Received: 7 May 2025 | Accepted: 14 November 2025

DOI: 10.1002/2688-8319.70163

Ecological Solutions and Evidence

RESEARCH ARTICLE

Great minds map alike: Citizen and expert distribution models of schistosome snail hosts in rural west Uganda

Noelia Valderrama-Bhraunxs[1,2] | Tine Huyse[2] | Emiel van Loon[3] | Anton Van Rompaey[1] | Ronald Twongyirwe[4] | Liesbet Jacobs[3]

[1]Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium

[2]Department of Biology, Royal Museum for Central Africa, Tervuren, Belgium

[3]Institute for Biodiversity and Ecosystems Dynamics, Amsterdam, The Netherlands

[4]Department of Environment and Livelihoods Support Systems, Mbarara University of Science and Technology, Mbarara, Uganda

Correspondence
Noelia Valderrama-Bhraunxs
Email: noeliadelcarmen.valderramabhraunxs@kuleuven.be

Funding information
Belgian Directorate General for Development, Cooperation and Humanitarian Aid, Grant/Award Number: XM-DAC-2-10-3853; Fonds Wetenschappelijk Onderzoek, Grant/Award Number: 11L3223N and 11L3225N

Handling Editor: Norris Erhabor

Abstract

1. Schistosomiasis is a parasitic disease that affects over 250 million people worldwide, with the majority living in rural areas of sub-Saharan Africa. The parasite relies on freshwater snails of the genus *Biomphalaria* as intermediate hosts. Mapping snail distribution is vital for identifying disease transmission hotspots. However, expert-led monitoring is often constrained by limited resources and restricted access to remote areas, highlighting the need for scalable and cost-effective alternatives.
2. This study evaluates the effectiveness of citizen science in predicting *Biomphalaria* spp. presence by comparing models built from expert - and citizen-collected data. We tested two scenarios: the first one assumed perfect detection and focused on environmental and geomorphological predictors, while the second accounted for imperfect detection to explore discrepancies between citizen observations and expert-derived detection probabilities.
3. In the perfect detection scenario, the expert and citizen models identified site type and NDVI as significant environmental predictors of snail presence. Although both models demonstrated low marginal R^2 values (~16%–17%), indicating limited explanatory power of broad-scale environmental predictors, conditional R^2 values exceeded 65%, suggesting that fine-scale, site-specific habitat characteristics are critical determinants of *Biomphalaria* spp. presence. For the imperfect detection scenario, the expert model and the citizen observations showed minimal discrepancies, primarily explained by individual observer variability and differences in sampling effort. Increased sampling effort consistently reduced false negatives and led to unexpected observations of snail presence by the citizens (i.e. observed presence in sites predicted unsuitable by the expert model).
4. *Practical implication*. Our findings demonstrate that citizen science data, when properly structured and statistically accounted for bias and errors, can generate ecological modelling outputs comparable to those based on expert-led surveys. We highlight the importance of accounting for observer variability, providing calibrated training and optimizing sampling strategies to enhance data quality. This

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2025 The Author(s). *Ecological Solutions and Evidence* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

study presents a transferable and cost-efficient framework for participatory ecological monitoring in resource-limited and undersampled regions.

KEYWORDS
Biomphalaria, citizen science, data validation, habitat suitability, schistosomiasis, snail-borne diseases

1 | INTRODUCTION

Schistosomiasis is a parasitic disease that affects over 250 million people worldwide (Lo et al., 2022). It represents a significant public health challenge, particularly in rural sub-Saharan Africa where safe water and sanitation are scarce (Okoro et al., 2023). Chronic infection causes anaemia, cognitive impairment and stunted growth in children, and in adults leads to liver fibrosis, bladder cancer and infertility (Gryseels et al., 2006). Humans become infected through dermal contact with freshwater containing cercariae, the free-swimming infective larvae of *Schistosoma* spp., which are released by infected freshwater snails (Ross et al., 2002). Parasite eggs excreted in human urine or faeces hatch in freshwater into miracidia (ciliated larvae) that infect compatible snail hosts (Gryseels et al., 2006). Within the snail, miracidia develop into sporocysts that multiply asexually and ultimately produce cercariae, which are shed back into the water (Warren, 1978). Consequently, in settings with limited sanitation and heavy reliance on natural water sources, water contamination and ongoing exposure to schistosomiasis are sustained (Stothard et al., 2017).

In the Ugandan context, intestinal schistosomiasis due to *Schistosoma mansoni* has been documented across 65% of the country's 146 districts, with lakeshore communities often exceeding 50% prevalence, whereas urogenital schistosomiasis due to *S. haematobium* has been reported in fewer than 3% of the districts (Adriko et al., 2024). No urogenital schistosomiasis has been reported from our study area in western Uganda (southern Lake Albert) (Adriko et al., 2024). Given that *Biomphalaria* snails act as intermediate hosts for *S. mansoni* (Morgan et al., 2001), mapping their distribution is key for identifying potential transmission hotspots and guiding targeted disease control strategies (Stensgaard et al., 2006; Tumusiime et al., 2024; Wang et al., 2024). However, obtaining fine-scale *Biomphalaria* spp. distribution data in remote, resource-limited areas is challenging, as traditional expert-led monitoring is often not feasible due to logistical constraints (Shiff, 2017).

To address these data gaps, citizen science—consisting of 'actively involving the general public in scientific endeavours generating new knowledge or understanding' (ECSA, 2015)—offers a practical complement to expert-led monitoring. It enables local communities to collect high-resolution data and improve vector monitoring (Brees et al., 2021; Tumusiime et al., 2024). In Western Uganda, the ATRAP (Action Towards Reducing Aquatic Snail Parasitic Diseases) project operationalized this model by engaging community members in repeated sampling of snails intermediate hosts, providing the data used in this study. When combined with information on environmental and geomorphological drivers of snail presence, citizen-generated datasets can enhance disease surveillance. Nevertheless, method-related biases remain a concern (e.g. spatial/temporal sampling bias, uneven effort and misidentification) arising from opportunistic sampling or contributions by non-specialists (Callaghan et al., 2021; Elbroch et al., 2011; Kosmala et al., 2016). Although previous work has demonstrated that citizen science can reliably capture presence–absence data for snail hosts (Tumusiime et al., 2024), evidence is limited on how such biases might affect predictive models and ecological interpretations (Bird et al., 2014).

This study addresses these gaps by evaluating the potential impact of data quality on predicting *Biomphalaria* snail occurrence, using expert-based data as a reference. We assess (i) whether model outputs derived from citizen and expert data are comparable, and if so, (ii) which environmental and geomorphological variables most strongly structure snail distribution. Additionally, we examine the role of detection-related factors such as observer identity, sampling consistency, timing, duration, density and frequency, in determining the accuracy of citizen science observations. Finally, we explore the added value that citizen science can provide for vector surveillance, particularly in remote, resource-limited settings.

2 | METHODS

2.1 | Study area

The data for this study were collected from the Lake Albert region in western Uganda (Figure 1a), a landscape marked by substantial altitudinal variation (620–1400 m a.s.l.) along the Great Rift Valley. After applying data filters (Section 2.2.1.1), 56 sampling units—hereafter water contact sites (WCS)—were retained for analysis. These sites spanned four site types: springs, streams, wetlands and lake shorelines (Figure 1b). A detailed site description can be consulted in Tumusiime et al. (2024) Supporting Information (S1).

