Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A fine-grained multitrophic biodiversity and environmental dataset for understanding global change impacts in drylands

Ana Morales-González

moralesglezana@gmail.com

Estación Biológica de Doñana-CSIC

Walter R. Jubber

University of the Witwatersrand

Daniela Abegg

University of Zurich

Dominik M. Behr

University of Zurich

Tharina L. Bird

University of Pretoria

Christian Deschodt

University of Pretoria

Andrea Fuller

University of the Witwatersrand

Guillermo Gómez-Peña

Estación Biológica de Doñana-CSIC

Ken C. Graf

University of Zurich

Charles R. Haddad

University of the Free State

Sonja Huber

Estación Biológica de Doñana-CSIC

Marcin Jan Kamiński

Museum and Institute of Zoology-PAS

Masondo Mbongiseni

Kalahari Research Trust

Arpat Ozgul

University of Zurich

Candela Yáñez da Silva

Estación Biológica de Doñana-CSIC

Maria Paniw

Estación Biológica de Doñana-CSIC

---

data-descriptor

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11088411/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A fine-grained multitrophic biodiversity and environmental dataset for understanding global change impacts in drylands

Ana Morales-González[1,2 †], Walter R. Jubber[3,4 †], Daniela Abegg[5], Dominik M. Behr[5], Tharina L. Bird[6,7], Christian Deschodt[6], Andrea Fuller[3], Guillermo Gómez-Peña[1], Ken C. Graf[5], Charles R. Haddad[8], Sonja Huber[1], Marcin Jan Kamiński[9], Masondo Mbongiseni[4], Arpat Ozgul[5], Candela Yáñez da Silva[1], Maria Paniw[1]

[1] Department of Conservation Biology and Global Change, Estación Biológica de Doñana-CSIC, Sevilla, Spain.

[2] Institute of Nature Conservation, Polish Academy of Sciences, Kraków, Poland.

[3] Brain Function Research Group, Department of Physiology, Faculty of Health Sciences, University of the Witwatersrand, Parktown 2193, South Africa.

[4] Kalahari Research Trust, Kuruman River Reserve, Van Zylsrus 8467, Northern Cape, South Africa.

[5] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland.

[6] Department of Zoology and Entomology, University of Pretoria, P/B X28, Hatfield, South Africa.

[7] DITSONG: General Entomology, National Museum of Natural History, Pretoria 0002, South Africa.

[8] Department of Zoology & Entomology, University of the Free State, Bloemfontein 9301, South Africa.

[9] Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland.

†These authors share first authorship.

Corresponding author: Ana Morales González (moralesglezana@gmail.com) and Maria Paniw (maria.paniw@ebd.csic.es)

Abstract

Drylands harbour unique biodiversity that is increasingly threatened by climatic extremes and ongoing global change. Yet, biodiversity data, particularly for terrestrial soil invertebrates, remain extremely scarce. Here, we present a fine-grained multitrophic biodiversity and environmental database from the southern Kalahari, South Africa, comprising twelve interoperable datasets collected since 2023 across a study area of 7,294 ha spanning contrasting macrohabitats and land-use types. The core dataset contains abundance records for 394 above- and below-ground invertebrate species and morphospecies, collected through monthly live trapping. Complementary datasets include invertebrate biomass, vegetation composition, camera-trap records of 39 vertebrate species, more than 5.2 million soil temperature and moisture measurements, and spatial metadata. Shared spatial and temporal identifiers enable integrated analyses across trophic levels, environmental conditions, and spatial scales. The resource provides a growing baseline for investigating species activity patterns and biodiversity responses to land use, climatic variability, and global change in dryland ecosystems.

Background & Summary

Drylands cover more than 40% of the global landmass and support a high level of biodiversity[1,2]. Continued climate change, including increased temperature extremes and unpredictable rainfall patterns that result in more frequent drought-heatwave compound events[3], severely affects both above-ground[4–6] and below-ground species[7],

50 with these impacts expected to persist into the future. While vertebrate responses to
51 environmental temperature and moisture variation are increasingly studied, soil
52 invertebrates remain a key yet understudied component of biotic communities globally,
53 especially in drylands[8]. Terrestrial soil invertebrates spend all or part of their life cycle
54 below ground, and vary widely in body size from small-bodied mesofauna (often
55 defined as 0.2 mm to 2 mm in length) to macrofauna (greater than 2 mm in length)[9,10].
56 In dryland environments, certain macrofauna species, such as the darkling beetle
57 *Stridulomus sulcicollis* (Péringuey, 1885), can reach up to 80 mm in length and are
58 among the largest darkling beetles recorded worldwide (Koch, 1955)[11]. Soil
59 invertebrates respond strongly to fluctuations in soil moisture and temperature[12,13]. In
60 addition, in drylands, soil invertebrates provide essential ecosystem services and act as
61 soil engineers by physically, chemically and biologically modifying their
62 environment[14], primarily through activities such as burrowing, nest construction, and
63 recycling organic matter. These activities in turn accelerate litter decomposition,
64 improve soil structure through bioturbation (increasing soil pore space, improving
65 aeration and drainage), and support nutrient cycling[15–17]. Furthermore, soil invertebrates
66 are primary food sources for a number of dryland vertebrates, such as meerkats,
67 aardvarks, and pangolins, linking below-ground biodiversity to above-ground food
68 webs[18–20]. Therefore, in the face of environmental change, the need to assess
69 spatiotemporal variation in soil fauna biodiversity and how soil fauna is linked to
70 above-ground biodiversity is increasingly recognised[21,22].
71 To understand dryland biodiversity responses to environmental change requires
72 datasets that integrate multiple trophic levels, capture both above-ground and below-
73 ground communities, and record environmental conditions at fine temporal resolution.
74 Here we present a coordinated, multi-trophic biodiversity and environmental dataset

collected at the Kalahari Research Centre, Northern Cape, South Africa, a dryland system characterised by a mean annual rainfall of 299 mm (calculated from 2010 to 2026 using weather data collected on site), sandy nutrient-poor soils, and strong seasonal variation in temperature and moisture[23]. Data collection began in August 2023 and is ongoing, spanning a mosaic of land-use types, including natural veld, supporting only indigenous wildlife native to the southern Kalahari; integrated rotational grazing, in which cattle (bulk grazer) rotate alongside wild game; and mixed-species rotational grazing, in which small livestock (sheep, goats) and large livestock (cattle, horses, llamas) rotate alongside a mixed wild game community. These different management areas allow comparisons of biodiversity across contrasting anthropogenic disturbance gradients.

We integrated invertebrate, vertebrate, vegetation, and soil microclimate data within a single, spatially explicit framework that supports a range of research applications. The dataset can be used to examine how land-use type and seasonal variation in soil moisture and temperature influence the composition and biomass of both above-ground and below-ground invertebrate communities. In addition, the data can be used to explore how those patterns propagate to vertebrate communities that depend on invertebrates as prey, linking the database to the long-term research on the Kalahari meerkats at the same study site. Concurrent microclimate records enable analyses of biotic responses to temperature and moisture at ecologically relevant timescales. A subset of the data (November 2024 – March 2025) has been used to understand differences in soil fauna diversity caught in pitfalls vs subterranean traps[23]. As the time series extends, the dataset will additionally support assessments of interannual variability and longer-term trends in community composition, providing a baseline for evaluating future changes in dryland systems.

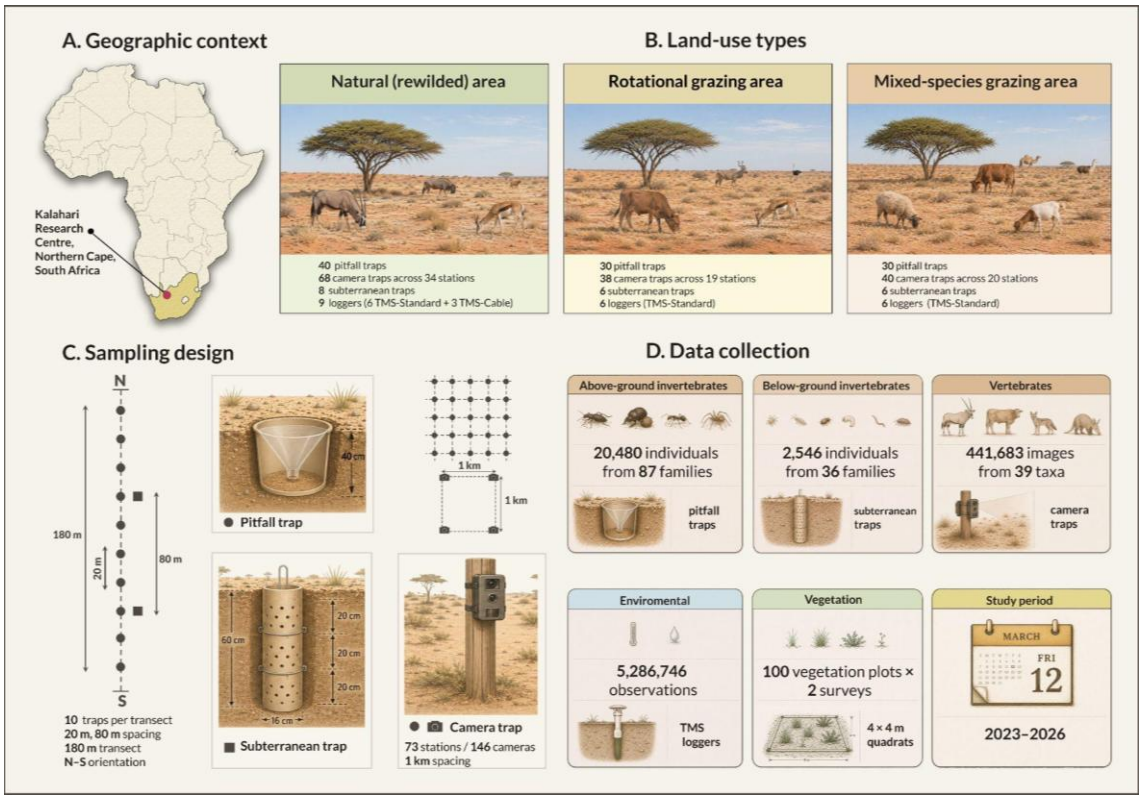


Figure 1. **Overview of the study design and data collection framework.** (**1**) Geographic location of the Kalahari Research Centre in the Northern Cape, South Africa. (**2**) The three land-use types included in the study (natural, rotational grazing, and mixed-species grazing) and the distribution of sampling devices within each. (**3**) Sampling design for pitfall traps, subterranean traps, vegetation surveys and camera traps. (**4**) Summary of the biodiversity and environmental datasets generated, including above- and below-ground invertebrates, vertebrates, soil microclimate, vegetation, and the study period.

Methods

Data management and workflows

We collected, processed, and managed all data following reproducible research practices. We processed all datasets using workflows developed in R version 4.2.3[24]. Raw data, cleaned datasets, metadata, and processing scripts are publicly available through open repositories (Table 1). The database comprises twelve complementary datasets that can be linked through shared identifiers (Fig. 2).

Table 1. Summary of the datasets included in the Kalahari multitrophic biodiversity and environmental database. The database comprises twelve complementary datasets describing invertebrate communities, vertebrate communities, vegetation composition, soil microclimate, macrohabitats, land-use types, and sampling locations collected at the Kalahari Research Centre, South Africa, between 2023 and 2026. For each dataset, we report the number of observations, taxa, and sampling units, together with the names of the files corresponding to raw data, cleaned data, and processing scripts as applicable, which are available at[25]. All datasets can be linked through shared identifiers to facilitate integrated analyses across trophic levels, environmental conditions, and spatial scales.

Dataset no.	Dataset name	Number of observations	Number of species/ morpho species	Number of sample units	File names for raw data, clean data and code (available at[25])
i	Invertebrate species composition	9,705	394	100 pitfall traps and 20 subterranean traps	'Invertebrate_diversity_FEB26.xlsx' (raw data) 'Dataset_i_clean.xlsx' (clean data) '01_clean_dataset_i_invertebrates.R' (cleaning code) '02_sum_dataset_i_invertebrates.R' (summary code)
ii	Invertebrate biomass for pitfall traps	15,708	—	100 pitfall traps	'stand_biomass' folder (raw data) 'Dataset_ii_clean.xlsx' (clean data) '04_clean_dataset_ii_invertebrates.R' (cleaning code) '06_sum_dataset_ii_invertebrates.R' (summary code)
iii	Invertebrate biomass for subterranean traps	605	—	20 subterranean traps	'Sub_biomass' folder (raw data) 'Dataset_iii_clean.xlsx' (clean data) '07_clean_dataset_iii_invertebrates.R' (cleaning code) '08_sum_dataset_iii_invertebrates.R' (summary code)
iv	Vegetation species composition	992	34	100 4x4 quadrats	'vegetation_data.xlsx' (raw data) 'Dataset_iv_clean.xlsx' (clean data) '09_clean_dataset_iv_vegetation.R' (cleaning code) '10_sum_dataset_iv_vegetation.R' (summary code)
v	Vertebrate species composition	441,683	39	71 camera trap stations*	'Dataset_v_clean.csv.zip' (clean data) '11_sum_dataset_v_vertebrates.R' (summary code)
vi	Deployment dates for loggers	21	—	21 loggers	logger_deployment.xlsx (raw data) Dataset_vi_clean.csv (clean data) '03_clean_dataset_vi&vii_soil.R' (cleaning code) '05_sum_dataset_vii_soil.R' (summary code)
vii	Soil temperature and moisture	5,286,746	—	21 loggers	'soil_data_1', 'soil_data_2', 'soil_data_3' folders (raw data) 'Dataset_vii_clean.csv.gz' (clean data) '03_clean_dataset_vi&vii_soil.R' (cleaning code) '05_sum_dataset_vii_soil.R' (summary code)
viii	Macrohabitat polygons	12,270	—	study area	'Dataset_viii_clean.gpkg' (clean data)' '12_sum_dataset_viii_macrohabitat.R' (summary code)
ix	Pitfall trap locations	100	—	study area	'Dataset_ix_clean.csv' (clean data)'
x	Subterranean trap locations	20	—	study area	'Dataset_x_clean.csv' (clean data)'

xi	Camera trap locations	71	—	study area	'Dataset_xi_clean.csv' (clean data)'
xii	Land-use type polygons	10	—	study area	'Dataset_xii_clean.gpkg' (clean data)

* Originally, 73 stations were deployed, each equipped with two camera traps. In the current version of the dataset, data from two stations were excluded because of camera trap malfunctions.

