A correction article relating to this paper has been published and can be found at https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.1140.

CITIZEN SCIENCE:
THEORY AND PRACTICE

Are the Data Buggy? Sources and Magnitudes of Error in an Arthropod Citizen Science Project

RESEARCH ARTICLES

ALLEN H. HURLBERT
ELLEN WEAVER
*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Citizen science projects are a key source of biodiversity monitoring data that can help answer research questions that span broad spatial and temporal scales. For any given project, however, it is necessary to assess how reliably the data can be used for scientific inference. Here, we evaluate the sources and magnitudes of error associated with data from *Caterpillars Count!*, a citizen science project in which participants conduct foliage surveys using a protocol that requires them to detect and identify arthropods, and estimate their length. An analysis of data collected from an online virtual survey game complements previous work suggesting that participants can reliably detect arthropods on survey branches. Using both the online game as well as expert-verified photos, we find that participants can accurately identify arthropods to order or suborder ~90% of the time. Identifications of spiders, caterpillars, and ants are especially reliable, while less familiar groups such as sawfly larvae, daddy longlegs (harvestmen), and true bugs are less frequently recognized. "Aphid" was the most common classification submitted erroneously and mistakenly assigned to a wide range of smaller arthropods. In general, identification accuracy increased with arthropod size and with participant experience. Finally, we find that arthropod length estimates are highly accurate when participants use a ruler, and nearly as accurate if they simply know the width of their thumbnail for reference. These findings reveal the types of training materials and messaging that will further enhance data quality, and suggest that these data are broadly useful for examining spatial and temporal patterns of arthropod occurrence.

CORRESPONDING AUTHOR:
Allen H. Hurlbert
University of North Carolina, US
hurlbert@bio.unc.edu

KEYWORDS:
insects; misidentification; accuracy; participant self-efficacy; data quality; measurement error

TO CITE THIS ARTICLE:
Hurlbert, A.H. and Weaver, E. 2026. Are the Data Buggy? Sources and Magnitudes of Error in an Arthropod Citizen Science Project. *Citizen Science: Theory and Practice,* 11(1): 33, pp. 1–14. DOI: https://doi.org/10.5334/cstp.1012

A correction article relating to this paper has been published and can be found at https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.1140.

Many citizen science projects (also known as community or participatory science projects) facilitate the collection of geographically widespread biodiversity observations over time, resulting in datasets that are far broader in spatial and temporal extent than what could be generated through traditional means (e.g., Sullivan et al. 2009; Schwartz et al. 2012; Di Cecco et al. 2021). Such datasets are increasingly valuable as scientists attempt to understand the underlying drivers of rapidly changing ecosystems from continental to global scales. Despite the obvious advantages of tapping into a large pool of interested participants from around the world, citizen science presents potential challenges (Dickinson et al. 2010; Fraisl et al. 2022; Johnston et al. 2023). While some of these challenges relate to volunteer recruitment and retention, and overall project sustainability, the greatest challenge in terms of generating new scientific knowledge is ensuring data reliability and accuracy.

Thus far, the studies that have assessed citizen science data quality have reported mixed results, which is to be expected given the diverse range of citizen science projects and data types evaluated, as well as the variable methods used in assessment. While data quality may be influenced by aspects of project design, degree of training provided, and level of professional oversight (Brown and Williams 2019), the nature of the actual participant tasks are important as well. For projects that monitor biodiversity, those tasks typically involve some combination of (1) documenting the presence of biological taxa, (2) providing taxonomic identifications or classifications of those taxa, (3) enumerating those taxa, and (4) recording data on organismal traits like size, stage, or color. Each of these tasks include their own different potential biases and sources of error.

Error or bias associated with documenting the presence of biological taxa stems from two sources: the nature of the observation process and issues of detectability. With respect to the observation process itself, some projects involve participants providing incidental observations (e.g., a photo uploaded to iNaturalist), whereas others involve structured or semi-structured observations (Kelling et al. 2019). Incidental observations are prone to bias related to the spatial and temporal activity patterns of citizen scientist participants, and towards conspicuous or charismatic taxa (Di Cecco et al. 2021; Díaz-Calafat et al. 2024). However, even with structured or semi-structured observation protocols where participants report checklists of taxa and associated data on survey effort, participants may differ in their ability to detect inconspicuous or difficult-to-identify taxa (Kelling et al. 2015). Additionally, taxa themselves may vary in detectability (e.g., a conspicuous singing, bright, red cardinal as compared with a quiet, drab, and secretive sparrow), such that the number of organisms reported may more closely reflect the true number present for some taxa over others. This is a potential issue with any ecological study, but heterogeneity in species detectability may be even greater for non-experts (Isaac et al. 2014).

Taxonomic identification is another potential source of error in citizen science data collection, and the risk for error is exacerbated when the project does not involve the collection of any supplementary evidence (e.g., photo, video, audio) to support the identification or the validation of that evidence by outside experts. The accuracy of species level identifications by citizen science participants has been found to vary from as low as 30% (lichens; Munzi et al. 2023) to more than 90% (mammals; Kosmala et al. 2016 and Swanson et al. 2016), with substantial interspecific variation in identification accuracy in some cases (e.g., Crall et al. 2011). Identification accuracy tends to be higher when asking for classification at coarser taxonomic levels. Volunteers accurately identified trees to species ~74% of the time on average, while they accurately identified trees to genus 91–96% of the time (Bloniarz 1996). Ants were identified accurately to species only 70% of the time, but to genus 91% and to family 99% of the time (Arcos 2025). Aquatic macroinvertebrates were identified to family 91% of the time but to order 99% of the time (von Gönner et al. 2023). It is always incumbent on any project to assess the accuracy of identifications being asked of participants, but these results suggest that projects with research questions that can be addressed using coarser taxonomic information may be fairly robust to data quality concerns related to misidentification.

Some biodiversity monitoring projects require participants to estimate abundance, density, or percent cover of the taxa they are reporting. Where the accuracy of such estimates has been evaluated, there again have been mixed results. Volunteer estimates of seal and sea lion counts from drone imagery (typically up to 150 individuals) were within 5–9% of verified values, with a tendency toward underestimation (Wood et al. 2021). Citizen science observers tended to underestimate abundance patterns of flowers and both under- and overestimate abundance of insects in a phenology monitoring project (Feldman et al. 2018). In contrast, citizen science estimates of bee abundance (Kremen et al. 2011) and the density of various foliage arthropod groups (Hurlbert et al. 2019) were strongly correlated with expert or specimen-derived estimates. When the number of organisms to be counted is on the order of a few dozen or less, as in short duration bird counts, abundance estimates tend to be unbiased and with low error (Bonter and Cooper 2012; Farr et al. 2023). Although some methodological and statistical approaches exist to account for some sources of error associated with estimating abundance and occupancy (Kéry et al. 2010;

Bird et al. 2014; Altwegg and Nichols 2019), the suitability of those approaches depends upon the details of the data structure and observation process.