2.2 | Data

2.2.1 | Snail data

We analysed reports collected by the ATRAP (Action Towards Reducing Aquatic Snail Parasitic Diseases) project in western Uganda. Twenty-five citizen scientists (CS) sampled 76 WCS from March 2020

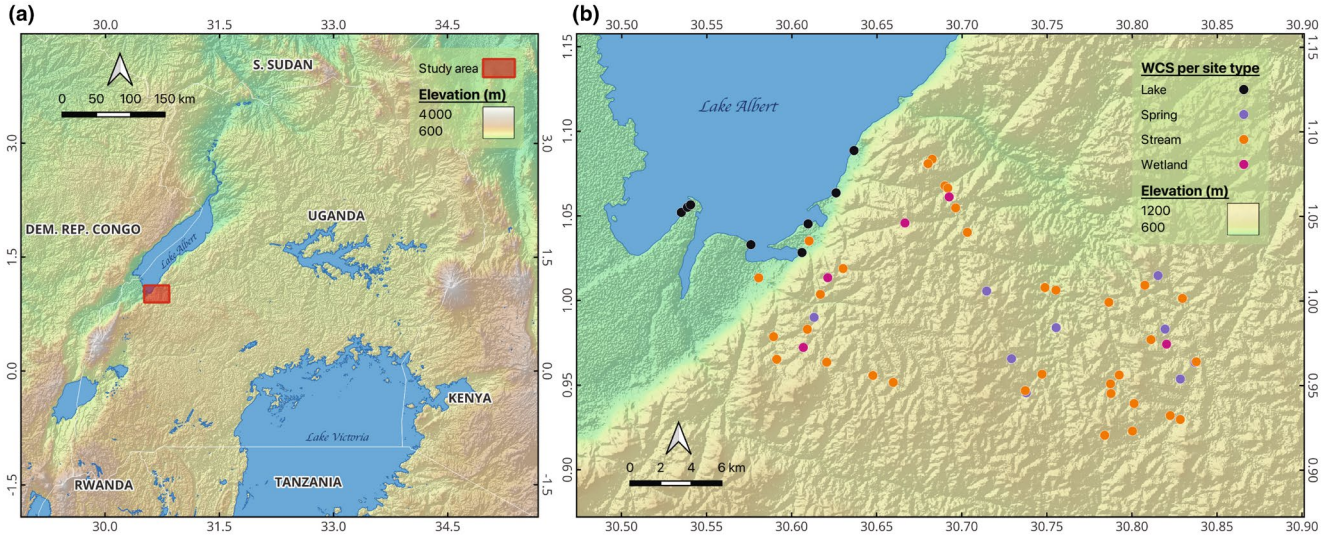


FIGURE 1 (a) Map of Uganda showing the study area. (b) Distribution of water contact sites by site type included in the study.

to February 2023; weekly sampling was encouraged but voluntary, consistent with citizen science principles. We adopted a flexible schedule to sustain participation in remote communities and to respect participants' livelihoods, while recording visit-level metadata (observer, date, duration) so that variable effort could be evaluated in the analysis. Sites were selected after a preliminary survey of locations with frequent risky water practices (e.g. swimming, water-fetching, washing and bathing) and included sites where *Biomphalaria* snails were present and sites where they were not detected during initial exploratory fieldwork (Brees et al., 2021) to avoid presence-only bias and enable presence–absence modelling to study the environmental and geomorphological drivers of snail distribution.

CS reports consisted of photographs of snails intermediate hosts classified to genus: *Biomphalaria*, *Bulinus* or *Radix*, with other snails grouped as 'pool'. Photos could contain multiple individuals and when more than a genus/category was encountered at a site, separate images were submitted per genus/category. For each visit, CS also recorded counts per genus/category from a timed search (scooping) targeting 30 min (Brees et al., 2021), although the duration varied between visits and CS. Start and stop times, observer identity, date and coordinates were recorded, and reports between 20 and 60 min were retained for analysis. These reports were submitted to a central server that could subsequently be accessed by the researchers. A researcher validated both the genus identification and snail counts as submitted by the CS. A malacologist ('the expert') sampled the same WCS monthly from July 2020 to May 2022 with a fixed 30-min scooping search per visit. The expert sampling interval overlapped with, but was shorter than, the CS sampling period (Mar 2020–Feb 2023) (Figure S1). CS received initial training on genus-level identification with photo guides, the timed-search protocol, metadata entry and personal protective equipment (PPE) requirements (e.g. gloves, boots). Annual refresher trainings reiterated safety, and the protocol specified not to sample if PPE was unavailable; ATRAP project coordinators then supplied replacement gear. More information about the sampling protocol can be consulted in Brees et al. (2021) and Tumusiime et al. (2024).

This study focused exclusively on reports of *Biomphalaria* spp., the only known intermediate host of the schistosome responsible for human schistosomiasis in this region of Uganda (Adriko et al., 2018). Due to the difficulty of distinguishing between species, identification was limited to the genus level, which is sufficient for epidemiological relevance, as all *Biomphalaria* spp. present in the area can host *S. mansoni* (Morgan et al., 2001). We used *Biomphalaria* spp. occurrence (presence–absence) for this first comparison of environmental and geomorphological drivers using citizen science and expert datasets. Although infected-snail abundance is more directly linked to transmission, infection status was not measured, and the available snail count records are observer-dependent (Tumusiime et al., 2024). Abundance modelling would therefore require additional assumptions beyond the scope of this study. Accordingly, we used snail occurrence as a conservative indicator of habitat suitability.

2.2.1.1 | *Data filters*

Of the 6570 CS reports, 959 were excluded due to reported inability to sample, citing reasons such as weather, illness or lack of motivation. Further exclusions included reports outside the sampling period (4), inaccurate GPS readings (accuracy >5 m) (547) and coordinates located more than 200 m from designated WCS (416). Reports lacking *Biomphalaria* spp. data (62), containing misidentifications (32) and with sampling durations outside, the 20–60-min range (359) were also excluded. When multiple entries occurred on the same day at a site, only the first was retained (50). To account for potential variability in snail detection, only sites with at least 20 surveys were included (132), resulting in 62 WCS. Only WCS sampled by both the expert and CS were retained, resulting in 3631 valid reports and 56 WCS sampled by 22 CS (Figure S2).

The expert submitted 1382 reports of which 26 were excluded due to fewer than 10 observations per WCS. Some WCS had <10

expert visits because they were added after project initiation, retired early or were temporarily inaccessible/unsafe (flooding/dry-out, land-access) and during periods of COVID-19 travel constraints. This threshold was established because the expert sampled each water contact site once a month, less frequently than the expected weekly sampling by the CS, resulting in 1356 reports. To ensure comparability between the expert and CS data, only reports from common WCS that met the set standards were included. This led to 1073 valid reports from 56 WCS (Figure S2).

2.2.2 | Static geomorphological variables

Geomorphological variables for the study region (Table S1) 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.). Four DEM tiles, each covering a 1° by 1° area around the study region, were downloaded and merged into a single raster. The raster was then reprojected to the target coordinate reference system (CRS: EPSG:32636–WGS84 36N) using the warp tool with a 30 m resolution and cubic B-spline resampling (Nonogaki et al., 2006). Finally, it was clipped to the relevant catchment area, defined using the HydroBASINS dataset (Hybas Lake 7 sub-basin, HYBAS_ID = '1071111290') (Lehner & Grill, 2013). All processing was conducted in QGIS.

Geomorphological variables were measured once per WCS and held constant across visits, whereas environmental covariates varied by site date.

Elevation (ELE): Elevation values were directly extracted from the Copernicus GLO-30 DEM.

Slope Steepness Factor (SSF): Calculated as the slope steepness (S) factor from the Universal Soil Loss Equation (RUSLE), following the methodology proposed by McCool et al. (1987). This factor measures the impact of slope gradient on soil erosion potential.

Flow accumulation (FA): Also referred to as upslope contributing area, quantifies the total overland flow passing through each raster cell, considering the number and area of upstream cells (Holmgren, 1994).

Topographic Index (TI): Indicates potential water accumulation zones and reflects the balance between gravitational forces driving water downslope, calculated as the natural logarithm of the ratio between the upslope contributing area (FA) and the local slope angle (tan(β)) (Quinn et al., 1991).

All derived variables—SSF, FA and TI—were computed using the *R.watershed* (Neteler et al., 2012) tool in QGIS.

2.2.3 | Environmental variables

Environmental variables (Table S2) were calculated for each water contact site but, and unlike static geomorphological variables, varied over time.

Precipitation (PP): Daily precipitation was obtained from CHIRPS (Funk et al., 2015) via Google Earth Engine (GEE). To capture intra-seasonal rainfall variability and its potential influence on snail presence, 7-day antecedent cumulative precipitation was calculated for each site-date using RStudio. Multi-day accumulation provides a more robust indicator of ecological responses than daily values, as it reflects sustained hydrological conditions (Schwartz et al., 2020). PP serves as our time-varying seasonal covariate; accordingly, we did not include separate month/season terms to avoid collinearity and preserve parsimony.