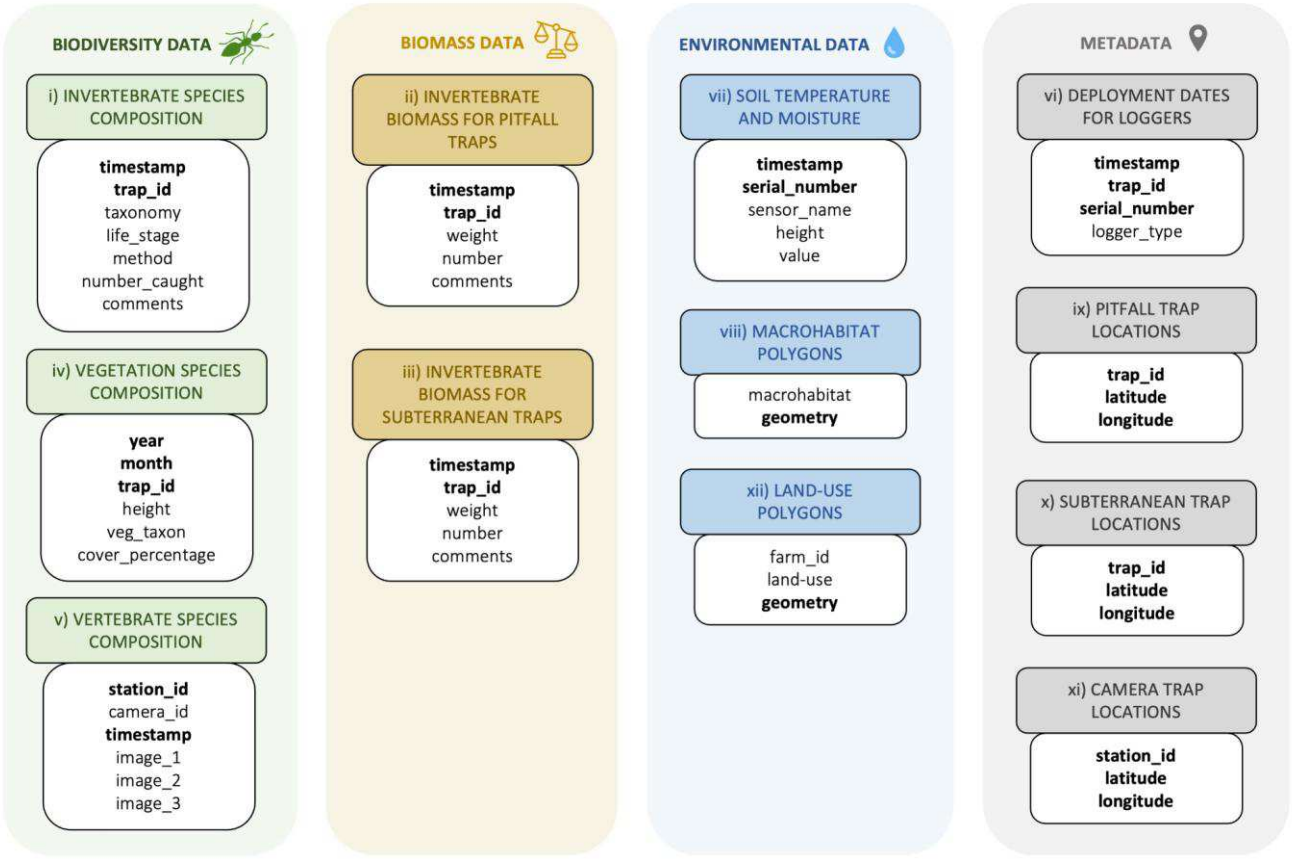


Figure 2. Relational structure of the Kalahari multitrophic biodiversity and environmental database. The twelve datasets are grouped into four thematic categories: biodiversity, biomass, environmental data, and metadata. Each box represents an individual dataset and depicts the variables included in that dataset. 'Taxonomy' represents multiple taxonomic fields (class, order, family, subfamily, tribe, genus, ands), which are recorded separately in the database but are summarised as a single term in the figure for simplicity. Variables shown in bold represent shared or linking variables that allow integration across datasets, including identifiers, temporal variables, spatial coordinates, and geometry, depending on the dataset. Variables shown in regular font are unique to each dataset. Complete variable definitions are provided in the section '*Data Records*' (Tables 2-12).

R packages

We used the R packages cowplot[26], dplyr[27], furrr[28], ggplot2[29], ggrepel[30], grid[24], hms[31], janitor[32], lubridate[33], magick[34], mgcv[35], myClim[36], patchwork[37], plotly[38], purrr[39], readr[40], readxl[41], skimr[42], stringdist[43], stringr[44], tidyr[45], tidyverse[46], and writexl[47].

Research site selection and basic site information

The study area is situated at the Kalahari Research Centre (KRC; -26.978615°, 21.832148°), a 7,294-hectare site in the southern Kalahari, Northern Cape, South Africa[18,48]. The area experiences hot summers, with rainfall typically from November to April (wet season), and cold winters from May to October (dry season). Mean annual rainfall averaged 299 mm from 2010 to 2026. In summer, the average daily air temperature is 30°C, with recorded maxima of 44°C. In winter, the average daily air temperature is 10°C, with recorded minima of -12°C[49,50]. Soils in the study area are typical of the southern Kalahari, characterised as moderately acidic to alkaline, nutrient-poor, and dominated by sand[51,52]. Macrohabitats include high and low duneveld found on red dune sands[53], white calcareous river terraces and floodplains, calcareous pan flats with sparse grass cover interspersed with short dwarf shrubs, and the diatomaceous fossil riverbed of the Kuruman River[54]. The contrast between the red dune sands of the duneveld and the white calcareous substrates of the pans, river terraces, and fossil riverbeds largely structures the main vegetation. The red duneveld is characterised by *Vachellia* savannah and open dune grassland, with woody species such as *Vachellia erioloba*, *V. haematoxylon*, *Boscia albitrunca*, *Senegalia mellifera*, and *Grewia flava*, and a grass layer dominated by *Schmidtia kalahariensis*, *Stipagrostis* spp., *Eragrostis lehmanniana*, and *Centropodia glauca*. The white calcareous macrohabitats support more open and patchy vegetation, including *Rhigozum trichotomum* shrubland, sparsely vegetated pans, and river-associated woodland. These areas are characterised by dwarf

shrubs such as *Justicia australis*, and riverbed trees and shrubs including *Vachellia erioloba*, *Senegalia mellifera*, and *Ziziphus mucronata*. Herbaceous and grass cover in the calcareous habitats is generally sparse and consists mainly of *Schmidtia kalahariensis*, *Stipagrostis obtusa*, *Sporobolus* spp., and a range of herbaceous forbs.

The study area is divided into three distinct land-use types (Fig. 1), each containing the main macrohabitats and vegetation: (1) Natural, spanning 3,220 hectares and supporting indigenous game species such as springbok (*Antidorcas marsupialis* (Zimmermann, 1780)), common duiker (*Sylvicapra grimmia* (Linnaeus, 1758)), eland (*Tragelaphus oryx* (Pallas, 1766)), blue wildebeest (*Connochaetus taurinus taurinus* (Burchell, 1823)), red hartebeest (*Alcelaphus buselaphus* (Pallas, 1766)), and oryx (*Oryx gazella* (Linnaeus, 1758)); (2) Rotational grazing farm area (cattle-game integration), covering 2,042 hectares and supporting both cattle and indigenous game species, including springbok, common duiker, kudu (*Tragelaphus strepsiceros* (Pallas, 1766)), and ostrich (*Struthio camelus* (Linnaeus, 1758)), managed through a rotational livestock farming approach[55,56]; and (3) Mixed-species grazing farm area (rotational grazing farm area with multiple livestock species-game integration), also 2,077 hectares, primarily used for mixed livestock farming with sheep, goats, camels, llama, and cattle, as well as various indigenous and non-indigenous game species: springbok, common duiker, blue wildebeest, lechwe (*Kobus leche* (Gray, 1850)), tsessebe (*Damaliscus lunatus lunatus* (Burchell, 1823)), roan (*Hippotragus equinus* (É. Geoffroy Saint-Hilaire, 1803)), and impala (*Aepyceros melampus melampus* (Lichtenstein, 1812)). The land-use areas are separated by game fencing (2.2 m to 2.4 m in height), including predator-proof fencing that extends one metre below ground and an additional 0.5 m above ground.

Experimental design

Invertebrate monitoring: Pitfall traps

Pitfall traps were used to sample above-ground invertebrates across the three land-use types. At each land-use type, three transect sites were randomly established along an elevational gradient, starting at a high-elevation red dune and extending downslope into calcareous soils. One additional transect site was established on a high dune in the Natural land-use type. All transects were oriented north to south, which is the dominant dune orientation at the study site; consistent orientation also reduced potential biases due to differences in slope position, aspect, and microclimatic conditions. Each transect site measured 180 m in length, with pitfall traps installed at 20 m intervals, resulting in 10 pitfall traps per transect site across 10 transect sites (n = 100). Each pitfall trap was installed by excavating a 400 mm-deep hole. The trap consisted of two-round plastic buckets (250 mm in diameter, 5 L volume), a base bucket embedded into the ground, and a second easily removable capture bucket, preventing soil collapse back into the hole. The capture bucket was placed flush with the soil surface within the base bucket, allowing unobstructed entry by surface-active invertebrates. Each pitfall trap was assigned a unique identifier, consisting of the transect site ID (O18, D15, H15, J13, J15, K18, L15, L17, P15, Q17) followed by the pitfall trap position within the transect (L01 - L10), e.g., Q17L01.

In each sampling session, approximately 20 mm of local substrate (sand) was added to the base of each container, with a moderate-sized piece of wood placed inside to provide shelter for captured organisms from direct sunlight. A transparent funnel with a 15 mm opening was inserted into each trap to facilitate the capture of invertebrates and reduce escape. Initial trap sampling was conducted over 48-hour sessions (from August 2023 to September 2023); however, due to logistical constraints, including

labour intensity, research assistant numbers, and fuel costs, the sampling regime was changed to 24-hour sessions (from October 2023 to January 2026). Transect sampling followed a repeated three-week cycle, five transect sites were sampled in the first week, the remaining five transect sites in the second week, followed by a rest week with no sampling. At the start of each session, the order in which transect sites were checked was randomised. Pitfall traps were opened (bucket lid removed), and invertebrates were collected from each trap at 4-hour intervals throughout the day, starting at 07:00, then 11:00, 15:00, and final collection at 19:00. The traps were left open overnight after sampling at 19:00. Final collection and closing (lid put back on) of the traps were made the next morning (around 07:00). For 48-hour sessions, the entire 24-hour cycle described was repeated once more following the same schedule. All collected specimens were transported to the laboratory, where they were identified and weighed, and then released at the site of their capture either in the cooler hours of the day for diurnal species or early evening for nocturnal species, at least 50 m from the trapping transect to avoid rapid recapture (see section '*Species identification and taxonomy*' for details). Information recorded from each sampling occasion is detailed in '*Dataset i*' and '*Dataset ii*'.

Predation in the pitfall traps was minimised by emptying the traps at regular intervals, and by providing a layer of sand that allowed trapped individuals to burrow into the sand. Initial descriptive statistics supported that we caught a variety of functional above-ground and soil-fauna types (e.g., predators, decomposers, herbivores). However, as predation could not be fully ruled out (which would bias our estimates of diversity), we implemented a mesh system (from 11 June 2025 onward) to separate small from large items in every second bucket along a transect. A wire mesh disk measuring 250 mm in diameter with an aperture size of 2 mm, designed to fit snugly in

the bucket, was installed in every even-numbered pitfall trap along a transect line (i.e. 2, 4, 6, 8, 10), whereas every odd-numbered pitfall was left without a mesh divider. The mesh disk allowed smaller invertebrates to reach the bottom of the pitfall trap, whereas larger invertebrates remained on the mesh disk. A piece of bark or another natural alternative was placed on the mesh disk to allow invertebrates to hide beneath and avoid desiccation. Although the mesh disk did not completely rule out predation by smaller invertebrates, it was successful in separating the larger predatory invertebrates (scorpions, ground beetles) from potential prey, which allowed us to measure the effect of predation on dry trapping. In the data, pitfall traps that incorporate mesh disks are indicated with a "D" at the end of the ID, e.g., Q17L02D.

Invertebrate monitoring: Subterranean traps

The subterranean traps were constructed from 160-mm-diameter polyvinyl chloride (PVC) pipes with a 4.5-mm wall thickness, each cut to 600 mm in length (Jubber et al., 2026). Handles were affixed to the top of each subterranean trap to assist with pulling the traps out of the sand. The base was sealed with a glued end cap to prevent invertebrates from escaping from the bottom. Two designs were used to compare sampling efficiency for hypogaeic species: the triple-stratified and the double-stratified subterranean traps. The triple-stratified subterranean trap was divided into three equal sections of 200 mm each, separated by metal tin plates supported by easy-to-remove steel wire pins. The upper part of each section (100 mm) was perforated with 14 mm diameter holes (24 to 28 holes) around the pipe circumference to allow invertebrate entry. The lower part of each section was not perforated and served as a capture area. The double-stratified trap consisted of two sections, each 300 mm long, separated by a single metal tin plate. The double-stratified design had the first 150 mm upper section of

each 300 mm section with perforation (same number of holes and diameter as the triple-stratified trap), followed by 150 mm of unperforated material below. An internal tube, as used in other tube-based subterranean traps, was not included because aeolian sands would likely pour into the tube cavity once the internal tube was removed.

Initial installation of the subterranean traps involved using a soil auger with a 250 mm diameter bit to excavate the holes. Each trap was inserted so that the open top (without end-cap or lid) was flush with the soil surface, allowing surface invertebrates to enter the subterranean trap (see [21]). The stratified sections of the trap were refilled with excavated sand (which was not defaunated) from the same excavation site up to ground level. The traps remained in place for one month to allow soil invertebrates to naturally colonise the traps before retrieval. After one month, the traps were removed from the soil, and sand from each level was sampled separately, aided by the internal metal divider plates that kept the sections distinct. The traps were then reinstalled at the exact same location (with the aid of a spade to reopen the hole if needed) and refilled with sand from the original site. The collected sand from each section was transported to a field laboratory and sifted through a standard kitchen sieve with a 0.2 mm aperture to extract predominantly macrofauna larger than the sieve openings.

We assigned a unique identifier to each subterranean trap. The identifier follows the format SQ17L05L01, starting with S to indicate a subterranean trap, followed by the transect site ID (O18, D15, H15, J13, J15, K18, L15, L17, P15, Q17), the stratification level of the subterranean trap (L01 for double-stratified, L05 for three-stratified), and the specific stratum where the specimen was collected (L01 - L02 for double-stratified, L01 - L03 for three-stratified).

