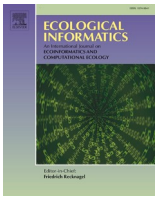
Ecological Informatics 100 (2026) 104127



Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



One size does not fit all: Citizen science models reveal context-specific drivers of schistosome host snails in sub-Saharan Africa

Noelia Valderrama-Bhraunxs [a,b], Larissa Bonifacio [b], Tine Huyse [b], Helena Greter [c,d], Jacob Mbaihondoum [e], Annour Adoum Batil [f], Emiel Van Loon [g], Anton Van Rompaey [a], Joule Madinga [h], Ronald Twongyirwe [i], Liesbet Jacobs [g,*]

[a] *Department of Earth and Environmental Sciences, KU Leuven, Belgium*
[b] *Department of Biology, Royal Museum for Central Africa, Belgium*
[c] *Swiss TPH, Switzerland*
[d] *University of Basel, Switzerland*
[e] *Alliance Sahélienne de Recherches Appliquées Pour le Développement Durable, Chad*
[f] *Institut de Recherche en Elevage Pour le Développement, Chad*
[g] *Institute for Biodiversity and Ecosystems Dynamics, University of Amsterdam, the Netherlands*
[h] *Biomedical department, University of Kikwit, Kikwit City, Congo*
[i] *Faculty of Agriculture, Environment and Veterinary Sciences, Mbarara University of Science and Technology, Uganda*

ARTICLE INFO

Keywords:
Citizen science
Schistosomiasis
Biomphalaria
Sub-Saharan Africa
Focal transmission
Intermediate host

ABSTRACT

Human schistosomiasis transmission depends on compatible freshwater snail hosts, yet fine-scale information on their occurrence remains limited in many endemic settings where routine malacological surveillance is logistically challenging. Citizen science can expand monitoring coverage and frequency, but ecological inference requires attention to observation processes, class imbalance and context dependence. We analysed visit-level *Biomphalaria* spp. presence/absence records from citizen-science networks in Uganda, the Democratic Republic of the Congo (DRC) and Chad. Using a harmonised workflow, we validated and filtered records, delineated monitoring sites, extracted environmental covariates, fitted country-specific binomial GLMMs and evaluated predictions using complementary cross-validation schemes. Across countries, *conditional* R^2 substantially exceeded *marginal* R^2, indicating that occurrence was strongly structured by persistent site-level heterogeneity not captured by the measured predictors. Retained predictors and effect sizes differed among countries, consistent with context-dependent host-environment relationships. A negative association between sampling experience and occurrence persisted after accounting for calendar time, indicating that participation history captured broader changes in the observation process than observer learning alone. Predictive performance depended strongly on the prediction target. Models showed substantial skill for repeated observations at previously monitored sites (Brier Skill Score: 0.47–0.54), but little or no skill at previously unsampled sites (−0.09 to −0.02). In the low-prevalence DRC and Chad datasets, probability-scale evaluation and calibration were highly important because high discrimination could coexist with predictions concentrated near zero. Overall, citizen science provided useful context-specific predictions of *Biomphalaria* spp. occurrence at established monitoring sites, supporting repeated fine-scale surveillance while cautioning against extrapolation beyond monitored networks.

1. Introduction

Human schistosomiasis remains a significant public health challenge in sub-Saharan Africa, particularly in rural areas where daily activities depend on open freshwater systems and access to sanitation is limited (Grimes et al., 2014). Transmission requires completion of an aquatic phase of the *Schistosoma* parasite life cycle: parasite eggs reaching freshwater hatch into miracidia that infect compatible freshwater snails, within which the parasite multiplies before cercariae are released into the water and infect humans during water contact (Gryseels et al.,

* Corresponding author.
E-mail address: l.h.jacobs@uva.nl (L. Jacobs).

https://doi.org/10.1016/j.ecoinf.2026.104127
Received 22 May 2026; Received in revised form 27 October 2026; Accepted 27 October 2026
Available online 31 October 2026
1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2006). For intestinal schistosomiasis caused by *S. mansoni*, *Biomphalaria* spp. are the obligate intermediate hosts and represent a necessary ecological component of transmission. Their occurrence is strongly influenced by local hydrology, waterbody characteristics, vegetation, temperature and land use, while transmission additionally depends on parasite presence and patterns of human-water contact (Kloos et al., 1997; Perez-Saez et al., 2016; Walz et al., 2015). Schistosomiasis transmission is consequently highly focal and dynamic, occurring where compatible snail hosts, suitable environmental conditions, parasite circulation and human exposure overlap in space and time.

This focality creates a surveillance challenge because intermediate host distributions can vary substantially over short distances and through time. Shorelines, streams, springs, wetlands and other water contact locations within the same landscape may differ markedly in their suitability for *Biomphalaria* spp., and these habitats can expand, contract or shift with rainfall, flooding, vegetation change and local disturbance (D. S. Brown, 1994; Perez-Saez et al., 2016; Woolhouse and Chandiwana, 1989). Malacological surveys provide high level of detail to estimate where these snails are; nevertheless, they constitute a logistical challenge in remote areas with limited resources; which are mainly the areas affected by schistosomiasis (Tumusiime et al., 2024). As a result, many highly endemic regions lack the needed information for guiding disease interventions and understanding drivers of risk.

Citizen science has emerged as a scalable and affordable complement to traditional surveillance methods for disease vectors (Palmer et al., 2017). Particularly, citizen science is a viable approach for obtaining high-resolution spatial and temporal snail occurrence data in low-resource areas where local communities have frequent contact with freshwater sources (Tumusiime et al., 2024; Valderrama-Bhraunxs, Huyse, et al., 2025). However, citizen science data are not only shaped by ecological drivers but also by observation processes, including observer experience, sampling effort and timing of observations (Courter et al., 2012; Kosmala et al., 2016). If these observation processes are not explicitly accounted for, models may confound ecological signals with artifacts of data collection (Valderrama-Bhraunxs, Huyse, et al., 2025). In addition, drivers behind species distributions may differ between different types of water bodies, climate zones and ecological settings. This context dependence is very relevant for schistosomiasis because transmission requires spatial and temporal overlap between infected definitive hosts, parasites and compatible intermediate host snails and their occurrence is strongly influenced by local hydrology, land use, and human water contact (Kloos et al., 1997; Perez-Saez et al., 2016).

With these implications, there is a growing need for transparent and reproducible workflows that allow researchers to use citizen science data effectively for intermediate host snail monitoring. To date, few studies have demonstrated how such workflows perform across multiple contrasting study systems using comparable citizen science protocols.

In this article we use citizen science observations from three contrasting regions in Uganda, the Democratic Republic of the Congo (DRC) collected through the ATRAP project and Chad collected through ParaSahel, with the aim (1) to identify environmental, geomorphological and anthropogenic conditions associated with snail presence while accounting for observation-related processes and (2) to delineate a coherent and harmonised analytical workflow for modelling presence/absence data from citizen science monitoring, and (3) examine how predictor support and model structure differ across Uganda, DRC and Chad, distinguishing components of the analytical workflow that can be applied consistently from ecological inference that remains context dependent.

2. Methods

2.1. Study areas and citizen science monitoring

This study analysed freshwater snail sampling reports collected through the ATRAP (Action Towards Reducing Aquatic Snail Parasitic Diseases) citizen science networks in Uganda and DRC (Brees et al., 2021; Valderrama-Bhraunxs, Bonifacio, Tumusiime, et al., 2025) and the ParaSahel (Parasite control in the Sahel) network in Chad. Monitoring was conducted in the Lake Albert region of Uganda, the Kimpese region of DRC and the south-eastern Lake Chad region of Chad (Fig. 1). These settings were included because the respective projects had established longitudinal citizen science snail monitoring networks there and generated sufficient repeated observations to support a harmonised analysis. Moreover, all three settings are epidemiologically relevant for schistosomiasis. In Uganda, intestinal schistosomiasis caused by *Schistosoma mansoni* is endemic in more than 65% of districts and can exceed 50% prevalence in lakeshore communities (Adriko et al., 2024). DRC has the third-highest schistosomiasis burden in sub-Saharan Africa, with both *S. mansoni and S. haematobium* widely endemic (Madinga et al., 2015). Chad is also characterised by strongly focal schistosomiasis transmission and limited malacological and epidemiological information (Moser et al., 2014). The three regions also differ substantially in hydrological and landscape characteristics relevant to freshwater snail habitat suitability, including hydrological connectivity, climatic conditions including annual precipitation and water body type (Tumusiime et al., 2024; Valderrama-Bhraunxs, Bonifacio, Tumusiime, et al., 2025). Elevation values from sampled locations (derived from a digital elevation model, see predictors section) show marked contrast among regions: Uganda spans between ~600 to 1300 m with a dominant high-elevation mode, DRC spans between ~250 to 460 m with a multi-modal distribution, and Chad shows a narrow range of ~280 to 285 m (Copernicus DEM - Global and European Digital Elevation Model | Copernicus Data Space Ecosystem, n.d.).

Within these study regions, monitoring was implemented through project-based citizen science networks in which community members were recruited and trained to conduct repeated intermediate host snail surveys. Each network comprised around 25 to 30 citizen scientists, and each participant sampled one to four locations with a suggested weekly frequency (Brees et al., 2021; Valderrama-Bhraunxs et al., 2025b). Sampling locations were selected to represent water contact sites where communities engage in water-related activities (Brees et al., 2021) associated with schistosomiasis exposure risk (e.g. bathing, swimming, washing, water collection) (Reitzug et al., 2023). Smartphones, protective equipment and compensation (to cover expenses for mobility and data bundles) were provided to support standardised reporting and reduce barriers to participation in rural settings (Ashepet et al., 2026; Benyei et al., 2023). Participants sampled for ~30 min and identified intermediate host snails to genus level (*Biomphalaria* spp., *Bulinus* spp. and *Radix* spp.), additional snail species were reported in a category labelled 'pool' (Brees et al., 2021; Valderrama-Bhraunxs et al., 2025b). After sampling, participants submitted observations through a standardised Kobo form (KoboCollect [Mobile app], 2026), including snail identification and abundance, date, time, observer ID, GPS coordinates and snail photographs. The submitted data were reviewed by expert malacologists and genus identification was verified or corrected, respectively. Species level discrimination often requires more detailed morphological and/or molecular confirmation (Palasio et al., 2019) and was therefore not attempted in this citizen science workflow.

In this study, we focused on *Biomphalaria* snails' presence/absence per visit. All the species of the genus *Biomphalaria* are epidemiologically relevant as an intermediate host of human intestinal schistosomiasis (Morgan et al., 2001), and a binary detection outcome is less sensitive than abundance to observer-dependent counting variability in citizen science reports (Tumusiime et al., 2024).

2.2. Analytical workflow overview

We implemented a harmonised workflow across the three countries to estimate associations between predictors and *Biomphalaria* spp. occurrence from citizen science reports, while accounting for

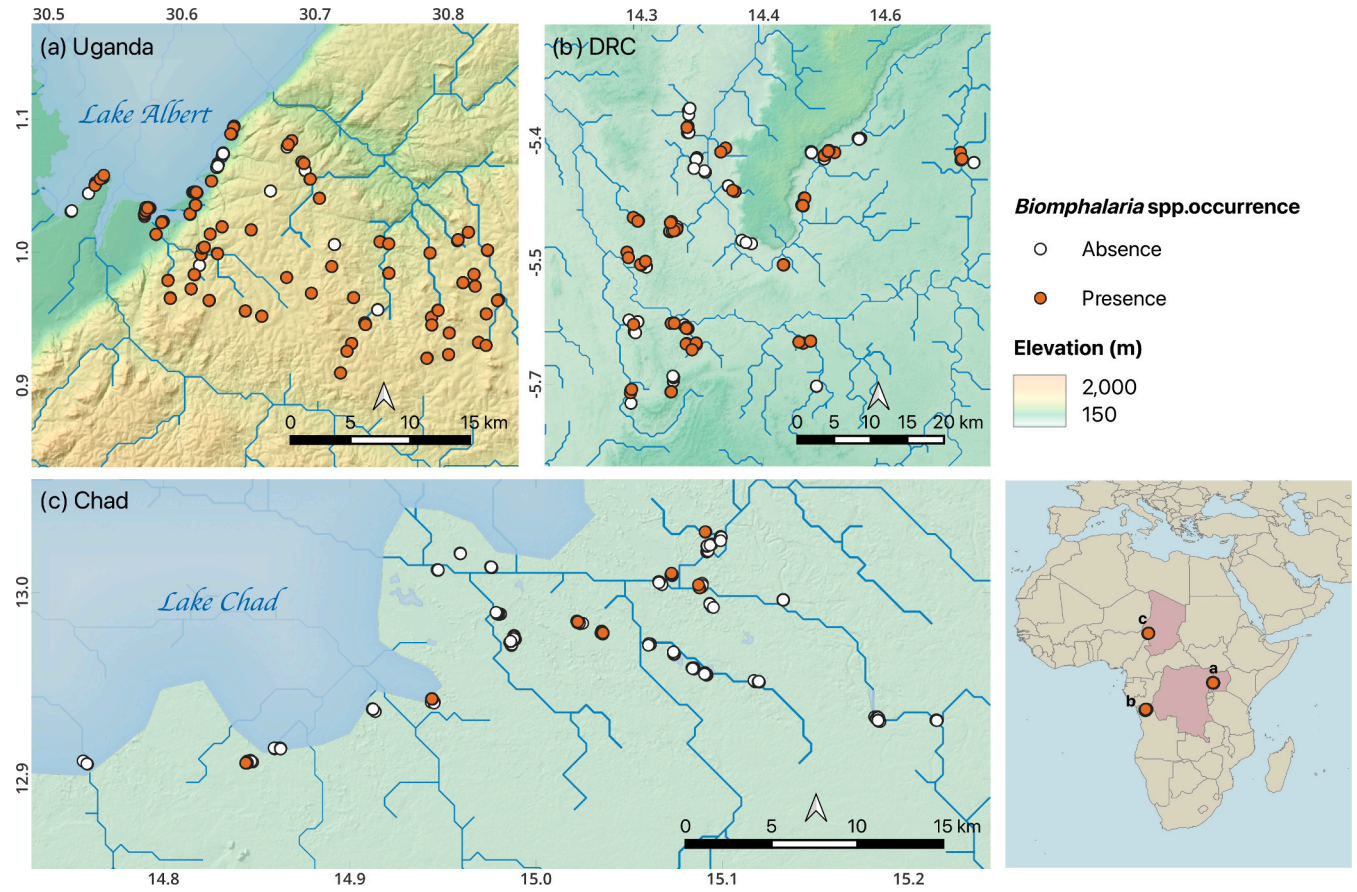


Fig. 1. Study regions and observed *Biomphalaria* spp. occurrence across three countries. Sampled sites are indicated as points (orange = presence, white = absence); presences are plotted on top. Background shading indicates elevation (m); with the hydrographic network and lakes in light blue (Lehner and Grill, 2013). Observed presence frequency was 53% (2534/4770) in Uganda, 16% (645/3934) in DRC and 7% (45/685) in Chad. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observation-related covariates and evaluating model reliability (Fig. 2). The workflow includes four stages: (i) pre-processing of the response variable (validation, filtering, and site delineation) and of predictors (covariate extraction, visual exploration, transformation and scaling), (ii) predictor diagnostics (collinearity checks), (iii) modelling (country-specific mixed models and model selection based on AIC and marginal R^2), and (iv) model diagnostics and evaluation (spatiotemporal dependence checks on residuals and within- site predictive performance).

