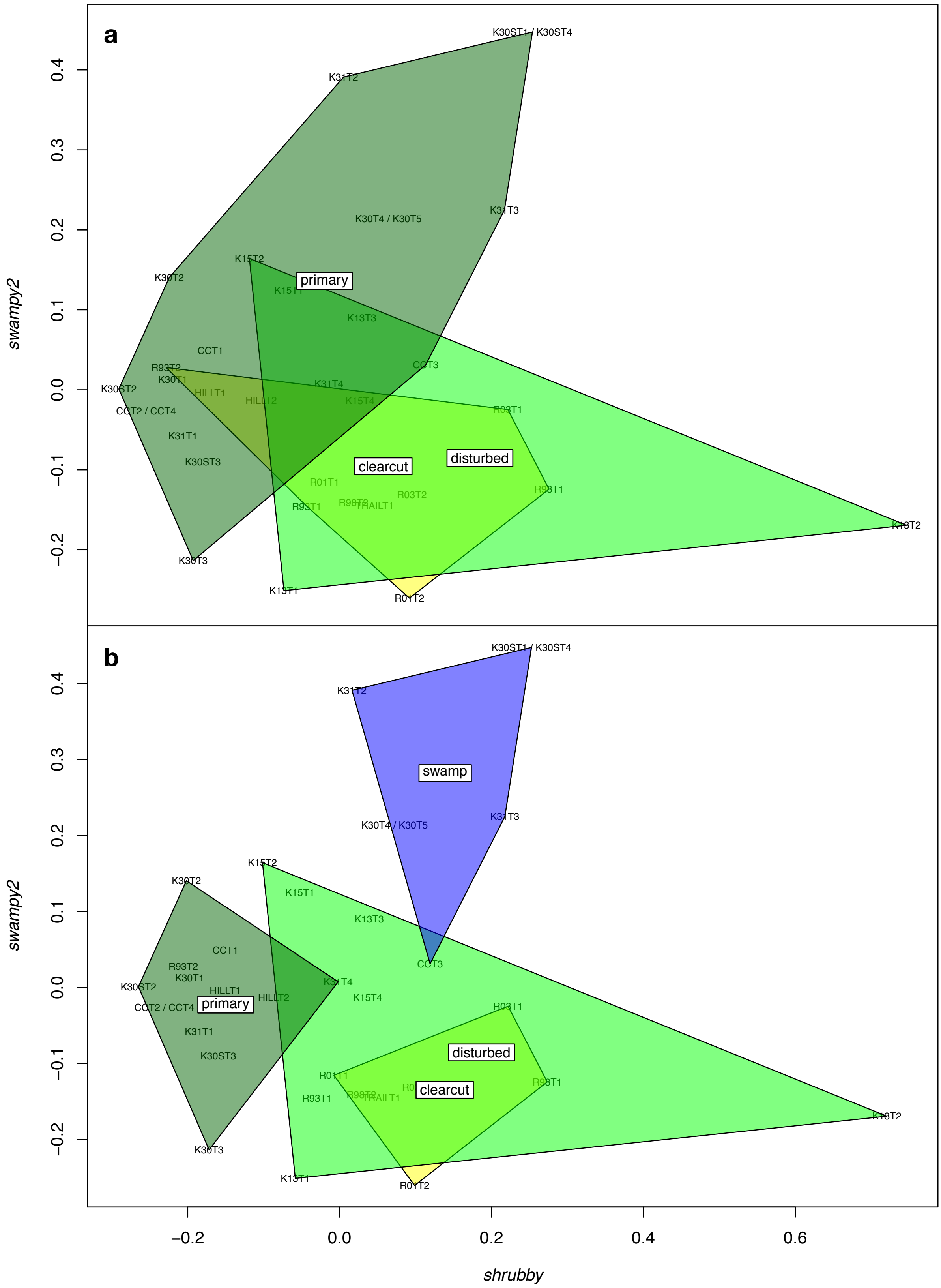

Figure 4. Visualisation of the vegetation surrounding the traps, used to get the forest type. Traps which are close together had similar vegetation in two 50 x 2.5 metre transects centred on the trap (Bray-Curtis distances based on tree counts, non-metric multidimensional scaling). We modified the original forest classification which was based on known successional history (**a**) by separating swampy primary forest and moving trap R93T1 to disturbed forest and R93T2 to primary forest (**b**).

Files ↑

0 ReadMe Uganda Malaise trapping 2014–2015 background data.pdf ↑

This file. Description of the files in the dataset "Uganda Malaise trapping 2014–2015 background data", and of the methods used to process them.