The sampling protocol was implemented in two stages between February and October 2024. Initially, a single triple-stratified subterranean trap was deployed at three

sites (J13, J15, and H15) within the Natural area as a pilot study from February to August 2024. In August 2024, the sampling effort at these three sites was expanded by adding a double-stratified subterranean trap, totalling six subterranean traps across the original three sites. Both trap designs were also deployed at three new sites: one in the Rotational grazing area (K18) and two in the Mixed-species grazing area (P15, Q17), resulting in six sampled sites. In October 2024, four additional sites were added (Natural - D15, Rotational grazing - L15 and L17, and Mixed-species grazing - O18), bringing the total to 10 sites across three land-use types. We continued our sampling efforts at all 10 sites until January 2026. At each transect site, the double- and triple-stratified subterranean traps were positioned along a second north-south transect line, 20 m east of the pitfall trap transect. The double-stratified trap was aligned with pitfall trap 1, and the triple-stratified trap was aligned with pitfall trap 5, giving 80 m of separation between the two subterranean traps to avoid spatial autocorrelation. Information recorded from each sampling occasion is detailed in '*Dataset i*' and '*Dataset iii*'.

Vegetation monitoring

We recorded the visually estimated percent cover of different plant species, as well as bare soil and dead vegetation, in a 4 × 4 m quadrat centred on each pitfall trap. Sampling was conducted twice, in June 2023 and February 2024. Information recorded from each sampling occasion is detailed in '*Dataset iv*'.

Mammal monitoring

Vertebrate communities were monitored using camera traps deployed in June 2023 at 73 sampling locations (hereafter "stations") spaced 1 km apart. Each station was

equipped with two camera traps. The current version of the camera-trap dataset covers June-November 2023. Subsequent records are currently being processed. In this version, data from two stations were excluded because of camera malfunctions.

Beginning in June 2023, a 1 km² grid of camera trap stations was established to assess the presence, movement, and utilisation of the three land-use types (Natural, Rotational grazing and Mixed-species grazing farm areas) by above-ground biodiversity (Fig. 1). Each station consisted of a treated wooden pole (1 m in length, 100 - 120 mm in diameter), with approximately 600 mm planted in the ground for stability and 400 mm extending above ground to facilitate the attachment of camera traps and, where necessary, Browning solar panels. Browning Recon Force Elite (5HP Ultra, 46MP) camera traps, equipped with 64GB SanDisk Extreme Pro 200MB/s SD cards (Ultra 3, Class 10), were deployed to capture images of species passing through the stations. Two camera traps were mounted at approximately 400 mm above ground level on each pole, facing slightly downward. Each camera trap was secured in a trail camera security box designed for Browning Recon Force models to protect the equipment. One camera trap faced north and the other south to record movement on both sides of the station and to minimise sun glare from the east and west, which could compromise image quality and the accuracy of both manual and AI-based species identification. In total, 73 stations (Natural = 34; Rotational = 19; Mixed-species = 20) and 146 camera traps were deployed across the study area, with stations spaced 1 km apart. Lithium-ion batteries powered the camera traps, and at the 34 stations in the Natural land-use type, Browning solar panels (BTC-SBP12, 178 mm x 152 mm x 51 mm) were connected to the camera traps as a pilot test. Two solar panels, one per camera trap, were secured atop the wooden pole above the camera traps, both solar panels facing north using 50 mm x 50

mm angle iron, and cabling was routed through plastic pipe or electrical trunking to protect against environmental and animal-related damage.

Each camera trap station was assigned a unique alphanumeric identifier corresponding to its position within the 1 km² sampling grid (e.g., E16, F12). Individual camera traps were identified using a camera-specific identifier. Camera traps were configured to capture photographs at 8 MP resolution, with a 5-second photo delay, a 5-shot rapid-fire upon motion detection at 0.3-second intervals, and blur reduction for the infrared flash. Each camera trap has a field of vision of 36 degrees and a capture detection range of 18 metres. Vegetation in front of the camera traps was trimmed using either a petrol trimmer or a hand slasher to reduce false triggers caused by wind movement. Each camera trap station was inspected monthly to check battery level, replace the SD card if necessary, and perform general maintenance on the camera trap and security box. Malfunctioning camera traps were identified and replaced, where possible, to minimise data loss. The area in front of the camera traps was also inspected to ensure vegetation remained sufficiently short to prevent unnecessary triggering. After collection, the images were downloaded from the SD cards, which were then cleared for reuse. The cleared SD cards were rotated back into use during subsequent maintenance visits. The processing and taxonomic classification of camera-trap images are described in '*Dataset v*'.

Soil monitoring

We deployed TOMST (https://tomst.com/web/en/systems/tms/tms-4/) TMS soil data loggers to measure soil and soil surface temperature and electrical conductivity at each pitfall trap transect. We used two types of TMS loggers, TMS-Standard and TMS-Cable loggers, each equipped with three temperature sensors (TMS_T1, TMS_T2,

TMS_T3) and one electrical conductivity sensor (TMS_moist), allowing us to record measurements at multiple soil depths and heights. TMS-Standard loggers recorded data every 15 minutes, measuring electrical conductivity in the upper soil layer (0 to 150 mm into the soil; TMS_moist) and temperature at −60 mm (TMS_T1, inside the soil), +20 mm (TMS_T2, near the soil surface), and +150 mm (TMS_T3, further away from the soil surface). TMS-Cable loggers also recorded data every 15 minutes, measuring electrical conductivity at −500 mm (TMS_moist) and temperature at −500 mm (TMS_T1 and TMS_T2) and +150 mm (TMS_T3). We installed two TMS-Standard loggers at each transect (except D15), placing one 1 m east of the first pitfall trap and the other 1 m east of the last pitfall trap. We installed one TMS-Cable logger at transects J13, J15, and H15, positioning it 0.5 m below ground between pitfall traps five and six. Manufacturer-stated specifications give a temperature accuracy of ±0.5 °C and a resolution of 0.0625 °C for the DS7505U+ sensors used in the TMS loggers[57]. Soil moisture is recorded as a unitless raw time-domain transmission signal, which can be interpreted as volumetric water content using a soil- or site-specific calibration curve [57].

Loggers were deployed in February 2024 (see '*Dataset vi*'), and we extracted the information recorded by each logger approximately every two months. We noted that the loggers began recording data once they left the factory. The procedure to remove information before deployment is detailed in *Technical Validation*. We downloaded the data using a laptop and a specialised adaptor. Each time we visited a transect, we checked whether the logger had been disturbed by animal activity. If the logger had been disturbed, for example, if an animal had exposed the logger by digging, we repositioned the logger and recorded any damage, as well as the logger ID. Information downloaded from each logger is detailed in '*Dataset vii*'.

392 **Invertebrate species identification and taxonomy**

393 Each invertebrate was identified to the lowest possible taxonomic level. If an
394 invertebrate was already part of the species catalogue at the Kalahari Research Centre, it
395 was released back onto the surface at the capture site. An individual from any new
396 species (for the study location) was collected, labelled, and preserved in ethanol. To
397 identify species, keys derived from the available literature were used[58–61], and relevant
398 experts were consulted (T.L., C.R., C, M.J. and see '*Acknowledgements*'). Some
399 invertebrates could not be identified; therefore, a representative of each collected
400 specimen was also preserved in sealable tubes containing 96% to 99% ethanol to
401 facilitate later DNA analysis. Voucher specimens were deposited in scientific natural
402 history collections. We employed a dry-capture design for both subterranean and pitfall
403 traps, in which invertebrates were not killed, and most captured specimens were
404 released (over 80%). Dry traps lessen soil invertebrate depletion, allowing for repeated
405 sampling, biomass estimation, and easier identification of species and morphospecies,
406 without negatively affecting local biodiversity[62,63].

407

408 **Dataset i: Invertebrate species composition**

409 We recorded the invertebrate community in each pitfall and subterranean trap on
410 each sampling occasion. For subterranean traps, we recorded the invertebrate
411 community per stratum. The sampling occasions spanned from August 2023 to April
412 2026 for pitfall traps and from February 2024 to January 2026 for subterranean traps.
413 Specifically, we identified individuals to the lowest possible taxonomic level and
414 recorded the number of individuals per taxon (class, order, family, subfamily, tribe,
415 genus, species) and the life stage. If no invertebrates were captured in a trap or stratum
416 during a sampling occasion, no record was created. New species were labelled as 'new

sp'. Specimens provisionally considered to belong to a new species, but whose identification remained uncertain and required confirmation, were labelled as 'cf. new sp'. Distinct species within the same genus that could not be assigned to described species were labelled as morphospecies 'sp. 1', 'sp. 2', etc. These specimens may also represent undescribed species.

Dataset ii: Invertebrate biomass for pitfall traps

We recorded the number of invertebrates and their collective total mass (using a Mettler balance) in each pitfall trap on each sampling occasion. Sampling occasions spanned from August 2023 to January 2026. If no invertebrates were captured in a trap during a sampling occasion, the number of individuals and total mass were recorded as zero.

Dataset iii: Invertebrate biomass for subterranean traps

We recorded the number of invertebrates and their collective total mass (using a Mettler balance) in every stratum of each subterranean trap on each sampling occasion. Sampling occasions spanned from March 2024 to January 2026, although sampling could not be conducted in some months (see '*Data Records*'). If no invertebrates were captured in a stratum during a sampling occasion, the number of individuals and total mass were recorded as zero.

Dataset iv: Vegetation species composition

We recorded the vegetation community around each pitfall trap on two sampling occasions: June 2023 and February 2024. Specifically, we measured mean vegetation

height (cm) and the percentage cover as the estimated proportion of ground surface occupied by bare soil, grass litter cover (i.e., cover of accumulated dead grass biomass lying on the ground surface, distinct from standing vegetation), and each standing vegetation taxon present within the quadrat, including grasses, shrubs, and trees. When identification to species level was not possible, plants were recorded at the finest taxonomic resolution achievable or classified into broader categories (unknown vegetation and unidentifiable stubs).

Dataset v: Vertebrate species composition

We recorded the vertebrate taxa captured by each camera trap at each sampling station from June to November 2023. This period corresponds to the data processed and included in the current version of the dataset. Camera trapping is ongoing.

The image-processing workflow began with filtering all images using MegaDetector v5a, an object-detection model that identifies animals, humans, and vehicles in camera trap images but does not classify animals to the species level[64]. Only images assigned a probability of at least 10% of containing an animal were retained for further processing; the low threshold was chosen to reduce the risk of discarding true detections. The retained images were then imported into TrapTagger, a web-based platform for AI-assisted camera trap processing (WildEye Conservation, https://wildeyeconservation.org/documentation/). Within TrapTagger, images captured within one minute of each other were automatically grouped into clusters, and clusters that potentially contained more than one species were flagged. For single-species clusters, only the first image in each cluster was annotated, and the resulting annotation was applied to all images within that cluster. Flagged clusters were reviewed and annotated image by image to account for multiple individuals of a species passing a

camera. The available AI species-identification models were not suitable for the dataset, so all mammal records were identified manually to the species level whenever possible. Small rodents and bats could not be reliably identified to species level from camera-trap images and were therefore recorded as Rodents and Bats, respectively. As the primary focus was on mammals, birds and reptiles were recorded only as Birds and Reptiles, respectively.

Dataset vi: Deployment dates for loggers

We recorded the date (day, month, and year) and time of deployment for each logger. We also documented any damage observed to the logger (e.g., breakage or exposure caused by animal digging) during transect sampling and visits conducted to download data from February 2024 to March 2026.

Dataset vii: Soil temperature and moisture

We downloaded soil temperature and electric conductivity measurements recorded by each logger at different depths and heights every two months, spanning from February 2024 to March 2026. For TMS-Standard loggers, temperature was recorded at 60 mm in the ground (–60 mm), +20 mm, and +150 mm above ground, and electric conductivity at 0 mm to −150 mm. For TMS-Cable loggers, temperature was recorded at +150 mm and –500 mm and electric conductivity at –500 mm. We subsequently converted raw electric conductivity measurements into ecologically relevant soil moisture values, i.e., volumetric water content (see section '*Technical validation*').

Dataset viii: Macrohabitat polygons

We delineated macrohabitat types across the entire study area using the vegetation classifications of Van Rooyen et al. (see [54]), with spatial boundaries derived from a macrohabitat map originally developed by Haverkamp (see [65]). We standardised habitat classes in QGIS, and each sampling transect site was assigned to one of five macrohabitat categories: river terrace, low duneveld, calcareous pan flats, high duneveld, and calcareous *Rhigozum* shrubland (Jubber et al., 2026). Descriptions of dominant vegetation and substrate characteristics of each category are described below.

River terrace. The macrohabitat category includes the fossil riverbed together with the calcareous river terraces and floodplains. The landscape is dominated by *Vachellia erioloba* and *Schmidtia kalahariensis.* The calcareous terraces occur along the Kuruman River, an ancient diatomaceous riverbed supporting open woodland and savanna. Dominant woody species include *Rhigozum trichotomum*, *Boscia albitrunca*, *Senegalia mellifera*, and occasional *Lycium* spp., while *Ziziphus mucronata* is characteristic of the calcareous riverbed. The herbaceous layer is dominated by *Schmidtia kalahariensis*, *Stipagrostis obtusa*, *Chloris virgata*, and *Eragrostis* cf. *porosa*, with *Tribulus zeyheri* and *Heliotropium lineare* also widespread.

Low duneveld. The low duneveld macrohabitat comprises mixed *Vachellia* (*Acacia*) savanna on red dune sands, forming ecotonal zones with varying density. Dominant woody species are *Vachellia erioloba, V. haematoxylon, Boscia albitrunca, Grewia flava*, and *Senegalia mellifera*. The herbaceous layer is dominated by *Schmidtia kalahariensis, Eragrostis lehmanniana, Centropodia glauca, Stipagrostis uniplumis*, and *Stipagrostis ciliata,* with *Bulbostylis hispidula* widespread. Dwarf shrubs, including *Pollichia campestris* and *Lasiosiphon polycephalus*, together with herbs such as

Crotalaria sphaerocarpa, Sesamum triphyllum, Limeum fenestratum and *Cleome kalahariensis,* add seasonal diversity.

Calcareous pan flats. The macrohabitat comprises calcrete pans and drainage lines on whitish, alkaline, silty soils, with sparse vegetation[52]. Grasses include *Sporobolus coromandelianus*, *S. rangei*, *Eragrostis echinochloidea*, and *Enneapogon desvauxii*. Dominant shrubs include *Justicia australis* and *Rhigozum trichotomum,* with *Tetraena* spp., *Salsola* spp., and *Acrotome inflata* also present.