2.3. Sampling unit, site definition and outcome

The unit of sampling was visit (reporting instance) within a site. For Uganda and the DRC, records were collected between March 2020 and April 2023 and are publicly available through GBIF (Valderrama-Bhraunxs et al., 2025a; Valderrama-Bhraunxs et al., 2025b). For Chad, sampling occurred between November 2023 and November 2024. Prior to filtering (see section 2.5), the datasets comprised 6570 (Uganda), 6891 (DRC) and 1323 (Chad) sampling instances. Because sampling locations were not fixed across all networks and multiple visits could occur within a small area, we delineated distinct sampled 'sites' (site_uid) by clustering visits by GPS coordinates using density-based spatial clustering (DBSCAN) (Hahsler et al., 2019). DBSCAN was applied independently within each country after filtering (removing invalid GPS/date/time records), using projected coordinates (UTM country specific zones) and a common parametrisation across countries. Parameters were calibrated for the case of Uganda and applied to DRC and Chad as in Uganda sites were consistently defined and revisit patterns consistently matched the predefined sites - allowing to validate an automatic clustering method that can be applied to all countries (Valderrama-Bhraunxs et al., 2025c). To assess the robustness of applying this neighbourhood radius to DRC and Chad, we repeated the site delineation using eps values of 150, 191.7 and 250 m while keeping minPts = 5 and the ≥10 visits per site criterion constant, and quantified agreement in cluster assignments using the Adjusted Rand Index (Table A1).

For each labelled site, we calculated the centroid of all visits' coordinates and measured the distance from each visit to that centroid. We then removed unusually distant points using the median absolute deviation (MAD) rule using the function *mad* from the package *stats* in R (R Core Team, 2025), retaining visits with distances (d) £ median (d) + 3*MAD(d). The DBSCAN neighbourhood radius (eps) was set to the 95th percentile of the retained centroid distances, with a small 10% buffer to be conservative with residual GPS inaccuracy and the centroid approximation. We verified the Uganda delineation by comparing DBSCAN clusters to the site labels before fixing parameters for all countries (minPts = 5; eps = 191.7 m). Clusters were retained only if they contained ≥10 visits; each retained cluster defined a site identifier (site_uid). The response variable was *Biomphalaria* spp. presence at a visit, coded as a Bernoulli outcome (bio_pres = 1 if *Biomphalaria* spp. individuals >0, otherwise 0). Thus, bio_pres = 0 indicates that no *Biomphalaria* spp. were detected during that visit and does not establish true absence from the site. Although sites were visited repeatedly, the monitoring programme was designed for longitudinal surveillance, and visits were not organised as replicate surveys within predefined closure periods required to estimate occurrence state and detection probability separately.

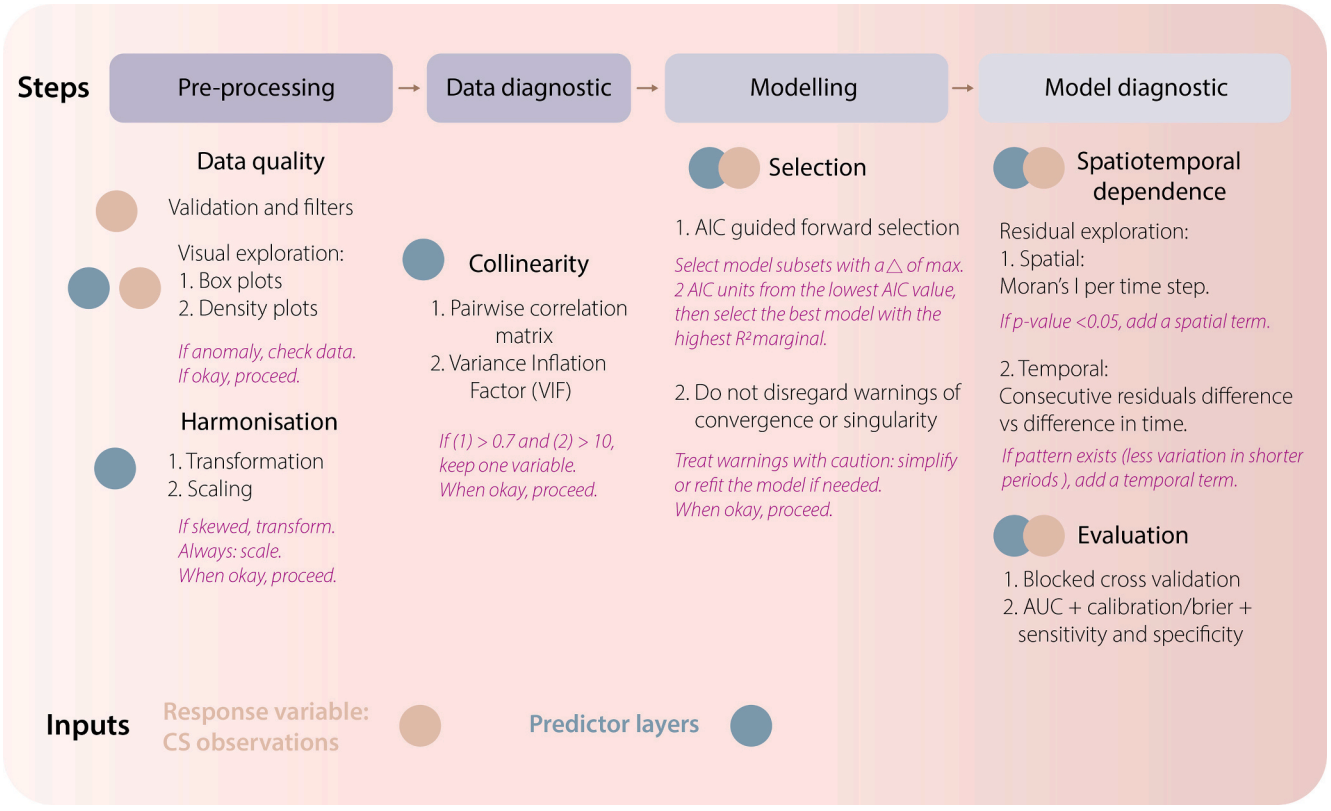


Fig. 2. Analytical workflow for modelling citizen science data. Presence/absence observations and predictor layers were processed through four steps: pre-processing, data diagnostics, modelling, model diagnostics and evaluation. Beige dots indicate steps using citizen science observations and blue dots indicate steps using predictor layers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Predictors

Predictors were grouped into six sets: geomorphological, climate, vegetation, land cover, anthropogenic and observation-related variables. Definitions, units, spatial and temporal resolution, and extraction geometry are reported in Table A2. Predictors were defined at two levels: 'visit-level' predictors were indexed to each sampling instance (site_uid per date), and could vary across visits within a site, whereas site-level predictors were summarised to a single value per site_uid (constant within site across visits in the model).

Geomorphological variables (site-level). Geomorphological predictors were derived from the Copernicus GLO-30 Digital Elevation Model (DEM) at a 1 arc-second resolution (Copernicus DEM - Global and European Digital Elevation Model | Copernicus Data Space Ecosystem, n.d.). For each site_uid, we built a convex hull around the visit coordinates assigned to the site and summarised raster values within this polygon using the 90th percentile (q90), following Valderrama-Bhraunxs et al., 2025c. Derived predictors included flow accumulation, topographic index and the slope steepness factor.

Climate variables (visit-level). Daily precipitation from CHIRPS (Funk et al., 2015) was aggregated as a 7-day antecedent cumulative rainfall to represent short-term hydrological conditions (Schwartz et al., 2020) relevant to snail habitats. Surface temperature was obtained from MODIS land-surface temperature products (Wan et al., 2021). Missing values in daily temperature were estimated using linear interpolation via the *imputeTS* package (Moritz and Bartz-Beielstein, 2017) in R. Precipitation and temperature were extracted at the site_uid centroid for each sampling date.

Vegetation variables (visit-level). Vegetation was represented using two NDVI products capturing complementary spatial and temporal supports. NDVI from Planet Labs (Planet Labs PBC, 2023) was indexed to each visit date by summarising (q90) NDVI within the site-specific convex hull polygon (Valderrama-Bhraunxs et al., 2025c). NDVI from MODIS [Product: MOD13Q1] (Didan et al., 2015) was indexed to each visit by extracting the mean NDVI within a 500 m buffer around the visit location. For each sampling event, candidate composites were searched within the 30 days preceding the sampling date, including the sampling date itself. The NDVI value was taken from the temporally closest composite that contained at least one valid pixel within the buffer. Missing Planet NDVI values were imputed using linear interpolation in R via the *imputeTS* package (Moritz and Bartz-Beielstein, 2017). We did not interpolate MODIS NDVI because MOD13Q1 is already a 16-day composite product and interpolation would impose additional smoothing beyond the product's own temporal aggregation.

Land cover variables (site-level). Land cover composition was derived within a 1 km buffer around each site centroid using Dynamic World (C. F. Brown et al., 2022) class histograms converted to proportional cover. Extraction windows were chosen based on periods with complete data availability (Uganda and Chad: 5-month windows; DRC: 1-year windows). Because land cover was extracted for multiple time windows, we summarised class proportions to a site-level median across available windows before modelling to avoid conflating observations density with apparent land cover variability. The 1 km buffer radius was selected as a harmonised local accessibility scale, supported by evidence indicating that rural households in Uganda commonly travel an average of 0.8 km to reach safe water (Naiga et al., 2015; Uganda Bureau of Statistics, 2010), acknowledging that this choice may not be equally appropriate across all contexts.

Anthropogenic variables (site-level). Population counts were derived from WorldPop (Tatem, 2017) layers within the same 1 km buffer. We calculated pop_rate_change as (population_2020 − population_2000) / population_2000 per site.

Observation-related variables (visit-level). We included (i) time of day at sampling (sampling time) to account for potential detectability

changes through day hours and (ii) sampling experience as a proxy for participation/observer learning over time. Sampling time was transformed into circular covariates: SAMTsin = sin(2π × t/86400) and SAMTcos = cos(2π × t/86400) (Valderrama-Bhraunxs et al., 2025c), where t is the sampling time in seconds since midnight (86,400 s). Sampling experience was defined as the cumulative number of submissions made by an observer up to and including the current visit. Higher values indicate a longer participation history within the monitoring program, which may increase sampling proficiency and hence detection likelihood (Kosmala et al., 2016).

2.5. Pre-processing: data quality and harmonisation

Data validation and filtering. Data quality control combined automated entry checks (form logic and field constraints) with verification of genus identification based on the reported photographs. Validation procedures followed ATRAP protocols described in Valderrama-Bhraunxs et al., 2025b and Brees et al. (2021). Reports were excluded based on pre-specified criteria (e.g. invalid dates/timestamps, inaccurate GPS readings, implausible sampling durations and records flagged during photo verification). The retained sample sizes and filtering rules are reported in Table A3.

Harmonisation, transformations and scaling. Predictors were extracted and aggregated to their target supports (centroid / convex hull / 1 km buffer; Table A2). To improve numerical stability and after a visual exploration, skewed variables were transformed using ln (x + 1) (Fig. A1), and all continuous predictors were standardised within each country (z-scoring). For exploratory checks, predictor distributions were inspected at the site level for static predictors and at the visit level for time varying predictors to avoid plots being dominated by heavily sampled sites. After the predictor-specific interpolation procedures described above, observations lacking a valid value for any predictor in the country-specific candidate set were excluded before model fitting, ensuring that all candidate models within a country were compared using the same observations. No additional imputation was performed.

2.6. Predictor diagnostic: collinearity

Because correlations among predictors can differ by context, collinearity was assessed separately within each country. We first inspected pairwise Pearson correlations, using $|\rho| > 0.7$ to identify potentially redundant predictor pairs (Dormann et al., 2013). Multivariate collinearity was subsequently assessed using variance inflation factors (VIF), using a threshold of VIF > 10 to indicate problematic collinearity (Zuur et al., 2009). Predictor exclusion was based on the combined evidence from these diagnostics and on conceptual redundancy among variables.

2.7. Statistical modelling workflow

We fitted separate generalised linear mixed models (GLMMs) for each country to estimate associations between predictors and *Biomphalaria* spp. occurrence while accounting for repeated sampling at the same site (Bolker et al., 2009). For visit *i* at site *j*:

$$y_{ij} \sim Bernoulli\left(p_{ij}\right)$$

$$logit\left(p_{ij}\right) = \beta_0 + \sum_k \beta_k X_{k.ij} + b_j$$

Where $X_{k.ij}$ are fixed effects from the predictor sets described in Section 2.4, and b_j is a site-specific random intercept. Models were fitted using the *glmer* function from the lme4 package (Bates et al., 2015).

2.7.1. Model selection

Candidate predictors (n) were defined a priori after preprocessing and collinearity screening. For each country, we evaluated all subsets of candidate fixed effects $(2^n - 1)$ and defined the supported set using DAIC ≤2 (Burnham and Anderson, 2004). Because sampling time was represented by circular terms (time_sin and time_cos), only models containing both components or neither were eligible for selection. Among eligible models within the supported set, we selected the model with the highest *marginal* R^2. A site-level random intercept (site_uid) was retained in all models.

2.7.2. Sensitivity and supplementary analysis

To assess whether unexplained heterogeneity reflected primarily differences between sites versus differences between observers, we refitted each selected representative model using a random intercept reflecting observers (recordedBy) instead of the site level random intercept (site_uid). Models including site and observer as a random nested effect were attempted but were numerically unstable (non-convergence and singular fits) and are therefore not reported.

Because sampling experience increased over the sampling period and was correlated with Julian day, we conducted a targeted sensitivity check on the selected models by replacing the raw experience term with a time-adjusted sampling experience index. Specifically, we regressed sampling experience on Julian day, $sampling_experience_i = \alpha + \beta^* JulianDay_i + e_i$. Here (e_i) represents variation in participation not explained by the linear progression of Julian date. We then refitted the selected models replacing raw sampling experience with exp_resid and including Julian date as a separate fixed effect. For Chad, where an abrupt shift was evident towards the end of the monitoring period, we additionally repeated the analysis after excluding the final four months to assess the sensitivity of the experience and temporal effects to this flood-affected period.

2.7.3. Model diagnostics and predictive evaluation

Model adequacy was assessed using DHARMa simulation-based scaled residuals (Hartig, 2025). Spatial autocorrelation was tested with a one-sided Moran's I (Moran, 1950), and computed on centred residuals (scaled residuals - 0.5) within each week and across cumulative distance thresholds from 1 to 10 km. Weeks were indexed as the number of weeks since the first sampling date in each dataset. Spatial weights were row-standardised, so that for each observation the weights assigned to its neighbours sum to 1 (Anselin and Bera, 1998). To account for multiple testing across week-distance combinations, the resulting *p-values* were adjusted within each country using the Benjamini-Hochberg false discovery rate procedure (Benjamini and Hochberg, 1995); combinations with the *adjusted p-value* < 0.05 were flagged as evidence of residual spatial dependence. When residual spatial dependence was detected, we refitted the model including a spatial smooth with the function *bam* from the package *mgcv* (S. Wood, 2011) to account for remaining spatial structure. Temporal autocorrelation implies that residuals closer in time tend to be more similar than residuals from observations separated by longer lags (Bence, 1995). Therefore, we explored within-site temporal structure by pairing each observation with the immediately preceding visit at the same site. For each consecutive pair, we calculated the days between visits and the absolute difference between centred DHARMa residuals. We then plotted residual differences against elapsed time (per site) to evaluate whether residual similarity decreased with increasing time between visits.

Predictive performance was evaluated using two complementary five-fold cross-validation schemes representing different prediction targets. First, known-site cross-validation evaluated prediction of new observations at previously monitored sites. Within each site_uid, observations were partitioned into five folds; models were trained on four folds and evaluated on the held-out fold. Second, site-blocked cross-validation evaluated prediction at previously unsampled sites. Entire sites were assigned to folds, with approximate stratification according to whether a site contained any presence records. Models were fitted after excluding all observations from the held-out sites, and predictions for these sites used fixed effects only because no site-specific random

intercept was available.