1 Processed data ↑

Folder containing the processed data. This is the main data of the dataset, and is ready to be used as it is. It contains GPS, vegetation, weather, trap and sample data from Kibale National Park 2014–2015, and has been created from the files in 2 Raw data by the R script 3 Process raw data.R.

Seven CSV files: Uganda GPS tracks.csv, Uganda GPS waypoints.csv, Uganda vegetation.csv, Uganda vegetation matrices.csv, Uganda weather data.csv, Uganda trap data.csv, Uganda sample data.csv.

Uganda GPS tracks.csv ↑

GPS tracks of our movements in Kibale National Park. These include trap emptying trips and general exploration of the forest. Collected by a Garmin GPS which was on most of the time, though there are several gaps where the batteries ran out or memory was full.

Same as the raw data in "GPS tracks 181130.csv" (2 Raw data).

CSV file with each track point on a separate row.

Columns:

track	Name of the track. Often automatically generated.
lat	Latitude in decimal degrees (WGS 84)
lon	Longitude in decimal degrees (WGS 84)
ele	Elevation in metres (WGS 84)
Y	Year (UTC)
M	Month (UTC)
D	Day (UTC)
H	Hours (UTC)
min	Minutes (UTC)
sec	Seconds (UTC)

Uganda GPS waypoints.csv ↑

GPS waypoints from Kibale National Park. These include trap sites, places were we hand-netted insects, and other places of interest. The notes are mostly what was written in Uganda, and are not guaranteed to be accurate or correctly spelt.

Same as the raw data in "GPS waypoints 181130.csv" (2 Raw data).

CSV file with each waypoint on a separate row.

Columns:

wpt	Name of the waypoint. Often just a number.
lat	Latitude in decimal degrees (WGS 84)
lon	Longitude in decimal degrees (WGS 84)
ele	Elevation in metres (WGS 84)
Y	Year (UTC)
M	Month (UTC)
D	Day (UTC)
H	Hours (UTC)
min	Minutes (UTC)
sec	Seconds (UTC)
notes	Notes on the waypoint

Uganda vegetation.csv ↑

Data on the vegetation surrounding 34 Malaise traps at Kibale National Park. These have been collected in a 5 x 5 metre square centred on each trap (presence / absence of plant species), and in two 50 x 2.5 metre transects centred on each trap (tree species of at least 5cm diameter at breast height, including dead logs of over 5 cm diameter).

Similar to the raw data in "Uganda vegetation 181130.csv" (2 Raw data), but with the addition of columns *.r* onwards and the data has been cleaned. Also, unidentified plants in the raw data (see columns *species*, *unidentified*) have been removed, and any 100 x 2.5 metre transects have been pruned to 50 x 2.5 metres.

CSV file with each plant species (5 x 5 m) or plant individual (50 x 2.5 m) on a separate row.

This data is used in Uganda vegetation matrices.csv, Uganda trap data.csv and Uganda sample data.csv.

Columns:

site	Name of the site
trap	Name of the Malaise trap
trap.heading	Direction in which the trap head, and one of the 50 x 2.5 m transects, was pointing. In degrees from north, rounded to the nearest ten degrees.

canopy.open	Percent of the sky visible from the ground (not covered by canopy). Measured with a spherical densiometer (Model-C, Robert E. Lemmon Forest Densiometers).
d	Day (UTC+3) on which the data was collected
m	Month (UTC+3) in which the data was collected
y	Year (UTC+3) in which the data was collected
date	Date when the data was collected in Finnish format d.m.Y (UTC+3)
transect	Transect name. One of 0 (5 x 5 m square), 1 (25 x 5 m in trap head direction), 2 (25 x 5 m 90° from trap head direction), 3 (25 x 5 m 180° from trap head direction), 4 (25 x 5 m 270° from trap head direction). See "Methods".
species	Name of the species. Anything in LARGE CAPS is not a formal species name. There are several names which do not match a known species name (we used a mix of Rutooro, English and Latin when identifying plants, and it sometimes shows). These have been marked with square brackets, e.g. [?]. Plants which couldn't be identified at all include the text UNIDENTIFIED in their name, these have been cleaned from the processed data. Plants were identified by Richard Sabiiti.
dist	Distance from the centre of the trap in metres. Rounded to the nearest metre.
dbh	Diameter at breast height in centimetres. For dead trees, diameter measured at approximately the middle of the log. Rounded to one decimal.
decay	Level of decay of dead logs. One of 1 (barely dead), 2 (light decay), 3 (clearly decaying). In practice, 1 and 2 were ambiguous and hard to classify objectively in the forest. It may be worth joining 1 and 2 in analyses.
notes	Notes on this species or plant individual. Often includes the length of the transect (distance from trap centre to end).
identified	TRUE if the plant has been identified to species or other taxon. FALSE otherwise. The identification does not need to be a formal species name. Useful for filtering out plants which could not be identified. Unidentified plants have been filtered from the processed data.
.r	Inverse distance to the trap centre, 1/*dist*. In units of 1/metres. Anything closer than one metre to the trap has been marked as 1. Useful in analyses for giving a greater weight (larger number) to plants that are near the trap.
dbh2	Diameter at breast height (in metres) squared, *dbh/100*dbh/100*. In units of m^2. Useful in analyses for giving a greater weight to large trees.
dbh3	Diameter at breast height (in metres) cubed, *dbh/100*dbh/100*dbh/100*. In units of m^3. Useful in analyses for giving a greater weight to large trees.
dbh2.r	Diameter at breast height squared divided by the distance to the trap, *dbh2*.r*. In units of metres. Useful in analyses for giving a greater weight to large trees near the trap.
dbh3.r	Diameter at breast height cubed divided by the distance to the trap, *dbh*3**.r*. In units of m^2. Useful in analyses for giving a greater weight to large trees near the trap.