High duneveld. The category comprises *Vachellia haematoxylon* savanna on red dune sands, occurring furthest from the Kuruman River. The herbaceous layer is dominated by *Stipagrostis amabilis*, with *Schmidtia kalahariensis*, *Stipagrostis ciliata*, *Centropodia glauca*, *Stipagrostis uniplumis,* and *Eragrostis lehmanniana* also common. Dwarf shrubs, including *Hermannia tomentosa* and *Plinthus sericeus,* creepers such as *Requienia sphaerosperma* and *Limeum arenicolum,* and prostrate cucurbits including *Citrullus lanatus* and *Acanthosicyos naudinianus* are locally abundant.

Calcareous *Rhigozum* shrubland. The macrohabitat comprises dense thickets of *Rhigozum trichotomum* on white calcareous sands, forming patches sufficiently extensive to be recognised as a distinct macrohabitat. These patches support *Schmidtia kalahariensis*, *Eragrostis lehmanniana*, *Aristida* spp., and *Tribulus zeyheri*. The expansion of *Rhigozum*, likely driven by overgrazing, results in dense, unpalatable thickets with low browse value despite the dominance of *Rhigozum*[54]

Dataset ix: Pitfall trap locations

We recorded the coordinates of each pitfall trap location in WGS84 (EPSG:4326) geographic coordinates, expressed in decimal degrees (latitude and longitude).

Dataset x: Subterranean trap locations

We recorded the coordinates of each subterranean trap location in WGS84 (EPSG:4326) geographic coordinates, expressed in decimal degrees (latitude and longitude).

Dataset xi: Camera trap locations

We recorded the coordinates of each camera trap location in WGS84 (EPSG:4326) geographic coordinates, expressed in decimal degrees (latitude and longitude).

Data Records

The data outputs consist of raw and clean data files for (i) invertebrate species composition, (ii) invertebrate biomass for pitfall traps, (iii) invertebrate biomass for subterranean traps, (iv) vegetation species composition, (v) vertebrate species composition, (vi) deployment dates for loggers, (vii) soil temperature and moisture, (viii) macrohabitat polygons, (ix) pitfall trap locations, (x) subterranean trap locations, (xi) camera trap locations, (xii) land-use type polygons. The final clean data files are available on Zenodo[25].

Data organisation and structure

Cleaned datasets are provided using the naming convention *Dataset_nr_clean*, where *nr* corresponds to the Roman dataset number in Table 1. Depending on dataset size and structure, files are distributed in either Excel (*.xlsx*; datasets i-iv), comma-separated (*.csv*; dataset vi), or compressed comma-separated (*.csv.gz*; dataset vii) formats. The datasets are structured to facilitate integration across the database through a set of shared or linking variables, while each dataset also contains variables specific to its contents (Fig. 2). Depending on the dataset, these linking variables include identifiers, temporal variables, spatial coordinates, and geometry. The code necessary to access the raw data and produce cleaned datasets are available on Zenodo (see Table 1). Detailed comments embedded within the scripts document validation procedures, identified issues, quality-control decisions, and data-processing steps. The section '*Technical Validation*' in the paper summarises these data cleaning procedures. The repository also contains the code used for data summarisation and figure generation. These scripts follow a numbered naming convention that reflects the processing workflow and dataset to which they apply. Scripts prefixed with *clean* perform data validation and cleaning procedures, whereas scripts prefixed with *sum* generate summary statistics and figures reported in the manuscript. Together, these scripts provide a complete and reproducible workflow from raw data to cleaned datasets and derived outputs.

Dataset i: Invertebrate species composition

The invertebrate species composition dataset reports the abundance of 394 potential species and morphospecies across 100 pitfall traps and 20 subterranean traps (Table 2). The total abundance was 23,025 individuals (including 20,480 individuals

584 from 87 families recorded in pitfall traps and 2,545 individuals from 36 families
585 recorded in subterranean traps), of which 100% were identified to order, 99% to family,
586 94% to genus, and 89% to species. Some specimens were not identified to species level,
587 either because identification was not possible or because the identification is still
588 pending. Moreover, some specimens are labelled with the qualifier *cf.* (Latin *confer*,
589 meaning 'compare with'), indicating that the specimen closely resembles the named
590 taxon but requires further comparison against reference material before the
591 identification can be confirmed. In total, we recorded 197 species, 193 genera, 94
592 families, 22 orders, and 5 classes. The mean number of species and morphospecies per
593 trap and sampling occasion (2.5 - 97.5 percentiles) was 21 (13 - 30) for pitfall traps and
594 4 (1 - 8) for subterranean traps. The most common species in pitfall traps (Fig. 3A) were
595 hotrod ants *Ocymyrmex barbiger* (Emery, 1886) (detected in all pitfalls, i.e., 100),
596 and the tortoise darkling beetles *Zophosis* (*Dignathosis*) *arcana* (Koch, 1958) (99
597 pitfalls) and *Z.* (*Heliophosis*) *subcordata* (Deyrolle, 1867) (99 pitfalls), whereas in
598 subterranean traps (Fig. 3B), they were the small darkling beetle *Calaharena*
599 *dutoiti* (Koch, 1963) (detected in 16 pitfalls out of 20) and termite *Allodontermes*
600 *schultzei* (Silvestri, 1908) (10 pitfalls). The most abundant species overall (Fig. 3C, D)
601 were the hotrod ants *O. barbiger* (4,917 individuals) and fierce ants *Tetramorium*
602 *solidum* (Emery, 1886) (4,148 individuals).
603
604 **Table 2. Data dictionary for the invertebrate species composition (dataset i).** The
605 dataset reports the abundance of 394 species and morphospecies across 120 pitfalls. For
606 each variable, we provide its description, type, range or levels, units, and type of
607 measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement

timestamp	Date and time when the invertebrates were collected from the pitfall / subterranean trap stratum	datetime	2023-08-02 11:25:00 – 2026-04-22 19:46:00	ISO 8601 (UTC)	recorded
trap_id	Pitfall ID / subterranean trap stratum ID where the invertebrates were collected.	character	e.g., Q17L01, SQ17L05L01		recorded
class	Taxonomic class of the invertebrates	character	e.g., Insecta		identified
order	Taxonomic order of the invertebrates	character	e.g., Hymenoptera		identified
family	Taxonomic family of the invertebrates	character	e.g., Formicidae		identified
subfamily	Taxonomic subfamily of the invertebrates	character	e.g., Formicinae		identified
tribe	Taxonomic tribe of the invertebrates	character	e.g., Camponotini		identified
genera	Taxonomic genera of the invertebrates	character	e.g., *Camponotus*		identified
species	Taxonomic specific epithet of the invertebrates	character	e.g., *fulvopilosus*		identified
life_stage	Life stage of the invertebrates	categorical	6 levels: adult, juvenile, larva, nymph, pupa, and egg sac		identified
method	Pitfall type	categorical	3 levels: pitfall, double-stratified subterranean, and three-stratified subterranean		recorded
number_caught	Number of individuals of a given taxon (at the finest taxonomic level identified) on a sampling occasion	integer	1 – 284		counted
comments	Notes or remarks related to the dataset record	character	e.g., poor condition		recorded

608

609

610

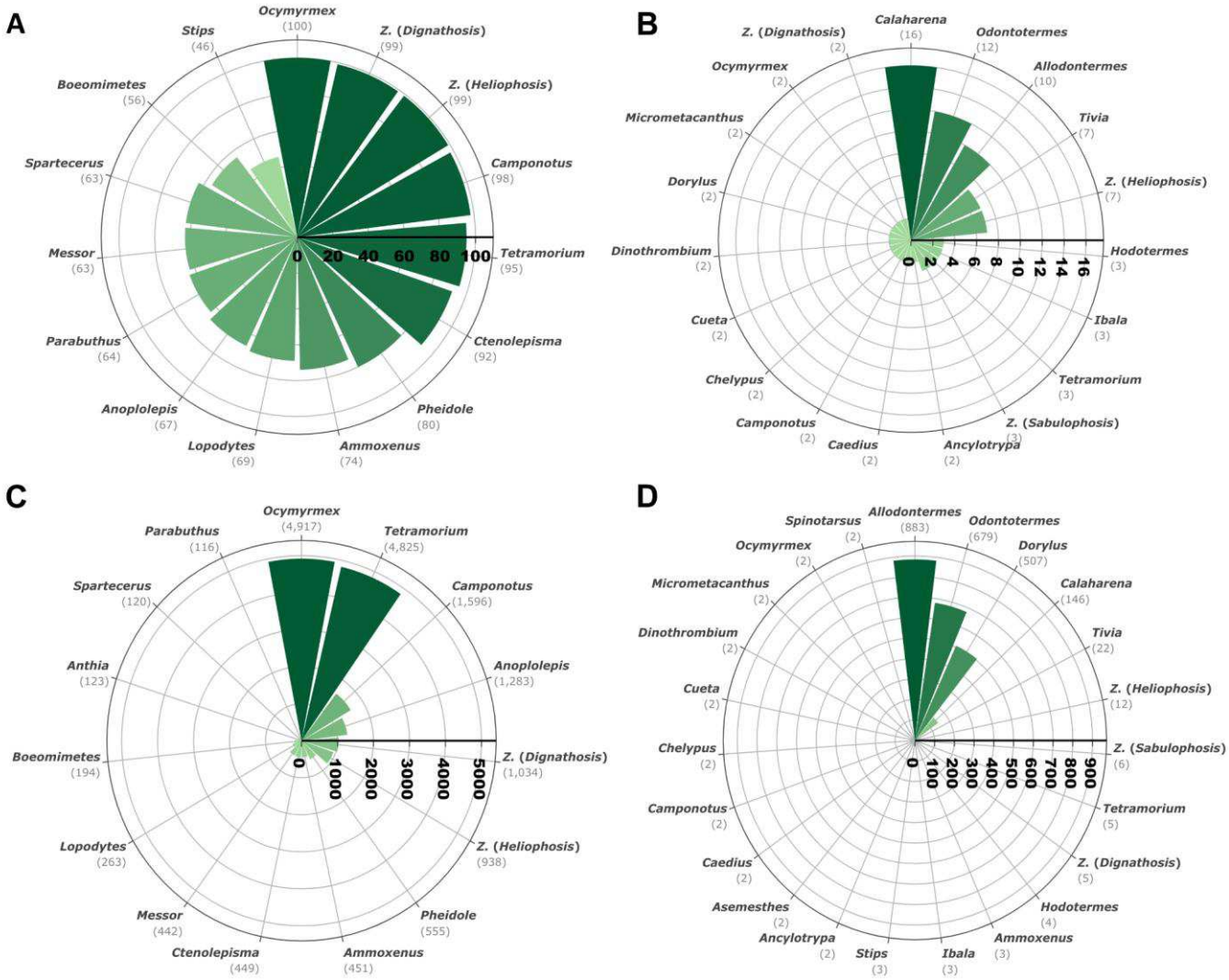


Figure 3. Most widespread and most abundant invertebrate genera in pitfall and subterranean traps. Genera ranked by (**A**) number of pitfall traps in which each genus was detected, (**B**) number of subterranean traps in which each genus was detected, (**C**) number of individuals in pitfall traps, and (**D**) number of individuals in subterranean traps. For all plots, the top 15 genera are shown, including ties at the 15th rank. Values are represented as radial bar plots, where bar length indicates the metric value and colour intensity increases with magnitude. The abbreviation "*Z.*" refers to the genus *Zophosis* Latreille, 1802.

Dataset ii: Invertebrate biomass for pitfall traps.

The invertebrate biomass dataset for pitfall traps reports trap-level abundance and total biomass from 15,708 sampling occasions across 100 pitfall traps (Table 3). In total, 49,396 individuals were recorded, corresponding to a collective biomass of 2,688 g. Empty traps, i.e., sampling occasions in which no invertebrates were captured, and both abundance and biomass were recorded as zero, represented 29% of all observations. The mean number of individuals per trap and sampling occasion (2.5 - 97.5 percentiles) was 3.3 (0 - 16 individuals), with a median of 2 individuals. The mean total biomass per trap and sampling occasion was 0.2 g (0 g –1.1 g), with a median of 0.006 g. Temporal variation in abundance and biomass across the study period is shown in Fig. 4.

Table 3. Data dictionary for the invertebrate biomass for pitfall traps (dataset ii). The dataset reports trap-level abundance and total biomass from 15,708 sampling occasions across 100 pitfall traps. For each variable, we provide its description, type, range or levels, units, and type of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
timestamp	Date and time when invertebrates were collected from the pitfall	datetime	2023-08-02 11:00:00 – 2026-01-14 10:33:00	ISO 8601 (UTC)	recorded
trap_id	Pitfall ID where the invertebrates were collected	character	e.g., Q17L01		recorded
weight	Collective total biomass of all invertebrates captured in the pitfall	numeric	0 – 69.6	g	measured
number	Total number of invertebrates captured in the pitfall	integer	0 – 405	individuals	counted
comments	Notes or remarks	character	e.g., one unknown spider		recorded

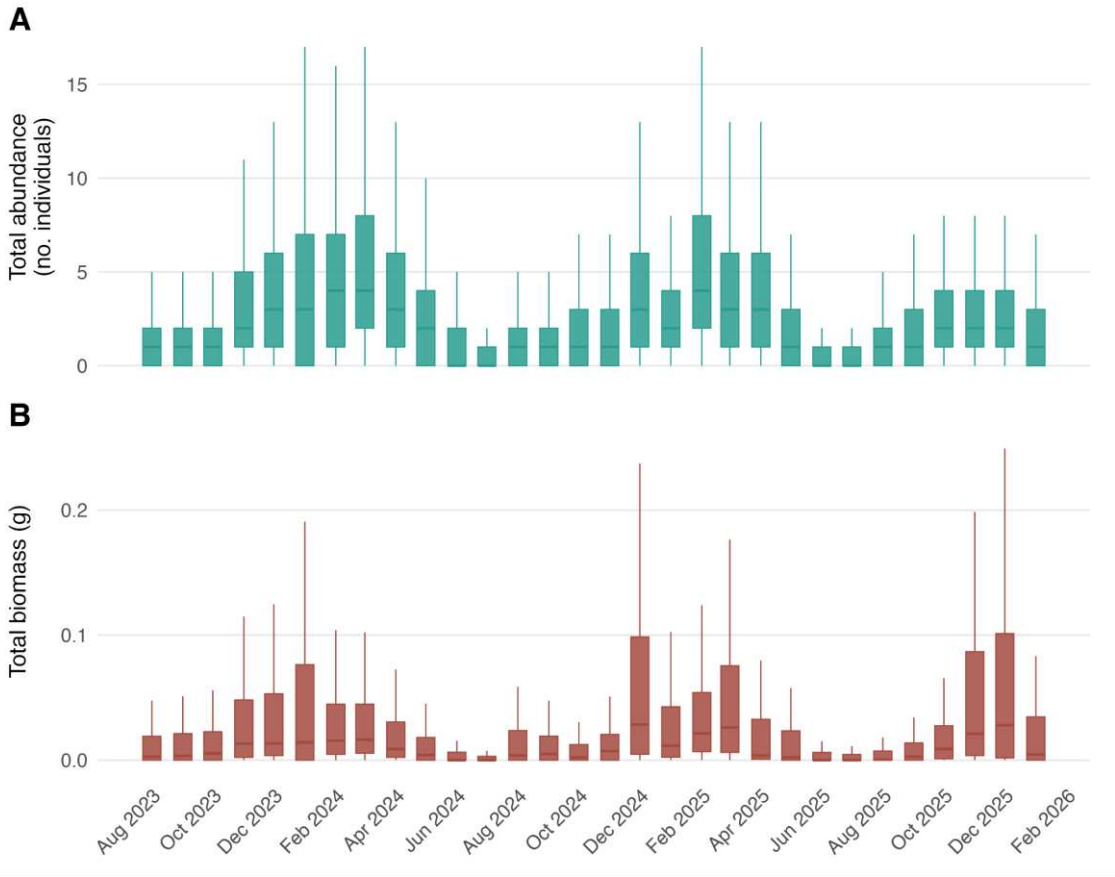


Figure 4. Temporal variation in invertebrate abundance and biomass in pitfall traps. Monthly distributions of (**A**) total abundance (number of individuals) and (**B**) total biomass (g) recorded per pitfall trap and sampling occasion across the study period.