Performance was assessed primarily with the Brier score (Brier, 1950). To account for differences in prevalence among countries and folds, model performance was compared with a prevalence-only reference prediction based on the training data and expressed as the Brier Skill Score (BSS), $\frac{1-BS_{model}}{BS_{ref}}$ (Murphy 1973). Calibration of site-blocked predictions was evaluated by comparing mean predicted probabilities with observed occurrence frequencies across five equal-frequency bins (Steyerberg et al., 2010). ROC-AUC (Hanley and McNeil, 1982) was retained as a secondary measure of discrimination.

Analyses were conducted in R and Python. Spatial processing and covariate extraction in QGIS and Google Earth Engine. All scripts for site clustering, covariate extraction, model fitting, diagnostics and cross-validation are referenced in the Data Availability statement.

3. Results

3.1. Data structure and site level imbalance

After filtering and DBSCAN-based site delineation, the final datasets comprised 4770 visits across 70 sites in Uganda, 3934 visits across 75 sites in DRC, and 685 visits across 29 sites in Chad. Site-level occurrence balance differed markedly among countries (Fig. 3). Absences predominated at 28 of 29 sites in Chad (96.6%) and 66 of 75 sites in DRC (88.0%), compared with 29 of 70 sites in Uganda (41.4%). Absence-only sites were also more frequent in Chad (22/29; 75.9%) and DRC (32/75; 42.7%) than in Uganda (10/70; 14.3%).

3.2. Collinearity screening

Pairwise correlation screening and VIF diagnostics were conducted separately per country (Table A4, Fig. A2). Topographic index (TI) and flow accumulation showed consistent redundancy across countries, and, given their conceptual dependence, TI was removed from the common candidate set. In Uganda, elevation and proportion water were strongly correlated, but they did not co-occur in any model within the ΔAIC ≤2 supported set.

3.3. Final country models differed in structure and random effect dominates the fit

Final GLMMs selected under ΔAIC ≤2 differed in which predictor sets were retained across countries (Fig. 4a). Coefficient estimates were also stable across the ΔAIC ≤2 supported model sets (Fig. A3): all displayed predictors retained the same effect direction across the models in which they occurred, with high similarity between-model variation in DRC and Chad, while Uganda showed greater variation in the strength of support for some predictors. In models with a site-level random intercept (site_uid), conditional R^2 substantially exceeded marginal R^2 (Table 1), indicating that unmeasured site-level heterogeneity accounted for a large share of variation beyond the fixed covariates. Replacing the site-level random intercept with an observer-level random intercept, while holding the fixed effect structure constant, reduced conditional R^2 in Uganda (0.72 to 0.42) and DRC (0.79 to 0.57), with little change in Chad (0.98 in both specifications; Table A5). AIC was also higher under the observer specification in all three countries, supporting site-level grouping as the more appropriate representation of residual heterogeneity.

Sampling experience increased mechanically over the monitoring timeline and was strongly correlated with Julian date in all three countries (ρ > 0.7) (Fig. A4). We therefore refitted the selected models using a time-adjusted experience index (exp_resid), representing variation in cumulative experience not explained by linear calendar-time progression (see section 2.7.2). Time-adjusted experience remained negatively associated with occurrence in Uganda (OR = 0.96 per +0.1; 95% CI 0.93–0.99; $p = 0.014$), DRC (OR = 0.95; 0.92–0.99; $p = 0.006$) and Chad (OR = 0.70; 0.53–0.93; $p = 0.015$) (Fig. A5). Julian date was negatively associated with occurrence in DRC (OR = 0.96 per 30 days; 0.95–0.98; $p < 0.001$) and Chad (OR = 0.47; 0.28–0.78; $p = 0.003$) but was not supported in Uganda ($p = 0.196$).

In Chad, the late monitoring period coincided with increased rainfall, reduced reporting during flooded conditions and near-zero observed prevalence (Fig. A6). After excluding the flood-affected final

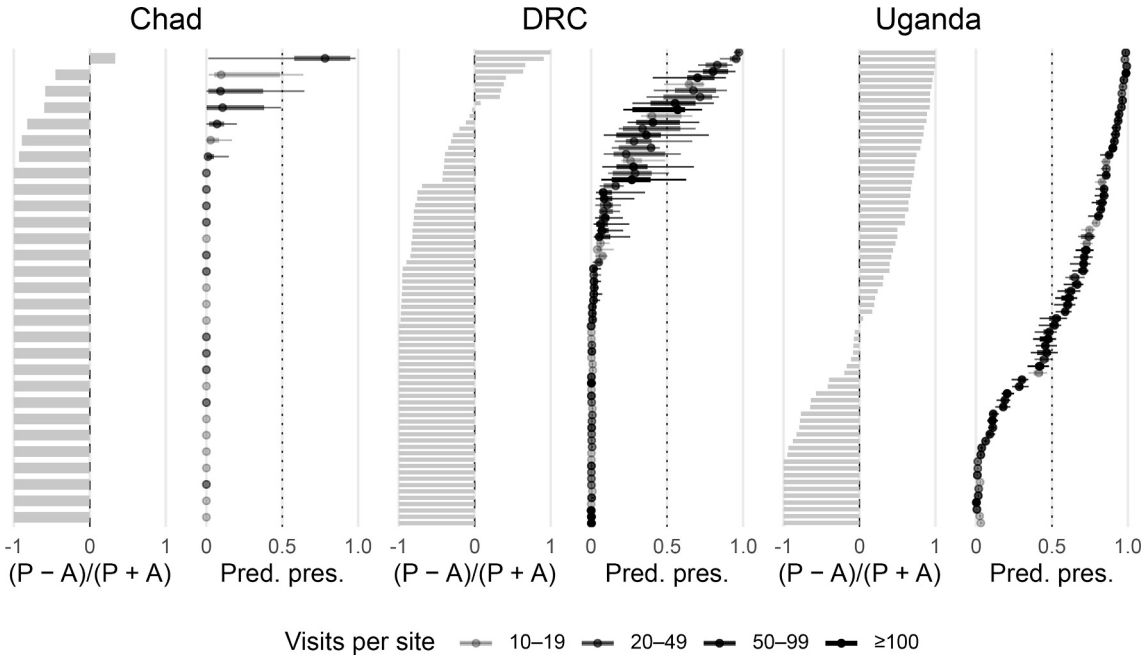


Fig. 3. Site-level class imbalance and predicted occurrence across countries. For each country (Chad, DRC and Uganda), the left panel shows the observed presence/absence balance per site, calculated as (P-A)/(P + A) (−1: all absences; 1: all presences; 0: perfect balance). The right panel summarised the distribution of the predicted probabilities of presence for the same sites (point: median; thick line: interquartile range; thin line: 5–95% range; dashed line: 0.5). Sites are ordered within each country by observed balance. Point opacity indicates the number of visits per site (lighter: fewer visits, darker: more visits; legend show reference values).

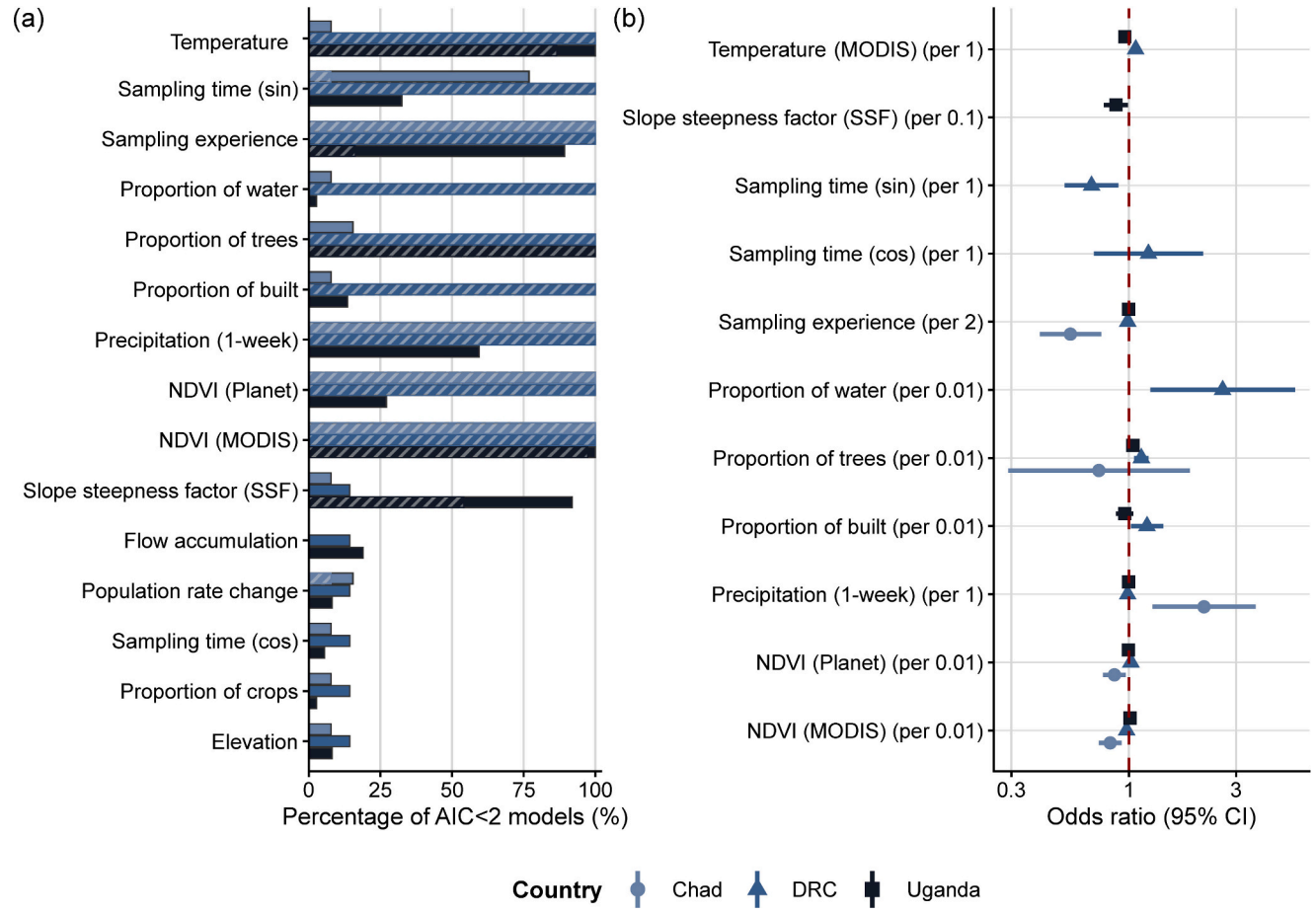


Fig. 4. Predictor selection and odds ratios across country-specific *Biomphalaria* spp. occurrence models. (a) Percentage of AIC-ranked models retaining each predictor, with striped sections indicating significant effects. (b) Odds ratios with 95% confidence intervals for predictors retained in the selected representative model per country.

Table 1
Fixed effect structure, random effect term, marginal and conditional R^2, and AIC of the selected representative model with a site random effect in each country.

Country	Fixed effects	Random effects	R^2m	R^2c	AIC
Uganda	Climate: temp_modis, pp_oneweek | Geomorphology: SS | Land cover: prop_trees_med, prop_built_med | Vegetation: NDVI_Modis, NDVI_Planet| Observation: sampling_experience		0.15	0.72	4072.22
DRC	Climate: pp_oneweek, temp_modis | Land cover: prop_water_med, prop_trees_med, prop_built_med | Vegetation: NDVI_Planet, NDVI_modis | Observation: sampling_experience, time_sin, time_cos	(1 | site_uid)	0.17	0.79	1960.9
Chad	Climate: pp_oneweek | Land cover: prop_trees_med | Vegetation: NDVI_Planet, NDVI_modis | Observation: sampling_experience		0.14	0.98	173.6

months, the negative association with raw sampling experience remained supported ($p = 0.006$). It also persisted after separating sampling experience from calendar time ($p = 0.010$), whereas Julian date was no longer supported ($p = 0.135$; Table A6). These results indicate that the late flood-affected period contributed to the overall temporal decline in Chad but not completely explained the negative sampling-experience association.

3.4. Diagnostics and predictive performance

3.4.1. Residual structure

Simulation-based residual diagnostics provided little evidence of systematic within-site temporal dependence (Fig. 5a), with mean absolute residual differences between consecutive visits showing no consistent increase at short lags. In contrast, residual spatial autocorrelation differed by country. Moran's I computed on weekly aggregated residuals was generally close to zero in Uganda and Chad but showed positive spatial autocorrelation in DRC (Fig. 5b).

In a sensitivity analysis, adding a spatial smoothing term reduced these peaks in DRC (Fig. A8), but the signal was not persistent across weeks and did not change the selected fixed effects. All coefficients retained their direction, and inferential support was unchanged for nine of the ten terms (Fig. A9). The largest proportional change occurred for proportion of water (23.6%), which remained positive and supported. Planet NDVI was the only term whose 95% confidence interval changed from narrowly excluding zero to narrowly overlapping zero. Given this similarity in the results, we proceeded with the less complex model in the core modelling workflow.

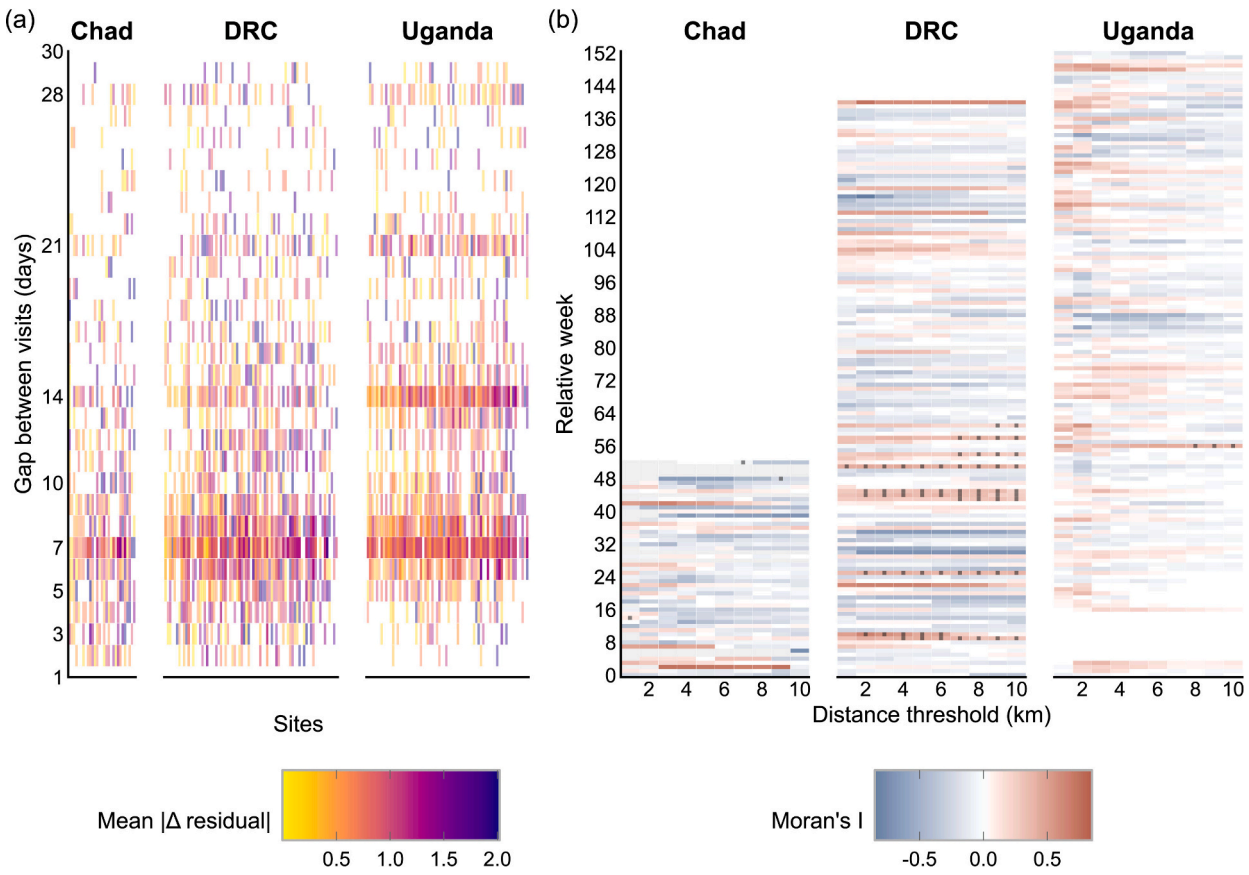


Fig. 5. Residual structure diagnostics (a) Within-site temporal structure shown as the mean absolute change in scaled residuals between consecutive site visits as a function of the elapsed time (days). Each tile represents one site-lag combination (yellow colours: smaller residual changes; purple colours: larger residual changes). (b) Weekly spatial autocorrelation of model residuals quantified by Moran's I across a range of distance threshold (km). Colours indicate Moran's I (blue: negative, red: positive); black dots mark week-distance combinations with statistically significant autocorrelation after multiple testing correction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4.2. Predictive performance at known and unseen sites

Known-site cross-validation showed substantial improvement over prevalence-only prediction in all three countries (Fig. 6a). Pooled Brier skill was 0.47 in Uganda, 0.47 in DRC and 0.54 in Chad, with corresponding ROC-AUC values of 0.89, 0.93 and 0.95. Thus, models predicted new observations at sites already represented during model fitting considerably better than a prevalence-only benchmark.