Surface temperature (ST): Daily surface temperature data were extracted from MODIS (Wan et al., 2021) through GEE and converted from Kelvin to Celsius. Temperature significantly affects *Biomphalaria* snail dynamics, including growth, fecundity, egg viability and survival (Kalinda et al., 2017).

Normalized difference vegetation index (NDVI): NDVI values were derived from high-resolution (3 m) Planet imagery (Planet Labs PBC, 2023) using red, green, blue and near-infrared (NIR) bands (Protocol S1). After mosaicking the images by date, NDVI values were calculated. The 90th percentile of NDVI values within convex hull polygons containing all sampling points per site was used to emphasize peak vegetation density, as higher NDVI values correspond to denser terrestrial vegetation (Fung & Siu, 2000).

Temporal gaps in NDVI and ST were addressed using linear and cubic interpolation via the *imputeTS* package (Moritz & Bartz-Beielstein, 2017) in R.

2.2.4 | Detection variables

Detection variables describe aspects of the sampling process and individual performance CS performance that may influence snail detection likelihood.

Citizens ID (ID): Each citizen scientist was assigned a unique identifier to assess potential unmeasured individual effects on snail detection.

Consistency (CONS): Defined as the ratio of valid monthly submissions to the expected frequency (1 report per site per week or 4 reports per month per site). Higher consistency was hypothesized to enhance snail detection due to increased familiarity and observation skills (Dickinson et al., 2010; Parrish et al., 2019).

Sampling time (SAMT): Time of sampling may influence detectability, as *Biomphalaria* snails' activity varies with light conditions (Rotenberg et al., 1989). Time data were transformed into circular variables using sine and cosine functions: $SAMT_{sin} = \sin(2\pi \times (SAMT/86400))$ and $SAMT_{cos} = \cos(2\pi \times (SAMT/86400))$, where SAMT represents the time of sampling in seconds since midnight within a 24-h period (86,400 s). This accounts for daily rhythmic patterns, with peaks at 6 AM ($SAMT_{sin}$) and midnight ($SAMT_{cos}$) and minimum values at 6 PM ($SAMT_{sin}$) and noon ($SAMT_{cos}$).

Sampling duration (SDU): Longer sampling times increase the likelihood of encounters and therefore detection probability (Gu & Swihart, 2004). Shorter durations near the 20-min minimum were

expected to yield lower observations than those closer to the 60-min maximum.

Cumulative sampling frequency (CSF): The total number of sampling events per citizen scientist across the study, with higher counts later in the period potentially reflecting increased proficiency and therefore detection likelihood (Kosmala et al., 2016).

Sampling density (SDE): Represents relative sampling intensity, calculated as the ratio of citizen to expert visits per site.

2.3 | Statistical modelling

To evaluate whether citizen-generated data yields a similar understanding of biophysical drivers of snail occurrence as expert-generated data, we compared model outputs from both sources.

We implemented two modelling scenarios (Figure 2). The first assumed perfect detection, considering only environmental and geomorphological variables as drivers of snail presence. This approach tested whether key predictors of snail presence were consistent between citizen and expert models.

The second scenario accounted for imperfect detection, assessing whether detection variables explained differences between the 'ideal' expert-based detection probabilities and the actual CS snail observations. Specifically, we examined how these variables influenced the *probabilities of false negatives* (defined as the expert model's probability of presence for cases where an absence is reported by the CS) and the *probabilities of unexpected observations* (defined as the expert model's probability of absence for cases where a presence is reported by the CS).

2.3.1 | Perfect detection scenario

This scenario focused solely on geomorphological and environmental predictors of snail presence/absence, excluding potential detection biases. To compare predictors identified by CS and the expert, we fit separate models for each dataset. Generalized linear mixed-effect models (GLMMs) were applied to account for repeated sampling at WCS and to evaluate the relationship between predictors and the binary response variable (Bolker, 2015): snail presence or absence. Given the binary nature of the response variable, a binomial distribution with a logit link function was used:

$$\begin{aligned}\text{logit}\left(P\left(Y_{ij}=1\right)\right)=&\beta_0+\beta_1\text{ELE}_i+\beta_2\text{SSF}_i+\beta_3\text{FA}_i+\beta_4\text{TI}_i+\beta_5\text{PP}_{ij}\\&+\beta_6\text{ST}_{ij}+\beta_7\left(\text{Site type}_i\times\text{NDVI}_{ij}\right)+\left(1|\,\text{WCS}\right)\end{aligned}\quad(1)$$

In Equation (1), $P\left(Y_{ij}=1\right)$ represents the probability of snail presence detected by the expert (SPex_{ij}) or the CS (SPcs_{ij}) at a given sampling instance (ij). Geomorphological predictors ($\text{ELE}_i, \text{SSF}_i, \text{FA}_i, \text{TI}_i$) vary across space ($i$), while environmental predictors ($\text{PP}_{ij}, \text{ST}_{ij}, \text{NDVI}_{ij}$) vary across both space (i) and time (j). The intercept β_0 represents the baseline log-odds of snail presence, and the coefficients β_1 to β_7 represent the fixed effects for each variable. The interaction term $\left(\text{Site type}_i\times\text{NDVI}_{ij}\right)$ captures how vegetation influences snail presence across different site types (lake, spring, stream, wetland). The random effect $(1|\,\text{WCS})$ accounts for variability in baseline log-odds across different WCS.

2.3.2 | Imperfect detection scenario

For imperfect detection, we assumed that the expert's best-fit model represents the 'true' probability of snail presence. We applied the best expert model to the citizen science data and compared the model-derived probabilities with the observed data from CS. The differences between the expert model's predictions and the CS observations were categorized into two cases: probability of *false negatives* and probability of *unexpected observations*.

False negatives (FN_{ij}) were defined as cases where the expert model predicted a snail probability of presence, but the CS reported an absence. *Unexpected observations* (UO_{ij}) were cases where the CS reported a snail presence, but the expert model predicted a probability of absence.

A GLMM with a beta distribution was fitted for both FN_{ij} and OU_{ij} separately, as the probabilities are bounded between 0 and 1 (Johnson, 2013), using the following model:

$$\begin{aligned}\text{logit}\left(P\left(Y_{ij}\right)\right)=&\beta_0+\beta_1\text{ID}_i+\beta_2\text{CONS}_{ij}+\beta_3\text{SAMT_sin}_{ij}\\&+\beta_4\text{SAMT_cos}_{ij}+\beta_5\text{SDU}_{ij}+\beta_6\text{CSF}_{ij}\\&+\beta_7\text{SDE}_i+\left(1|\,\text{WCS}\right)\end{aligned}\quad(2)$$

In Equation (2), $P\left(Y_{ij}\right)$ represents the expected value of the probability of false negatives (FN_{ij}) or the probability of unexpected observations (UO_{ij}) which ranges between 0 and 1, for a given sampling event (ij). The variable ID_i represents the CS ID and SDE_i denotes the sampling density, both of which vary across space (i). The variables $\text{CONS}_{ij}, \text{SAMT_sin}_{ij}, \text{SAMT_cos}_{ij}, \text{SDU}_{ij}, \text{CSF}_{ij}$ are detection variables that vary across space (i) and time (j). The intercept β_0 is the baseline log-odds when all the fixed effects are zero, and β_1 to β_7 represent the fixed effects for each predictor. The random effect $(1|\,\text{WCS})$ accounts for variability across WCS.

2.3.3 | Model selection and diagnostics

To minimize redundancy among predictors, we calculated a Spearman's correlation matrix (threshold $\rho>0.7$) for all numerical predictors within each scenario and case, retaining only one variable per highly correlated pair. Multicollinearity was further assessed using the generalized variance inflation factor (GVIF) and the adjusted GVIF (considering degrees of freedom) for categorical variables, with a threshold of 10 (Zuur et al., 2009). These steps were performed using the *vif* function from the *car* package (Fox & Weisberg, 2019) in R.