Uganda vegetation matrices.csv ↑

Community matrices of the vegetation surrounding 34 Malaise traps at Kibale National Park. These are based on the data in "Uganda vegetation.csv". There is no new data here, this file is just for convenience to save the effort of creating community matrices from scratch. See "Uganda vegetation.csv" for more detailed information on the data.

CSV file with five community matrices, each with plant species in rows and traps in columns.

Columns:

species	Name of the plant species. Anything in LARGE CAPS is not a formal species name. There are several names which do not match a known species name (we used a mix of Rutooro, English and Latin when identifying plants, and it sometimes shows). These have been marked with square brackets, e.g. [?]. Plants were identified by Richard Sabiiti.
plant.matrix	Which community matrix this row belongs to. One of transect0 (5 x 5 m presence/absence data), live (live trees dbh≥5cm in two 50 x 2.5 m transects centred on each trap), or dead1/dead2/dead3 (dead trees of diameter≥5cm in two 50 x 2.5 m transects centred on each trap, decay level 1/2/3).
Remaining 34 columns	34 columns, one for each Malaise trap around which the vegetation was measured. The numbers are 1 or 0 (presence/absence) for transect0, and sum of *dbh2.r* (amount of wood, weighted so that large trees near the trap are favoured) for all other plant matrices.

Uganda weather data.csv

↑

Weather data from Kibale National Park. These have been collected daily at the Makerere University Biological Field Station. The station had two thermometers (one in the open, one under the tree canopy) and one rain gauge. Weather data has probably been collected at about 9 am (UTC+3) each day, but there is some uncertainty about this and it may have varied. Contains the weather from 1.8.2014 – 4.9.2015.

Similar to the raw data in "Uganda weather 181130.csv" (2 Raw data), but with the addition of columns *rain15* onwards.

CSV file with each day's weather on a separate row.

This data is used in Uganda sample data.csv.

Columns:

d	Day (UTC+3)
m	Month (UTC+3)
Y	Year (UTC+3)
date	Date in Finnish format d.m.Y (UTC+3)
rain	Rain during the past 24 hours. In millimetres, rounded to one decimal.
minC.open	Minimum temperature during the past 24 hours. In degrees Celsius, rounded to the nearest degree. There were two thermometers, one in the trees and one in the open, this is almost certainly from the open. The original data had mixed up columns *minC.open*, *maxC.open*, *minC.forest* and *maxC.forest* (see notes in the Uganda research diary, 16.9.2015).
maxC.open	Maximum temperature during the past 24 hours. In degrees Celsius, rounded to the nearest degree. There were two thermometers, one in the trees and one in the open, this is almost certainly from the open. The original data had mixed up columns *minC.open*, *maxC.open*, *minC.forest* and *maxC.forest* (see notes in the Uganda research diary, 16.9.2015).
minC.forest	Minimum temperature during the past 24 hours. In degrees Celsius, rounded to the nearest degree. There were two thermometers, one in the trees and one in the open, this is almost certainly from the trees. The original data had mixed up columns *minC.open*, *maxC.open*, *minC.forest* and *maxC.forest* (see notes in the Uganda research diary, 16.9.2015).