Performance declined markedly when entire sites were withheld. Site-blocked Brier skill was close to or below zero in Uganda (−0.02), DRC (−0.09) and Chad (−0.04), indicating no improvement over the prevalence-only benchmark for occurrence at previously unsampled sites. Discrimination also decreased, with ROC-AUC values of 0.60, 0.59 and 0.54, respectively. Calibration of the site-blocked predictions showed substantial departures from the 1:1 relationship, with particularly narrow predicted-probability ranges in Chad and DRC (Fig. 6b, see axis). These results indicate that the models were substantially more informative for repeated observations at monitored sites than for extrapolation to new sites, consistent with the large contribution of site-level heterogeneity to model fit reported above.

4. Discussion

4.1. One size does not fit all sites: snail microhabitats are highly dynamic

Across all sites in the three countries, *conditional* R^2 far exceeded *marginal* R^2, indicating that a large share of the modelled variation in *Biomphalaria* spp. occurrence was associated with persistent differences among sites rather than with the measured fixed effects. The site-level random intercept can therefore be interpreted as a site baseline that absorbs microhabitat characteristics that are difficult to represent with the available covariates (e.g. depth, substrate, aquatic vegetation structure, and local hydrodynamics), as well as practical constraints such as site accessibility. Supporting this interpretation, replacing the site random intercept (site_uid) with an observer random intercept (recordedBy) did not increase explained variation: *conditional* R^2 dropped substantially in Uganda and DRC, and was similar only in Chad (Table 1, Table A5). Moreover, in a previous expert-only analysis involving a single observer, the site random effect still dominated model fit (Valderrama-Bhraunxs et al., 2025c), reinforcing the interpretation that the high *conditional* R^2 primarily reflects genuine site-level heterogeneity, and not necessarily observer differences.

Imperfect detection is inherent to ecological field surveys and is not unique to citizen science. The reliability of the observation process used here has been evaluated previously in the Ugandan ATRAP network using expert-collected data. In *Great minds map alike* (Valderrama-Bhraunxs et al., 2025c), citizen- and expert-derived models showed strong agreement in the ecological patterns associated with *Biomphalaria* spp. occurrence, while an additional imperfect-detection analysis found generally small discrepancies between expert-derived occurrence probabilities and citizen observations. The remaining discrepancies were primarily associated with observer identity and sampling effort. However, this validation was conducted in Uganda and should not be interpreted as equivalent independent validation of the DRC and Chad networks.

This strong site dominance is biologically plausible because *Biomphalaria* spp. populations are controlled by fine-scale and temporally

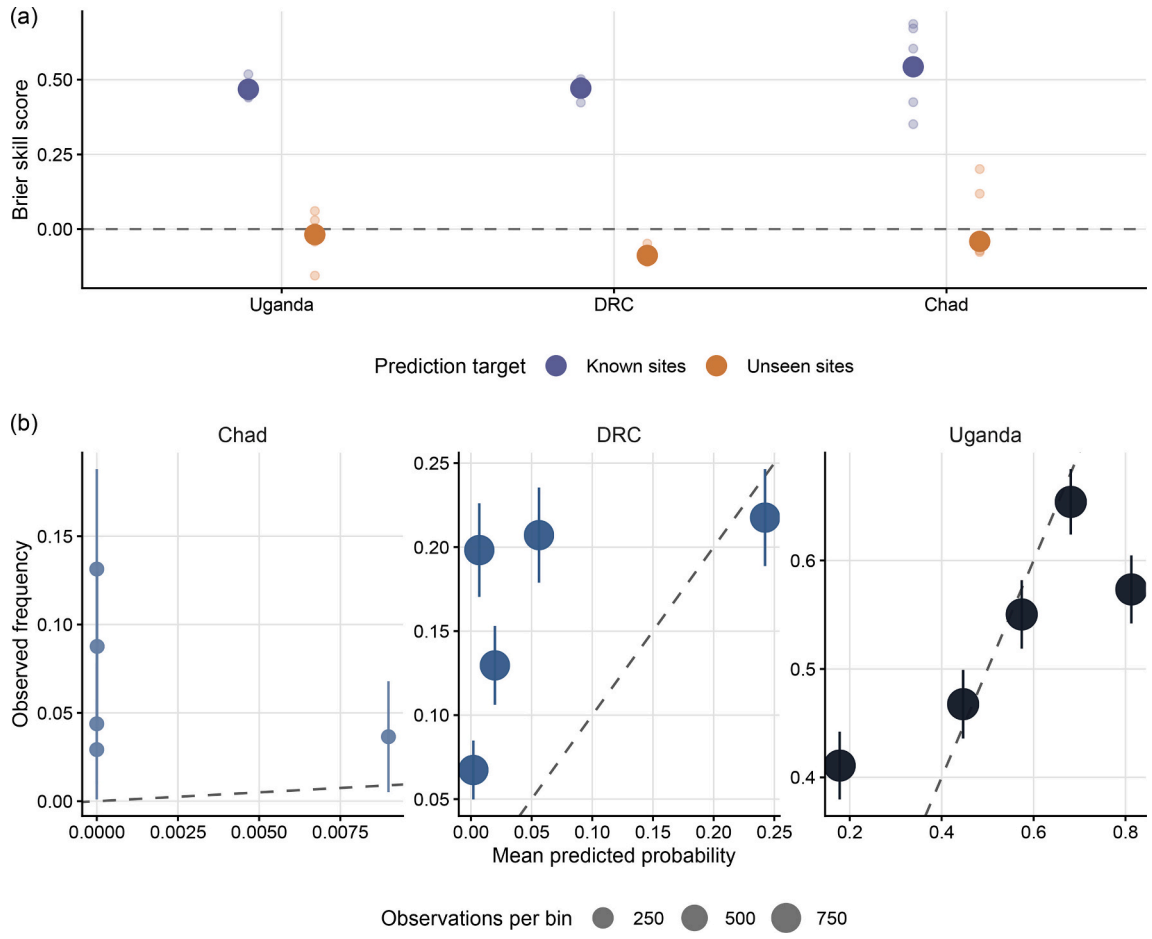


Fig. 6. Cross-validated predictive performance at known and unseen sites. (a) Brier skill relative to a prevalence-only benchmark; small points show folds and large points pooled estimates. Values >0 indicate improvement, 0 equivalent performance and < 0 poorer performance. (b) Calibration of unseen sites predictions across five equal-frequency bins; point size indicates observations per bin and the dashed line perfect calibration.

variable microhabitat conditions: water depth, substrate types, vegetation, and water flow (Utzinger and Tanner, 2000); which can vary over meters and days, often below the effective resolution of remotely sensed predictors. Field studies describe highly patchy distributions linked to aquatic vegetation and short local persistence, with seasonal peaks and redistribution during rainy periods; together, these dynamics can generate periodic windows of host availability (Woolhouse and Chandiwana, 1989). This patchiness also helps to explain why single malacological surveys often underperform: intensive sampling is required to characterize highly clustered, temporally variable snail presence, but that level of effort is usually infeasible for routine control programs (C. L. Wood et al., 2019).

The retained environmental associations provide additional ecological context to this site-specificity. Temperature was negatively associated with *Biomphalaria* spp. occurrence in Uganda but positively associated in DRC, and precipitation also differed in direction among countries, consistent with evidence that *Biomphalaria* spp. responses to climatic and hydrological conditions vary across ecological settings (Perez-Saez et al., 2016). NDVI showed similar context dependence. In our previous analysis around Lake Albert, its association with *Biomphalaria* spp. varied among habitat types, having a negative relationship with snail occurrence around lake sites, probably because NDVI near freshwater can represent a composite signal of terrestrial, aquatic vegetation and changing water extent (Valderrama-Bhraunxs et al., 2025c).

Some of these differences may also reflect variation in the *Biomphalaria* species represented within each genus-level dataset. For example, *B. sudanica* and *B. stanleyi* can occur within the same broader landscape yet differ in their associations with shoreline habitat, water depth and seasonal flooding (Iacovidou et al., 2026). Because our observations were identified at genus level, these mechanisms cannot be separated directly, and the country-specific coefficients should therefore be interpreted as context-dependent associations rather than general ecological responses of *Biomphalaria* spp. Importantly, resolving all of this fine-scale ecological heterogeneity was not necessary for useful prediction at monitored sites: the site random intercept captured persistent local differences that were not represented by the available predictors, and known-site cross-validation showed substantial predictive skill. The sharp decline in performance at previously unsampled sites, however, indicates that this unresolved heterogeneity becomes important when extrapolating beyond the monitored network.

Patchy and dynamic host distributions are a key ingredient of the characteristic focal transmission of schistosomiasis because exposure is concentrated around particular water-contact locations and varies across space and time even within broadly suitable areas (Aula et al., 2021). *Biomphalaria* spp. presence should not be treated as equivalent to transmission, however, it is a necessary condition for disease risk and a proxy for potential transmission; whereas actual transmission additionally depends on parasite presence, snail-parasite compatibility, and human-water contact behaviour (Aula et al., 2021; Stothard et al., 2017).

In conventional surveillance programs, where decisions are often based on *snapshot* surveys, this disease/host ecology creates a practical problem: a one-time monitoring can easily become outdated when

snails' microhabitats appear, disappear, or relocate over short distances and periods. This matters for intermediate host control because focal molluscicide is still part of WHO recommendations for schistosomiasis control and its effectiveness depends on targeting the right habitats at the right time (WHO, 2022), yet routine programs often rely on infrequent surveys that can quickly become outdated. Repeated monitoring of a defined set of human-water contact sites, therefore, can be more operationally useful than isolated surveys, because it provides an updated representation of where host snails are currently present. In other words, for a focal disease such as schistosomiasis, citizen science is most valuable as a fine-scale monitoring approach that maintains an updated picture of where intermediate host occurrence persists or re-emerges.

4.2. One size does not fit all periods: changing conditions and program dynamics

Sampling experience, defined here as cumulative number of submissions made by an observer, increased throughout the monitoring period. Unexpectedly, it was negatively associated with *Biomphalaria* spp. occurrence in all three networks. Under our original hypothesis, greater experience would be expected to increase (or at least not decrease) the probability of detecting snails under comparable ecological and sampling conditions. This expectation is consistent with citizen science studies showing that observation skills can improve with continued participation and that more experienced observers tend to report species at higher rates (Johnston et al., 2018; Kelling et al., 2015). However, our presence/absence outcome represents observed detection during a visit and does not identify latent occupancy: a non-detection cannot establish that organisms were truly absent from the sampled site (MacKenzie et al., 2017).

Cumulative sampling experience should therefore be interpreted as a component of the observation process, not a direct measure of observer proficiency. Because cumulative experience necessarily increased as monitoring progressed, it could also capture temporal changes unrelated to observer learning. Raw experience was strongly correlated with calendar time in all three countries ($\rho > 0.7$, Fig. A4), yet the negative association with time-adjusted experience remained supported in all three settings.

The temporal context nevertheless differed among countries. In Chad, the final monitoring period coincided with exceptional flooding (*Chad Situation Report #6 - December 2024 - Chad | ReliefWeb.*, n.d.), reduced reporting and a sharp decline in observed snail occurrence (Fig. A6). The original Chad model showed a strong negative Julian-date effect; however, after excluding the flood-affected final months, Julian date was no longer supported, whereas the negative time-adjusted experience association persisted. Thus, flooding contributed to the overall temporal decline but did not account for the experience association itself. In DRC, both Julian date and time-adjusted experience remained negatively associated with occurrence, indicating an overall temporal decline together with additional variation associated with participation history. In Uganda, time-adjusted experience remained negatively associated with occurrence without evidence of an independent linear decline with Julian date.

The recurrence of the experience association across settings does not identify its underlying mechanism. Several components of observation-related processes may contribute simultaneously: observer experience, previous encounters with the focal species and day of year have all been linked to non-detection probabilities in citizen science data (Jacobusse and Jongejans, 2026). Repeated visits also do not necessarily sample the exact same microhabitat (August et al., 2020), which is particularly relevant here because sites represented spatial clusters of repeated visits, allowing the precise area sampled within a site to vary among visits. At the same time, access constraints can alter the areas searched (Mair and Ruete, 2016; Sicacha-Parada et al., 2021); while local *Biomphalaria* spp. abundance can vary seasonally and during extreme events (Iacovidou et al., 2025), directly affecting encounter probability (Tumusiime et al., 2024).

Participation patterns provide an additional layer of complexity. Reporting frequency declined through time in both DRC and Uganda (Fig. A7), showing that monitoring intensity changed as the programs progressed. Similar dynamics have been described in other citizen science programs, where participation may peak early and later become concentrated among a smaller subset of contributors (Fischer et al., 2021). Motivation can also evolve: novelty and learning may help initiate engagement, whereas sustained participation is more strongly supported by perceived impact, recognition, and timely feedback (Rotman et al., 2012).

Separating these processes requires information that describes sampling independently of cumulative participation. Learning could be evaluated through repeated standardised identification or search exercises, or simultaneous surveys involving observers with different experience levels under comparable conditions. Visit-level effort, observer history, previous encounters with the focal taxon and seasonal conditions can improve the characterisation of detection (Jacobusse and Jongejans, 2026; Johnston et al., 2018, 2023). Hydrological conditions, extreme events and temporary access restrictions would help identify changes shared across observers. Precise visit locations and basic microhabitat information, including shoreline position, water depth, aquatic vegetation and substrate, would help identify shifts in the habitat actually searched within a nominal site. These additions are highly relevant for *Biomphalaria* spp., given the strong fine-scale structure of its habitat use.

4.3. One size does not fit all contexts: scaling requires workflow transfer, not predictor transfer

In this study, *transfer* means applying the citizen science workflow that was first set up in Uganda to other countries (DRC and later Chad). Core elements included shared field sampling guidance, training and support, reporting with required metadata and photos, and a common modelling, evaluation and diagnostic framework, while practical details, especially site definition, were adapted to fit local landscapes and logistics. The main lesson is that workflows can be transferred, but predictor relationships and reporting patterns depend on the context, since each setting shapes snail ecology and how observations are done.