Finally, some biodiversity monitoring projects ask participants to provide ancillary trait data for the organisms they observe, including life stage (Prysby and Oberhauser 2004), phenophase (Feldman et al. 2018; Salomé-Díaz et al. 2023), and, most commonly, size or length estimates (Hurlbert et al. 2019; May et al. 2019; Ostwald et al. 2025). Direct measurement of specimens (or images of specimens) by citizen scientists can be quite reliable, with the average error in bee length estimates found to be only 2.3% (Ostwald et al. 2025) and the average error in biomass estimates derived from tree circumference measurements to be only 1.1% (Harrison et al. 2020). Estimating the length of organisms based purely on visual inspection rather than direct measurement can be more difficult. Average error rate for volunteers visually estimating shark length during cage-diving was approximately 15% (average accuracy of 49 cm for sharks with median length of 325 cm; May et al. 2019), which was more than double the average error rate of experts. No studies of which we are aware have evaluated the accuracy of arthropod length estimates based purely on visual inspection.

While the published studies discussed above provide a rough sense of the reliability of these four distinct tasks associated with citizen science–based biodiversity monitoring, error rates will clearly vary based on the specific details of project tasks and methodologies (Kosmala et al. 2016; Aceves-Bueno et al. 2017; Brown and Williams 2019). It is therefore incumbent on researchers to understand error and bias specific to each individual citizen science project, which can contribute to a fuller understanding of the reliability of such tasks more generally. Here, we evaluate the sources and magnitudes of error associated with *Caterpillars Count!*, a project in which citizen science participants conduct foliage surveys that require all four of the tasks described above: detecting, identifying, and enumerating arthropods, and estimating their length (Hurlbert et al. 2019). Previous work showed that estimates of arthropod density by citizen scientists were very similar to estimates by trained scientists in this dataset (Hurlbert et al. 2019). The data from this project have been used to explore large-scale patterns of arthropod abundance and phenology (Di Cecco and Hurlbert 2022; Di Cecco et al. 2023), and the effects of urbanization (Roper-Edwards and Hurlbert 2024) and non-native plant species (Whitener and Hurlbert 2026) on foliage arthropod communities. An evaluation of the other sources of error in *Caterpillars Count!* data will reveal which types of analyses and conclusions are likely to be most reliable, and will help provide effective recommendations for more accurate data collection in the future. Here, we assess the ability of participants to detect arthropods on visually complex backgrounds, evaluate the arthropod orders and suborders that can and cannot be reliably identified, and quantify the accuracy of arthropod length estimates with and without the use of measuring aids.

METHODS

CATERPILLARS COUNT! PROTOCOL

A *Caterpillars Count!* monitoring site consists of at least 10 tagged branches of woody plant species that have been chosen in advance by the site creator. On these tagged branches, participants conduct foliage surveys, recording the number of individual arthropods found, their identity to taxonomic order or suborder, and their length in millimeters. Participants follow one of two survey protocols as determined by the site creator: (1) visual surveys involve the visual inspection of 50 leaves and their associated petioles, twigs, and branches, while (2) beat sheet surveys involve striking the branch with a stick to dislodge arthropods onto a white fabric beat sheet held below. The act of detecting arthropods on a white sheet is expected to be easier than detecting them on a branch, especially for arthropods that are smaller or cryptically patterned. Nevertheless, previous work showed similar estimates of density for most arthropod groups for each of the two methods (Hurlbert et al. 2019).

During survey submission, users are instructed to classify any arthropods found as: "ants," "aphids and psyllids," "bees, wasps, and sawflies," "beetles," "caterpillars," "daddy longlegs," "flies," "grasshoppers, crickets," "leafhoppers and cicadas," "moths, butterflies," "spiders," "true bugs," "other," and "unidentified" (Table 1). If selecting "bees, wasps, and sawflies," users can check an additional box to indicate "sawfly larvae," which phenotypically resemble caterpillars. For each arthropod found on a survey, participants have the option to upload a supporting photo, which is automatically submitted to iNaturalist for experts to verify the arthropod's identification (Campbell et al. 2023).

VIRTUAL SURVEY GAME

The *Caterpillars Count!* project website offers a Virtual Survey Game as a training tool for participants to practice arthropod identification and survey protocols (https://caterpillarscount.unc.edu/virtualSurvey). The online game asks users to (1) find arthropods on the background image of a tree branch, and for each arthropod found they must (2) identify it to order or suborder as defined above, and (3) estimate its length in millimeters given a ruler shown

COMMON NAME	TAXONOMIC NAME	TOTAL RECORDS	VERIFIED PHOTOS
ants	Formicidae	21,369	1,489
aphids, psyllids, and allies	Hemiptera: Sternorrhyncha	12,439	450
bees, wasps, and sawflies	Hymenoptera (adults)	6,474	533
beetles	Coleoptera	30,921	3,602
caterpillars	Lepidoptera (larvae)	21,755	2,717
daddy longlegs (harvestmen)	Opiliones	3,945	124
flies	Diptera	26,656	1,272
grasshoppers, crickets	Orthoptera	5,880	683
leafhoppers and cicadas	Hemiptera: Auchenorrhyncha	16,158	2,338
moths, butterflies	Lepidoptera (adults)	2,914	325
sawfly larvae	Hymenoptera: Symphyta (larvae)	2,391	274
spiders	Araneae	33,943	3,317
true bugs	Hemiptera: Heteroptera	9,138	1,270

Table 1 Common and taxonomic names of arthropod groups recorded in the *Caterpillars Count!* dataset, their prevalence by total number of records, and the number of photos for which identifications have been confirmed on iNaturalist.

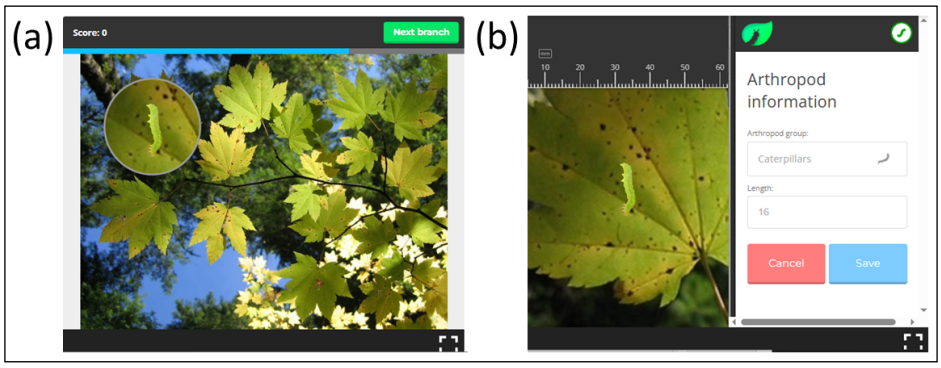


Figure 1 Screenshots of the Caterpillars Count! Virtual Survey Game. **(a)** An image of a branch on which 2–8 arthropods are randomly sized and placed. The circular inset shows a magnified image wherever the user places their cursor. **(b)** After clicking on an arthropod on the branch, the user is prompted to identify it using a dropdown list, and then estimate its body length in millimeters. After clicking Save, the user is brought back to the full branch image to continue searching until they feel they have found all of the arthropods or time runs out (blue bar in panel a).

at the top of the screen (Figure 1). A single gameplay presents sets of two to eight arthropods on each of three successive branch images of increasing visual complexity. To incentivize task completion, search time is limited to one minute per branch, but the timer stops when a user clicks on an arthropod to classify and measure it. The number, identity, size, axis of rotation, and location of arthropods on the branch are randomized with each gameplay from a library of 81 arthropod images. After entering an identification and length estimate for each arthropod, the user is provided immediate feedback on the true arthropod group and actual length. When a user plays while logged in to their account, three subscores are saved to the database: the percent of arthropods found, percent of correct identifications, and a length accuracy score.