Predictor variables were standardized (mean-centred and scaled by standard deviation) across the CS and expert dataset (including

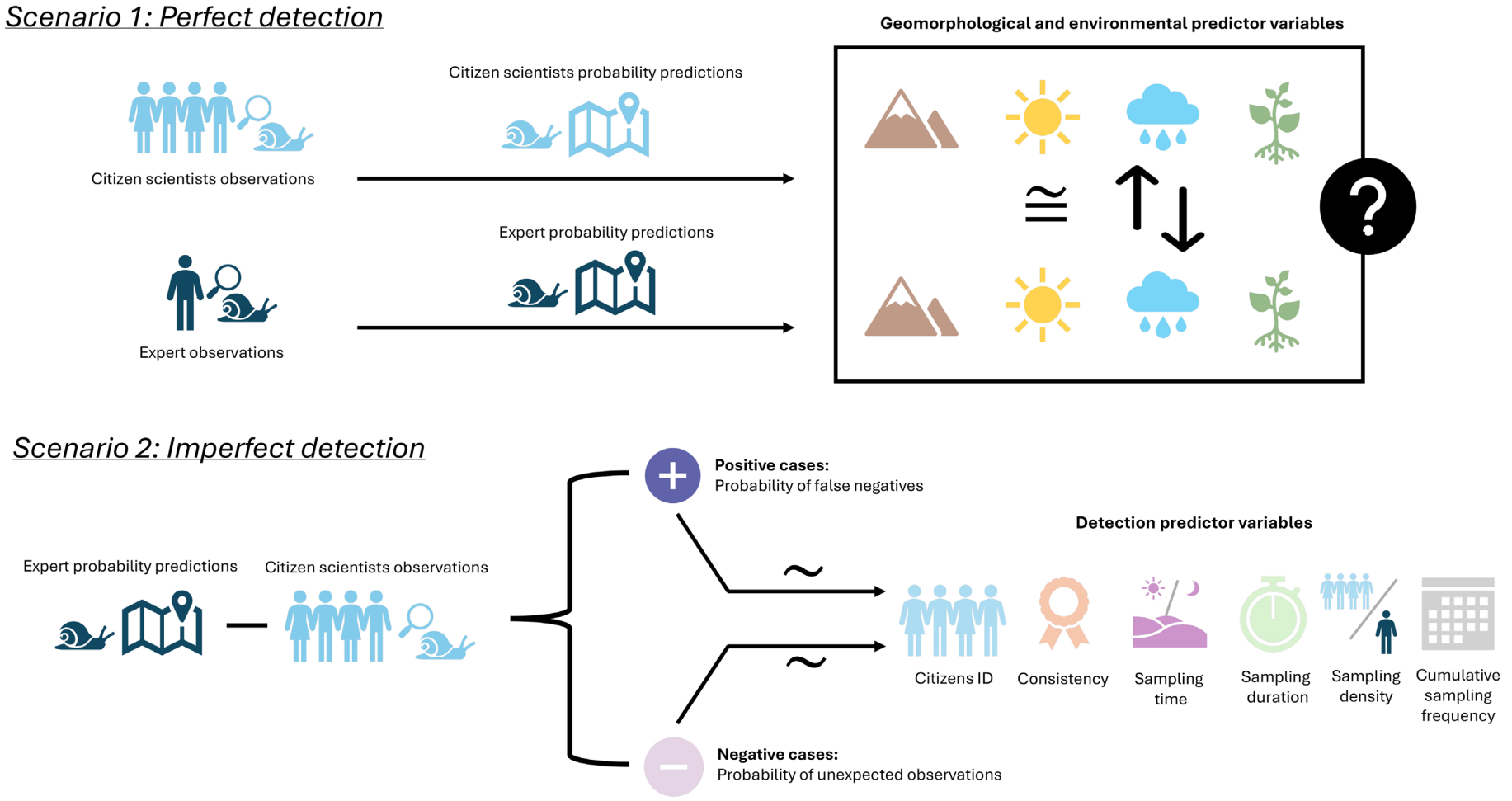


FIGURE 2 Overview of modelling scenarios.

geomorphological and environmental predictors) and for each case of the imperfect detection scenario (false negatives and unexpected detection) to improve model convergence and facilitate the comparison of effect sizes (Schielzeth, 2010).

For the perfect detection scenario, we used the *glmer* function from the *lme4* package (Bates et al., 2015) to fit binomial models. For imperfect detection, we applied the *glmmTMB* function from *glmmTMB* (Brooks et al., 2017), modelling probabilities with a beta distribution (Johnson, 2013).

All possible predictor combinations were tested, and model performance was assessed using Akaike's information criterion (AIC). Models within 2 AIC units of the best-performing model were considered equally valid (Burnham & Anderson, 2002). Final models were selected based on AIC, as well as marginal and conditional R^2 values, indicating variance explained by fixed effects and by both fixed and random effects, respectively.

After model selection, we performed a likelihood ratio test (ANOVA) (Kaufmann & Schering, 2014) to assess whether adding the NDVI:Site type interaction term significantly improved model fit, based on the hypothesis that NDVI, as a proxy of vegetation, influences snail presence differently across site types (Ofulla et al., 2013).

Finally, we compared models based on the full CS dataset and a temporally aligned subset (2805 reports) corresponding to the expert sampling period (Figure S1) to test sensitivity to sampling coverage.

3 | RESULTS

3.1 | Overview of model selection and diagnostics

3.1.1 | Diagnostics

In the perfect detection scenario, a Spearman's correlation matrix showed a strong positive correlation ($\rho > 90\%$) between the topographic index (TI) and flow accumulation (FA) in both datasets (Figure S3). Due to this high collinearity, we retained FA for further analysis, as it provided a more straightforward interpretation. In the imperfect detection scenario, no highly correlated predictors pairs were identified (Figure S4). Furthermore, none of the numerical or categorical predictors exceeded the GVIF threshold of 10 in either scenario, indicating no concerns with multicollinearity (Tables S3 and S4).

3.1.2 | Model selection

A total of 508 models were computed, representing all possible combinations of seven predictor variables ($2^7 - 1$), across four outcomes: perfect detection for both expert and CS datasets, and imperfect detection for the probability of false negatives and unexpected observations.

Table 1 summarizes the best-performing models for each scenario and dataset, detailing selected predictors, AIC and R^2 values. Complete model summaries are provided in Table S5 (perfect detection) and Table 2 (imperfect detection).

3.2 | Perfect detection scenario

The perfect detection scenario compared *Biomphalaria* snail presence/absence models by the expert and CS, based on geomorphological and environmental variables. Although the best models for citizens and experts did not share the same structure (Table 1), the most parsimonious models (AIC difference < 2), considered comparable between them, were evaluated to assess predictor frequency and significance (Figure 3a). Site type was included in all the best models, while NDVI appeared in fewer expert models (3 out of 15) but was consistently present in the CS models. Both site type and NDVI were significant predictors in both datasets and the interaction between site type and NDVI was consistently significant in CS models (Figure 3a).

Despite differences in effect size and significance, consistent trends in direction and uncertainty were observed across both datasets (Figure 3b). While the inclusion of the site-NDVI interaction term did not significantly improve the expert model ($p = 0.147$), it significantly improved the CS model fit ($p = 0.011$), with significant effects for spring ($p = 0.042$) and stream ($p = 0.003$) sites.

3.3 | Imperfect detection scenario

3.3.1 | Overview of observed probabilities

The observed probabilities for false negatives and unexpected observations were predominantly near zero, indicating a strong

TABLE 1 Selected models for the perfect and imperfect detection scenarios.

Scenario	Model	Selected predictors [partial R^2]	AIC value	Marginal R^2	Conditional R^2
Perfect detection	Expert $(SPex_{ij})$	ELE [0.021], PP [0.001], Site type × NDVI [0.004], Site type [0.139], NDVI [0.005]	978.8	0.17	0.65
	Citizen scientists $(SPcs_{ij})$	FA [0.010], SS [0.034], Site type × NDVI [0.002], Site type [0.098], NDVI [0.002]	3087.7	0.16	0.70
Imperfect detection	False negatives (FN_{ij})	SAMTsin [0.001], SAMTcos [0.001], CSF [0.002], ID [0.423], SDE [0.053]	−5349.8	0.59	0.99
	Unexpected observations (UO_{ij})	SAMTsin [0.0], CSF [0.018], ID [0.716], SDE [0.034]	−8571.6	0.75	0.98

TABLE 2 Odd ratios for predictor variables in the imperfect detection scenario.