maxC.forest	Maximum temperature during the past 24 hours. In degrees Celsius, rounded to the nearest degree. There were two thermometers, one in the trees and one in the open, this is almost certainly from the trees. The original data had mixed up columns *minC.open*, *maxC.open*, *minC.forest* and *maxC.forest* (see notes in the Uganda research diary, 16.9.2015).
collector	Name of the person who collected the data
rain15	Smoothed rainfall in millimetres. Average rainfall during a time interval of 15 days centred on this day (i.e. this day ± one week). Days for which there is no weather data (before 1.8.2014 / after 4.9.2015) are not included in the interval.
max15.f	Smoothed maximum temperature in degrees Celsius. Average daily maximum temperature (*maxC.forest*) during a time interval of 15 days centred on this day (i.e. this day ± one week). Days for which there is no weather data (before 1.8.2014 / after 4.9.2015) are not included in the interval.
min15.f	Smoothed minimum temperature in degrees Celsius. Average daily minimum temperature (*minC.forest*) during a time interval of 15 days centred on this day (i.e. this day ± one week). Days for which there is no weather data (before 1.8.2014 / after 4.9.2015) are not included in the interval.
rainy	TRUE if it rained during the past 24 hours, FALSE if it did not. Very light rain (less than 0.05mm) counts as FALSE.
season	The season. One of dry (dry or less wet season) or wet (wet or rainy season). This is an estimate of when Kibale National Park generally has wet and dry seasons. The estimate matches that of Hartter et al. (2012, https://doi.org/10.1371/journal.pone.0032408) which is based on climate data.

Uganda trap data.csv

↑

Data on the 34 Malaise trap sites in Kibale National Park. This includes data on the trap's location, vegetation, forest type and the total sampling effort at each trap site. Also, see 5 Trap pictures for pictures of the trap sites.

Similar to the raw data in "Uganda trap data 181130.csv" (2 Raw data), but with the addition of columns *n* onwards.

CSV file with each trap on a separate row.

Columns:

trap	Name of the Malaise trap. Traps whose name contains a slash were moved by a few metres at some point (e.g. K30ST1 / K30ST4 was re-erected after being trampled by elephants).
site	Name of the site
lat	Latitude in decimal degrees (WGS 84)
lon	Longitude in decimal degrees (WGS 84)
ele	Elevation in metres (WGS 84)
notes	Description of the habitat near the trap. Includes a general description of the forest compartment or area, a longer description of the trap site, and the original description of the trap site written in Uganda (often full of typos and errors).
n	Number of living trees (diameter at breast height of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. See "Uganda vegetation.csv".
d2	Total diameter squared of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. See "Uganda vegetation.csv".

d3	Total diameter cubed of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. See "Uganda vegetation.csv".
d2r	Total diameter at breast height squared divided by the distance to the trap, of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. See "Uganda vegetation.csv".
d3r	Total diameter at breast height cubed divided by the distance to the trap, of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. See "Uganda vegetation.csv".
n.dead1	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 1. See "Uganda vegetation.csv".
n.dead2	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 2. See "Uganda vegetation.csv".
n.dead3	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 3. See "Uganda vegetation.csv".
n.dead	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. All levels of decay. See "Uganda vegetation.csv".
d2.dead1	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 1. See "Uganda vegetation.csv".
d2.dead2	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 2. See "Uganda vegetation.csv".
d2.dead3	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 3. See "Uganda vegetation.csv".
d2.dead	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. All levels of decay. See "Uganda vegetation.csv".
d3.dead1	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 1. See "Uganda vegetation.csv".
d3.dead2	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 2. See "Uganda vegetation.csv".
d3.dead3	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 3. See "Uganda vegetation.csv".
d3.dead	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. All levels of decay. See "Uganda vegetation.csv".
d2r.dead1	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 1. See "Uganda vegetation.csv".
d2r.dead2	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 2. See "Uganda vegetation.csv".
d2r.dead3	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 3. See "Uganda vegetation.csv".
d2r.dead	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. All levels of decay. See "Uganda vegetation.csv".