For the 83 users who played the online game at least two times while logged in, we compared between subscore categories using paired Wilcoxon signed rank tests to establish the relative accuracy of the three tasks. Within each subscore category, we also used paired Wilcoxon tests to compare scores from the first gameplay to the user's best score from their multiple plays in order to document users' ability to improve at a given task. We considered "first" and "best" scores of the virtual game to reflect 1) how well a naïve user performs at the three tasks, and 2) how good a user could possibly be at those tasks after repeated game play. This was an opportunistic analysis of

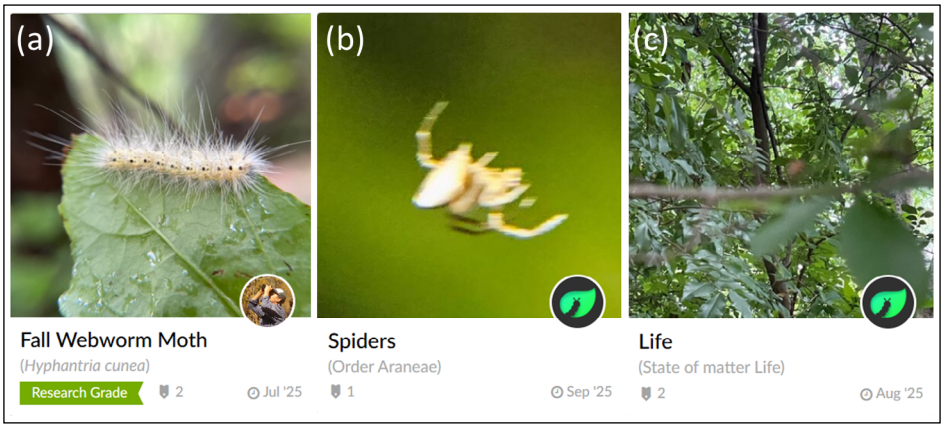


Figure 2 Examples of photo observations submitted with *Caterpillars Count!* surveys that get shared with iNaturalist. **(a)** A high-quality photo allowing consensus identification to species on iNaturalist. Photo credit: Allen Hurlbert. **(b)** A photo with sufficient details allowing confirmation of arthropod order. Photo credit: Kathleen Shea. **(c)** A poor-quality photo representing a *Caterpillars Count!* identification that cannot be confirmed. Photo credit: Alyssa Smith.

game scores rather than a study, so the number of game plays varies substantially by user. Of these users, two thirds played just 2–4 times, while one user played 39 times.

IDENTIFICATION ACCURACY ON SURVEYS

Of 205,934 total arthropod records submitted through 12 December 2025, approximately 20% of submissions are accompanied by a photo taken by the user, and 48% of photo observations have a consensus or majority identification on iNaturalist (i.e., at least 2 verified identifications and agreement by at least 2/3 of identifiers; Figure 2a). While some photos that have not received a verified identification will eventually obtain identifications in the future, some are poor in quality for which identifications will never be possible (Figure 2c). For the 13 focal arthropod groups listed above, we calculated two identification error rates. The false positive rate is the percentage of reports of a given arthropod group that turned out to be another group based on iNaturalist identifications. The false negative rate is the percentage of verified arthropods of a given group that were incorrectly reported by *Caterpillars Count!* users as some other group.

In addition, we calculated false negative rates as a function of the reported length of the arthropod. We fit generalized linear models predicting the binary correct or incorrect classification as a function of arthropod length using a binomial error distribution and link function for each arthropod group.

Finally, we assessed the degree to which identification errors decrease over time as a user submits more photo observations. For this analysis we wanted to focus specifically on the arthropod groups that are commonly misidentified, thus we excluded the more easily distinguishable groups of spiders, ants, and caterpillars (see Results). For each user with at least 20 photo observations that had been identified on iNaturalist (n = 79 users), we arranged the observations in order of submission and ran a generalized linear mixed-effects model predicting classification agreement as a function of submission order with user ID as a random effect. We confirmed that the model ran with no convergence issues, and that there was no overdispersion or outliers.

LENGTH ACCURACY

Forty-four first year undergraduate students at the University of North Carolina at Chapel Hill were asked to estimate the lengths of 6 pinned arthropod specimens ranging in length from 6 to 31 mm. Students were randomly assigned to one of three treatment groups: Group 1 was asked to estimate lengths without the use of any aids, Group 2 was first allowed to measure the width of their thumbnail with a ruler, and then was asked to estimate lengths using only their thumb as a guide, and Group 3 was allowed to use a ruler to help estimate lengths. For each participant and arthropod specimen, we calculated the absolute deviation in millimeters of the estimate from the true length, as well as the percent deviation (absolute deviation divided by the true length). Absolute and percent deviation were compared between treatment groups using a Wilcoxon test.

RESULTS

VIRTUAL SURVEY GAME

When playing the virtual survey game for the first time, the median percentage of arthropods found on the branch image was 66.7%, the median identification accuracy of those found was 80%, and the median length accuracy was 74.5% (Figure 3). However, after subsequent attempts

(mean = 5 total gameplays, median = 3), all three subscores increased ($p < 10^{-10}$) to 92.9% of arthropods found, 100% identification accuracy, and 83.9% length accuracy, indicating significant improvement in all three skills with additional practice. Identification accuracy was found to be statistically greater than both the rate of finding arthropods ($p = 0.0004$) and length accuracy ($p < 10^{-10}$). The rate of finding arthropods in the background image was also substantially greater than the accuracy of arthropod length estimation ($p = 0.0004$; Figure 3).

IDENTIFICATION ACCURACY OF FIELD SURVEYS

Across the 18,121 photos submitted that were reported to be among the 13 focal arthropod groups, users were correct in their classifications 90.2% of the time. However, error rates varied substantially across arthropod groups (Figure 4). "Aphid" submissions were found to have the highest rate of false positives (41.5%) of all arthropod groups, with many of these submissions turning out to be small beetles, leafhoppers, true bugs, or other orders (Figure 4a). Incorrect "bees and wasps" submissions (22.6%) were most often actually flies, and incorrect "true bugs" submissions (21.6%) were most often actually beetles (Figure 4a). "Caterpillars" had a 7.6% false positive rate, but these were almost entirely sawfly larvae, which are visually and functionally similar. "Ants," "spiders," and "sawfly larvae" submissions were the most reliable with less than 5% of submissions of these arthropod groups labeled incorrectly.

The false negative rate represents the proportion of arthropods that truly belonged to a particular group based on verified photos that were incorrectly reported by *Caterpillars Count!* users as being in a different group. Sawfly larvae were the group most commonly reported as something else, with 60% of verified sawfly larvae observations submitted as caterpillars (Figure 4b). Daddy longlegs (also known as harvestmen, order Opiliones) were reported as spiders nearly 40% of the time. True bugs had a 29.4% false negative rate with "beetles" as the most common incorrect classification. Leafhoppers, flies, beetles, and grasshoppers/crickets all had low false negative rates of 8% or less, while ants, spiders, and caterpillars were the least often misidentified arthropods in the field (≤ 1%).