The contrast among regions makes this easier to understand. In Uganda, monitoring focused on relatively stable locations around Lake Albert with multiple site types (lake shores, wetlands, streams, and springs) with more balanced snail presence/absence data (Tumusiime et al., 2024; Valderrama-Bhraunxs et al., 2025b). In DRC, sampling was concentrated along the Lukunga River corridor and recorded more absence-dominated data (Valderrama-Bhraunxs et al., 2025b). The episodic spatial autocorrelation observed in DRC may partly reflect spatially structured conditions along this corridor, including local hydrology, water quality or other microhabitat characteristics not represented by the available predictors. Monitoring in Chad took place in a strongly seasonal lakeshore system with pronounced wet/dry contrasts (Pham-Duc et al., 2020), where shifting access and moving water margins also influenced where sampling could be conducted within monitored areas; *Biomphalaria* spp. detections were also very rare. These differences matter because they shape what a *site* represents in practice, how effort is distributed through time, and how environmental covariates should be interpreted. For example, weekly precipitation may primarily reflect habitat conditions in one setting, but in another it may also capture periods when sites become difficult to reach or when sampling shifts within a site footprint.

For scaling intermediate host monitoring, these results indicate that the transferable value lies less in exporting predictor coefficients than in applying a workflow that maintains comparability while accommodating local conditions. Across the three settings, the main sources of uncertainty identified in this study arose from site definition and fine-scale

habitat heterogeneity, changing access and environmental conditions, observer and participation processes, and differences in occurrence prevalence. Based on these findings, Fig. 7 summarises the corresponding practical safeguards that can be incorporated into citizen science monitoring workflows. These include defining sites as spatial footprints, access and effort metadata, photo verification for presences, and periodic checks of presence/absence imbalance with multiple observers help prevent ecological inference from being confounded by detection opportunity.

5. Conclusion

Across the three citizen science networks evaluated here, *Biomphalaria* spp. occurrence was strongly structured by site-level heterogeneity, while the importance and direction of measured environmental predictors varied among settings. This reinforces that ecological relationships identified from citizen-generated monitoring data should be interpreted in their local context. The models nevertheless showed substantial predictive skill for repeated observations at previously monitored sites, supporting the value of longitudinal surveillance of defined human-water contact locations, while their limited performance at previously unsampled sites cautions against extrapolation beyond the monitored network.

Temporal patterns further showed that cumulative 'experience' is not a simple indicator of learning, it is correlated with time, and it can absorb changes in accessibility, motivation, effort and the microhabitat patch that is sampled. For scaling, the transferable value lies in a shared workflow that explicitly considers the main sources of uncertainty (Fig. 7). With these safeguards in place, citizen science can provide a scalable, repeated information on intermediate host occurrence for schistosomiasis surveillance in rural sub-Saharan Africa, complementing conventional surveillance where coverage and sampling frequency remain limited.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT 5.5 to improve text readability and support code optimisation, specifically debugging and refactoring. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Noelia Valderrama-Bhraunxs: Writing – review & editing, Writing – original draft, Software, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Larissa Bonifacio:** Writing – review & editing, Data curation. **Tine Huyse:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Helena Greter:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jacob Mbaihondoum:** Writing – review & editing, Supervision. **Annour Adoum Batil:** Writing – review & editing, Supervision. **Emiel Van Loon:** Writing – review & editing, Supervision, Conceptualization. **Anton Van Rompaey:** Writing – review & editing, Supervision. **Joule Madinga:** Writing – review & editing, Supervision. **Ronald Twongyirwe:** Writing – review & editing, Supervision. **Liesbet Jacobs:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

This work was supported by the Belgian Directorate General for Development, Cooperation and Humanitarian Aid (XM-DAC-2-10-3852,

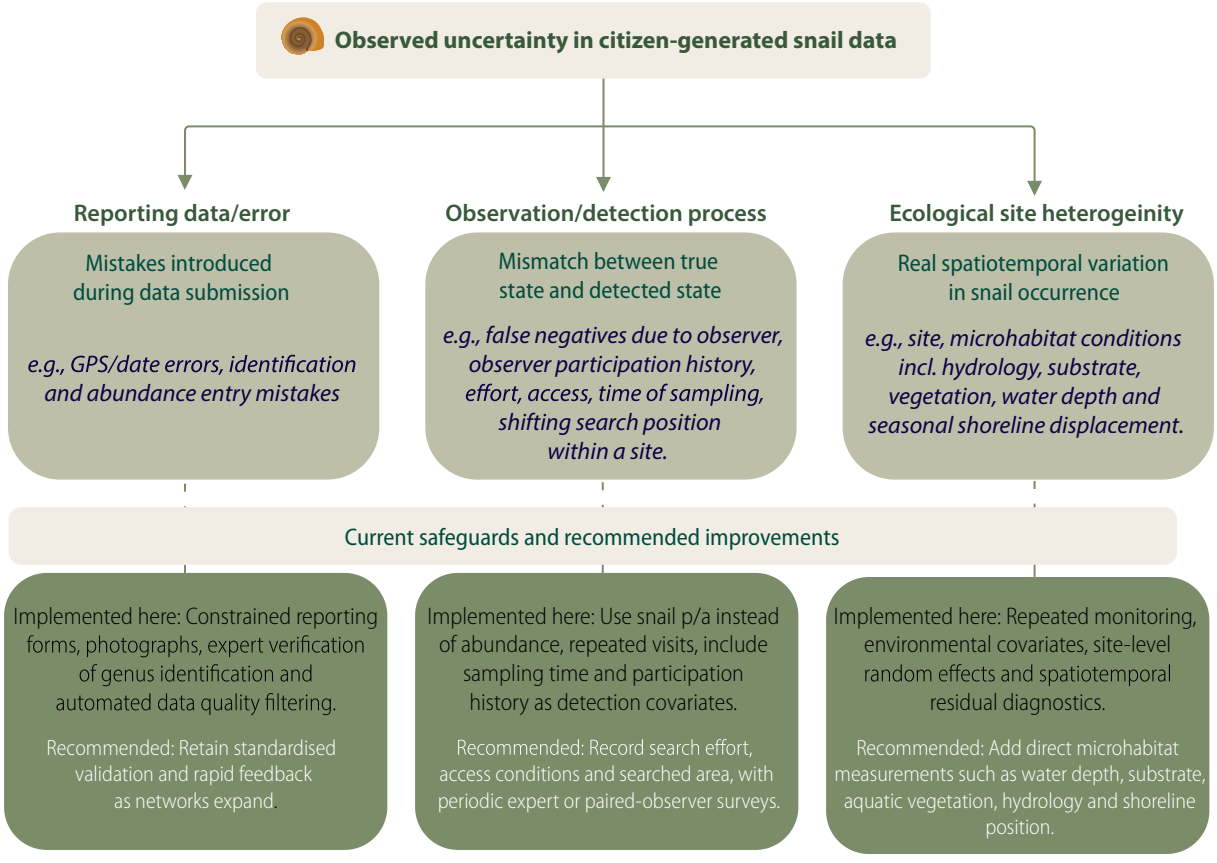


Fig. 7. Sources of uncertainty in citizen-generated snail occurrence data and corresponding safeguards.

XM- DAC-2-10-3853), the Swiss Network for International Studies (SNIS C22062) and the Fonds Wetenschappelijk Onderzoek FWO (11L3223N and 11L3225N) for NVB's PhD fellowship.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all the citizen scientists from the ATRAP and ParaSahel network for their invaluable contributions: In Uganda: Alinda Hassan, Chotum Friday, Atanasi Marisel, Opio Isingoma, Ategeka Augustine, Bahemuka Bob, Masereka Haruna, Sebakara Fobius, Nyamahunge Imelda, Businge Zabron, Tumusiime Janet, Nsenga David, Night Marygoret, Kyaligonza Noeline, Mwesige Robert, Barisigara Gard, Tusingwire Henry, Natabi Specioza, Nakingi Rose, Bamuturaki Charles, Kamukama Josias, Nuwagaba Emmanuel, Tweheyo Julius, Bahungirehe Crezestom, Ategeka Rogers, Unimu Hadijah, and Fuwarinyo Richard. In DRC: Bulezi Jephirin, Dianzenza Héritier, Diatuke Dinayame, Kiangebeni Mbuta, Dikizekiyo Paul, Kalumpuniko Siluafundiswa, Kavungu Zolan- donga, Kinsengwa Lukwikilu, Landu Lusevakueno, Landu Tazi, Lubanzadio Adèle, Lufuma Lwakanda, Luntadila Dayina, Lusevakweno N'sanda, Luvuma Swamina, Luzolo Matufwene, Mansweki Makuntima, Mantantu Muanga, Mayawila, Mabanza Nlandu, Mena Ntonzi, Ndongala Gracia, Nginamawu Luvevola, Nkumbu Ngina, Nsilulu Zanzambi, Nzinga Emmanuel, Puabu Lusimana, Vantour Kinuani, and Zoladio Sienge. In Chad: Younouss Mahamat, Goudja Saleh, Abdoulaye Mahamat, Mahamat Adam, Djou François, Idriss Alawi, Abdoulaye Awan, Roasngar Raphaél, Djibrine Kamanda, Abakar Oumar Abakar, Mohamat Saleh, Ibrahim Moussa, Abakar Adam, Abdraman Ramadan, Idriss Abakar, Moustapha Hassan, Tom Idriss, Abakar Issa, Djidda Abdraman, Idriss Bineye, Moussa Abdoulaye, Kelli Hisseine, Mbaionbé Obed, Souleyemane Sénoussi, Abdoulaye Alhadje Obi, Mahamat Kelli Abakar, Abbass Idriss, Haoua Issa, Saleh Mbang, Hassan Ali Soumaila. We also would like to acknowledge Julius Tumusiime, Daisy Namirembe, Germain Kapour, Theo Emboni, Vanessa Djetouroum, Roland Naingam and Ali Abba Abakar for their support to the citizen science networks.

Appendix A. Appendices

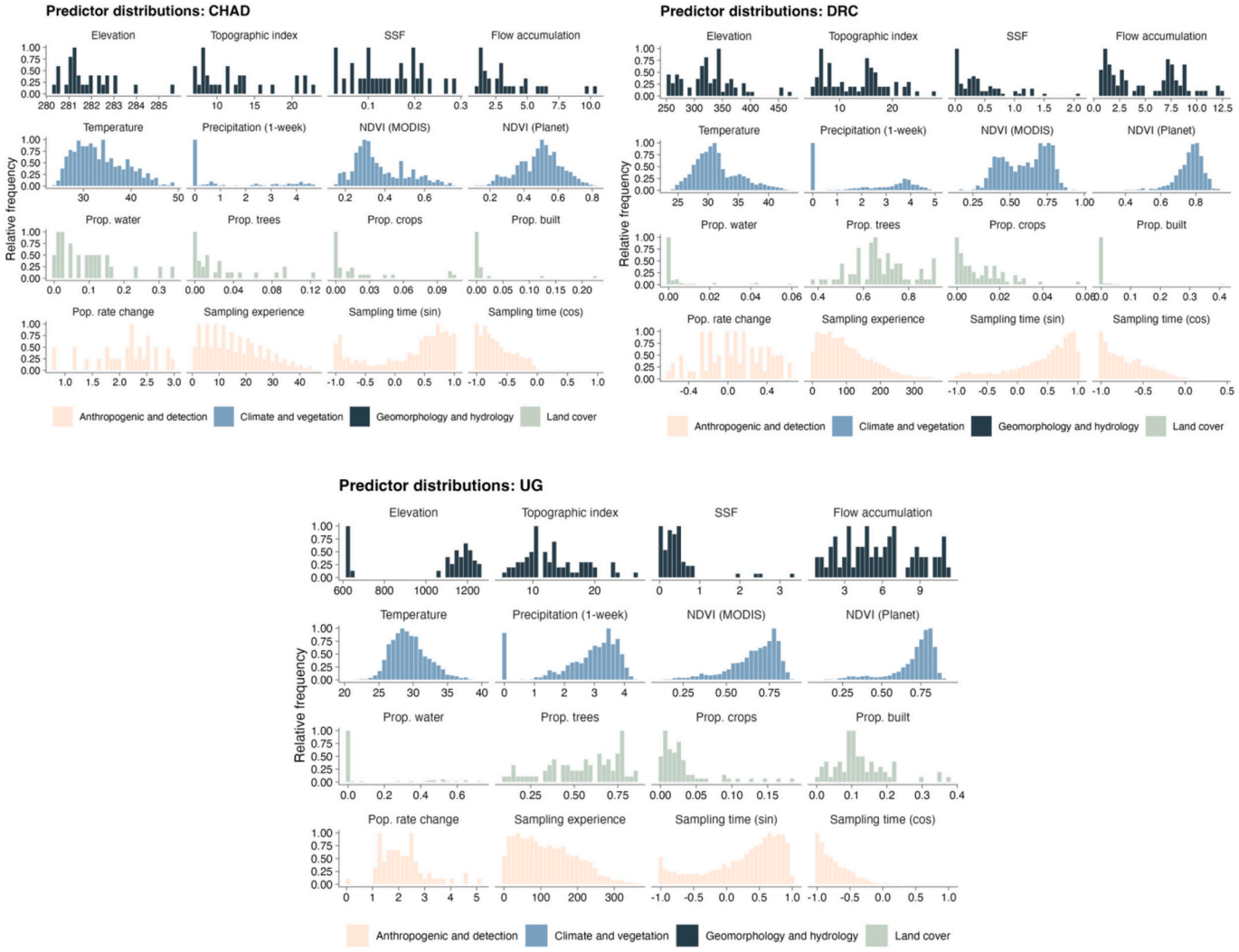


Fig. A1. Distributions of predictor variables by country (Chad, DRC and Uganda) after preprocessing. Histograms show relative frequencies for geomorphological, climatic, vegetation, land-cover, anthropogenic and observation covariates. Flow accumulation was log-transformed as ln (x + 1) for visualisation.