d3r.dead1	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 1. See "Uganda vegetation.csv".
d3r.dead2	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 2. See "Uganda vegetation.csv".
d3r.dead3	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 3. See "Uganda vegetation.csv".
d3r.dead	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. All levels of decay. See "Uganda vegetation.csv".
canopy.open	Percent of the sky visible from the ground (not covered by canopy). Measured with a spherical densiometer (Model-C, Robert E. Lemmon Forest Densiometers).
grassy	What kind of vegetation was found in a 5 x 5 m square centred on the trap. Smaller values are more closed canopy forest, larger values more grassy. Based on a NMDS ordination of presence/absence data in the 5 x 5 m square. See "Methods".
swampy	What kind of vegetation was found in a 5 x 5 m square centred on the trap. Smaller values are higher elevation, larger values are more swampy. Based on a NMDS ordination of presence/absence data in the 5 x 5 m square. See "Methods".
shrubby	What kind of trees (dbh ≥ 5 cm) were found in two 50 x 2.5 m transects centred on the trap. Smaller values are more full grown large trees, larger values more shrubby vegetation. Based on a NMDS ordination of tree counts in the transects. See "Methods".
swampy2	What kind of trees (dbh ≥ 5 cm) were found in two 50 x 2.5 m transects centred on the trap. Smaller values are less swampy forest (or denser trees), larger values more swampy forest (or sparse trees). Based on a NMDS ordination of tree counts in the transects. See "Methods".
forest.1	The forest type around the trap. One of hill (higher elevation & gravelly soil), slope (most common elevation & red clay), or water (swampy areas). Based on impressions gained in the forest. This forest classification is very subjective and may not work well in analyses.
forest.2a	The forest type around the trap. One of primary (barely or never logged), disturbed (40-50% of the trees cut approx. 1969), clearcut (former plantations clearcut 1993-2003), or farm (outside the forest in agricultural land). Based on the known successional history of the area the trap was in.
forest.2b	The forest type around the trap. One of forest (either unlogged or partially logged), clearcut (former plantations clearcut 1993-2003), or farm (outside the forest in agricultural land). Same as *forest.2a* except primary and disturbed forest has been combined. Based on the known successional history of the area the trap was in.
forest.3	The forest type around the trap. One of primary, swamp (waterlogged primary forest), disturbed, clearcut, or farm (outside the forest in agricultural land). Similar to *forest.2a*, but swamp has been separated from other primary forest and two traps are in a different forest type (R93T1 in disturbed and R93T2 in primary). Based on the vegetation near the trap and the known successional history of the area the trap was in. See "Methods".
forest.gradient.1	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut in 2003), 2 (former plantation clearcut in 2001), 3 (former plantation clearcut in 1998), 4 (former plantation clearcut in 1993), 5 (disturbed forest approx. 50% logged in 1968-69), 6 (disturbed forest approx. 40% logged in 1968-69), 7 (primary forest 2-3 trees / km^2 logged before 1970), or 8 (primary forest with no known logging). Based on the known successional history of the area the trap was in. Numbering the gradient 0–8 is subjective and may not work well in analyses.

forest.gradient.2a	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut in 2003), 2 (former plantation clearcut in 2001), 3 (former plantation clearcut in 1998), 5 (disturbed forest approx. 50% logged in 1969-69), 6 (disturbed forest approx. 40% logged in 1968-69), 7 (primary forest with minimal disturbance), or 8 (primary forest with no known logging). Similar to *forest.gradient.1*, but without the former plantation clearcut in 1993 since its two traps are in a different forest type (R93T1 in 5, R93T2 in 7). Based on the vegetation near the trap and the known successional history of the area the trap was in (see "Methods"). Numbering the gradient 0–8 is subjective and may not work well in analyses.
forest.gradient.2b	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut 1998-2003), 2 (disturbed forest), 3 (primary forest). Same as *forest.gradient.2a* except at a coarser resolution. Based on the vegetation near the trap and the known successional history of the area the trap was in (see "Methods"). Numbering the gradient 0–3 is subjective and may not work well in analyses.
tdiff	Sampling effort at the trap site. In trap days (the total number of days there was a Malaise trap at the site).

Uganda sample data.csv ↑

Data on 876 insect samples, collected by 34 Malaise traps September 2014 – September 2015 in Kibale National Park. This includes data on the trap site, the weather, and the date and time when the sample was collected. Also includes all the data in the Kotka Collection Management System (downloaded 14.9.2016, see e.g. http://mus.utu.fi/ZMUT.CCT1-141022).

Similar to the raw data in "Uganda sample data 181130.csv" (2 Raw data), but with the addition of columns *n* onwards.

CSV file with each sample on a separate row.

Columns:

MYNamespaceID	The first half of the sample's ID in the Kotka database. Always "utu:ZMUT" for samples of the Zoological Museum of the University of Turku.
MYObjectID	Name of the sample. The name consists of the trap name followed by the year, month and day.
MZDateEdited	The date and time when the sample was last edited in Kotka
MYCollectionID	Which museum collection the sample belongs to. Same for all the samples.
MYDatasetID.0.	Which Malaise trap the sample came from. In Kotka format (this has been made human-readable in column *trap*).
MYDatasetID.1.	Which site the sample came from. In Kotka format (this has been made human-readable in column *site*).
MYDatasetID.2.	Kotka dataset which the sample belongs to. Same for all the samples ("Uganda 2014-2015 Malaise samples", http://tun.fi/GX.1997).
MYDatasetID.3.	Kotka dataset which the sample belongs to. Same for all the samples ("Uganda 2014-2015", http://tun.fi/GX.1998).
MYDatasetID.4.	Kotka dataset which the sample belongs to. Blank unless the sample was damaged in some way. Damaged samples may have been trampled by an elephant, eaten by a tree mouse, or otherwise do not represent a normal catch. This has been made human-readable in column *damaged*.
MYStatus	"Ok" for all samples