For many arthropod groups, the probability that users incorrectly identified the group decreased for specimens with larger length estimates (Figure 5). This was particularly true for true bugs ($p < 10^{-24}$), leafhoppers ($p < 10^{-8}$), and bees and wasps ($p < 10^{-4}$), but also for moths and beetles as well ($p < 0.01$ for both groups). In contrast, there was no clear effect of length on error rate for grasshoppers and crickets, flies, daddy longlegs, caterpillars, ants, and spiders. For those last three groups, error rates were extremely low regardless of length, while for daddy longlegs, error rates were high regardless of length. Curiously, aphid misclassification rates increased with length ($p < 0.01$), although 88% of aphid observations were reported to be only 3 mm or less.

For the groups where error rates declined with arthropod length, the length at which error rates fell below a 10% misclassification threshold varied by group (Figure 5, *dashed red line*). True bug error rates did not fall below this threshold until specimens were at least ~15 mm in length, whereas for leafhoppers and beetles, specimens as small as ~5 mm were reliably classified at this level. Moths and

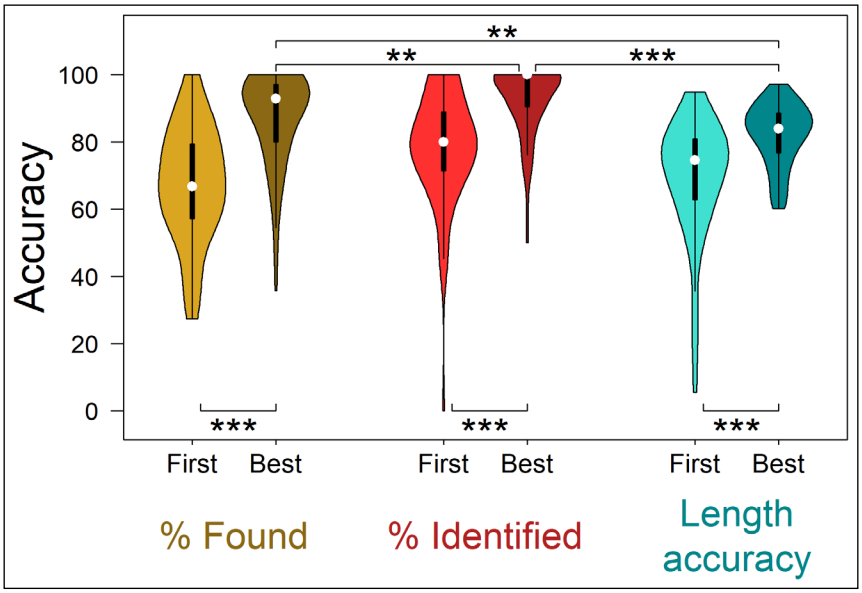


Figure 3 Distributions of the first and best scores of the three Virtual Survey Game subscores: finding arthropods on the branch image, correctly identifying them to order or suborder, and accurately estimating their length in millimeters. ** $p < 0.001$; *** $p < 10^{-10}$.

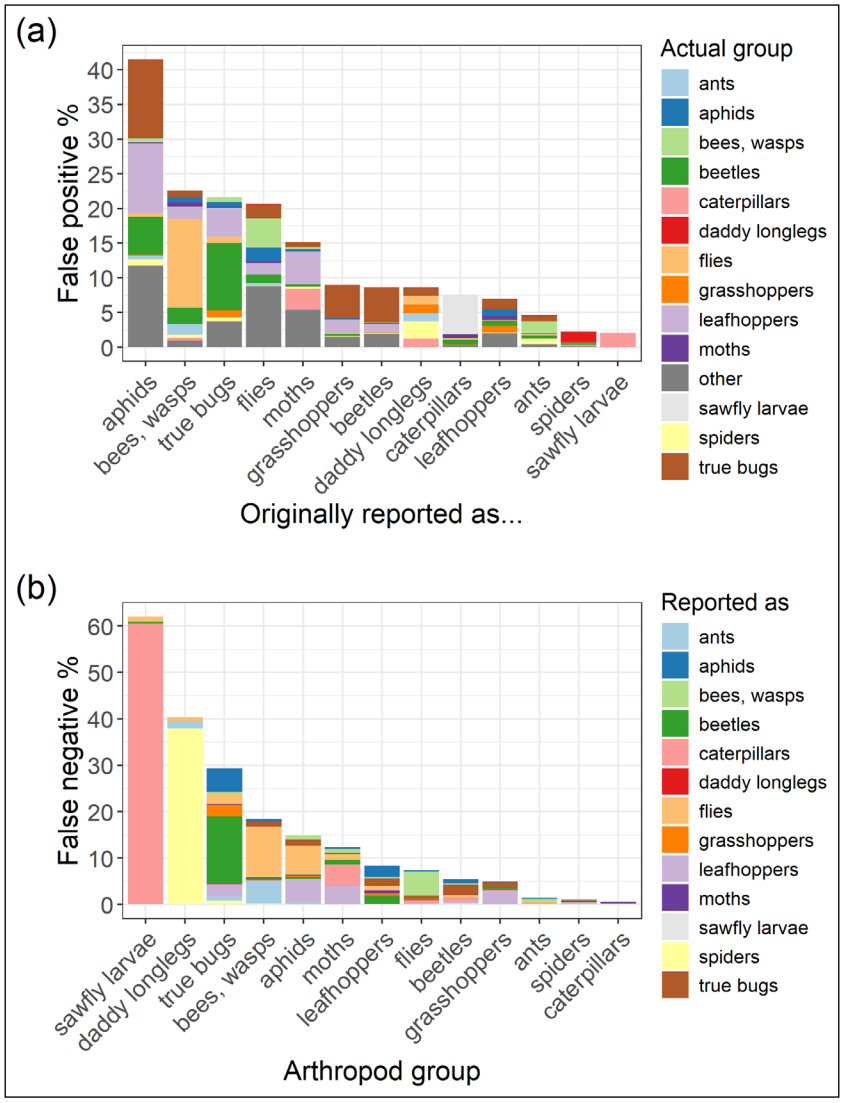


Figure 4 (a) The proportion of reports of a given arthropod group that were incorrectly classified as other groups (false positives). **(b)** The proportion of verified arthropods of a given group that were incorrectly reported as some other group (false negatives).

bees and wasps were intermediate (12–13 mm) in the size at which error rates fell below 10%.

Finally, there were 78 users who had submitted at least 20 photos of the ten most difficult to distinguish arthropod groups (i.e., excluding the easy to identify caterpillars, spiders, and ants). Users were less likely to make classification errors with continued participation, as indexed by the cumulative number of photos submitted by a user at the time of submission ($p = 0.01$; Figure 6). The average false negative rate for these difficult to identify groups fell from 13.5% to 12% on average after submitting 100 photos, and below 10% on average after submitting 300 photos.

LENGTH ESTIMATION ACCURACY

When estimating the length of pinned arthropod specimens without any aids, students tended to underestimate length (median deviation -3 mm, median relative deviation -33%; Figure 7). When students were allowed to measure their thumbnail width prior to estimating arthropod lengths, deviations were much less extreme (median deviation -1 mm, median relative deviation -9%; $p < 0.001$ for both comparisons). Length estimation accuracy for students that were allowed to use a ruler was much higher than the control group (median deviation -0.5 mm, median relative deviation -1.6%; $p < 10^{-7}$ for both), but indistinguishable from the group that used their thumb as a reference ($p > 0.4$ for both comparisons).