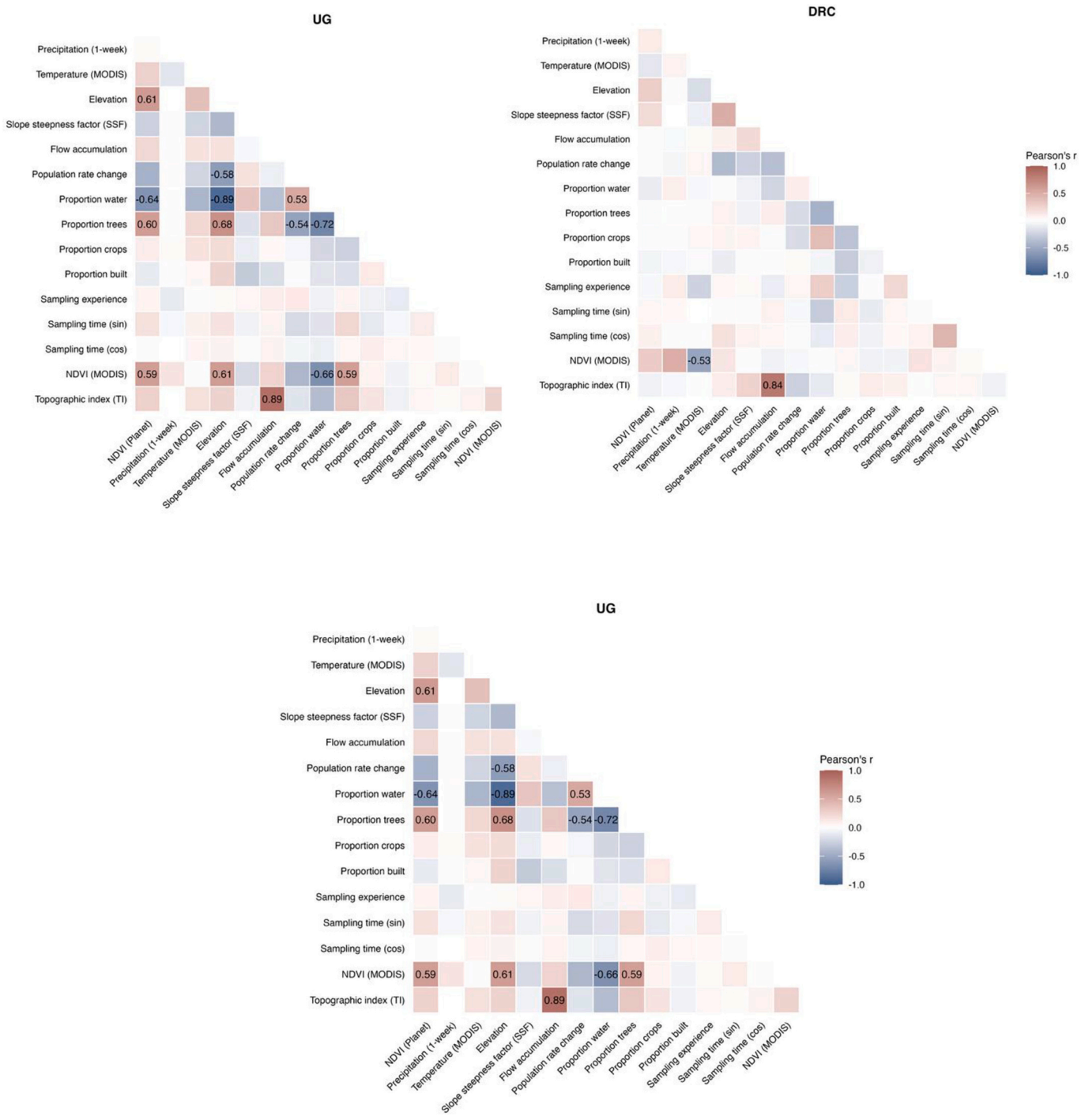


Fig. A2. Pearson correlation matrices of predictor variables by country, used to assess pairwise collinearity before modelling.

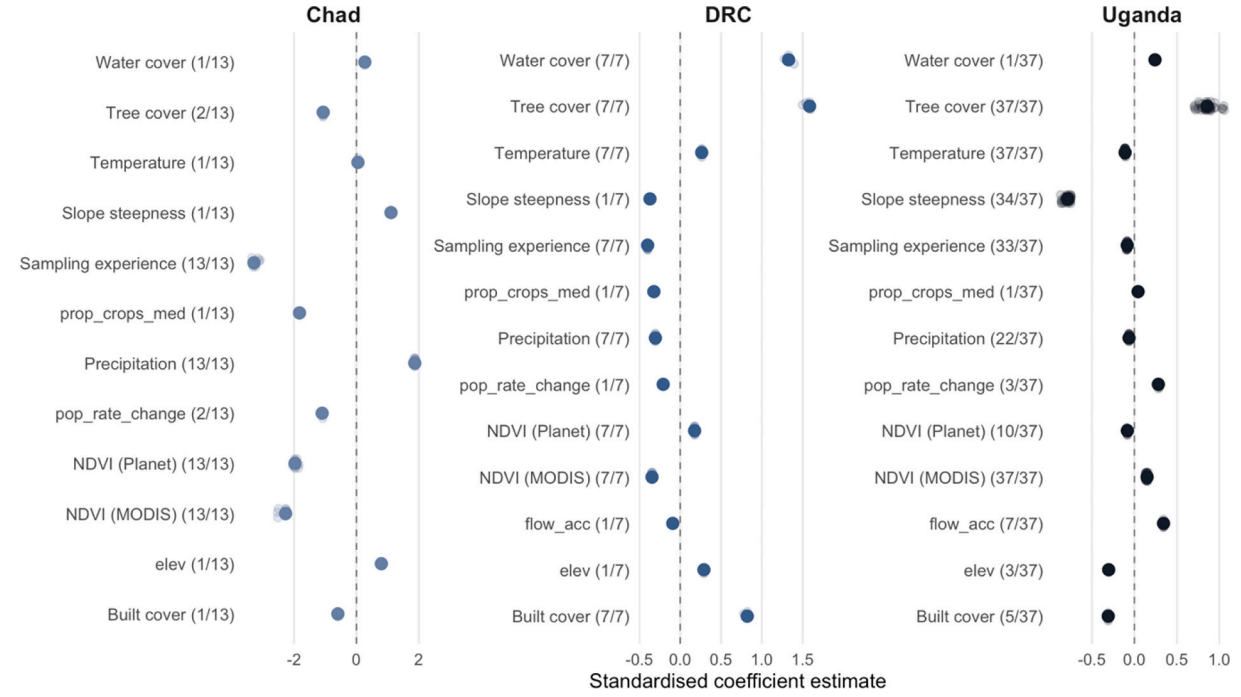


Fig. A3. Stability of fixed effect estimates across the ΔAIC ≤2 supported model sets. Translucent points show standardised coefficient estimates from individual supported models and larger solid points indicate the median estimate across models.
Values in parentheses indicate the number of supported models containing each predictor relative to the total number of supported models for that country.

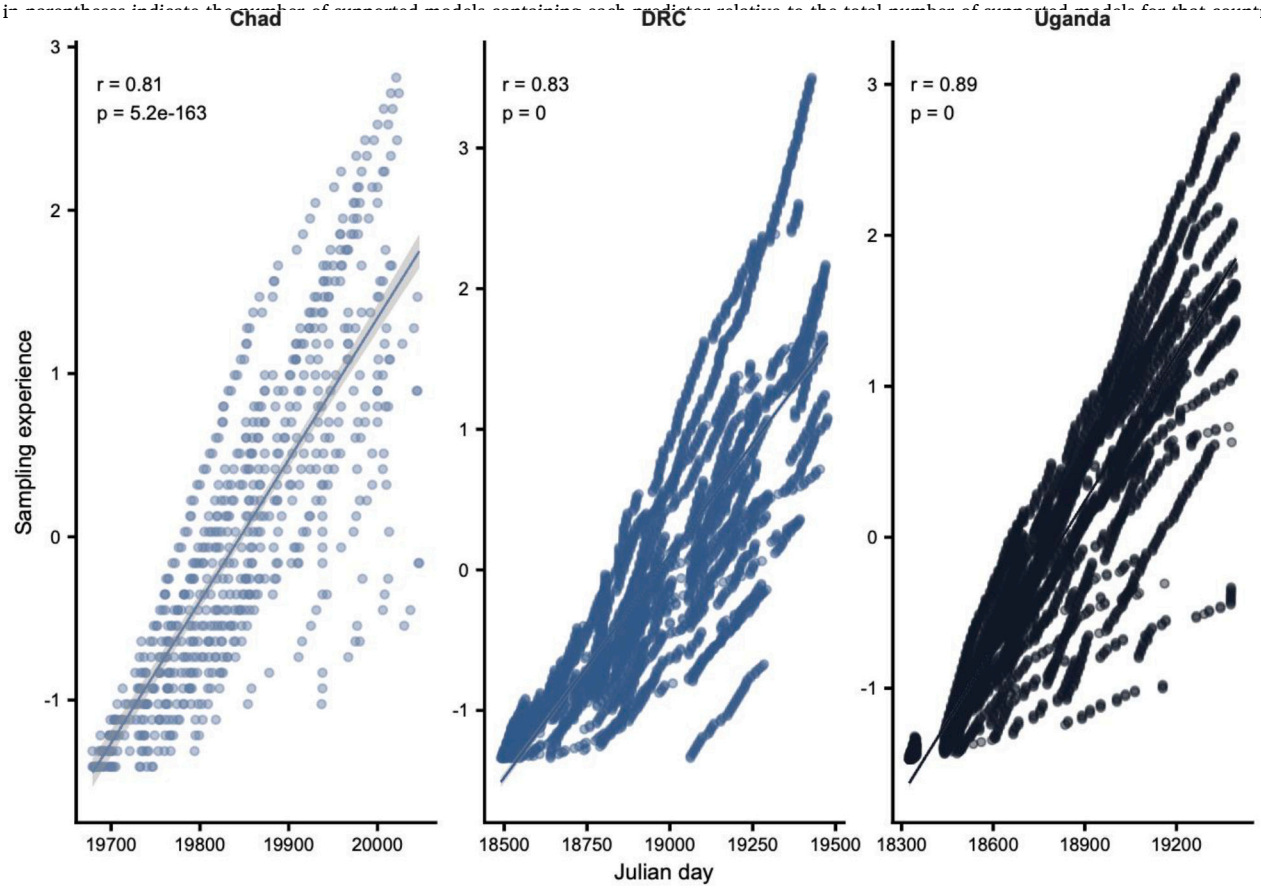


Fig. A4. Country-specific relationships between sampling experience and Julian date, with Pearson correlation coefficients and fitted linear trends.

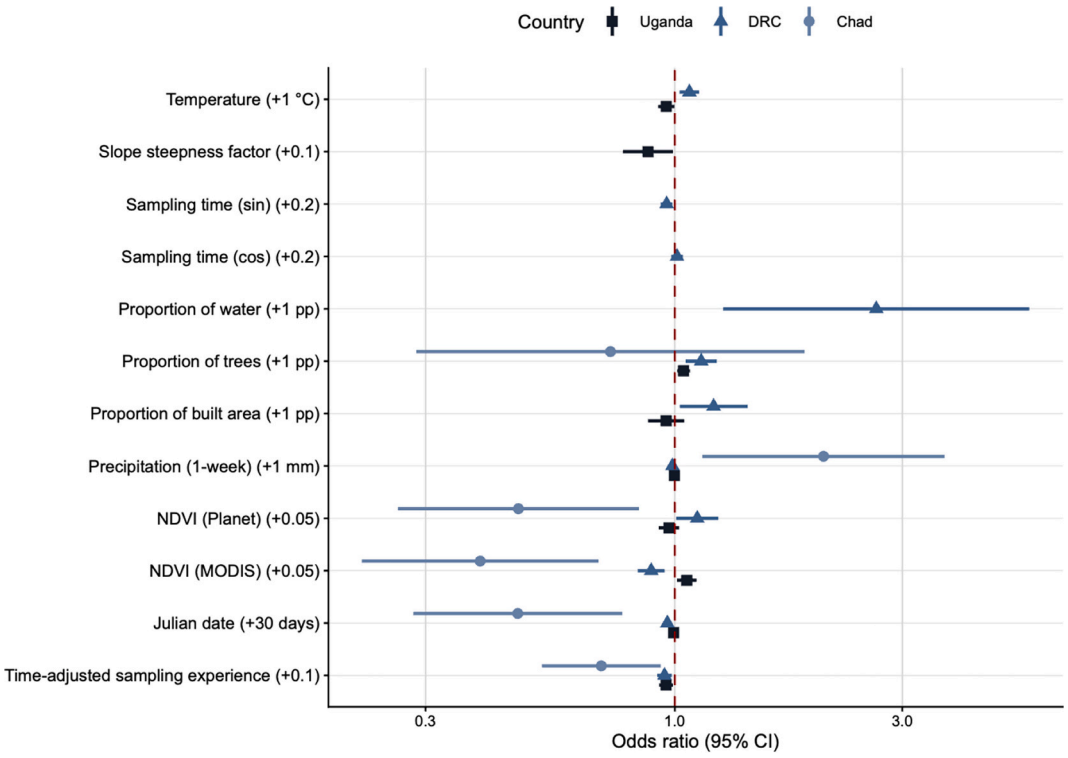


Fig. A5. Odds ratios (95% CI) from country-specific sensitivity models including Julian date and time-adjusted sampling experience.

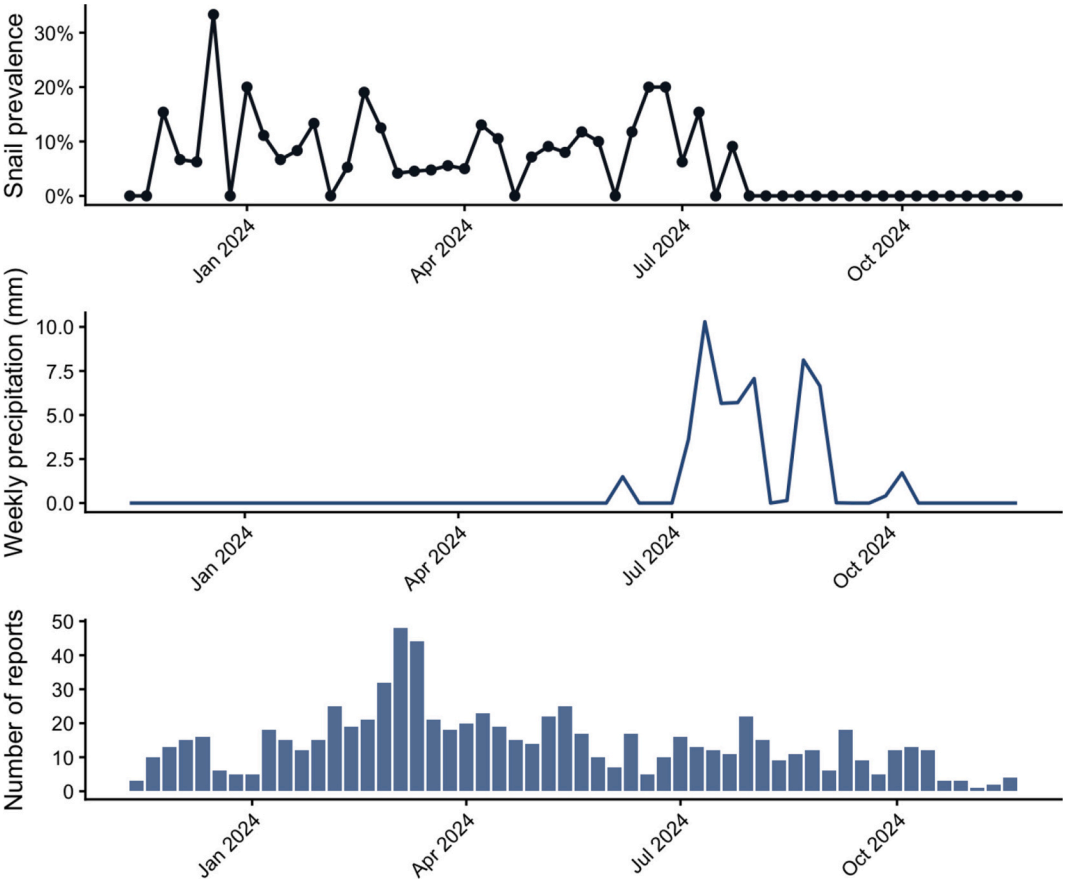


Fig. A6. Time series from Chad showing snail prevalence, weekly precipitation and number of reports, highlighting the decline in snail prevalence during the flood period, alongside extreme precipitation and reduced reporting.