MYNotes	Notes on the sample. The verbatim notes written in Uganda may contain errors.
MYVerificationStatus	"Ok" for all samples.
[MYSeparatedTo.X]	[35 columns] Kotka IDs of specimens that have been separated from the sample. Typically these are insect (e.g. ichneumonid) individuals. Valid as of 14.9.2016, and does not include recently separated specimens which do not yet have a Kotka ID.
MYPreservation	How the sample is preserved. All the samples are kept in approximately 80% ethanol.
MYLanguage	"english" for all samples
MYEditNotes	Blank for all samples.
MYOriginalSpecimenID	Name of the sample. The name consists of the trap name followed by the year, month and day.
MYGathering.0..MYLeg..0.	Who collected the sample. Same for all samples.
MYGathering.0..MYDateBegin.	Date when the sample started to be collected (UTC+3). In format dd.mm.YYYY
MYGathering.0..MYDateEnd.	Date when the sample stopped being collected (UTC+3). In format dd.mm.YYYY
MYGathering.0..MYDateVerbatim.	Time period when the sample was collected (UTC+3). In format "dd.mm.YYYY hh:(min)(min) - dd.mm.YYYY hh:(min)(min)". The hours and minutes are mainly based on GPS track data, but some are more rough estimates.
MYGathering.0..MYSamplingMethod.	How the sample was collected. Same for all samples.
MYGathering.0..MYLatitude.	Latitude in decimal degrees (WGS 84). Also to be found in column *lat*.
MYGathering.0..MYLongitude.	Longitude in decimal degrees (WGS 84). Also to be found in column *lon*.
MYGathering.0..MYCoordinateSystem.	Coordinate system. Same (WGS84) for all samples.
MYGathering.0..MYAlt.	Elevation in metres (WGS 84). Also to be found in column *ele*.
MYGathering.0..MYCoordinateSource.	How the coordinates were obtained. Same for all samples.
MYGathering.0..MYCountry.	Country where the sample was collected. Same for all samples.
MYGathering.0..MYAdministrativeProvince.	District where the sample was collected. Same for all samples.
MYGathering.0..MYMunicipality.	Municipality where the sample was collected. Same for all samples.
MYGathering.0..MYLocality.	Locality where the sample was collected. Same (Kibale National Park) for all samples, though the FARM samples were actually some distance outside.
MYGathering.0..MYLocalityDescription.	Description of where the trap was. In format "Site XX, Malaise trap XXXX, Kanyawara".
MYGathering.0..MYHabitatDescription.	Description of the habitat near the trap. Includes a general description of the forest compartment or area, a longer description of the trap site, and the original description of the trap site written in Uganda (often full of typos and errors). Same as the notes in Uganda trap data.csv.
MYGathering.0..MYCollectingEventName.	Name of the collecting event. Same for all samples.
MYGathering.0..MYUnit..0..MYRecordBasis.	What kind of record this is in the Kotka database (specimen, observation..). Same for all samples.
MYGathering.0..MYUnit..0..MYNotes.	Notes on the sample in the Kotka database. Same for all samples.

MYGathering.0..MYUnit..0..MYCount.	Number of individuals in the sample. Same (numerous) for all samples.
MYGathering.0..MYUnit..0..MYIdentification..0..MYTaxonRank.	What rank the sample has been identified to. Same (kingdom) for all samples, since the samples contain a diverse mix of animals.
MYGathering.0..MYUnit..0..MYIdentification..0..MYTaxon.	What taxon the sample belongs to. "Animalia" for all samples, since the samples contain a diverse mix of animals.
MYGathering.0..MYUnit..0..MYIdentification..0..MYAuthor.	Author of the taxon. Same for all samples.
MYGathering.0..MYUnit..0..MYIdentification..0..MYPreferredIdentification.	Kotka field used to specify which taxon identification is preferred. Same for all samples.
trap	Name of the Malaise trap the sample comes from. Traps whose name contains a slash were moved by a few metres at some point (e.g. K30ST1 / K30ST4 was re-erected after being trampled by elephants).
site	Name of the site
damaged	Blank unless the sample was damaged in some way. Damaged samples may have been trampled by an elephant, eaten by a tree mouse, or otherwise do not represent a normal catch.
lat	Latitude in decimal degrees (WGS 84)
lon	Longitude in decimal degrees (WGS 84)
ele	Elevation in metres (WGS 84)
n	Number of living trees (diameter at breast height of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. See "Uganda vegetation.csv".
d2	Total diameter squared of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. See "Uganda vegetation.csv".
d3	Total diameter cubed of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. See "Uganda vegetation.csv".
d2r	Total diameter at breast height squared divided by the distance to the trap, of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. See "Uganda vegetation.csv".
d3r	Total diameter at breast height cubed divided by the distance to the trap, of living trees (dbh ≥ 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. See "Uganda vegetation.csv".
n.dead1	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 1. See "Uganda vegetation.csv".
n.dead2	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 2. See "Uganda vegetation.csv".
n.dead3	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. Decay level 3. See "Uganda vegetation.csv".
n.dead	Number of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. All levels of decay. See "Uganda vegetation.csv".
d2.dead1	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 1. See "Uganda vegetation.csv".
d2.dead2	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 2. See "Uganda vegetation.csv".