	False negatives			Unexpected observations		
Predictor	OR	95% CI	*p*-value	OR	95% CI	*p*-value
Sampling time (sin)	1.02	1.00, 1.04	0.027	–	–	–
Sampling time (cos)	1.01	1.00, 1.03	0.059	–	–	–
Cumulative sampling frequency	0.96	0.91, 1.01	0.120	1.10	1.06, 1.14	<0.001
Sampling duration	–	–	–	1.00	0.99, 1.01	0.800
Sampling density	0.03	0.00, 0.73	0.031	3.55	1.25, 10.00	0.030
ID						
1	0.00	0.00, 0.03	0.006	–	–	–
3	1.37	0.03, 57.9	0.870	4.72	0.65, 34.1	0.120
4	21.10	0.06, 7512	0.310	–	–	–
5	0.00	0.00, 0.08	<0.001	16.30	3.52, 75.2	<0.001
6	–	–	–	695.00	96.1, 5029	<0.001
7	0.00	0.00, 0.03	0.001	83.50	5.54, 1258	0.001
8	0.00	0.00, 0.01	<0.001	26.70	1.44, 496	0.028
9	0.00	0.00, 0.07	<0.001	2.42	0.51, 11.5	0.270
10[a]	–	–	–	–	–	–
14	0.00	0.00, 0.04	0.003	401.00	17.0, 9444	<0.001
15	0.01	0.00, 0.08	<0.001	9.54	1.89, 48.1	0.006
16	–	–	–	2.61	0.56, 12.1	0.220
17	–	–	–	5.36	1.46, 19.7	0.011
18	0.02	0.00, 0.28	0.003	23.90	6.24, 91.7	<0.001
19	–	–	–	39.70	1.13, 1390	0.042
20	–	–	–	164.00	5.53, 4846	0.003
21	0.02	0.00, 0.19	<0.001	19.70	4.88, 79.6	<0.001
22	0.00	0.00, 0.01	0.002	633.00	23.8, 16846	<0.001
23	0.01	0.00, 0.09	<0.001	9.51	2.43, 37.2	0.001
24	0.00	0.00, 0.00	<0.001	36.10	0.86, 1512	0.060
25	0.00	0.00, 0.00	<0.001	–	–	–

Note: Significant *p*-values (<0.05) highlighted in red.

Abbreviations: CI, confidence interval; OR, odds ratio.

[a]ID 10 was used as a reference since it had the highest sampling frequency.

alignment between expert-derived predictions and the CS observations (Figure 4).

3.3.2 | False negatives (FN_{ij})

A one SD increase in the sine ($SD_{SAMP_sin} = 0.55$) transformation of sampling time was associated with a 2% higher probability of false negatives ($p = 0.027$). In contrast, a one SD increase in sampling density ($SD_{SDE} = 1.52$), the ratio of citizen scientist to expert sampling efforts per site, reduced false negatives' likelihood by 97% ($p = 0.031$). False negatives' probabilities varied substantially across CS IDs (80%, $p \leq 0.050$), with some IDs showing around 99% lower probabilities compared to ID 10, which served as the reference level due to its high sampling effort (Table 2).

3.3.3 | Unexpected observations (UO_{ij})

A one SD increase in cumulative sampling frequency ($SD_{CSF} = 75.49$), the number of sampling events per citizen scientist, increased unexpected observations odds by 10% ($p \leq 0.001$). Similarly, a one SD increase in sampling density ($SD_{SDE} = 1.18$) increased the likelihood of unexpected observations by 355% ($p = 0.03$). Unexpected observations probabilities varied across CS IDs (76%, $p \leq 0.05$), with some IDs showing 5.36 to 695 times higher probabilities compared to ID 10 (Table 2).

3.4 | Additional analysis with reduced citizen scientist dataset

Using the reduced CS dataset aligned with the expert sampling period, we observed a general decrease in statistical significance across

scenarios, though effect directions remained consistent (Figure S5; Tables S6 and S7). Notably, cumulative sampling frequency, which was non-significant in the extended false negatives model, became highly significant (OR = 1.16, $p \leq 0.001$). A one SD increase (88.62 sampling instances) corresponded to a 16% rise in false negatives detection probability, though its partial R^2 remained low (0.001).

4 | DISCUSSION

4.1 | *Great minds map alike*: Citizen scientists models replicate expert-derived *Biomphalaria* spp. suitability models

Our results reveal strong congruence between CS and expert models, with both identifying site type and NDVI as significant predictors of *Biomphalaria* spp. presence (Figure 3). This alignment reinforces growing evidence that citizen science can reliably complement and, in many cases, replicate expert-driven ecological modelling (Dickinson et al., 2010, 2012; Fraisl et al., 2022; Mair & Ruete, 2016).

A closer examination of these patterns, indicates that NDVI exhibited a habitat-dependent relationship with snail presence, showing positive associations in streams and springs, negative associations in lake sites and no significant effects in wetlands (Table S5). These site-specific differences likely reflect how vegetation is captured across different waterbodies in our study area. In small streams and springs, where water cover is typically low and interference of water pixels is minimal, NDVI primarily reflects terrestrial vegetation which may provide shading. Reduced light exposure due to shading could facilitate snail detection, as *Biomphalaria* snails tend to be less active during direct and prolonged light exposure (Rotenberg et al., 1989). Conversely, in lake and wetland sites, NDVI likely represents a composite signal influenced by water surface fluctuations, aquatic and terrestrial vegetation. The negative

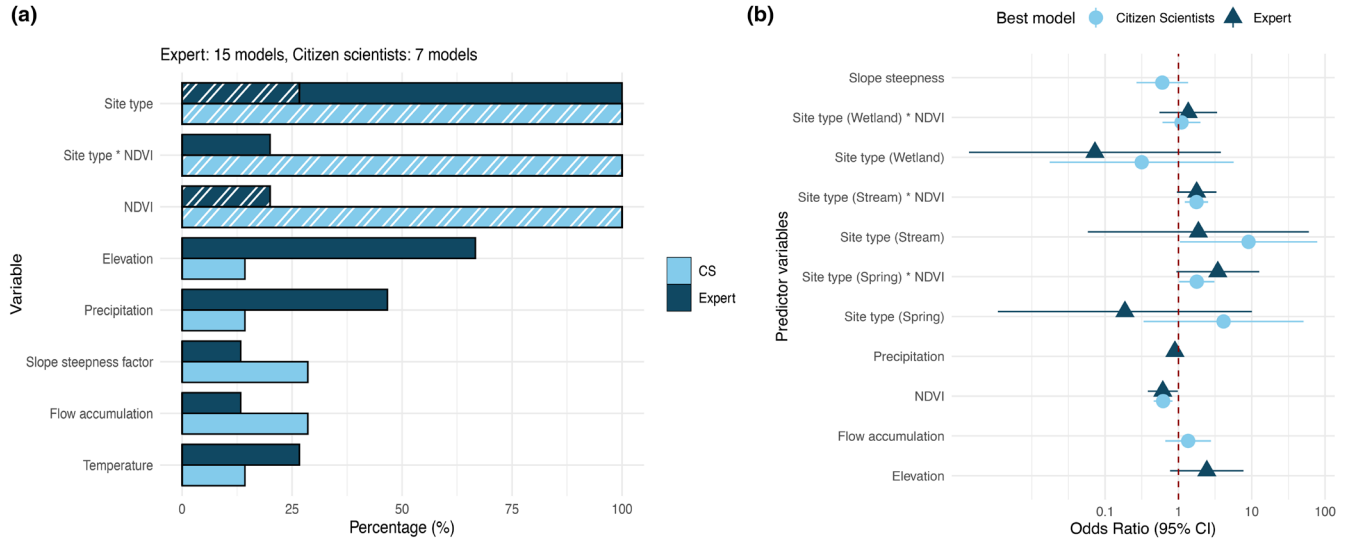


FIGURE 3 (a) Summary of predictor variables frequency in top models by AIC; the stripes indicate how frequent the predictors were significant. (b) Graphic representation of standardized odd ratios for the best models derived from the expert and citizen scientists' datasets.