DISCUSSION

We used three different sources of information to evaluate the ability of citizen scientist participants to reliably find,

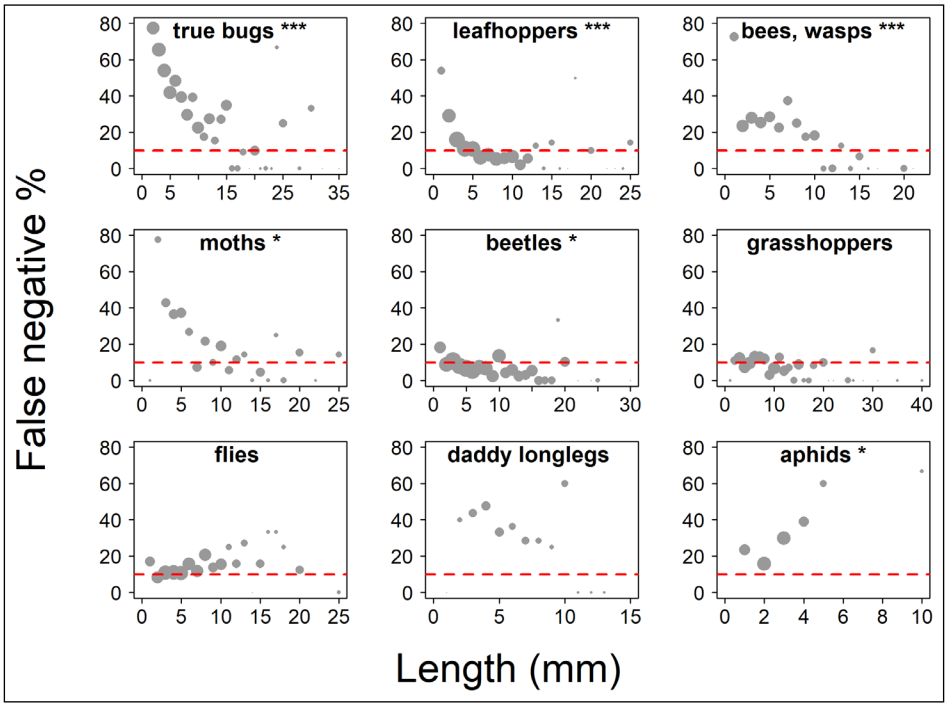


Figure 5 The proportion of arthropods of a given group as verified on iNaturalist that were incorrectly reported to *Caterpillars Count!* as some other group (false negative %) as a function of the reported length of the specimen. Symbol size reflects the number of specimens reported at that length. Dashed red line indicates the 10% misidentification rate for reference. *** $p < 0.0001$; * $p < 0.01$ based on generalized linear models.

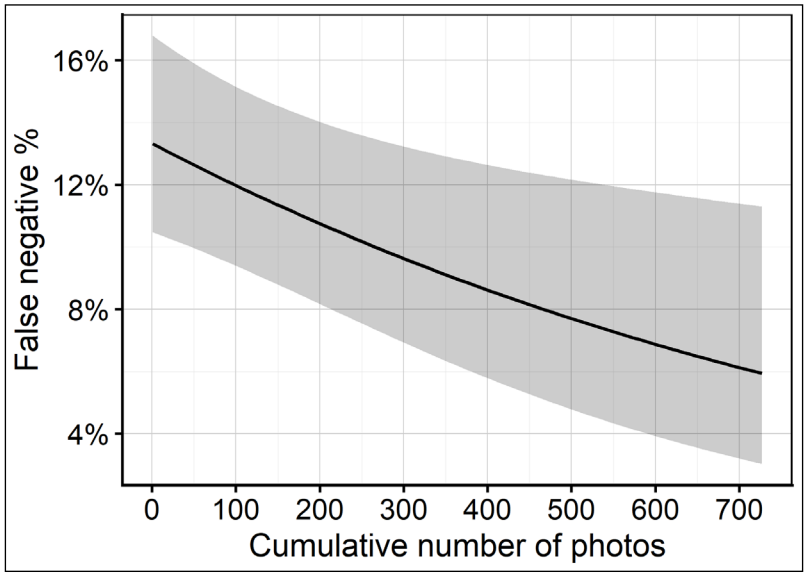


Figure 6 The probability of submitting an incorrectly identified photo observation (false negative %) decreases with the number of photos submitted by a user ($p = 0.01$). The shaded area indicates the 95% confidence interval. The error rate is based on the collective classification accuracy of true bugs, hoppers, bees and wasps, moths, beetles, grasshoppers and crickets, flies, daddy longlegs, and aphids, but excludes caterpillars, ants, and spiders, for which accuracy is reliably high.

identify, and estimate the length of different groups of foliage arthropods. One of those sources, the *Caterpillars Count!* virtual survey game, appears effective in both training and assessing volunteer participants in data collection tasks. The other two sources—independent identification of participant-submitted photos and an experiment in the accuracy of arthropod length estimation—enabled the assessment of participant skills in a more real-world context. We found that participants generally were able to perform these tasks well, with the biggest challenge being the identification of certain unfamiliar arthropod taxa. In the section that follows, we discuss each of the tasks in

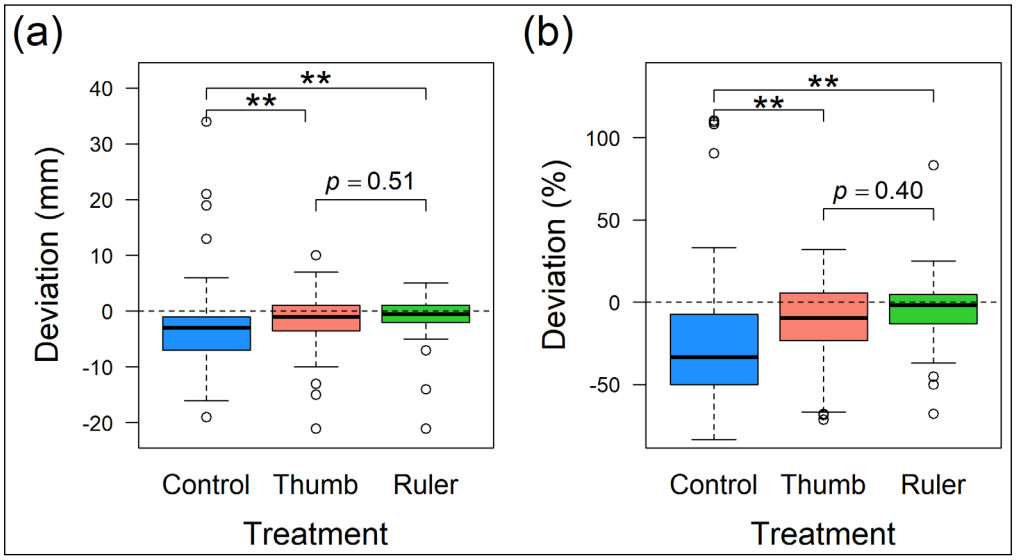


Figure 7 Comparison of the absolute **(a)** and relative **(b)** deviations in arthropod length estimates from the true values for three treatment groups: students who estimated length without the use of aids (control, blue, n = 18), students who were allowed to measure their thumbnail and then use that as a reference when estimating arthropod length (thumb, salmon, n = 12), and students who were allowed to use a ruler directly when estimating lengths (ruler, green, n = 14). ** $p < 0.0003$.

turn, consider opportunities for improved training, and potential caveats when analyzing *Caterpillars Count!* data.