d2.dead3	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 3. See "Uganda vegetation.csv".
d2.dead	Total diameter squared of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. All levels of decay. See "Uganda vegetation.csv".
d3.dead1	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 1. See "Uganda vegetation.csv".
d3.dead2	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 2. See "Uganda vegetation.csv".
d3.dead3	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. Decay level 3. See "Uganda vegetation.csv".
d3.dead	Total diameter cubed of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^3. All levels of decay. See "Uganda vegetation.csv".
d2r.dead1	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 1. See "Uganda vegetation.csv".
d2r.dead2	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 2. See "Uganda vegetation.csv".
d2r.dead3	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. Decay level 3. See "Uganda vegetation.csv".
d2r.dead	Total diameter squared divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of metres. All levels of decay. See "Uganda vegetation.csv".
d3r.dead1	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 1. See "Uganda vegetation.csv".
d3r.dead2	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 2. See "Uganda vegetation.csv".
d3r.dead3	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. Decay level 3. See "Uganda vegetation.csv".
d3r.dead	Total diameter cubed divided by the distance to the trap, of dead trees (diameter of at least 5 cm) in two 50 x 2.5 metre transects centred on the trap. In units of m^2. All levels of decay. See "Uganda vegetation.csv".
canopy.open	Percent of the sky visible from the ground (not covered by canopy) at the trap. Measured with a spherical densiometer (Model-C, Robert E. Lemmon Forest Densiometers).
grassy	What kind of vegetation was found in a 5 x 5 m square centred on the trap. Smaller values are more closed canopy forest, larger values more grassy. Based on a NMDS ordination of presence/absence data in the 5 x 5 m square. See "Methods".
swampy	What kind of vegetation was found in a 5 x 5 m square centred on the trap. Smaller values are higher elevation, larger values are more swampy. Based on a NMDS ordination of presence/absence data in the 5 x 5 m square. See "Methods".
shrubby	What kind of trees (dbh ≥ 5 cm) were found in two 50 x 2.5 m transects centred on the trap. Smaller values are more full grown large trees, larger values more shrubby vegetation. Based on a NMDS ordination of tree counts in the transects. See "Methods".

swampy2	What kind of trees (dbh ≥ 5 cm) were found in two 50 x 2.5 m transects centred on the trap. Smaller values are less swampy forest (or denser trees), larger values more swampy forest (or sparse trees). Based on a NMDS ordination of tree counts in the transects. See "Methods".
forest.1	The forest type around the trap. One of hill (higher elevation & gravelly soil), slope (most common elevation & red clay), or water (swampy areas). Based on impressions gained in the forest. This forest classification is very subjective and may not work well in analyses.
forest.2a	The forest type around the trap. One of primary (barely or never logged), disturbed (40-50% of the trees cut approx. 1969), clearcut (former plantations clearcut 1993-2003), or farm (outside the forest in agricultural land). Based on the known successional history of the area the trap was in.
forest.2b	The forest type around the trap. One of forest (either unlogged or partially logged), clearcut (former plantations clearcut 1993-2003), or farm (outside the forest in agricultural land). Same as *forest.2a* except primary and disturbed forest has been combined. Based on the known successional history of the area the trap was in.
forest.3	The forest type around the trap. One of primary, swamp (waterlogged primary forest), disturbed, clearcut, or farm (outside the forest in agricultural land). Similar to *forest.2a*, but swamp has been separated from other primary forest and two traps are in a different forest type (R93T1 in disturbed and R93T2 in primary). Based on the vegetation near the trap and the known successional history of the area the trap was in. See "Methods".
forest.gradient.1	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut in 2003), 2 (former plantation clearcut in 2001), 3 (former plantation clearcut in 1998), 4 (former plantation clearcut in 1993), 5 (disturbed forest approx. 50% logged in 1968-69), 6 (disturbed forest approx. 40% logged in 1968-69), 7 (primary forest 2-3 trees / km^2 logged before 1970), or 8 (primary forest with no known logging). Based on the known successional history of the area the trap was in. Numbering the gradient 0–8 is subjective and may not work well in analyses.
forest.gradient.2a	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut in 2003), 2 (former plantation clearcut in 2001), 3 (former plantation clearcut in 1998), 5 (disturbed forest approx. 50% logged in 1969-69), 6 (disturbed forest approx. 40% logged in 1968-69), 7 (primary forest with minimal disturbance), or 8 (primary forest with no known logging). Similar to *forest.gradient.1*, but without the former plantation clearcut in 1993 since its two traps are in a different forest type (R93T1 in 5, R93T2 in 7). Based on the vegetation near the trap and the known successional history of the area the trap was in (see "Methods"). Numbering the gradient 0–8 is subjective and may not work well in analyses.
forest.gradient.2b	The forest type around the trap as a successional gradient from farmland to primary forest. One of 0 (agricultural land outside the forest), 1 (former plantation clearcut 1998-2003), 2 (disturbed forest), 3 (primary forest). Same as *forest.gradient.2a* except at a coarser resolution. Based on the vegetation near the trap and the known successional history of the area the trap was in (see "Methods"). Numbering the gradient 0–3 is subjective and may not work well in analyses.
rain.sum	Amount of rain during the sampling period. In millimetres, rounded to one decimal. Based on weather measurements made at 24 hour intervals, see "Uganda weather data.csv".
rain.t	Number of rainy days during the sampling period. Based on weather measurements made at 24 hour intervals, and includes partial days if the sampling started or ended during a rainy day. Days with at least 0.05 mm of rain count as rainy. See "Uganda weather data.csv".
tdiff	Length of the sampling period in days
rain.mean	Average daily rainfall during the sampling period. In units of mm / day.
rain.t.mean	Average number of rainy days during the sampling period, *rain.t* / *tdiff*.