Dataset iii: Invertebrate biomass for subterranean traps

The invertebrate biomass dataset for subterranean traps reports stratum-level abundance and total biomass from 605 sampling occasions across 50 subterranean trap strata, comprising 20 strata from 10 double-stratified and 30 strata from 10 triple-stratified subterranean traps (Table 4). In total, 2,290 individuals were recorded,

corresponding to a collective biomass of 30.6 g. Empty strata represented 57.2% of all sampling occasions. The mean number of individuals per stratum and sampling occasion (2.5-97.5 percentiles) was 3.8 (0-30.8 individuals), with a median of zero individuals. The mean total biomass per stratum and sampling occasion was 0.05 g (0 g-0.4 g), with a median of zero g. Temporal variation in abundance and biomass across the study period for double- and triple-stratified traps is shown in Fig. 5.

Table 4. Data dictionary for the invertebrate biomass for subterranean traps (dataset iii). The dataset reports stratum-level abundance and total biomass from 605 sampling occasions across 50 subterranean trap strata derived from 20 subterranean traps (10 double-stratified and 10 triple-stratified). For each variable, we provide its description, type, range or levels, units, and type of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
timestamp	Date and time when invertebrates were collected from the subterranean trap stratum	datetime	2024-03-28 00:00:00 – 2026-01-08 12:01:00	ISO 8601 (UTC)	recorded
trap_id	Subterranean trap stratum ID where the invertebrates were collected	character	e.g., SQ17L05L01		recorded
weight	Collective total biomass of all invertebrates captured in the subterranean trap stratum	numeric	0 – 3.29	g	measured
number	Total number of invertebrates captured in the subterranean trap stratum	integer	0 – 284	individuals	counted
comments	Notes or remarks	character	e.g., one burrowing skink		recorded

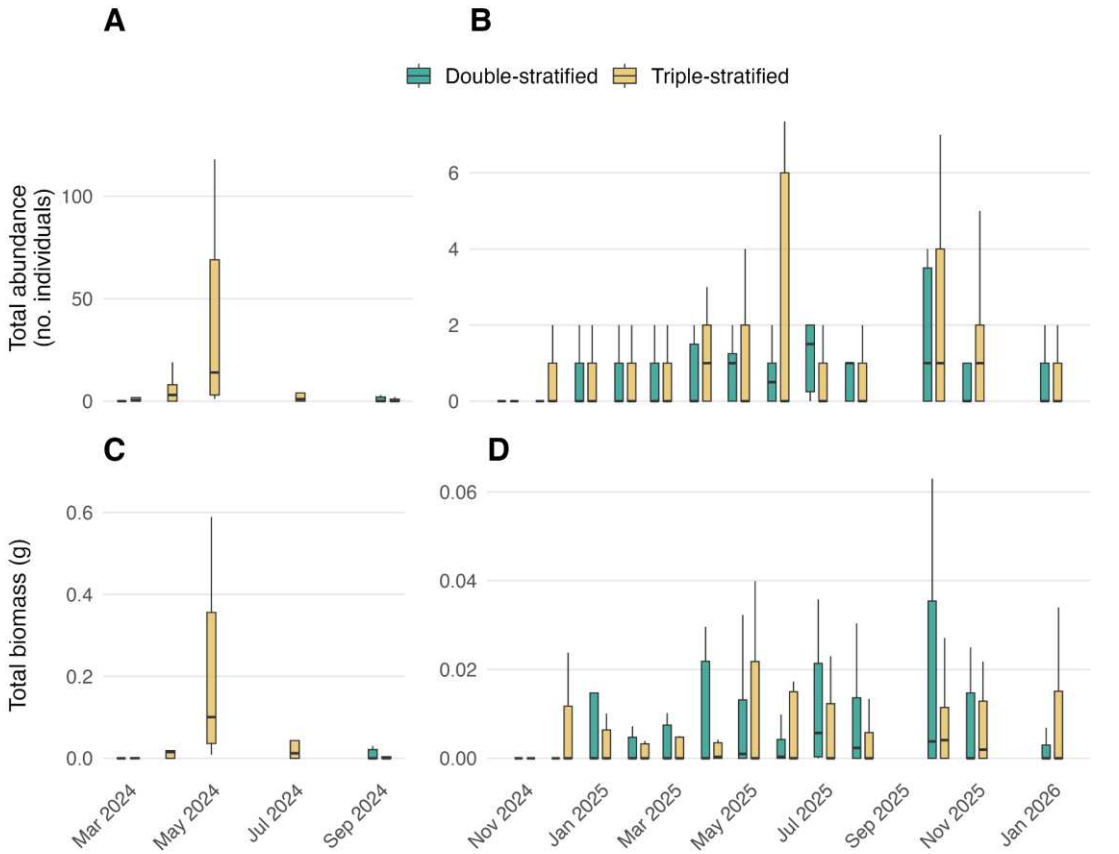


Figure 5. Temporal variation in invertebrate abundance and biomass in subterranean trap strata for double- and triple-stratified traps. Monthly distributions of total abundance (number of individuals) (**A, B**) and total biomass (g) (**C, D**) per subterranean trap stratum and sampling occasion across the study period, where **A** and **C** correspond to the initial deployment phase, during which subterranean traps were progressively established across sites, and **B** and **D** correspond to the full monitoring phase, when all 10 sites were sampled.

Dataset iv: Vegetation species composition

The vegetation species composition dataset reports the percentage cover of bare soil, dead grass, and vegetation around each of the 100 pitfall traps. Specifically, vegetation was surveyed within the 4 × 4 m quadrat centred on each of the 100 pitfall

traps, surveyed on two sampling occasions (June 2023 and February 2024; Table 5). Per sampling occasion, mean (2.5 – 97.5 percentiles) vegetation cover was 42.2% (3.9 – 90.3%), whereas mean bare soil and dead grass cover were 53.9% (10 – 96%) and 7.1% (0 – 40.4%), respectively. Across sampling events with vegetation present, mean vegetation height was 65.3 cm (20 cm –117 cm). In total, we recorded 34 vegetation taxa identified to species (32) or genus (2) level, one unidentified vegetation category (*unknown vegetation*), and one category representing unidentifiable plant remains (*unidentifiable stubs*). Across sampling events with vegetation present, the mean number of vegetation species recorded per quadrat was 3.22 (1 species – 6.07 species). The most common species were *Schmidtia kalahariensis* (detected in 84 quadrats), *Tribulus zeyheri* (57 quadrats), and *Rhigozum trichotomum* (36 quadrats). The species with the highest mean cover when present were *Stipagrostis amabilis* (29.1%, 0.5-71.1%), *Vachellia mellifera* (25.4%, 1-90%), and *Rhigozum trichotomum* (24.8%, 2.4-53%). Temporal variation in vegetation cover and composition across the two sampling occasions is shown in Fig. 6.

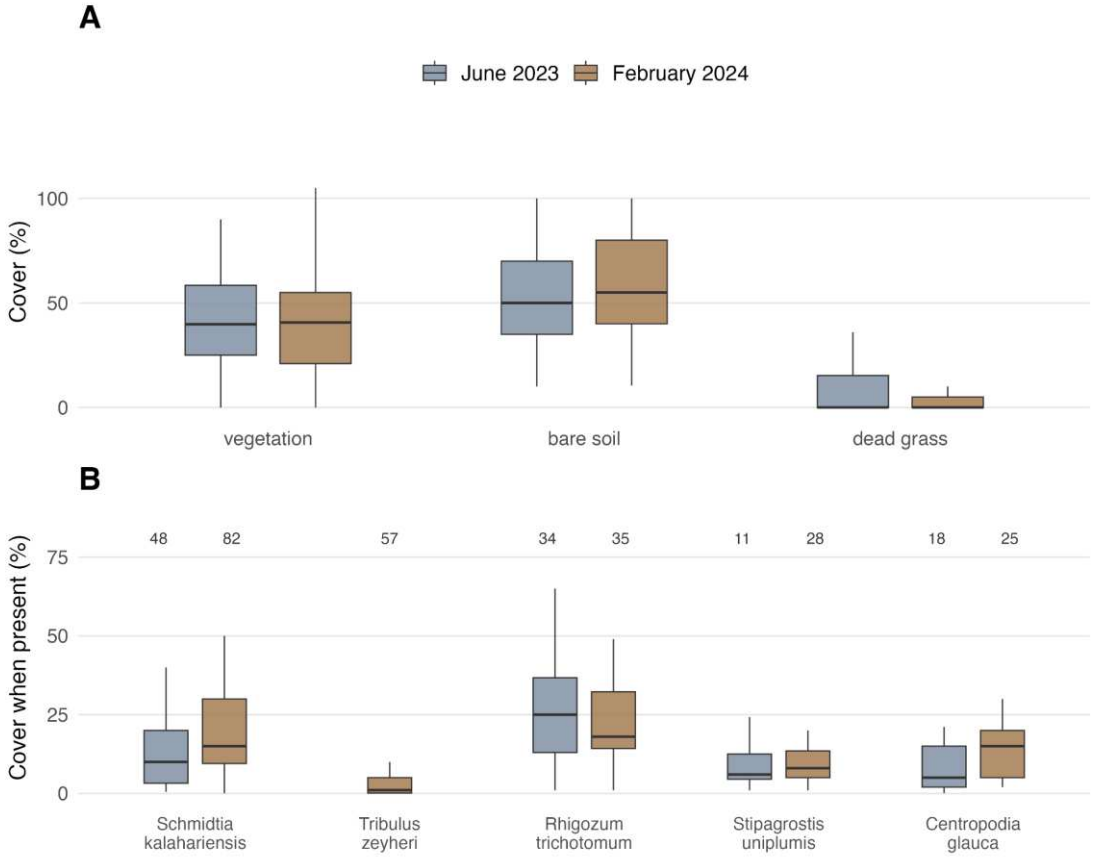
Table 5. Data dictionary for vegetation species composition (dataset iv). The dataset reports the percentage cover of bare soil, dead grass, and vegetation species recorded around 100 pitfall traps on two sampling occasions (June 2023 and February 2024). For each variable, we provide its description, type, range or levels, units, and type of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
year	Year when vegetation was recorded	integer	2023, 2024	year	recorded
month	Month when vegetation was recorded	integer	2, 6	month	recorded
trap_id	Pitfall ID around which vegetation was recorded	character	e.g., D15L09		recorded
height	Vegetation height recorded within the sampling quadrat	numeric	15-170	cm	visually estimated

veg_taxon	Vegetation taxon or cover category recorded within the sampling quadrat	character	e.g., *Schmidtia kalahariensis*, *Verbesina encelioides*, unknown vegetation, unidentifiable stubs, bare soil, dead grass		identified
cover_percentage	Percentage cover of the vegetation taxon or cover category within the sampling quadrat	numeric	0.1-100	%	visually estimated

699

700



701

702 **Figure 6. Temporal variation in vegetation cover around pitfall traps. (A)**

703 Distribution of vegetation cover (including identified species, unknown vegetation and

704 unidentifiable stubs), bare soil cover, and dead grass cover (%) recorded during the June

705 2023 and February 2024 sampling occasions. Vegetation cover includes pitfall traps

706 where vegetation cover was 0% within the sampled 4x4 quadrant. (**B**) Distribution of

707 cover (%) for the five most widespread identified plant species. Boxplots were

708 calculated using only quadrats where the corresponding species was present. Numbers

709 above boxplots indicate the number of quadrats in which each species was recorded on
710 each sampling occasion.
711
712 **Dataset v: Vertebrate species composition**
713 The dataset comprises 441,683 image records of 39 vertebrate species collected
714 from 71 sampling stations during June-November 2023 (Table 6). Of these images,
715 1.5% consisted of either detections that could not be reliably identified because of poor
716 image quality but may represent vertebrates (classified as 'unknown'), or images
717 containing humans and/or vehicles, with or without livestock (classified as
718 'vehicle/human/livestock'). The remaining records included 12 wild ungulate species,
719 19 wild non-ungulate mammal species, four domestic mammal categories (cattle, sheep,
720 horse, and camel), and four higher-level taxonomic categories (rodents, bats, birds, and
721 reptiles). Records containing rodents, bats, birds, or reptiles were not identified to the
722 species level.
723 The mean number of vertebrate taxa (species or broad identification categories)
724 detected per sampling station (2.5th-97.5th percentiles) was 16.9 (9.3-22). The most
725 frequently detected wild ungulate species (Fig. 7A) was *Antidorcas marsupialis* (31%
726 of image records), whereas the wild ungulate species recorded at the highest number of
727 sampling stations (Fig. 7B) was *Sylvicapra grimmia* (68 of 71 stations). Among wild
728 non-ungulate mammal species, *Lepus* spp. was the most frequently detected (3.8% of
729 image records; Fig. 7C), while *Otocyon megalotis* (Desmarest, 1822) occurred at the
730 highest number of sampling stations (70 stations; Fig. 7D). Among domestic mammal
731 categories, sheep was the most frequently recorded (14% of image records), and cattle
732 was detected at the highest number of sampling stations (39 stations). Birds were

recorded at all 71 sampling stations and accounted for 4.9% of all image records. Rodents were detected at 20 stations (0.1% of image records), reptiles at seven stations (0.1% of image records), and bats at a single station (one image record).