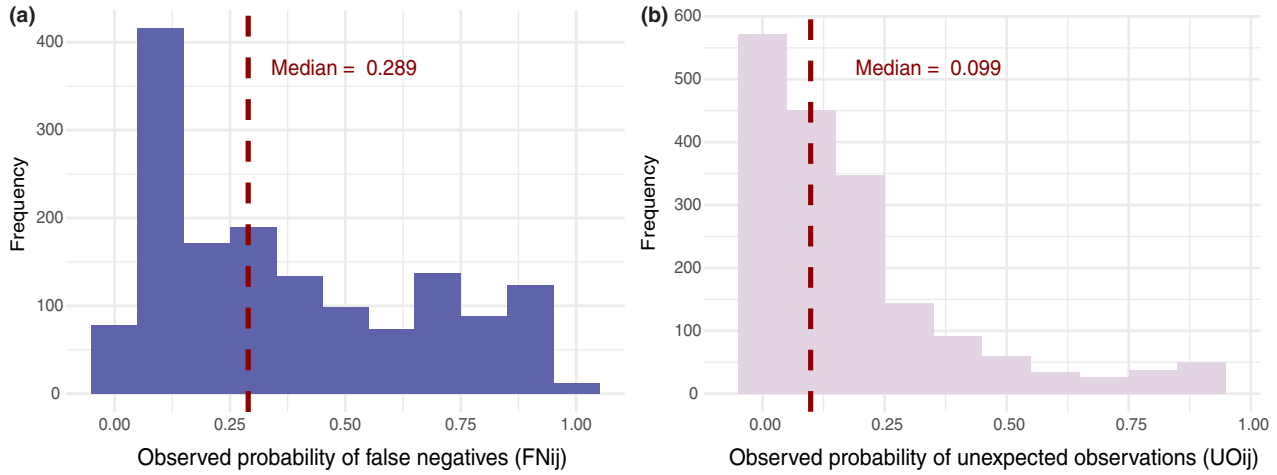


FIGURE 4 Histogram of imperfect detection: (a) probability of false negatives, (b) probability of unexpected observations.

association between NDVI and *Biomphalaria* spp. presence in lake sites may suggest that water expansion favours snail colonization. Previous studies have reported that intermediate host snails can benefit from flooding conditions, which create additional breeding areas (Guo et al., 2021; Qiu et al., 2022). However, these studies also emphasize the importance of considering short and long-term effects of such extreme events, as their impact on snail populations may vary over time.

4.2 | Broad-scale variables do not accurately predict the distribution of *Biomphalaria* snails—likely due to their patchy distribution

Both CS and expert models identified NDVI and site type as key predictors, but their predictive power was limited, with marginal R^2 values of 17% for the expert model and 16% for the CS model. Notably, these values were remarkably close, further confirming the consistency between predictions from both datasets. However, the low R^2 values indicate that broad-scale predictors alone do not fully explain *Biomphalaria* spp. distribution.

In contrast, the relatively high conditional R^2 values (>65%) across both datasets suggest that site-specific characteristics play a dominant role, regardless of the data source. This finding aligns with research demonstrating that freshwater invertebrates are highly dependent on microhabitat features such as flow velocity, water depth and substrate type (Vaessen & Houbrechts, 2022). For *Biomphalaria* spp., suitable habitats are typically associated with shallow water depths (2–4 cm) and substrates like plant detritus or bedrock, while stable water levels in man-made structures further support snail presence by fostering algal growth (Utzinger & Tanner, 2000). These habitat characteristics shape the patchy distribution of *Biomphalaria* snails, which in turn reinforces the focal nature of schistosomiasis transmission and prevalence (Stothard et al., 2017; Walz et al., 2015).

Given the importance of these fine-scale habitat characteristics, future efforts should focus on identifying relevant environmental variables that can be robustly and safely collected through citizen science initiatives. Integrating measurements such as substrate composition, water depth and aquatic vegetation cover could improve our understanding of *Biomphalaria* snails' distribution and refine habitat suitability models.

4.3 | Optimizing citizen science: Accounting for observer variability and sampling effort

Despite common concerns over variability in CS observations (Earp et al., 2022; Kosmala et al., 2016), our results indicate minimal discrepancies between the 'optimal' expert model and the CS data (Figure 4). However, individual variability among CS participants emerged as the strongest driver of these minor differences in snail detection, followed by sampling density and cumulative sampling frequency (Table 2).

Inter-observer variability was the primary source of detection differences (Table 1, partial R^2). Some CS recorded absences when the expert model predicted a presence, while others recorded presences that were not predicted by the expert model. Such variability is well documented in citizen science datasets (Kosmala et al., 2016) and likely reflects differences in prior experience, motivation and search effort (Kosmala et al., 2016). Although often viewed as a limitation, observer variability does not inherently compromise data quality when appropriate statistical methods, such as GLMMs with random effects, are used to account for it (Bird et al., 2014), as demonstrated in this study. Practically, future programs could further reduce variability by standardizing effort with a timer and mandatory start/stop times; providing more frequent calibration/refresher trainings; scheduling occasional paired resurveys (two observers at the same WCS within a limited period) to quantify inter-rater variability (Kosmala et al., 2016) and organizing regular expert–CS joint surveys for feedback (Freitag et al., 2016; Kosmala et al., 2016).

Sampling effort also influenced detection patterns particularly in reducing false negatives. The probabilities of false negatives decreased substantially with higher sampling density. This emphasizes the critical role of intensified sampling in detecting patchily distributed species, particularly for non-expert observers (Mair & Ruete, 2016). In the reduced dataset, cumulative sampling frequency was significant, with an additional 87 sampling instances leading to a 16% increase in false negatives (Table S7). However, since including this term did not improve the marginal R^2, these results should be interpreted with caution. Overall, our findings confirm that sampling density, an indicator of sampling effort, plays a key role in reducing false negatives, though its impact is secondary to individual observer differences.

The probability of unexpected observations increased with sampling density and cumulative sampling frequency (Table 2), consistent with the idea that frequent engagement with a task improves observation skills (Kosmala et al., 2016) and that repeated site visits enhance occupancy estimation (Mackenzie et al., 2005). The ability of CS to detect snails where the expert model predicted low suitability suggests that fine-scale, frequent sampling can reveal ecological patterns that broad-scale expert models may overlook. Similar advantages are seen in vector surveillance: in Mosquito Alert (Južnic-Zont et al., 2022), participants contributed a vast number of real-time mosquito records that enabled early warning systems, which were not feasible with traditional data collection methods. Large, structured volunteer datasets can therefore complement expert surveys by improving spatiotemporal coverage (Kelling et al., 2019).

While increasing sampling effort enhances detection, resource efficiency remains a critical consideration. Previous work in this network estimated that CS data collection costs were 7.6 times lower than expert-driven surveys, even when accounting for PPE and annual refresher trainings (Tumusiime et al., 2024). However, despite the cost's advantages, there is a clear trade-off between effort and marginal gains: in our case, 75 additional sampling instances resulted in only a 10% increase in unexpected observations (Table 2).