The most difficult task to directly assess is the ability of participants to locate arthropods on a branch, as it requires comparing what participants report to the true number of arthropods on survey branches which is typically unknowable. However, in the online game, the number of arthropods present is known, and users that played two or more times were able to find a median 92.9% of arthropods on the visually complex branch image background. This is a substantial increase from finding only 66.7% when playing the game for the first time, which likely reflects a combination of improved searching ability with practice as well as becoming more familiar with the artificial game environment. The ability of participants to detect arthropods on actual survey branches has also been assessed indirectly by comparing density estimates (number of arthropods per survey branch by arthropod group) between citizen scientists and trained scientists at the same monitoring site. Hurlbert et al. (2019) not only documented similar density estimates between citizen scientists and trained scientists, but also between visual surveys (where failure to detect cryptic arthropods might be more pronounced) and beat sheet surveys (where arthropods should be easy to detect on a white background). In addition to the online game, activities that could help train participants in thorough visual searches include having them survey branches on which a trainer has added known "bugs" using a marker or small stickers, especially in less conspicuous locations such as on the underside of leaves or twigs, and where multiple leaves overlap. Ultimately, the most straightforward way for participants to improve their arthropod detection ability will be conducting more surveys in the field. Finding ways to document exactly how much and how rapidly detection ability increases over the course of participation would also be valuable.

The ability of participants to correctly identify arthropods to order or suborder is one of the most critical skills of the projects, and we were able to assess it through both the online game as well as through independent verification of participant-submitted photos. In the online game, arthropod identification was the task at which participants performed best, with a median accuracy of 100% after playing two or more times. Although the actual arthropod images used were randomly selected with each gameplay, they do come from a finite pool of 81 images (spanning the focal groups mentioned), and therefore it is possible that users learn the identity of specific images that show up repeatedly with multiple plays. Increasing the size of this pool of potential images would reduce the likelihood of this source of bias. Still, we found that 90% of actual survey photos were correctly identified by *Caterpillars Count!* participants based on classifications assigned independently on iNaturalist. Unsurprisingly, commonly known and morphologically distinct groups such as ants, spiders, and caterpillars were extremely reliably identified (failure to correctly identify them <1% of the time).

Our analysis revealed three less well-known groups that users frequently failed to identify correctly: sawfly larvae, daddy longlegs, and true bugs. Sawfly larvae (suborder Symphyta) are long, soft-bodied herbivores that look very similar to lepidopteran caterpillars, and indeed 60% of sawfly

larvae were reported as caterpillars. Although taxonomically distinct, sawfly larvae serve similar ecological roles to caterpillars as both herbivores and food for avian and other predators such that many studies effectively lump them together (e.g., Clem and Held 2015; Narango et al. 2020; Pan et al. 2021). The order Opiliones, frequently referred to as harvestmen or daddy longlegs, were unfamiliar to a significant fraction of participants who reported them as spiders ~40% of the time. While also in class Arachnida, daddy longlegs tend to be generalist omnivores or scavengers, and to the extent that they are predators, their prey are typically quite small (Pinto-da-Rocha et al. 2007). Being functionally distinct from the spiders for which they are confused is a greater potential problem for ecological analyses (i.e., compared to mistaking sawfly larvae for caterpillars); however, they are relatively rare in the *Caterpillars Count!* dataset (0.6% of photos compared with 16.9% of photos for spiders). True bugs (order Hemiptera, suborder Heteroptera) are the third most common group that participants failed to recognize consistently (~30% false negatives). As a morphologically diverse group, it is not surprising that they were mistaken for a diverse set of taxa, especially beetles but also aphids, hoppers, orthopterans, and even flies. As with a number of groups, this problem was most pronounced for small arthropods where diagnostic features are potentially difficult to distinguish.

Aphids, psyllids, and allies (suborder Sternorrhyncha) were the group that was suggested in error most frequently (false positive rate, ~40%). Given that 98% of reported aphids were 5 mm or less in length, it appears that participants often assigned a label of "aphid" to small arthropods about which they were otherwise unsure. Clearly, *Caterpillars Count!* data should not be used to infer anything about patterns of aphid occurrence in the absence of photographic evidence. Adult bees and wasps (order Hymenoptera) and flies (order Diptera) were both suggested in error ~20% of the time, and were frequently mistaken for each other. The small size of these visibly winged individuals, and the fact that the majority of adult Hymenoptera were not the striped yellow-and-black members of Apidae and Vespidae with which the general public might be most familiar, contribute to these identification errors. What participants suggested as flies (often around 3 mm) frequently turned out to be bark lice (order Psocodea) that are not mentioned in *Caterpillars Count!* training materials. True bugs were also suggested in error ~20% of the time, often for what were actually beetles or leafhoppers.

We found that participants' ability to identify arthropods in the field does increase modestly over time. Some improvement is expected purely from increased experience with the arthropods and the survey protocols (Kosmala et al. 2016; Meschini et al. 2021; Bianco et al. 2023), but feedback to the participants is also important to improve their identification skills (Peter et al. 2021). On a weekly basis, any photos submitted by *Caterpillars Count!* participants that received a consensus identification on iNaturalist at the order/suborder level result in an email to the observer letting them know whether their original classification of the photo was confirmed or mistaken, and summarizing the participant's overall identification accuracy to date. Beginning in 2025, a user dashboard on the project website allows participants to see how their identification accuracy varies by arthropod group, and therefore which groups they are most frequently misidentifying. As the proportion of users including photos with their survey submissions continues to increase (in 2025, 40% of surveys had at least one photo), not only will a greater fraction of *Caterpillars Count!* observations be verified, but users will continue to receive the feedback that should increase identification accuracy moving forward. Future work will establish for which arthropod groups participants are able to improve identification accuracy the most.

The third task asked of participants in the project is the estimation of arthropod length. After playing the online game multiple times, this was the task on which participants performed least well (median 84% length accuracy) and the task that displayed the least amount of improvement from users' first game play. While a ruler is displayed at the top of the game screen, users are not able to use it to measure arthropods directly and are limited to conducting a visual comparison between the arthropod and the ruler. However, when allowed to use a ruler directly, university students demonstrated the ability to estimate the length of specimens with little bias (median deviation -0.5 mm) and 79% of measurements within 2 mm of the true value. Asked to estimate length without any aids, subjects had a tendency toward underestimation (median deviation -3 mm), but this bias was reduced when they had recently measured a reference length like the width of a thumbnail. This information suggests that the simple messaging to *Caterpillars Count!* users about knowing their own thumbnail width could increase the accuracy of the project's length data, although length estimation accuracy in the field may be lower for mobile arthropods compared with the pinned specimens used in this evaluation.