Table 6. Data dictionary vertebrate species composition (dataset v). The dataset reports image records from camera traps across 71 sampling stations. For each variable, we provide its description, type, range or levels, units, and type of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
station_id	Sampling station ID where the image was recorded	character	e.g., E16		recorded
camera_id	Camera trap ID that recorded the image	character	e.g., 054		recorded
timestamp	Date and time when the image was recorded	datetime	2023-06-07 17:21:49 – 2023-11-29 14:12:07	ISO 8601 (UTC)	recorded
image_1	Species or identification category present in the image	character	e.g., *Oryx gazella*		identified
image_2	Additional species or identification category present in the image, if applicable	character	e.g., *Connochaetes taurinus taurinus*		identified
image_3*	Additional species or identification category present in the image, if applicable	character	e.g., *Damaliscus pygargus phillipsi*		identified

* No image contained more than three species or identification categories.

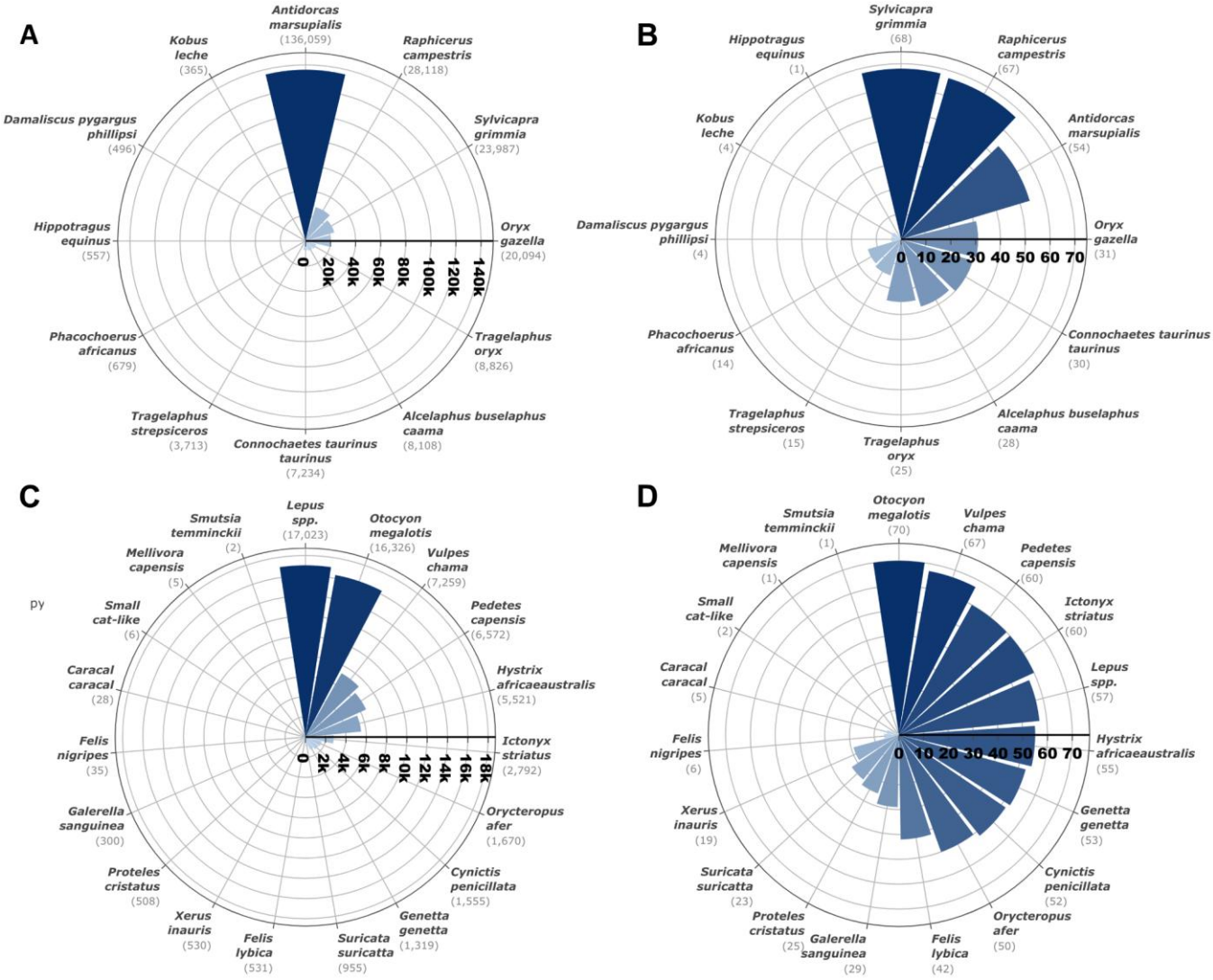


Figure 7. Most widespread and most frequently recorded vertebrate species in camera traps. Species ranked by (**A**) total number of image records for wild ungulate species, (**B**) number of camera stations at which each wild ungulate species was detected, (**C**) total number of image records for wild non-ungulate mammal species, and (**D**) number of camera stations at which each wild non-ungulate mammal species was detected. Values are represented as radial bar plots, where bar length indicates the metric value and colour intensity increases with magnitude.

749

750 **Dataset vi: Deployment dates for loggers**

751 The dataset records the date, time and pitfall adjacent to the deployment of 21 loggers
752 across 9 transects (Table 7).

753

754 **Table 7. Data dictionary for deployment dates for loggers (dataset vi).** The dataset
755 reports the date, time and pitfall adjacent to the deployment of 21 loggers across 9
756 transects. For each variable, we provide its description, type, range or levels, units, and
757 type of measurement. Variables that can be used to link datasets are highlighted in Fig.
758 2.

Variable	Description	Type	Range or levels	Units	Measurement
timestamp	Date and time when the logger was deployed	datetime	2024-02-08 18:30:00 – 2024-02-14 19:02:00	ISO 8601 (UTC)	recorded
trap_id	Pitfall ID adjacent to the logger. For cable loggers, which are positioned between two pitfalls, only the first one is reported	character	e.g., J13L05		recorded
serial_number	Serial ID of the logger	character	95139712 – 95145259		defined
logger_type	Type of TMS logger	categorical	2 levels: standard and cable		recorded

759

760 **Dataset vii: Soil temperature and moisture**

761 The soil temperature and moisture dataset contains a time series of soil
762 temperature and moisture measured at multiple depths and heights using sensors from
763 21 loggers deployed across 9 transects (Table 8). A total of 5,286,746 observations were
764 collected across all sensors and depths. Soil temperature values ranged from -10.9 to
765 60°C, while moisture ranged from 0 to 0.233 volumetric water content (VWC). Time
766 series showed clear seasonal dynamics in soil temperature across all depths, with higher

values during summer and lower values during winter (Fig. 8). Day-night fluctuations were also evident, particularly in above surface sensors, whereas deeper sensors showed reduced variability. Moisture displayed episodic peaks corresponding to discrete events, followed by gradual declines over time, resulting in higher temporal variability compared to temperature measurements (Fig. 8). These patterns were consistent across loggers but varied in magnitude, likely depending on local conditions. Overall, the observed temporal patterns highlight strong seasonal signals in temperature and more irregular, event-driven dynamics in soil moisture, with variability decreasing with depth.

Variable	Description	Type	Range or levels	Units	Measurement
timestamp	Date and time when the soil measurement was taken	datetime	2024-02-08 18:30:00 – 2026-03-07 15:45:00		recorded
serial_number	Serial ID of the logger	character	95139712 – 95145259		defined
sensor_name	Sensor name by manufacturer	categorical	4 levels: TMS-1, TMS-2, TMS-3 (temperature sensors) and VWC_moist (soil moisture sensor)		recorded
height	Depth or height at which the sensor recorded the measurement	character	-500 to +150	mm	recorded
value	Measured variable value	numerical	-10.9 – 60 (temperature); 0 – 0.233 (moisture)	ºC (temperature); VWC (volumetric water content)	recorded

Table 8. Data dictionary for soil temperature and moisture (dataset vii). The dataset reports the temperature and moisture values measured over time by a total of 84 sensors (4 per logger) across 21 loggers. For each variable, we provide its description, type, range or levels, units, and type of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

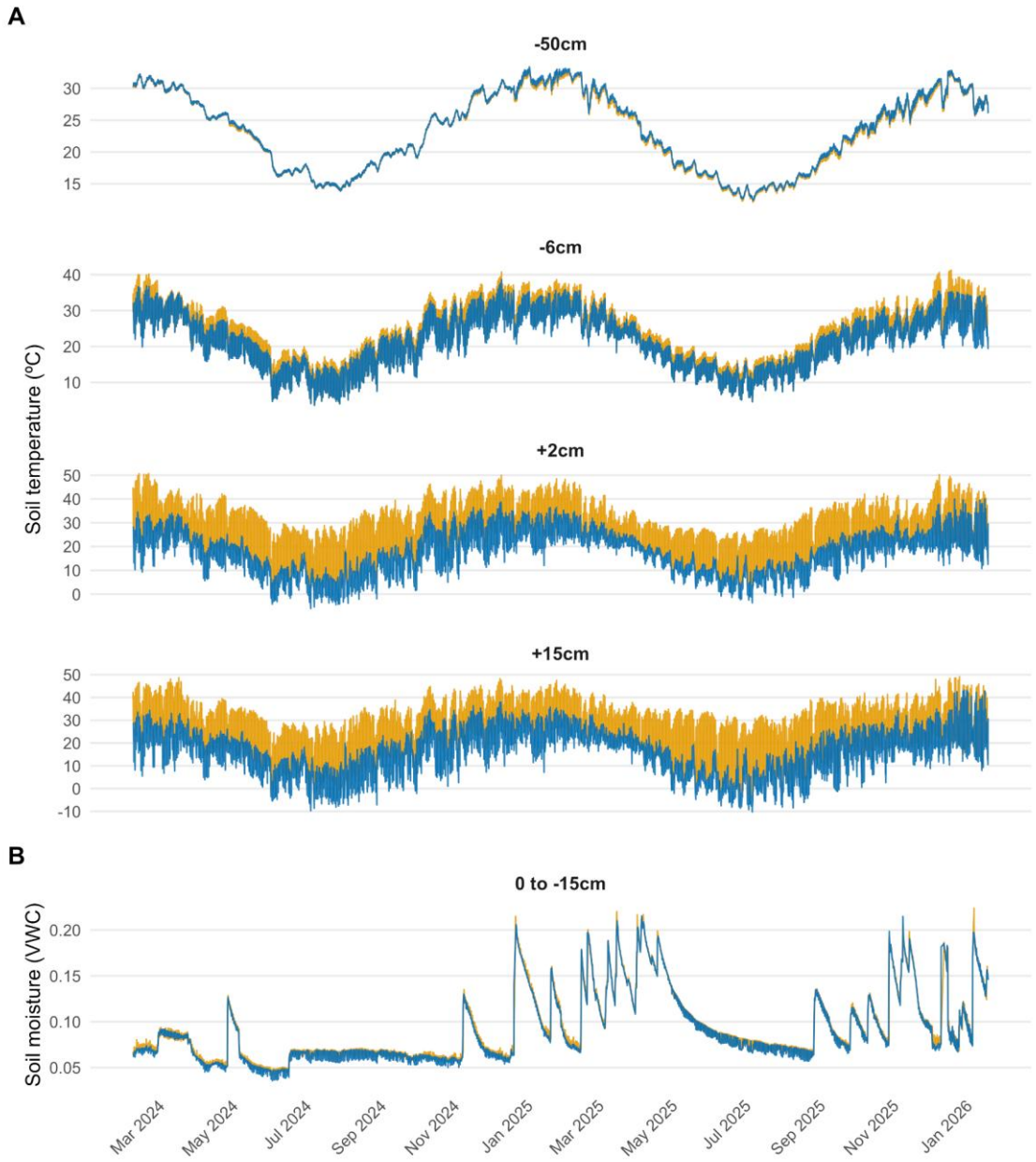


Figure 8. Temporal dynamics of soil temperature and moisture measured at different depths at a representative transect (J15). (**A**) Soil temperature (°C) recorded at four depths (-50, -6, +2 and +15 cm). (**B**) Soil moisture (VWC) measured at 0 to -15 cm. Colours indicate day (orange) and night (blue) measurements. Data correspond to standard logger 95145246 for all depths except -50 cm, which corresponds to cable logger 95139712.

Dataset viii: Macrohabitat polygons

The macrohabitat polygon dataset delineates the macrohabitat types present across the study area (7,294 ha; Table 9). We identified five macrohabitat types (Fig. 9). Macrohabitat polygons from the surrounding landscape are also included to provide spatial context. Within the study area, the most extensive macrohabitat was low duneveld (3,195 ha, 44%), followed by high duneveld (2,679 ha, 37%), combined river terrace and the diatomaceous fossil riverbed (737, 10%), and calcareous *Rhigozum* shrubland (673 ha, 9%). The smallest macrohabitat was white calcareous pan flats (52 ha, <1%). In comparison across land-use types, the Natural area (3,220 ha) had high duneveld as the most extensive coverage (1,202 ha, 37%), followed by low duneveld (1,180 ha, 37%), river terrace and floodplains (554 ha, 17%), *Rhigozum* shrubland (269 ha, 8%), and the smallest land coverage was white calcareous pan flats (14 ha, <1%). The Rotational grazing management area (2,042 ha) was dominated by low duneveld (1,518 ha, 74%), followed by *Rhigozum* shrubland (202 ha, 10%), high duneveld (156 ha, 8%), river terrace and floodplains (140 ha, 7%), and the smallest coverage was white calcareous pan flats (26 ha, 1%). The final land-use type was the Mixed-species rotational grazing area (2,077 ha), which had high duneveld as the most extensive coverage (1,321 ha, 64%), followed by low duneveld (497 ha, 24%), *Rhigozum* shrubland (202 ha, 10%), river terrace and floodplains (43 ha, 2%), and the smallest coverage was again white calcareous pan flats (12 ha, <1%).