4.4 | General recommendations for citizen science projects in vector surveillance

Projects should be co-designed with local stakeholders and participants to align objectives with local needs, build ownership and trust and ensure feasible protocols (De Agustin Camacho et al., 2023; Ind & Coates, 2013). Co-design should also specify safety rules, PPE requirements and ethical considerations. Effort and evidence should ideally be standardized (e.g. a ~30-min timed search with recorded start/stop times, observer ID, GPS accuracy threshold) and photos per category should meet a pre-specified minimum pixel resolution to ensure valid and verifiable records. If flexible scheduling yields patchy data that limit inference, agree with participants on minimal visit frequency (e.g. biweekly or monthly), communicate it clearly and record any deviation while retaining flexibility for safety and access. The sampling design should include both known presence and initial non-detection sites to avoid presence-only bias (Phillips et al., 2009). Periodic refresher trainings should be combined with simple and timely feedback to sustain engagement (Brees et al., 2021; Freitag et al., 2016; Kosmala et al., 2016). To limit inter-observer variability, occasional paired resurveys, observer rotations (Kosmala et al., 2016) and expert–CS joint surveys (co-visits to a subset of sites) should be foreseen. Data curation should include systematic photo validation and predefined exclusion criteria. Analyses should be effort-aware, modelling observer effects and reporting the explained variance. Targeted expert resurveys in high-uncertainty areas can increase data quality, and where appropriate, expert-citizen data-integration approaches can further improve inference. To further inform disease risk assessments, these designs should be paired with bias-corrected abundance (accounting for effort and observer) and targeted infection/shedding assays.

Taken together, these design, calibration and validation steps reduce inter-observer variability and yield accurate, effort-aware citizen science datasets at lower cost, providing a practical complement to expert surveys for vector surveillance in remote settings.

AUTHOR CONTRIBUTIONS

Noelia Valderrama-Bhraunxs, Liesbet Jacobs, Emiel van Loon and Tine Huyse conceived the idea and designed methodology. Noelia Valderrama-Bhraunxs handled data curation, software development and wrote the original draft. Tine Huyse managed project administration, while Liesbet Jacobs, Emiel van Loon, Tine Huyse, Anton Van Rompaey and Ronald Twongyirwe provided supervision. All authors contributed to writing, reviewing and editing the manuscript.

ACKNOWLEDGEMENTS

We thank all the citizen scientists from the ATRAP network for their invaluable contributions: 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. We also would like to acknowledge Julius Tumusiime for contributing to the expert collection data.

FUNDING INFORMATION

This work was supported by the Belgian Directorate General for Development, Cooperation and Humanitarian Aid (XM-DAC-2-10-3853) and FWO Fonds Wetenschappelijk Onderzoek (11L3223N and 11L3225N) for NVB's PhD fellowship.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2688-8319.70163.

DATA AVAILABILITY STATEMENT

Data available at https://doi.org/10.6084/m9.figshare.28778069 (Valderrama-Bhraunxs et al., 2025).

ORCID

Noelia Valderrama-Bhraunxs https://orcid.org/0000-0002-1804-6715
Emiel van Loon https://orcid.org/0000-0002-8895-0427

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. *Parasite Epidemiology and Control*, *27*, e00394. https://doi.org/10.1016/J.PAREPI.2024.E00394

Adriko, M., Tinkitina, B., Tukahebwa, E. M., Standley, C. J., Stothard, J. R., & Kabatereine, N. B. (2018). Data on the pre-MDA and post MDA interventions for *Schistosoma mansoni* and *Schistosoma haematobium* in a co-endemic focus in Uganda: 1951–2011. *Data in Brief*, *20*, 991–998. https://doi.org/10.1016/J.DIB.2018.08.200

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154. https://doi.org/10.1016/J.BIOCON.2013.07.037

Bolker, B. M. (2015). Linear and generalized linear mixed models. *Ecological Statistics*, 309–333. https://doi.org/10.1093/ACPROF:OSO/9780199672547.003.0014

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 Science: Theory and Practice*, *6*(1), 1–13. https://doi.org/10.5334/CSTP.388

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Burnham, K. P., & Anderson, D. R. (2002). Multimodal inference. *Sociological Methods & Research*, *33*, 261–304. https://www.nhbs.com/model-selection-and-multi-model-inference-book

Callaghan, C. T., Poore, A. G. B., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J. L., Vergés, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *Bioscience*, *71*(1), 55–63. https://doi.org/10.1093/BIOSCI/BIAA131

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

De Agustin Camacho, A., Petegem, W. V. A. N., Droog, M. D. E., & Jacobs, L. (2023). Context—Goal—Method—Outcome: Alignment in citizen science project design and its relation to supporting the United Nations sustainable development goals. *Citizen Science: Theory and Practice*, *8*(1). https://doi.org/10.5334/CSTP.570

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 149–172. https://doi.org/10.1146/ANNUREV-ECOLSYS-102209-144636/CITE/REFWORKS

Earp, H., Vye, S., Bohn, K., Burrows, M., Chenery, J., Dickens, S., Foster, C., Grist, H., Lamont, P., Long, S., Morrall, Z., Pocklington, J., Scott, A., Watson, G., West, V., Jenkins, S., Delany, J., & Sugden, H. (2022). Do you see what I see? Quantifying inter-observer variability in an intertidal marine citizen science experiment. *Citizen Science: Theory and Practice*, *7*(1). https://doi.org/10.5334/CSTP.483

ECSA. (2015). *Ten principles of citizen science*. ECSA.

Elbroch, M., Mwampamba, T. H., Santos, M. J., Zylberberg, M., Liebenberg, L., Minye, J., Mosser, C., & Reddy, E. (2011). The value, limitations, and challenges of employing local experts in conservation research. *Conservation Biology*, *25*(6), 1195–1202. https://doi.org/10.1111/J.1523-1739.2011.01740.X

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression*. Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P. Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, *2*(1), 1–20. https://doi.org/10.1038/s43586-022-00144-4

Freitag, A., Meyer, R., & Whiteman, L. (2016). Strategies employed by citizen science programs to increase the credibility of their data. *Citizen Science: Theory and Practice*, *1*(1), 2. https://doi.org/10.5334/CSTP.6

Fung, T., & Siu, W. (2000). Environmental quality and its changes, an analysis using NDVI. *International Journal of Remote Sensing*, *21*(5), 1011–1024. https://doi.org/10.1080/014311600210407

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. *Scientific Data*, *2*. https://doi.org/10.1038/SDATA.2015.66

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

Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, *116*(2), 195–203. https://doi.org/10.1016/S0006-3207(03)00190-3

Guo, S. Y., Li, L., Zhang, L. J., Li, Y. L., Li, S. Z., & Xu, J. (2021). From the one health perspective: Schistosomiasis japonica and flooding. *Pathogens*, *10*(12), 1538. https://doi.org/10.3390/PATHOGENS10121538

Holmgren, P. (1994). Multiple flow direction algorithms for runoff modelling in grid-based elevation models: An empirical evaluation. *Hydrological Processes*, *8*(4), 327–334. https://doi.org/10.1002/HYP.3360080405

Ind, N., & Coates, N. (2013). The meanings of co-creation. *European Business Review*, *25*(1), 86–95. https://doi.org/10.1108/09555341311287754

Johnson, R. W. (2013). Applications of the beta distribution. *ArXiv Preprint*. http://arxiv.org/abs/1307.6437

Južnic-Zont, Ž., Sanpera-Calbe, I., Eritja, R., Palmer, J. R. B., Escobar, A., Garriga, J., Oltra, A., Richter-Boi, A., Schaffner, F., Della Torre, A., Miranda, M. Á., Koopmans, M., Barzon, L., Ferre, F. B., Alarcón-Elbal, P. M., González, M. A., Puig, M. A., Bakran-Leb, K., Balatsos, G., … Zittra, C. (2022). Mosquito alert: leveraging citizen science to create a GBIF mosquito occurrence dataset. *GigaByte*, *2022*, gigabyte54. https://doi.org/10.46471/GIGABYTE.54

Kalinda, C., Chimbari, M., & Mukaratirwa, S. (2017). Implications of changing temperatures on the growth, fecundity and survival of intermediate host snails of schistosomiasis: A systematic review. *International Journal of Environmental Research and Public Health*, *14*(1), 80. https://doi.org/10.3390/IJERPH14010080

Kaufmann, J., & Schering, A. (2014). *Analysis of variance ANOVA*. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.STAT06938

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., Hochachka, W. M., Julliard, R., Kraemer, R., & Guralnick, R. (2019). Using semistructured surveys to improve citizen science data for monitoring biodiversity. *Bioscience*, *69*(3), 170–179. https://doi.org/10.1093/BIOSCI/BIZ010

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *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. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/HYP.9740