This evaluation of the accuracy of citizen science data is important for understanding whether and how those data can be reliably used for scientific analysis. The stated goal of the *Caterpillars Count!* project is to understand how the abundance and phenology of foliage arthropods vary over broad geographic gradients and are potentially impacted by global change. Previous work highlighted the ability of citizen scientists to provide estimates of density

and phenology that were similar to the estimates based on trained scientists (Hurlbert et al. 2019). Here, we have additionally shown that citizen scientists perform well at identifying arthropods to order or suborder overall. Analyses focused on familiar and visually distinctive groups such as ants, caterpillars, and spiders are on solid ground, while the data are less reliable for aphids, flies, daddy longlegs, hymenopterans, and true bugs. That said, distinguishing features for some of these groups become more apparent with larger size, and for some groups like true bugs (whose adults typically have a noticeable triangle or X shape on the dorsum), using a minimum length threshold (e.g., at least 10 or 15 mm) may increase the reliability of abundance and phenology estimates.

Arthropod lengths are collected in order to estimate arthropod biomass using allometric equations specific to each taxonomic group (Sample et al. 1993; Sabo et al. 2002; Gruner 2003). Biomass estimates derived in this way may serve as useful first approximations, however, more work is needed to validate such estimates. Within-group variability in allometric relationships (Johnston and Cunjak 1999; Versluijs et al. 2023) together with the propagation of uncertainty in length estimates implies that analyses of arthropod biomass from *Caterpillars Count!* data will presumably result in noisier estimates of biomass than of density or occurrence. Size distributions could be more directly evaluated in terms of length rather than biomass as has been done in other large-scale arthropod analyses (e.g., Stork and Blackburn 1993; Novotný and Kindlmann 1996). Nevertheless, encouraging participants to increase their ability to estimate length accurately through training materials and the online survey game should improve the quality of this aspect of the *Caterpillars Count!* dataset.

CONCLUSIONS

This study adds to the growing evidence that citizen science participants are able to collect reliable biodiversity monitoring data using standardized protocols (Kosmala et al. 2016), including being able to accurately identify arthropods to order. We were able to evaluate this directly for *Caterpillars Count!* because of the ability to independently verify arthropod identifications for the subset of observations with photos, and because of an online survey game that helps participants practice the three necessary skills associated with conducting foliage surveys. Thus, in addition to well-designed survey procedures and proper training for participants (Brown and Williams 2019), we encourage all citizen science projects to incorporate strategies such as those described here for the assessment of data quality, bias, and uncertainty.

DATA ACCESSIBILITY STATEMENT

All data and code underlying this publication are available in the Github repository https://github.com/hurlbertlab/caterpillars-on-plants, and will be published to Zenodo upon acceptance.

ETHICS AND CONSENT

This study was determined to be Exempt by the University of North Carolina IRB #19-2679.

ACKNOWLEDGEMENTS

We thank the thousands of volunteers that have participated in the *Caterpillars Count!* project over the years and the dedicated natural resource managers that help oversee monitoring sites.

FUNDING INFORMATION

Funding was provided by the National Science Foundation grant EF-1702708.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Ellen Weaver: conceptualization; formal analysis; writing – original draft, review, and editing. Allen Hurlbert: conceptualization; data curation; formal analysis; writing – revised draft, review, and editing; supervision; funding.

PUBLISHER'S NOTE

A correction article was published on 17/11/2026 that updates an incorrect link in the Data Accessibility Statement and provides a new link to the archived repository on Zenodo. For more information, see DOI: 10.5334/cstp.1140.

AUTHOR AFFILIATIONS

Allen H. Hurlbert orcid.org/0000-0002-5678-9907
University of North Carolina, US

Ellen Weaver
University of North Carolina, Chapel Hill, US

REFERENCES

Aceves-Bueno, E., Adeleye, A.S., Feraud, M., Huang, Y., Tao, M., Yang, Y., and **Anderson, S.E.** (2017) The accuracy of citizen science data: A quantitative review. *The Bulletin of the Ecological Society of America*, 98(4), 278–290. DOI: https://doi.org/10.1002/bes2.1336

Altwegg, R. and **Nichols, J.D.** (2019) Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, 10(1), 8–21. DOI: https://doi.org/10.1111/2041-210X.13090

Arcos, J. (2025) From citizen to scientist: Evaluating ant observation accuracy on a Mediterranean archipelago (p. 2025.02.08.637263). bioRxiv. DOI: https://doi.org/10.1101/2025.02.08.637263

Bianco, V.D., Peterman, K., Becker-Klein, R., Davis, C.M., Grover, A., and **Phillips, T.** (2023) Cross-project analysis of volunteers' scientific observation skills. *Citizen Science: Theory and Practice*, 8(1). DOI: https://doi.org/10.5334/cstp.628

Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., et al. (2014) Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144–154. DOI: https://doi.org/10.1016/j.biocon.2013.07.037

Bloniarz, D.V. (1996) The use of volunteer initiatives in conducting urban forest resource inventories [Ph.D., University of Massachusetts Amherst]. https://www.proquest.com/docview/304324221/abstract/5A91CE6EC4864F32PQ/1

Bonter, D.N. and **Cooper, C.B.** (2012) Data validation in citizen science: A case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, 10(6), 305–307. DOI: https://doi.org/10.1890/110273

Brown, E.D., and **Williams, B.K.** (2019) The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*, 33(3), 561–569. DOI: https://doi.org/10.1111/cobi.13223

Campbell, C.J., Barve, V., Belitz, M.W., Doby, J.R., White, E., Seltzer, C., Di Cecco, G., et al. (2023) Identifying the identifiers: How iNaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science. *BioScience*, 73(7), 533–541. DOI: https://doi.org/10.1093/biosci/biad051

Clem, C.S., and **Held, D.W.** (2015) Species richness of eruciform larvae associated with native and alien plants in the southeastern United States. *Journal of Insect Conservation*, 19(5), 987–997. DOI: https://doi.org/10.1007/s10841-015-9815-0

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J., and **Waller, D.M.** (2011) Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, 4(6), 433–442. DOI: https://doi.org/10.1111/j.1755-263X.2011.00196.x

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P., and **Hurlbert, A.H.** (2021) Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, 71(11), 1179–1188. DOI: https://doi.org/10.1093/biosci/biab093

Di Cecco, G.J., Belitz, M.W., Cooper, R.J., Larsen, E.A., Lewis, W.B., Ries, L., Guralnick, R.P., et al. (2023) Phenology in adult and larval Lepidoptera from structured and unstructured surveys across eastern North America. *Frontiers of Biogeography*, 15(1). DOI: https://doi.org/10.21425/F5FBG56346

Di Cecco, G.J., and **Hurlbert, A.H.** (2022) Caterpillar patterns in pace and time: insights from and contrasts between two citizen science datasets. In R. J. Marquis and S. Koptur (Eds.), Caterpillars in the Middle: Tritrophic Interactions in a Changing World (pp. 541–556). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-86688-4_17

Díaz-Calafat, J., Jaume-Ramis, S., Soacha, K., Álvarez, A., and **Piera, J.** (2024) Revealing biases in insect observations: A comparative analysis between academic and citizen science data. *PLOS ONE*, 19(7), e0305757. DOI: https://doi.org/10.1371/journal.pone.0305757

Dickinson, J.L., Zuckerberg, B., and **Bonter, D.N.** (2010) Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 149–172. DOI: https://doi.org/10.1146/annurev-ecolsys-102209-144636

Farr, C.M., Ngo, F., and **Olsen, B.** (2023) Evaluating data quality and changes in species identification in a citizen sciencebBird monitoring project. *Citizen Science: Theory and Practice*, 8(1). DOI: https://doi.org/10.5334/cstp.604