Table 9. Data dictionary for macrohabitat polygons (dataset viii). The dataset contains the spatial delineation of the five macrohabitat categories mapped across the study area, together with macrohabitat polygons from the surrounding landscape. For each variable, we provide its description, type, range or levels, units, and method of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
macrohabitat	Macrohabitat category assigned to each polygon	categorical	5 levels: Calcareous pan flats, Calcareous Rhigozum shrubland, High duneveld, Low duneveld, River terrace and Diatomaceous fossil riverbed		habitat mapping according to [21,54,65]
geometry	Polygon geometry representing the spatial extent of each macrohabitat patch	multipolygon		Decimal degrees (WGS84)	habitat polygons according to [21,65]

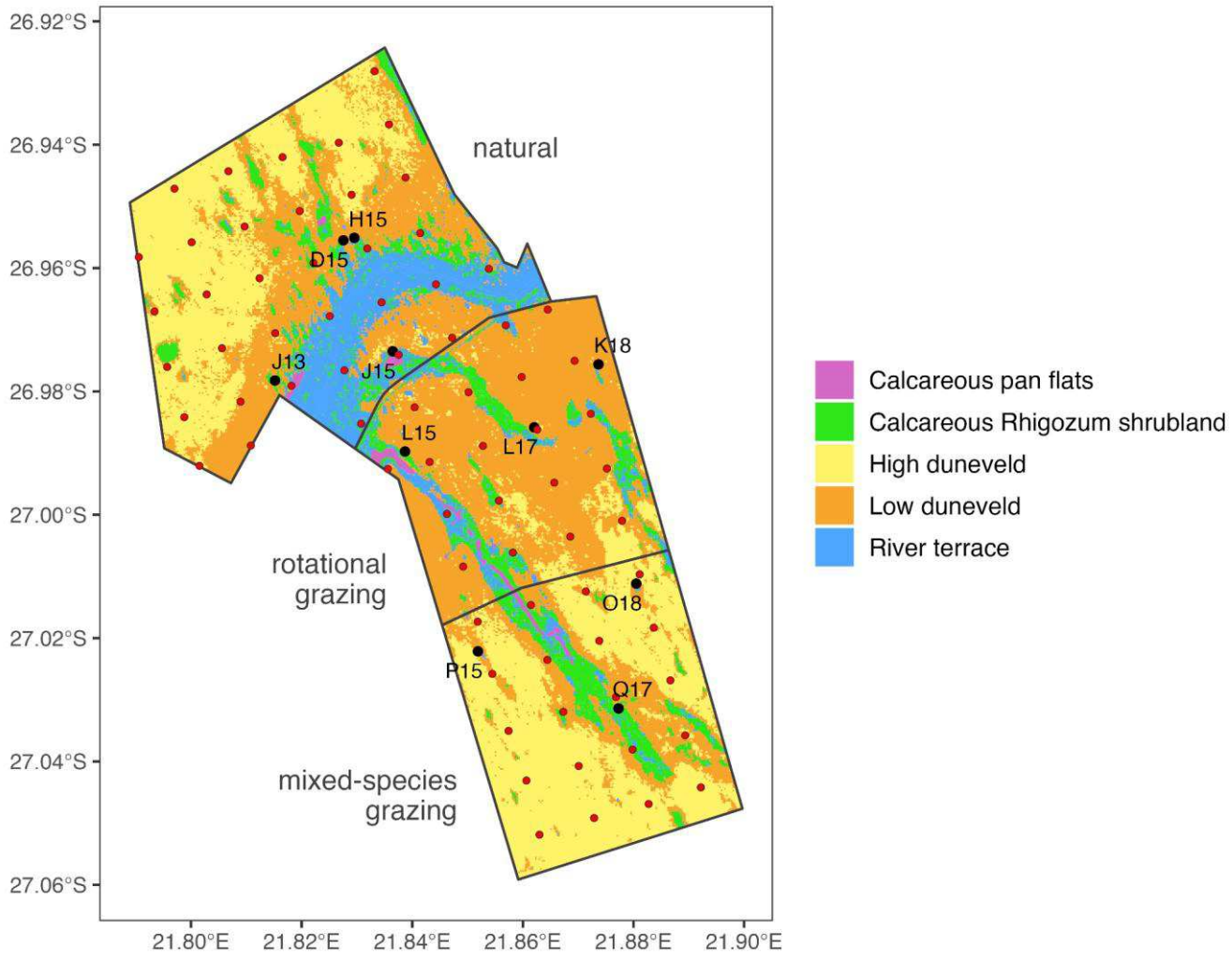


Figure 9. Macrohabitat map of the study area. Map showing the spatial distribution of the main macrohabitat types across the study area: calcareous pan flats, calcareous *Rhigozhum* shrubland, high duneveld, low dueveld, and river terrace (including the fossil riverbed and calcareous river terraces and floodplains). Grey polygons indicate the three land-use types or management regimes (natural, rotational grazing, and mixed-species grazing), which are separated by game fencing. Black points indicate the locations and IDs of the 10 sampling transects. Red points indicate the locations of the 73 camera trap stations originally deployed across the study area.

Dataset ix: Pitfall trap locations

The dataset records the coordinates of the 100 pitfall traps set across 10 transects (Fig. 9; Table 10).

Table 10. Data dictionary for pitfall trap locations (dataset ix). The dataset provides the geographic coordinates of the 100 pitfall traps deployed across ten transects. For each variable, we provide its description, type, range or levels, units, and method of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
trap_id	Pitfall trap ID	character	e.g., J13L05		recorded
latitude	Latitude of the pitfall trap location	numeric	-27.03 to -26.95	Decimal degrees (WGS84)	gps
longitude	Longitude of the pitfall trap location	numeric	21.81 to 21.88	Decimal degrees (WGS84)	gps

Dataset x: Subterranean trap locations

The dataset records the coordinates of the 20 subterranean traps set across 10 transects (Fig. 9; Table 11).

Table 11. Data dictionary for subterranean trap locations (dataset x). The dataset provides the geographic coordinates of the 20 subterranean traps deployed across nine transects. For each variable, we provide its description, type, range or levels, units, and method of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
trap_id	Subterranean trap ID	character	e.g., SQ17L01		recorded

latitude	Latitude of the subterranean trap location	numeric	-27.03 to -26.95	Decimal degrees (WGS84)	gps
longitude	Longitude of the subterranean trap location	numeric	21.81 to 21.88	Decimal degrees (WGS84)	gps

848

849 **Dataset xi: Camera trap locations**

850 The camera trap locations dataset records the coordinates of the 73 camera trap
851 stations originally deployed across the study area (Fig. 9; Table 12).

852

853 **Table 12. Data dictionary for camera trap locations (dataset xi).** The dataset
854 provides the geographic coordinates of the 73 camera trap stations originally deployed
855 across the study area. Data from stations J15 and J16 are not included in the current data
856 release because of camera trap malfunctions. For each variable, we provide its
857 description, type, range or levels, units, and method of measurement. Variables that can
858 be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
station_id	Sampling station ID	character	e.g., E16		recorded
latitude	Latitude of the sampling station	numeric	-27.05 to -26.92	Decimal degrees (WGS84)	gps
longitude	Longitude of the sampling station	numeric	21.78 to 21.89	Decimal degrees (WGS84)	gps

859
860

861 **Dataset xii: Land-use type polygons**

862 The land-use type polygons dataset delineates the three game-fenced areas
863 comprising the study area, each corresponding to a different land-use type (natural,
864 rotational grazing, and mixed-species grazing) (Fig. 9; Table 13). Boundaries of the

neighbouring farms are also included to provide the spatial context of the surrounding landscape.

Table 13. Data dictionary for land-use type polygons (dataset xii). The dataset contains the spatial delineation of the three land-use types comprising the study area, together with the boundaries of the seven neighbouring farms. For each variable, we provide its description, type, range or levels, units, and method of measurement. Variables that can be used to link datasets are highlighted in Fig. 2.

Variable	Description	Type	Range or levels	Units	Measurement
farm_id	Farm identifier corresponding to each polygon	factor	10 levels: KRR639, Rus en Vrede N, Rus en Vrede S (comprising the study area), and seven neighbouring farms		recorded
land_use	Land-use type (management regime) applied to each farm	factor	3 levels: natural, rotational grazing and mixed-species grazing; NA for neighbouring farms		recorded
geometry	Polygon geometry representing the spatial extent of each farm	multipolygon		Decimal degrees (WGS84)	Digitised farm boundaries

Technical Validation

Community data validation (dataset i)

The raw dataset was carefully checked and cleaned to ensure the accuracy of taxonomic and sampling information. Corrections included standardising variable names; formatting date and time information; reporting missing information as NA; removing records that miss critical data (e.g., number of specimens); harmonising trap identifiers and correcting typographical errors; standardising taxonomic names across all levels, including ‘cf.’ prefixes; inspecting outliers in specimen counts (extreme

values were biologically plausible, mostly ants and termites); ensuring records fell within the study period; and standardising life-stage labels. For taxonomic validation, see the section '*Methods*'.

Invertebrate biomass validation (dataset ii, iii)

The raw invertebrate biomass data were recorded across multiple files, each corresponding to a sampling session and trap type (pitfall or subterranean traps). We individually checked and cleaned each file before merging them into a single dataset for each sampling method (dataset ii: pitfall traps; dataset iii: subterranean traps). Data validation included standardising variable names; formatting date and time information; reporting missing information as NA; removing records with missing critical data (e.g., timestamp, trap identifier, or biomass values); harmonising trap identifiers and correcting typographical errors; ensuring records fell within the study period; and identifying duplicate records for manual inspection and removal. Outliers in abundance and biomass were inspected using dataset-specific thresholds informed by expected ranges of captured invertebrates and field experience. Extreme abundance values were biologically plausible (mostly ants and termites) and therefore retained. Records containing impossible values (e.g., negative biomass values) were removed. Unusually large biomass values were manually reviewed using field comments, original annotations, and expert assessment from the field team. Obvious transcription errors (e.g., misplaced decimal points or accidental inclusion of vertebrate or substrate mass) were corrected conservatively, whereas biologically plausible extreme values were retained.

907 **Vegetation species composition validation (dataset iv)**

908 The raw vegetation dataset was quality-checked and cleaned to ensure
909 consistency in sampling metadata, vegetation cover estimates, and taxonomic
910 information. Data validation included standardising variable names and formats;
911 harmonising pitfall trap identifiers and verifying that all records corresponded to valid
912 sampling locations; converting variables to consistent data types; and inspecting
913 missing values across all variables. Missing values in vegetation cover variables were
914 interpreted as absence records and replaced with zeros, whereas missing vegetation
915 height measurements were retained as NA. To assess the consistency of vegetation
916 cover estimates, we calculated total cover per sampling record as the sum of all
917 vegetation, bare soil, and dead grass cover values. Records with total cover slightly
918 below 100% were interpreted as minor discrepancies associated with visual field
919 estimates and were proportionally rescaled to ensure that total cover summed to 100%.
920 Records with total cover exceeding 100% were retained, as they reflected vertical
921 overlap among vegetation layers (e.g., herbaceous vegetation beneath shrubs or trees)
922 rather than data-entry errors. Vegetation height values were screened for unrealistic
923 visual estimations, and all detected extreme values were considered biologically
924 plausible and retained following manual inspection. Additional quality-control
925 procedures included checking for duplicate sampling records, verifying the temporal
926 consistency of sampling information, and harmonising vegetation taxon names.
927 Ambiguous or inconsistently labelled vegetation categories were standardised into a
928 common “unknown vegetation” category, whereas “unidentifiable stubs” were retained
929 separately. Finally, the dataset was converted to a standardised long format.

930

Soil data validation (dataset vi, vii)

The raw soil data were compiled from multiple files, merged into a single dataset, and subsequently quality-checked, cleaned, and filtered to ensure consistency and biologically realistic values. We followed the procedures by [36]. Measurements recorded prior to logger deployment were removed using deployment timestamps, as devices record continuously from the moment they are removed from their packaging. Data validation included removing duplicate records, retaining only loggers with metadata, and ensuring consistency between sensor identifiers and deployment information. Outliers were initially identified using sensor-specific thresholds, including values outside expected physical ranges and abrupt changes between consecutive measurements. Thresholds were defined according to sensor characteristics, allowing higher variability for raw soil moisture signals and more constrained changes for temperature sensors. Detected outliers were replaced with missing values (NA). Raw electric conductivity measurements were converted to ecologically relevant moisture values, i.e., volumetric water content (VWC), using temperature-corrected calibration functions, with sensor-specific configurations depending on logger type. Values recorded under frozen conditions were set to NA. To further detect anomalies not captured by threshold-based methods, we applied Generalised Additive Models (GAMs) to each time series, modelling cyclic seasonal (day-of-year) and diel (hour-of-day) patterns. Observations with residuals exceeding three standard deviations were flagged as potential anomalies. Visual inspection of time series plots confirmed that these points corresponded to sustained deviations from expected temporal patterns (e.g., sensor disturbances) and were removed from the dataset. Additional checks included verifying the consistency of measurement depths across logger types, standardising height labels, removing incomplete records, and ensuring a uniform data structure. After

cleaning, time series were visually inspected to confirm expected seasonal and diel patterns in soil temperature and event-driven variability in soil moisture, indicating that the dataset is suitable for further analysis.

Usage Notes

Data use and best practice notes. The datasets described in the paper are distributed under a CC BY license. Users are requested to cite both the data paper and the associated Zenodo repository when using any of the datasets, including future updates to the repository. We also welcome enquiries regarding data use and encourage users to contact the authors if collaboration is of interest.

Data Availability

The datasets generated in this study are publicly available on Zenodo[25].

Code Availability

The code that supports the reproducibility of this study is publicly available on Zenodo[25].

References

1. Gross, N. *et al.* Unforeseen plant phenotypic diversity in a dry and grazed world. *Nature* **632**, 808–814 (2024).
2. Lewin, A., Murali, G., Rachmilevitch, S. & Roll, U. Global evaluation of current and future threats to drylands and their vertebrate biodiversity. *Nat. Ecol. Evol.* **8**, 1448–1458 (2024).
3. Wang, C. *et al.* Drought-heatwave compound events are stronger in drylands. *Weather*

Clim. Extrem. **42**, 100632 (2023).