minC.mean	Average minimum temperature during the sampling period. In degrees Celsius. See column *minC.forest* in "Uganda weather data.csv".
maxC.mean	Average maximum temperature during the sampling period. In degrees Celsius. See column *maxC.forest* in "Uganda weather data.csv".
season	The season during which the sample was collected. Ranges from 0 (dry season) to 1 (wet season). Values in between are samples collected partly in the dry and partly in the wet season. See "Uganda weather data.csv".
d.start	Date when the sample started to be collected (UTC+3). In format YYYY-mm-dd.
t.start	Time when the sample started to be collected (UTC+3). In days from midnight [0-1]. This is usually based on GPS track data, but may be a more rough estimate.
d.end	Date when the sample stopped being collected (UTC+3). In format YYYY-mm-dd.
t.end	Time when the sample stopped being collected (UTC+3). In days from midnight [0-1]. This is usually based on GPS track data, but may be a more rough estimate.

2 Raw data ↑

Folder containing the raw data. This is the raw GPS, vegetation, weather, trap and sample data collected in Kibale National Park 2014–2015. It has been processed into the files in 1 Processed data by the R script 3 Process raw data.R.

Six CSV files. See the corresponding files in 1 Processed data for descriptions of the files.

3 Process raw data.R ↑

R script used to process the raw data. Reads the raw data files in 2 Raw data, cleans them up and adds new columns, and saves in 1 Processed data. To use the script, rename the working directory (and set *savefiles* to TRUE if you want to save the processed data). Much of the code is self-explanatory or is explained in the Methods.

The code is in R (https://www.R-project.org/), and also uses the package vegan (https://CRAN.R-project.org/package=vegan). Uses two functions in 4 R functions to get weather data.

4 R functions ↑

Two R functions used by 3 Process raw data.R. These make it easier to get the weather data for a specific time period, and are used to add weather data to each Malaise sample.

The code is in R (https://www.R-project.org/).

5 Trap pictures ↑

Pictures of the 34 Malaise traps used to collect insects at Kibale National Park. Each picture shows one trap site, and is named after the trap (with a prefix based on successional order). Also, see Uganda trap data.csv for descriptions of the trap sites.

6 Map

↑

Folder containing a map of the study site and the data used to draw it. The map shows the trap locations and forest types.

The map has been drawn with QGIS (https://www.qgis.org). The folder contains a qgs file for drawing the map (and a backup ~qgs file), a PDF version of the map, and a folder with the source data.

7 Uganda research diary 2014-2015 [cut version].rtf

↑

Research diary written during the 2014–2015 fieldwork in Kibale National Park. Contains information on where we walked in the forest, any problems encountered, and descriptions of methods. These may prove useful in tracking down the source of errors.

Much of this is verbatim as written in Uganda, and may contain errors. No guarantees on readability, especially for the first few weeks when I wrote at quite some length.

This is the cut version. Private information such as names or contact details has been deleted. Any deletions are marked with square brackets [] and the reason for the deletion is given.