Lo, N. C., Bezerra, F. S. M., Colley, D. G., Fleming, F. M., Homeida, M., Kabatereine, N., Kabole, F. M., King, C. H., Mafe, M. A., Midzi, N., Mutapi, F., Mwanga, J. R., Ramzy, R. M. R., Satrija, F., Stothard, J. R., Traoré, M. S., Webster, J. P., Utzinger, J., Zhou, X. N., … Garba, A. (2022). Review of 2022 WHO guidelines on the control and elimination of schistosomiasis. *The Lancet Infectious Diseases*, *22*(11), e327–e335. https://doi.org/10.1016/S1473-3099(22)00221-3

Mackenzie, D. I., Royle, J. A., Mackenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, *42*(6), 1105–1114. https://doi.org/10.1111/J.1365-2664.2005.01098.X

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

McCool, D. K., Brown, L. C., Foster, G. R., Mutchler, C. K., & Meyer, L. D. (1987). Revised slope steepness factor for the universal soil loss equation. *Transactions of ASAE*, *30*(5), 1387–1396. https://doi.org/10.13031/2013.30576

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*. https://doi.org/10.1017/S0031182001007703

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *R Journal*, *9*(1), 207–218. https://doi.org/10.32614/RJ-2017-009

Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open-source GIS. *Environmental Modelling and Software*, *31*, 124–130. https://doi.org/10.1016/J.ENVSOFT.2011.11.014

Nonogaki, S., Shiono, K., Masumoto, S., & Ekawa, M. (2006). An algorithm of surface estimation using cubic b-spline function for geological modeling. *International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*.

Ofulla, A. V., Adoka, S. O., Anyona, D. N., Abuom, P. O., Karanja, D., Vulule, J. M., Okurut, T., Matano, A. S., Dida, G. O., Jembe, T., & Gichuki, J. (2013). Spatial distribution and habitat characterization of schistosomiasis host snails in lake and land habitats of western Kenya. *Lakes & Reservoirs: Research & Management*, *18*(2), 197–215. https://doi.org/10.1111/LRE.12032

Okoro, O. J., Deme, G. G., Okoye, C. O., Eze, S. C., Odii, E. C., Gbadegesin, J. T., Okeke, E. S., Oyejobi, G. K., Nyaruaba, R., & Ebido, C. C. (2023). Understanding key vectors and vector-borne diseases associated with freshwater ecosystem across Africa: Implications for public health. *Science of the Total Environment*, *862*, 160732. https://doi.org/10.1016/J.SCITOTENV.2022.160732

Parrish, J. K., Jones, T., Burgess, H. K., He, Y., Fortson, L., & Cavalier, D. (2019). Hoping for optimality or designing for inclusion: Persistence, learning, and the social network of citizen science. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 1894–1901. https://doi.org/10.1073/pnas.1807186115

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Planet Labs PBC. (2023). *Planet application program interface: In space for life on Earth*. https://api.planet.com

Qiu, J., Han, D., Li, R., Xiao, Y., Zhu, H., Xia, J., Jiang, J., Han, Y., Shao, Q., Yan, Y., & Li, X. (2022). Satellite imagery-based identification of high-risk areas of schistosome intermediate snail hosts spread after flood. *Remote Sensing*, *14*(15), 3707. https://doi.org/10.3390/RS14153707

Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, *5*(1), 59–79. https://doi.org/10.1002/HYP.3360050106

Ross, A. G. P., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y., Williams, G. M., & McManus, D. P. (2002). Schistosomiasis. *New England Journal of Medicine*, *346*(16), 1212–1220. https://doi.org/10.1056/NEJMRA012396

Rotenberg, L., Jurberg, P., & Pieri, O. S. (1989). Relationship between light conditions and behavior of the freshwater snail *Biomphalaria glabrata* (Say). *Hydrobiologia*, *174*(2), 111–116. https://doi.org/10.1007/BF00014059/METRICS

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. https://doi.org/10.1111/J.2041-210X.2010.00012.X

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

Shiff, C. (2017). Why reinvent the wheel? Lessons in schistosomiasis control from the past. *PLoS Neglected Tropical Diseases*, *11*(10), e0005812. https://doi.org/10.1371/JOURNAL.PNTD.0005812

Stensgaard, A. S., Jørgensen, A., Kabatereine, N. B., Rahbek, C., & Kristensen, T. K. (2006). Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda. *Geospatial Health*, *1*(1), 93–104. https://doi.org/10.4081/GH.2006.284

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. *Infectious Diseases of Poverty*, *6*(1), 1–11. https://doi.org/10.1186/S40249-016-0215-9/FIGURES/3

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 Neglected Tropical Diseases*, *18*(4), e0012062. https://doi.org/10.1371/JOURNAL.PNTD.0012062

Utzinger, J., & Tanner, M. (2000). Microhabitat preferences of *Biomphalaria pfeifferi* and *Lymnaea natalensis* in a natural and a man-made habitat in southeastern Tanzania. *Memórias do Instituto Oswaldo Cruz*, *95*(3), 287–294. https://doi.org/10.1590/S0074-02762000000300002

Vaessen, Q., & Houbrechts, G. (2022). Microhabitats of a threatened freshwater mussel in gravel-bed rivers (Belgium). *10th International Conference on Geomorphology*, Coimbra, Portugal, 12–16 Sep 2022, *ICG2022*. https://doi.org/10.5194/ICG2022-400

Valderrama-Bhraunxs, N., Huyse, T., Van Loon, E., Van Rompaey, A., Twongyirwe, R., & Jacobs, L. (2025). *Integrated multisource modeling dataset on schistosome snail hosts in west Uganda: Snail occurrence, environmental, geomorphological, and detection variables*.

Walz, Y., Wegmann, M., Dech, S., Vounatsou, P., Poda, J. N., N'Goran, E. K., Utzinger, J., & Raso, G. (2015). Modeling and validation of environmental suitability for schistosomiasis transmission using remote sensing. *PLoS Neglected Tropical Diseases*, *9*(11), e0004217. https://doi.org/10.1371/JOURNAL.PNTD.0004217

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.

Wang, X. Y., Li, Q., Li, Y. L., Guo, S. Y., Li, S. Z., Zhou, X. N., Guo, J. G., Bergquist, R., Juma, S., Zhang, J. F., Yang, K., & Xu, J. (2024). Prevalence and correlations of schistosomiasis mansoni and schistosomiasis haematobium among humans and intermediate snail hosts: A systematic review and meta-analysis. *Infectious Diseases of Poverty*, *13*(1), 1–15. https://doi.org/10.1186/S40249-024-01233-0/FIGURES/3

Warren, K. S. (1978). The pathology, pathobiology and pathogenesis of schistosomiasis. *Nature*, *273*(5664), 609–612. https://doi.org/10.1038/273609A0;KWRD

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. https://doi.org/10.1007/978-0-387-87458-6

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Protocol to download Planet data.

Table S1: Summary of geomorphological variables.

Table S2: Summary of environmental variables.

Table S3: Generalized Variance Inflation Factor for the perfect detection predictor variables.

Table S4: Generalized Variance Inflation Factor for the imperfect detection predictor variables.

Table S5: Odd ratios for predictor variables in the perfect detection scenario.

Table S6: Odd ratios for predictor variables in the perfect detection scenario (reduced dataset).

Table S7: Odd ratios for predictor variables in the imperfect detection scenario (reduced dataset).

Figure S1: Time span of sampling instances for the citizen scientists and the expert, red box indicating overlapping period between expert and citizen scientists.

Figure S2: 56 common water contact sites.

Figure S3: Correlation matrix of geomorphological and environmental predictors for the (a) expert model and (b) the citizen scientist's model.

Figure S4: Correlation matrix of detection predictors for the (a) false negatives and (b) the unexpected observations.

Figure S5: Graphic representation of odd ratios for the best models derived from the expert, extended and reduced citizen scientist's datasets.

How to cite this article: Valderrama-Bhraunxs, N., Huyse, T., van Loon, E., Van Rompaey, A., Twongyirwe, R., & Jacobs, L. (2025). *Great minds map alike*: Citizen and expert distribution models of schistosome snail hosts in rural west Uganda. *Ecological Solutions and Evidence*, *6*, e70163. https://doi.org/10.1002/2688-8319.70163