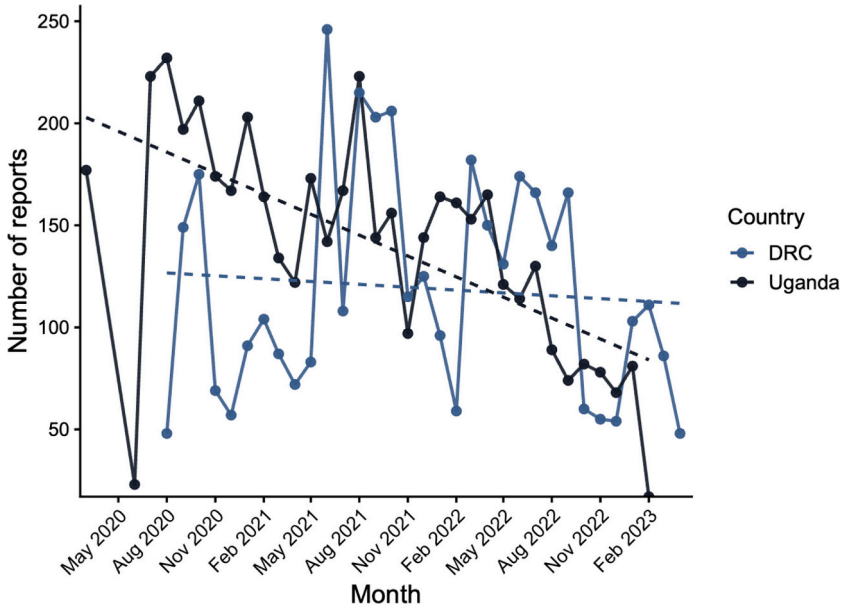


Fig. A7. Monthly reporting frequency in Uganda and the DRC, with fitted linear trends indicating a decline over time.

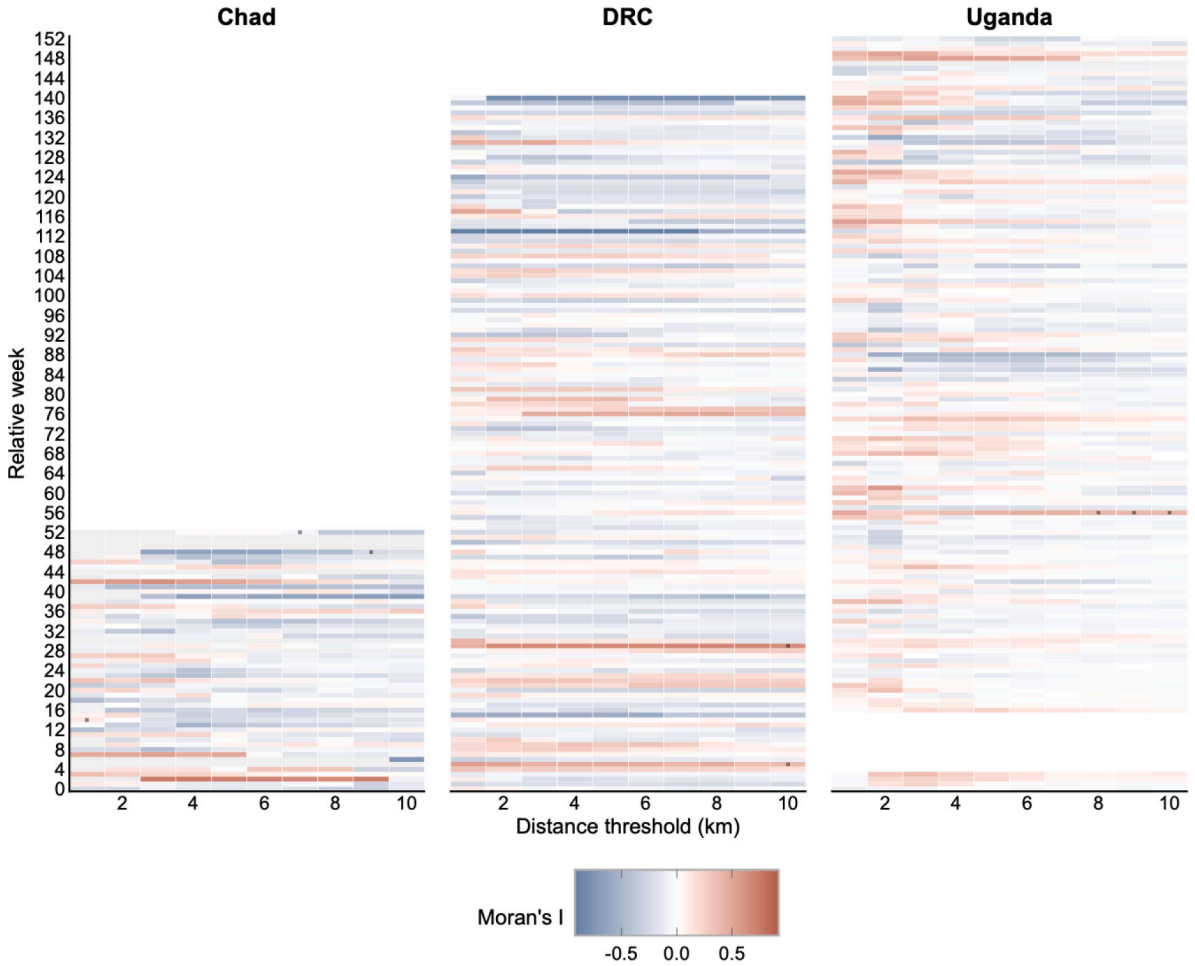


Fig. A8. Weekly spatial autocorrelation (Moran's I) across distance thresholds after inclusion of the spatial smoothing term, indicating reduced autocorrelation relative to Fig. 5b.

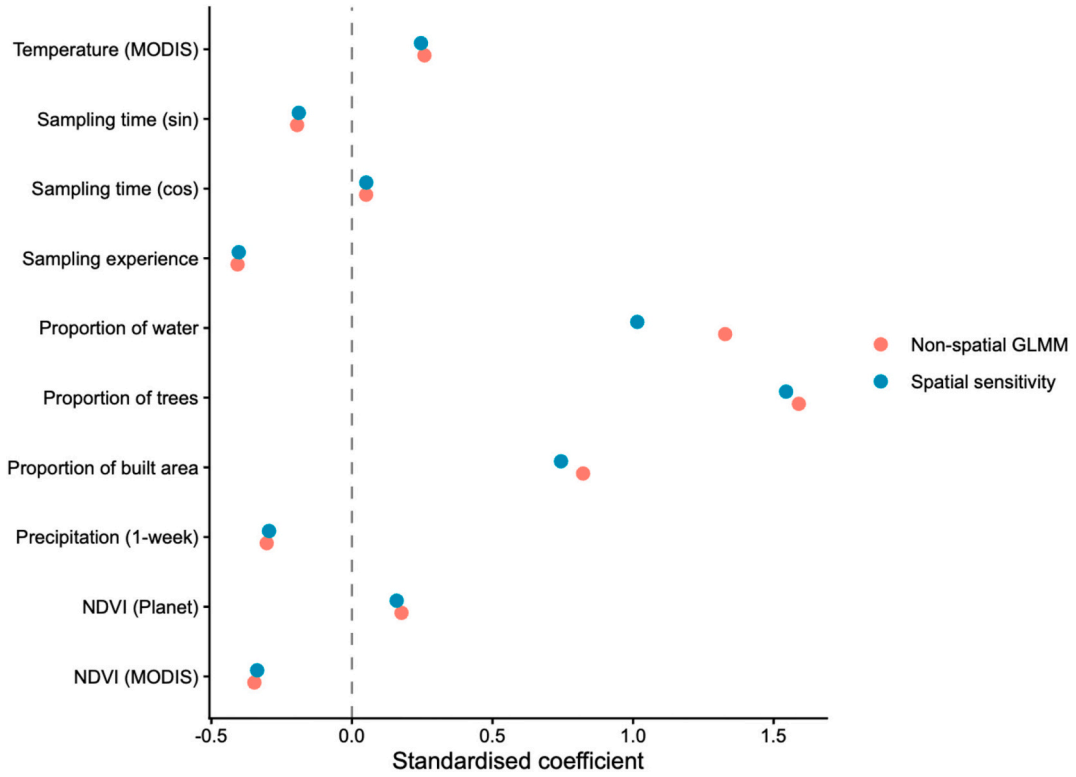


Fig. A9. Sensitivity of DRC fixed-effect estimates to residual spatial structure. Standardised coefficients from the selected non-spatial GLMM are compared with estimates from a sensitivity model including a two-dimensional spatial smooth.

Table A1. Sensitivity of DBSCAN site delineation to the neighbourhood radius (eps) in DRC and Chad.

Country	eps (m)	Retained sites	Retained visits	Baseline visits retained (%)	ARI
DRC	150	84	4784	99.7	0.985
DRC	**191.7**	**81**	**4798**	**100.0**	**1.000**
DRC	250	78	4803	100.0	0.964
Chad	150	33	728	97.2	1.000
Chad	**191.7**	**33**	**749**	**100.0**	**1.000**
Chad	250	32	762	100.0	0.979

The 191.7 m solution in grey was used as the reference. Adjusted Rand Index (ARI) measures agreement in cluster membership among visits retained in both the reference and alternative solutions.

Table A2. Summary of predictor variables, including definitions, units, resolution and extraction geometry.

Category	Covariate name	Definition	Units	Spatial resolution	Temporal resolution	Extraction geometry
Geomorphological	Elevation	Terrain elevation above sea level derived from Copernicus GLO-30 DEM; summarised per site as the 90th percentile of raster values	m	30 m	Static	Site convex hull, q90
	Flow accumulation	DEM-derived upslope contributing area; summarised per site as q90	Count of contributing cells	30 m	Static	Site convex hull, q90
	Topographic index	Topographic wetness index derived from DEM, representing the tendency for water accumulation as a function of upslope area and slope	Unitless	30 m	Static	Site convex hull, q90
	Slope steepness factor	Erosion-related slope factor	Unitless	30 m	Static	Site convex hull, q90
Climatic variables	Precipitation	Antecedent cumulative rainfall over the 7 days preceding each sampling date	mm	5 km	Daily aggregated to 7 days	Site centroid
	Temperature	Land surface temperature associated with each sampling date; missing daily values linearly interpolated	°C	1 km	Daily	Site centroid
Vegetation	NDVI Modis	MODIS NDVI assigned to each visit from the temporally closest valid composite within a 30-day precedent window; mean extracted per visit	Unitless NDVI (−1 to 1)	250 m	16-day composite	500 m buffer around visit
	NDVI Planet	Planet-derived NDVI indexed to visit date and summarised within site polygon; missing values linearly interpolated	Unitless NDVI (−1 to 1)	3–10 m	visit-indexed	Site convex hull, q90

(*continued on next page*)

(*continued*)

Category	Covariate name	Definition	Units	Spatial resolution	Temporal resolution	Extraction geometry
Land cover	Proportion of water	Proportional water cover derived from Dynamic World class histograms	Proportion (0–1)	10 m	Multi-date window, median to site level	1 km buffer around site centroid
	Proportion of crops	Proportional cropland cover derived from Dynamic World class histograms	Proportion (0–1)	10 m	Multi-date window, median to site level	1 km buffer around site centroid
	Proportion of trees	Proportional tree cover derived from Dynamic World class histograms	Proportion (0–1)	10 m	Multi-date window, median to site level	1 km buffer around site centroid
	Proportion of built	Proportional built-area cover derived from Dynamic World class histograms	Proportion (0–1)	10 m	Multi-date window, median to site level	1 km buffer around site centroid
Anthropogenic	Population rate change	Relative population change per site, calculated as (population_2020 − population_2000) / population_2000	Unitless ratio	100 m	Annual snapshots	1 km buffer around site centroid
Observation-related variables	Sampling experience	Cumulative number of submissions made by the observer up to and including the focal visit	Count	NA	Visit-level cumulative	Observer history
	Sampling time (time sin)	Circular transformation of sampling time: sin(2π × t / 86,400)	Unitless	NA	Visit timestamp	Visit metadata
	Sampling time (time cos)	Circular transformation of sampling time: cos(2π × t / 86,400)	Unitless	NA	Visit timestamp	Visit metadata

Table A3. Filtering steps for the multi-country analysis.

Filtering step	Uganda	DRC	Chad
Original dataset	6570	6891	1323
Excluding records without sampling instances	5611	6891*	1323*
Excluding dates outside the study period	5607	6862	1313
Excluding records with date difference > 7 days between 'today' and 'end'	5560	6514	1253
Excluding inaccurate GPS records**	5018	6438	1114
Retaining reports with sampling duration between 1 and 60 min***	4868	5813	900
Final analytical subset: valid *Biomphalaria* spp. reports in clusters with ≥10 visits	4770	3934	685

The three snail genera initially evaluated were *Biomphalaria* spp., *Bulinus* spp. and *Radix* spp.

To have more information about the previous filters check Valderrama-Bhraunxs, Bonifacio, Tumusiime, et al. (2025).

* Non-sampling instances were not recorded in the DRC and Chad.

** GPS accuracy threshold >5 m.

*** Sampling duration was not used as an additional predictor because, in the DRC, observers often started sampling before activating the timer, resulting in many 1-min records. Reports exceeding 60 min were nevertheless excluded as likely invalid.

Table A4. Variance inflation factors (VIF) for predictor variables in country-specific candidate model sets prior to collinearity screening.

Variable	DRC	Uganda	Chad
Precipitation (1-week)	1.72	1.07	2.63
Temperature	1.88	1.59	1.79
NDVI (Planet)	1.28	2.15	1.91
Elevation	1.7	7.71	1.64
Flow accumulation	4.33	6.5	2.42
Slope steepness factor	1.48	1.31	1.92
Topographic index (TI)	4.04	6.12	2.56
Population rate change	1.62	1.74	1.81
Proportion water	1.75	8.57	1.3
Proportion trees	1.71	5.09	1.32
Proportion crops	1.51	2.05	1.22
Proportion built	1.31	1.76	1.34
Sampling experience	1.32	1.13	2.32
Sampling time (sin)	1.37	1.1	1.11
Sampling time (cos)	1.33	1.04	1.13
NDVI (MODIS)	2.53	2.61	3.84

Table A5. Fixed effect structure, random effect term, marginal and conditional R^2, and AIC of the selected representative model with an observer random effect in each country.

Country	Fixed effects	Random effects	R^2m	R^2c	AIC
Uganda	Climate: pp_oneweek, temp_modis | Geomorphology: SS | Land cover: prop_trees_med, prop_built_med | Vegetation: NDVI_Planet, NDVI_Modis | Observation: sampling_experience	(1 | recordedBy)	0.2	0.42	5414.1
DRC	Climate: pp_oneweek, temp_modis | Land cover: prop_water_med, prop_trees_med, prop_built_med | Vegetation: NDVI_Planet, NDVI_Modis | Observation: sampling_experience, time_sin, time_cos		0.12	0.57	2723.74
Chad	Climate: pp_oneweek | Land cover: prop_trees_med | Vegetation: NDVI_Planet, NDVI_Modis | Observation: sampling_experience		0.15	0.98	174.99

Table A6. Reduced-period sensitivity model for Chad after excluding the flood-affected final months.

Chad model for the reduced period	Estimate	*p-value*
Intercept	−14.17	<0.001
Precipitation (1-week)	1.17	0.195
NDVI (Planet)	−1.58	0.062
Proportion of trees	−0.96	0.56
Time-adjusted sampling experience	−4.02	0.01
Julian date	−1.73	0.135
NDVI (MODIS)	−1.91	0.084

After excluding the flood-affected final months in Chad, the negative association with sampling experience remained supported ($p = 0.006$). The association also persisted after accounting for its correlation with calendar time ($p = 0.010$), indicating that it was not driven solely by the anomalous final period or by overall programme progression.

Data availability

Data available at https://doi.org/10.6084/m9.figshare.31943346 and https://doi.org/10.15468/t2vtd5 (Valderrama-Bhraunxs, Bonifacio, & Huyse, 2025a). Code available at https://github.com/ndcvb/Biomphalaria-citizenscience-subsaharanAfrica.

References

Adriko, M., Tinkitina, B., Arinaitwe, M., Tukahebwa, E.M., Mubangizi, A., Ortega, J.C., Zoure, H., Mwinzi, P.N., Kinvi, B., Djirmay, A.G., Njenga, S., Mazigo, H.D., 2024. Time to change implementation units for mass drug administration against schistosomiasis in Uganda: evidence from communities levels data validation and its implication in planning. Paras. Epidemiol. Contr. 27, e00394. https://doi.org/10.1016/J.PAREPI.2024.E00394.