4. Conradie, S. R., Woodborne, S. M., Cunningham, S. J. & McKechnie, A. E. Chronic, sublethal effects of high temperatures will cause severe declines in southern African arid-zone birds during the 21st century. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14065–14070 (2019).
5. Paniw, M., Maag, N., Cozzi, G., Clutton-Brock, T. & Ozgul, A. Life history responses of meerkats to seasonal changes in extreme environments. *Science* **363**, 631–635 (2019).
6. Fuller, A. *et al.* How dryland mammals will respond to climate change: the effects of body size, heat load and a lack of food and water. *J. Exp. Biol.* **224**, jeb238113 (2021).
7. Maestre, F. T. *et al.* Changes in biocrust cover drive carbon cycle responses to climate change in drylands. *Glob. Chang. Biol.* **19**, 3835–3847 (2013).
8. Cameron, E. K. *et al.* Global gaps in soil biodiversity data. *Nat Ecol Evol* **2**, 1042–1043 (2018).
9. Potapov, A. M. Multifunctionality of belowground food webs: resource, size and spatial energy channels. *Biol. Rev. Camb. Philos. Soc.* **97**, 1691–1711 (2022).
10. Lavelle, P. *et al.* Soil macroinvertebrate communities: A world-wide assessment. *Glob. Ecol. Biogeogr.* **31**, 1261–1276 (2022).
11. Kamiński, M. J. *et al.* A catalogue of the tribe Sepidiini Eschscholtz, 1829 (Tenebrionidae, Pimeliinae) of the world. *Zookeys* **844**, 1–121 (2019).
12. Sohlenius, B. & Boström, S. Annual and long-term fluctuations of the nematode fauna in a Swedish Scots pine forest soil. *Pedobiologia* **45**, 408–429 (2001).
13. Thakur, M. P. *et al.* Reduced feeding activity of soil detritivores under warmer and drier conditions. *Nat. Clim. Chang.* **8**, 75–78 (2018).
14. Jouquet, P., Dauber, J., Lagerlof, J., Lavelle, P. & Lepage, M. Soil invertebrates as ecosystem engineers: Intended and accidental effects on soil and feedback loops. *Appl. Soil Ecol.* **32**, 153–164 (2006).
15. Ruiz, S., Hallett, P. D. & Or, D. Bioturbation—Physical processes. in *Encyclopedia of Soils in the Environment* 100–114 (Elsevier, 2023).

16. Filser, J. *et al.* Soil fauna: key to new carbon models. *SOIL* **2**, 565–582 (2016).
17. Wall, D. H., Bardgett, R. D. & Kelly, E. Biodiversity in the dark. *Nat. Geosci.* **3**, 297–298 (2010).
18. Jubber, W. R., Manser, M. B. & Fuller, A. Feeding through the ages: Revisiting the diet of meerkats. *J. Arid Environ.* **227**, 105331 (2025).
19. Phakoago, M., Panaino, W., Maloney, S. K. & Fuller, A. Prey preference and dietary overlap between the aardvark and Temminck's pangolin in the kalahari. *J. Arid Environ.* **230**, 105422 (2025).
20. Pietersen, D. W., Symes, C. T., Woodborne, S., McKechnie, A. E. & Jansen, R. Diet and prey selectivity of the specialist myrmecophage, Temminck's ground pangolin. *J. Zool. (1987)* **298**, 198–208 (2016).
21. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
22. Nielsen, U. N. & Ball, B. A. Impacts of altered precipitation regimes on soil communities and biogeochemistry in arid and semi-arid ecosystems. *Glob. Chang. Biol.* **21**, 1407–1421 (2015).
23. Jubber, W. R., Fuller, A. & Paniw, M. Capturing the unseen: A low-cost method for stratified subterranean sampling of soil invertebrates in drylands. *Ecol. Solut. Evid.* **7**, (2026).
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2023).
25. Morales-González, A. *et al.* A comprehensive multi-trophic biodiversity and environmental dataset for understanding global change impacts in dryland ecosystems. Preprint at https://doi.org/10.5281/zenodo.18376602 (2026).
26. Wilke, C. O. Streamlined Plot Theme and Plot Annotations for 'ggplot2' [R package cowplot version 1.2.0]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=cowplot (2025).
27. A Grammar of Data Manipulation [R package dplyr version 1.2.1]. *Comprehensive R*

Archive Network (CRAN) https://CRAN.R-project.org/package=dplyr (2026).

28. Apply Mapping Functions in Parallel using Futures [R package furrr version 0.4.0]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=furrr (2026).

29. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Measurement (Mahwah NJ)* **17**, 160–167 (2019).

30. Slowikowski, K. Automatically Position Non-Overlapping Text Labels with 'ggplot2' [R package ggrepel version 0.9.8]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=ggrepel (2026).

31. Müller, K. Pretty Time of Day [R package hms version 1.1.4]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=hms (2025).

32. Firke, S. Simple Tools for Examining and Cleaning Dirty Data [R package janitor version 2.2.1]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=janitor (2024).

33. Grolemund, G. & Wickham, H. Dates and Times Made Easy with lubridate. *J. Stat. Softw.* **40**, 1–25 (2011).

34. Ooms, J. Advanced Graphics and Image-Processing in R [R package magick version 2.9.1]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=magick (2026).

35. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B Stat. Methodol.* **73**, 3–36 (2011).

36. Man, M. *et al. myClim*: Microclimate data handling and standardised analyses in R. *Methods Ecol. Evol.* **14**, 2308–2320 (2023).

37. Pedersen, T. L. The Composer of Plots [R package patchwork version 1.3.2]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=patchwork (2025).

38. Sievert, C. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. (Chapman

& Hall/CRC, Philadelphia, PA, 2020).

39. Functional Programming Tools [R package purrr version 1.2.2]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=purrr (2026).

40. Read Rectangular Text Data [R package readr version 2.2.0]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=readr (2026).

41. Read Excel Files [R package readxl version 1.5.0]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=readxl (2026).

42. Compact and Flexible Summaries of Data [R package skimr version 2.2.2]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=skimr (2026).

43. van der Loo, M. Approximate String Matching, Fuzzy Text Search, and String Distance Functions [R package stringdist version 0.9.17]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=stringdist (2026).

44. Wickham, H. Simple, Consistent Wrappers for Common String Operations [R package stringr version 1.6.0]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=stringr (2025).

45. Tidy Messy Data [R package tidyr version 1.3.2]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=tidyr (2025).

46. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

47. Ooms, J. Export Data Frames to Excel 'xlsx' Format [R package writexl version 1.5.4]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=writexl (2025).

48. Russell, A. F. *et al.* Factors affecting pup growth and survival in co-operatively breeding meerkats Suricata suricatta. *J. Anim. Ecol.* **71**, 700–709 (2002).

49. Paniw, M. *et al.* Higher temperature extremes exacerbate negative disease effects in a social mammal. *Nat. Clim. Chang.* **12**, 284–290 (2022).

50. Roffe, S. J., Fitchett, J. M. & Curtis, C. J. Investigating changes in rainfall seasonality across South Africa: 1987–2016. *Int. J. Climatol.* **41**, (2021).

51. Wang, L., D'Odorico, P., Ringrose, S., Coetzee, S. & Macko, S. A. Biogeochemistry of

Kalahari sands. *J. Arid Environ.* **71**, 259–279 (2007).

52. Van Rooyen, T. H. The soils of the Kalahari Gemsbok National Park. *Koedoe* **27**, 45–61 (1984).

53. van Rooyen, N. & van Rooyen, M. W. Vegetation of the south-western arid Kalahari: An overview. *Trans. R. Soc. S. Afr.* **53**, 113–140 (1998).

54. Van Rooyen, M. W., Van Rooyen, N., du P. Bothma, J. & Van den Berg, H. M. Landscapes in the Kalahari Gemsbok National Park, South Africa : original research. *Koedoe* (2008).

55. Kong, T. M. Understanding Land Management and Desertification in the South African Kalahari with Local Knowledge and Perspectives. (University of Arizona, 2012).

56. Reinhard, J. E., Geissler, K. & Blaum, N. Grass and ground dwelling beetle community responses to holistic and wildlife grazing management using a cross-fence comparison in Western Kalahari rangeland, Namibia. *J. Insect Conserv.* **26**, 711–720 (2022).

57. Wild, J. *et al.* Climate at ecologically relevant scales: A new temperature and soil moisture logger for long-term microclimate measurement. *Agric. For. Meteorol.* **268**, 40–47 (2019).

58. Borkent, A. & Sinclair, B. J. Key to Diptera families - larvae. in *Manual of Afrotropical Diptera. Volume 1. Introductory chapters and keys to Diptera families* (eds. Kirk-Spriggs, A. H. & Sinclair, B. J.) 357–405 (South African National Biodiversity Institute, Pretoria, 2017).

59. Dippenaar-Schoeman, A. S., Haddad, C. R., Foord, S. H. & Lotz, L. N. South African national survey of Arachnida photo identification guide: The Gnaphosidae of South Africa part 1 (A-D). version 1: 1-77 (2021). (2021) doi:10.5281/ZENODO.7197174.

60. Penrith, M.-L. Two new species of Stizopina (Coleoptera: Tenebrionidae: Opatrini) from Namaqualand, southern Africa, and the relationships between the psammophilous genera. *Annals of the Transvaal Museum* **33**, 353–363 (1984).

61. Slingsby, P. *The Ant of Southern Africa: The Ant Book for All*. (Slingsby Maps, Muizenburg, 2017).

62. Knapp, M., Knappová, J., Jakubec, P., Vonička, P. & Moravec, P. Incomplete species lists

produced by pitfall trapping: How many carabid species and which functional traits are missing? *Biol. Conserv.* **245**, 108545 (2020).

63. Montgomery, G. A., Belitz, M. W., Guralnick, R. P. & Tingley, M. W. Standards and best practices for monitoring and benchmarking insects. *Front. Ecol. Evol.* **8**, 579193 (2021).
64. Beery, S., Morris, D. & Yang, S. Efficient Pipeline for Camera Trap Image Review. *arXiv [cs.CV]* (2019) doi:10.48550/arXiv.1907.06772.
65. Harrison, N. D. *et al.* Behavioural change during dispersal and its relationship to survival and reproduction in a cooperative breeder. *J. Anim. Ecol.* **90**, 2637–2650 (2021).

22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2023).
24. Wilke, C. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.3, https://CRAN.R-project.org/package=cowplot (2024).
25. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr (2023).
26. Vaughan, D. & Dancho, M. furrr: Apply Mapping Functions in Parallel using Futures. R package version 0.3.1, https://CRAN.R-project.org/package=furrr (2022).
27. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, (2016).
28. Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.4, https://CRAN.R-project.org/package=ggrepel (2023).
29. Müller, K. hms: Pretty Time of Day. R package version 1.1.3, https://CRAN.R-project.org/package=hms (2023).
30. Firke, S. janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.2.1, https://CRAN.R-project.org/package=janitor (2024).
31. Grolemund, G. & Wickham, H. Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, **40**, 1-25, https://www.jstatsoft.org/v40/i03/ (2011).

32. Ooms, J. magick: Advanced Graphics and Image-Processing in R. R package version 2.8.2, https://CRAN.R-project.org/package=magick (2023).
33. Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B.* **73**, 3-36 (2011).
34. Man, M., Kalčík, V., Macek, M., Brůna, J., Hederová, L., Wild, J. & Kopecký, M. myClim: Microclimate data handling and standardised analyses in R. *Methods in Ecology and Evolution*, **14**, 2308-2320, https://doi.org/10.1111/2041-210X.14192 (2023).
35. Pedersen T. patchwork: The Composer of Plots. R package version 1.3.0, https://CRAN.R-project.org/package=patchwork (2024).
36. Sievert, C. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida (2020).
37. Wickham, H. & Henry, L. purrr: Functional Programming Tools. R package version 1.0.2, https://CRAN.R-project.org/package=purrr (2023).
38. Wickham, H., Hester, J. & Bryan, J. readr: Read Rectangular Text Data. R package version 2.1.4, https://CRAN.R-project.org/package=readr (2023).
39. Wickham, H. & Bryan, J. readxl: Read Excel Files. R package version 1.4.3, https://CRAN.R-project.org/package=readxl (2023).
40. Waring, E., Quinn, M., McNamara, A., Arino de la Rubia, E., Zhu, H. & Ellis, S. skimr: Compact and Flexible Summaries of Data. R package version 2.1.5, https://CRAN.R-project.org/package=skimr (2022).
41. van der Loo, M. The stringdist package for approximate string matching. *The R Journal*, **6**, 111-122, https://CRAN.R-project.org/package=stringdist (2014).
42. Wickham, H. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.1, https://CRAN.R-project.org/package=stringr (2023).
43. Wickham, H., Vaughan, D., & Girlich, M. tidyr: Tidy Messy Data. R package version 1.3.0, https://CRAN.R-project.org/package=tidyr (2023).
44. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R.,

Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. Welcome to the tidyverse. *Journal of Open Source Software*, **4**, 1686, https://doi.org/10.21105/joss.01686 (2019).

45. Ooms, J. writexl: Export Data Frames to Excel 'xlsx' Format. R package version 1.4.2, https://CRAN.R-project.org/package=writexl (2023).

Author contributions

A.MG.., WR.J. and M.P. performed data curation, methodology, validation, and preparation of the original manuscript draft. A.MG. implemented the code to process, synthesize and visualize the study data. WR.J., D.A., DM.B., TL.B., C.D., A.F., G.GP., KC.G., CR.H., S.H., MJ.K., M.M., A.O., C.YS.., and M.P. performed the investigation. M.P. and A.O. conceptualized, supervised, and acquired funding for the invertebrate and camera-trap data collection, respectively. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to thank the Kalahari Research Trust for having granted us access to the study area at the Kalahari Research Centre. Our thanks also extend to the University of the Witwatersrand (Department of Physiology). We would like to express our special thanks to John Irish, Lorenzo Pendrini, Ian Engelbrecht, Lizel Hugo Coetzee, and Jan Andries Neethling for verifying the identifications of the different invertebrate specimens in their respective fields of expertise. Alexis Watson, Kallie

Swarts, and Precious Lukhele assisted with collecting the invertebrates, sorting them and entering the data. Solifuges were deposited and sent to the Ditsong Museum in Pretoria, South Africa; various arachnids were deposited in the Bloemfontein Museum in the Free State, South Africa; the Tenebrionidae beetles were sent to the Museum and Institute PAS in Warsaw, Poland; and various orders of insects were donated to the Museum of Windhoek in Namibia.

Funding

We acknowledge the support provided by grant no. CNS2023-144460 (MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR) to A.MG. and M.P.; by grant no. 2024/55/D/NZ8/01512 (National Science Centre NCN, Poland) to A.MG.; and by grant no. RYC2021-033192-I (MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR) to M.P..

Additional information

Correspondence and requests for materials should be addressed to A.MG., WR.J., and M.P..