Anselin, L., Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Handbook of Applied Economic Statistics, pp. 237–259. https://doi.org/10.1201/9781482269901-36.

Ashepet, M.G., Mulmi, J., Michellier, C., Jacobs, L., Pype, K., Huyse, T., 2026. Citizen science principles in practice: lessons from Uganda and the Democratic Republic of Congo. Community Sci. 5 (1), e2025CSJ000149. https://doi.org/10.1029/2025CSJ000149.

August, T., Fox, R., Roy, D.B., Pocock, M.J.O., 2020. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. Sci. Rep. 10 (1), 11009-. https://doi.org/10.1038/s41598-020-67658-3.

Aula, O.P., McManus, D.P., Jones, M.K., Gordon, C.A., 2021. Schistosomiasis with a focus on Africa. Trop. Med. Infect. Dis. *6* (3), 109. https://doi.org/10.3390/tropicalmed6030109.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1). https://doi.org/10.18637/JSS.V067.I01.

Bence, J.R., 1995. Analysis of short time series: correcting for autocorrelation. Ecology 76 (2), 628–639. https://doi.org/10.2307/1941218.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat Methodol. 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Benyei, P., Skarlatidou, A., Argyriou, D., Hall, R., Theilade, I., Turreira-García, N., Latreche, D., Albert, A., Berger, D., Cartró-Sabaté, M., Chang, J., Chiaravalloti, R., Cortesi, A., Danielsen, F., Haklay, M., Jacobi, E., Nigussie, A., Reyes-García, V., Rodrigues, E., Woods, T., 2023. Challenges, strategies, and impacts of doing citizen science with marginalised and indigenous communities: reflections from project coordinators. Citizen Sci. Theory Pract. 8 (1). https://doi.org/10.5334/CSTP.514.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24 (3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008.

Brees, J., Huyse, T., Tumusiime, J., Kagoro-Rugunda, G., Namirembe, D., Mugabi, F., Nyakato, V., Anyolitho, M.K., Tolo, C.U., Jacobs, L., 2021. The potential of citizen-driven monitoring of freshwater snails in schistosomiasis research. Citizen Sci. Theory Pract. 6 (1), 1–13. https://doi.org/10.5334/cstp.388.

Brier, G., 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78 (1).

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data *9* (1), 251. https://doi.org/10.1038/s41597-022-01307-4.

Brown, D.S., 1994. Freshwater Snails Of Africa And Their Medical Importance. https://doi.org/10.1201/9781482295184.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33 (2), 261–304. https://doi.org/10.1177/0049124104268644.

Chad Situation Report #6 - December 2024 - Chad | ReliefWeb. (n.d.). Retrieved February 24, 2026, from https://reliefweb.int/report/chad/chad-situation-report-6-december-2024.

Copernicus DEM - Global and European Digital Elevation Model | Copernicus Data Space Ecosystem. (n.d.). Retrieved October 4, 2024, from https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM.

Courter, J.R., Johnson, R.J., Stuyck, C.M., Lang, B.A., Kaiser, E.W., 2012. Weekend bias in citizen science data reporting: implications for phenology studies. Int. J. Biometeorol. 57 (5), 715–720. https://doi.org/10.1007/s00484-012-0598-7.

Didan, K., Barreto Munoz, A., Solano, R., Huete, A., 2015. MODIS Vegetation Index User's Guide (MOD13 Series). http://vip.arizona.edu.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., Mcclean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Fischer, H., Cho, H., Storksdieck, M., 2021. Going beyond hooked participants: the nibble-and-drop framework for classifying citizen science participation. Citizen Sci.: Theory Pract. 6 (1). https://doi.org/10.5334/cstp.350.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. Sci. Data 2. https://doi.org/10.1038/SDATA.2015.66.

Grimes, J.E.T., Croll, D., Harrison, W.E., Rg Utzinger, J., Freeman, M.C., Templeton, M. R., 2014. The Relationship between Water, Sanitation and Schistosomiasis: A Systematic Review and Meta-analysis. https://doi.org/10.1371/journal.pntd.0003296.

Gryseels, B., Polman, K., Clerinx, J., Kestens, L., 2006. Human schistosomiasis. Lancet 368 (9541), 1106–1118. https://doi.org/10.1016/S0140-6736(06)69440-3.
Hahsler, M., Piekenbrock, M., Doran, D., 2019. Dbscan: fast density-based clustering with R. J. Stat. Softw. 91, 1–30. https://doi.org/10.18637/jss.v091.i01.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143. https://doi.org/10.1148/radiology.143.1.7063747.
Hartig, F., 2025. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models.
Iacovidou, M.A., Byaruhanga, A.M., Besigye, F., Nabatte, B., Kabatereine, N.B., Chami, G. F., 2025. Ecological niche stability of *Biomphalaria* intermediate hosts for *Schistosoma mansoni* under extreme flooding and seasonal change. BioRxiv, 2025.01.07.631649. https://doi.org/10.1101/2025.01.07.631649.
Iacovidou, M.A., Byaruhanga, A.M., Besigye, F., Nabatte, B., Kabatereine, N.B., Chami, G. F., 2026. Temporal variability and flooding influence the ecological niche of Biomphalaria intermediate hosts for Schistosoma mansoni in rural Uganda. Proc. R. Soc. B Biol. Sci. 293 (2063). https://doi.org/10.1098/rspb.2025.2083.
Jacobusse, G.W., Jongejans, E., 2026. Non-detection by citizen scientists modeled as a function of visit characteristics. Ecol. Model. 514, 111474. https://doi.org/10.1016/J.ECOLMODEL.2026.111474.
Johnston, A., Fink, D., Hochachka, W.M., Kelling, S., 2018. Estimates of observer expertise improve species distributions from citizen science data. Methods Ecol. Evol. 9 (1), 88–97. https://doi.org/10.1111/2041-210X.12838;WEBSITE:WEBSITE:BESJOURNALS;WGROUP:STRING:PUBLICATION.
Johnston, A., Matechou, E., Dennis, E.B., 2023. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. Methods Ecol. Evol. 14 (1), 103–116. https://doi.org/10.1111/2041-210X.13834;PAGE:STRING:ARTICLE/CHAPTER.
Kelling, S., Johnston, A., Hochachka, W.M., Iliff, M., Fink, D., Gerbracht, J., Lagoze, C., La Sorte, F.A., Moore, T., Wiggins, A., Wong, W.K., Wood, C., Yu, J., 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? PLoS One 10 (10), e0139600. https://doi.org/10.1371/JOURNAL.PONE.0139600.
Kloos, H., Fulford, A.J.C., Butterworth, A.E., Sturrock, R.F., Ouma, J.H., Kariuki, H.C., Thiongo, F.W., Dalton, P.R., Klumpp, R.K., 1997. Spatial patterns of human water contact and *Schistosoma mansoni* transmission and infection in four rural areas in Machakos District, Kenya. Soc. Sci. Med. 44 (7), 949–968. https://doi.org/10.1016/S0277-9536(96)00218-3.
KoboCollect [Mobile app], 2026. KoboToolbox.
Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. Front. Ecol. Environ. 14 (10), 551–560. https://doi.org/10.1002/FEE.1436.
Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27 (15), 2171–2186. https://doi.org/10.1002/HYP.9740.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2017. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence, Second edition, pp. 1–641. https://doi.org/10.1016/C2012-0-01164-7.
Madinga, J., Linsuke, S., Mpabanzi, L., Meurs, L., Kanobana, K., Speybroeck, N., Lutumba, P., Polman, K., 2015. Schistosomiasis in the Democratic Republic of Congo: a literature review. Parasit. Vectors 8 (1), 601. https://doi.org/10.1186/S13071-015-1206-6.
Mair, L., Ruete, A., 2016. Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLoS One 11 (1), e0147796. https://doi.org/10.1371/journal.pone.0147796.
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17. https://doi.org/10.2307/2332142.
Morgan, J.A.T., Dejong, R.J., Snyder, S.D., Mkoji, G.M., Loker, E.S., 2001. Schistosoma mansoni and Biomphalaria: past history and future trends. Parasitology 123 (SUPPL). https://doi.org/10.1017/s0031182001007703.
Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. R J. 9 (1), 207–218. https://doi.org/10.32614/RJ-2017-009.
Moser, W., Greter, H., Schindler, C., Allan, F., Ngandolo, B.N.R., Moto, D.D., Utzinger, J., Zinsstag, J., 2014. The spatial and seasonal distribution of *Bulinus truncatus*, *Bulinus forskalii* and *Biomphalaria pfeifferi*, the intermediate host snails of schistosomiasis, in N'Djamena, Chad. Geospat. Health 9 (1), 109–118. https://doi.org/10.4081/GH.2014.9.
Murphy, A.H., 1973. Hedging and skill scores for probability forecasts. JApMe 12 (1), 215–223. https://doi.org/10.1175/1520-0450(1973)012.
Naiga, R., Penker, M., Hogl, K., 2015. Challenging pathways to safe water access in rural Uganda: from supply to demand-driven water governance. Int. J. Commons 9 (1), 237. https://doi.org/10.18352/bmgn-lchr.480.
Palasio, R.G.S., Xavier, I.G., Chiaravalotti-Neto, F., Tuan, R., 2019. Diversity of *Biomphalaria* spp. freshwater snails and associated mollusks in areas with schistosomiasis risk, using molecular and spatial analysis tools. Biota Neotrop. 19 (4), e20190746. https://doi.org/10.1590/1676-0611-bn-2019-0746.
Palmer, J.R.B., Oltra, A., Collantes, F., Delgado, J.A., Lucientes, J., Delacour, S., Bengoa, M., Eritja, R., Bartumeus, F., 2017. Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes. Nat. Commun. 8 (1), 916. https://doi.org/10.1038/s41467-017-00914-9.
Perez-Saez, J., Mande, T., Ceperley, N., Bertuzzo, E., Mari, L., Gatto, M., Rinaldo, A., 2016. Hydrology and density feedbacks control the ecology of intermediate hosts of schistosomiasis across habitats in seasonal climates. Proc. Natl. Acad. Sci. USA 113 (23), 6427–6432. https://doi.org/10.1073/PNAS.1602251113/SUPPL_FILE/PNAS.1602251113.SD02.TXT.
Pham-Duc, B., Sylvestre, F., Papa, F., Frappart, F., Bouchez, C., Crétaux, J.F., 2020. The Lake Chad hydrology under current climate change. Sci. Rep. 10 (1), 5498. https://doi.org/10.1038/s41598-020-62417-w.
Planet Labs PBC, 2023. Planet Application Program Interface: In Space for Life on Earth. https://api.planet.com.
R Core Team, 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
Reitzug, F., Ledien, J., Chami, G.F., 2023. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. PLoS Negl. Trop. Dis. 17 (6), e0011377. https://doi.org/10.1371/journal.pntd.0011377.
Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., Jacobs, D., 2012. Dynamic changes in motivation in collaborative citizen-science projects. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, vol. 12, pp. 217–226. https://doi.org/10.1145/2145204.2145238.
Schwartz, N.B., Lintner, B.R., Feng, X., Powers, J.S., 2020. Beyond MAP: A guide to dimensions of rainfall variability for tropical ecology. Biotropica 52 (6), 1319–1332. https://doi.org/10.1111/BTP.12830.
Sicacha-Parada, J., Steinsland, I., Cretois, B., Borgelt, J., 2021. Accounting for spatial varying sampling effort due to accessibility in citizen science data: a case study of moose in Norway. Spatial Statist. 42, 100446. https://doi.org/10.1016/j.spasta.2020.100446.
Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology (Cambridge, Mass.) *21* (1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.
Stothard, J.R., Campbell, S.J., Osei-Atweneboana, M.Y., Durant, T., Stanton, M.C., Biritwum, N.K., Rollinson, D., Ombede, D.R.E., Tchuem-Tchuenté, L.A., 2017. Towards interruption of schistosomiasis transmission in sub-Saharan Africa: developing an appropriate environmental surveillance framework to guide and to support "end game" interventions. Infect. Dis. Poverty 6 (1), 1–11. https://doi.org/10.1186/S40249-016-0215-9/FIGURES/3.
Tatem, A.J., 2017. WorldPop, open data for spatial demography. Sci. Data *4* (*1*), 170004. https://doi.org/10.1038/sdata.2017.4.
Tumusiime, J., Valderrama Bhraunxs, N., Kagoro-Rugunda, G., Namirembe, D., Albrecht, C., Twongyirwe, R., Tolo, C.U., Jacobs, L., Huyse, T., 2024. Citizens can help to map putative transmission sites for snail-borne diseases. PLoS Negl. Trop. Dis. 18 (4), e0012062. https://doi.org/10.1371/journal.pntd.0012062.
Uganda Bureau of Statistics, 2010. National Household Survey 2009/2010.
Utzinger, J., Tanner, M., 2000. Microhabitat preferences of Biomphalaria pfeifferi and Lymnaea natalensis in a natural and a man-made habitat in southeastern Tanzania. Mem. Inst. Oswaldo Cruz 95 (3), 287–294. https://doi.org/10.1590/S0074-02762000000300002.
Valderrama-Bhraunxs, N., Bonifacio, L., Huyse, T., 2025a. The ATRAP Project: Citizens Scientists Reporting Snail Intermediate Host Occurrence in Uganda and the Democratic Republic of the Congo. In version 1.6.. GBIF.org. https://doi.org/10.15468/t2vtd5.
Valderrama-Bhraunxs, N., Bonifacio, L., Tumusiime, J., Kapour, G., Namirembe, D., Umba-Tolo, C., Kagoro-Rugunda, G., Mitashi-Mulopo, P., Mandinga, J., Jacobs, L., Huyse, T., 2025b. Snail intermediate host occurrence recorded by citizen scientists in rural Uganda and the Democratic Republic of the Congo. Gigabyte 2025, gigabyte162-0. https://doi.org/10.46471/gigabyte.162.
Valderrama-Bhraunxs, N., Huyse, T., van Loon, E., Van Rompaey, A., Twongyirwe, R., Jacobs, L., 2025c. Great minds map alike: citizen and expert distribution models of schistosome snail hosts in rural West Uganda. Ecol. Solut. Evid. 6 (4), e70163. https://doi.org/10.1002/2688-8319.70163.
Walz, Y., Wegmann, M., Dech, S., Raso, G., Utzinger, J., 2015. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. Parasit. Vectors 8 (1), 163. https://doi.org/10.1186/s13071-015-0732-6.
Wan, Z., Hook, S., Hulley, G., 2021. MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center.
WHO, 2022. WHO Guideline on Control and Elimination of Human Schistosomiasis. https://www.ncbi.nlm.nih.gov/books/NBK578392/.
Wood, C.L., Sokolow, S.H., Jones, I.J., Chamberlin, A.J., Lafferty, K.D., Kuris, A.M., Jocque, M., Hopkins, S., Adams, G., Buck, J.C., Lund, A.J., Garcia-Vedrenne, A.E., Fiorenza, E., Rohr, J.R., Allan, F., Webster, B., Rabone, M., Webster, J.P., Bandagny, L., De Leo, G.A., 2019. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. Proc. Natl. Acad. Sci. USA 116 (46), 23182–23191. https://doi.org/10.1073/PNAS.1903698116/SUPPL_FILE/PNAS.1903698116.SAPP.PDF.
Wood, S., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B Stat Methodol. 73 (1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x.
Woolhouse, M.E.J., Chandiwana, S.K., 1989. Spatial and temporal heterogeneity in the population dynamics of *Bulinus globosus* and *Biomphalaria pfeifferi* and in the epidemiology of their infection with schistosomes. Parasitology 98 (Pt(1)), 21–34. https://doi.org/10.1017/S0031182000059655.
Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R. Statistics for Biology and Health. https://doi.org/10.1007/978-0-387-87458-6.