Feldman, R.E., Žemaitė, I., and **Miller-Rushing, A.J.** (2018) How training citizen scientists affects the accuracy and precision of phenological data. *International Journal of Biometeorology*, 62(8), 1421–1435. DOI: https://doi.org/10.1007/s00484-018-1540-4

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

Gruner, D. (2003) Regressions of length and width to predict arthropod biomass in the Hawaiian Islands. *Pacific Science*, 57, 325–336. DOI: https://doi.org/10.1353/psc.2003.0021

Harrison, B., Martin, T.E., and **Mustari, A.H.** (2020) The accuracy of volunteer surveyors for obtaining tree measurements in tropical forests. *Ambio*, 49(1), 98–106. DOI: https://doi.org/10.1007/s13280-019-01147-4

Hurlbert, A., Hayes, T., McKinnon, T., and **Goforth, C.** (2019) Caterpillars Count! A citizen science project for monitoring foliage arthropod abundance and phenology. *Citizen Science: Theory and Practice*, 4(1), 1–12. DOI: https://doi.org/10.5334/cstp.148

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., and **Roy, D.B.** (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10), 1052–1060. DOI: https://doi.org/10.1111/2041-210X.12254.

Johnston, A., Matechou, E., and **Dennis, E.B.** (2023) Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, 14(1), 103–116. DOI: https://doi.org/10.1111/2041-210X.13834

Johnston, T.A., and **Cunjak, R.A.** (1999) Dry mass–length relationships for benthic insects: A review with new data from Catamaran Brook, New Brunswick, Canada. *Freshwater Biology*, 41(4), 653–674. DOI: https://doi.org/10.1046/j.1365-2427.1999.00400.x

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., et al. (2019) Using semistructured surveys to improve citizen science data for monitoring biodiversity. *BioScience*, 69(3), 170–179. DOI: https://doi.org/10.1093/biosci/biz010

Kelling, S., Johnston, A., Hochachka, W.M., Iliff, M., Fink, D., Gerbracht, J., Lagoze, C., et al. (2015) Can observation skills of citizen scientists be estimated using species accumulation curves? *PLOS ONE*, 10(10), e0139600. DOI: https://doi.org/10.1371/journal.pone.0139600

Kéry, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., and **Zbinden, N.** (2010) Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, 24(5), 1388–1397. DOI: https://doi.org/10.1111/j.1523-1739.2010.01479.x

Kosmala, M., Wiggins, A., Swanson, A., and **Simmons, B.** (2016) Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10), 551–560. DOI: https://doi.org/10.1002/fee.1436

Kremen, C., Ullman, K.S., and **Thorp, R.W.** (2011) Evaluating the quality of citizen-scientist data on pollinator communities. *Conservation Biology*, 25(3), 607–617. DOI: https://doi.org/10.1111/j.1523-1739.2011.01657.x

May, C., Meyer, L., Whitmarsh, S., and **Huveneers, C.** (2019) Eyes on the size: Accuracy of visual length estimates of white sharks, *Carcharodon carcharias*. *Royal Society Open Science*, 6(5), 190456. DOI: https://doi.org/10.1098/rsos.190456

Meschini, M., Machado Toffolo, M., Marchini, C., Caroselli, E., Prada, F., Mancuso, A., Franzellitti, S., et al. (2021) Reliability of data collected by volunteers: A nine-year citizen science study in the Red Sea. *Frontiers in Ecology and Evolution*, 9. DOI: https://doi.org/10.3389/fevo.2021.694258

Munzi, S., Isocrono, D., and **Ravera, S.** (2023) Can we trust iNaturalist in lichenology? Evaluating the effectiveness and reliability of artificial intelligence in lichen identification. *The Lichenologist*, 55(5), 193–201. DOI: https://doi.org/10.1017/S0024282923000403

Narango, D.L., Tallamy, D.W., and **Shropshire, K.J.** (2020) Few keystone plant genera support the majority of Lepidoptera species., 11(1), Article 1. DOI: https://doi.org/10.1038/s41467-020-19565-4

Novotný, V., and **Kindlmann, P.** (1996) Distribution of body sizes in arthropod taxa and communities. *Oikos*, 75(1), 75–82. DOI: https://doi.org/10.2307/3546323

Ostwald, M.M., Smith, C., Allen, J., Buetow, A., Manner, A.R., Guralnick, R., Goldsmith, C., et al. (2025) Leveraging community science to measure bee body size from museum specimens. *Ecology and Evolution*, 15(6), e71665. DOI: https://doi.org/10.1002/ece3.71665

Pan, X., Mizuno, T., Ito, K., Ohsugi, T., Nishimichi, S., Nomiya, R., Ohno, M., et al. (2021) Assessing temporal dynamics of predation and effectiveness of caterpillar visual defense using sawfly larval color and resting posture as a model. *Insect Science*, 28(6), 1800–1815. DOI: https://doi.org/10.1111/1744-7917.12884

Peter, M., Diekötter, T., Kremer, K., and **Höffler, T.** (2021) Citizen science project characteristics: Connection to participants' gains in knowledge and skills. *PLOS ONE*, 16(7), e0253692. DOI: https://doi.org/10.1371/journal.pone.0253692

Pinto-da-Rocha, R., Machado, G., and **Giribet, G.** (2007) Harvestmen: The Biology of Opiliones. Harvard University Press. DOI: https://doi.org/10.2307/j.ctv322v442

Prysby, M., and **Oberhauser, K.S.** (2004) Temporal and geographical variation in monarch densities: Citizen scientists document monarch population patterns. In K. S. Oberhauser and M. J. Solensky (Eds.), Monarch butterfly biology and conservation (pp. 9–20) Cornell University Press.

Roper-Edwards, I.R., and **Hurlbert, A.H.** (2024) Arthropod community composition in urban landscapes is shaped by both environmental filtering and dispersal limitation. *PLOS ONE*, 19(8), e0297507. DOI: https://doi.org/10.1371/journal.pone.0297507

Sabo, J.L., Bastow, J.L., and **Power, M.E.** (2002) Length–mass relationships for adult aquatic and terrestrial invertebrates in a California watershed. *Journal of the North American Benthological Society*, 21(2), 336–343. DOI: https://doi.org/10.2307/1468420

Salomé-Díaz, J., Golubov, J., Díaz-Segura, O., Ramírez-Gutiérrez, Ma. C., Sifuentes de la Torre, S., Koleff, P., Quintero, E., et al. (2023) Practice makes the expert: The importance of training volunteers in the generation of phenological data from photographs of biodiversity observation platforms. *PLOS ONE*, 18(3), e0282750. DOI: https://doi.org/10.1371/journal.pone.0282750

Sample, B.E., Cooper, R.J., Greer, R.D., and **Whitmore, R.C.** (1993) Estimation of insect biomass by length and width. *The American Midland Naturalist*, 129(2), 234–240. DOI: https://doi.org/10.2307/2426503

Schwartz, M.D., Betancourt, J.L., and **Weltzin, J.F.** (2012) From Caprio's lilacs to the USA National Phenology Network. *Frontiers in Ecology and the Environment*, 10(6), 324–327. DOI: https://doi.org/10.1890/110281

Stork, N.E., and **Blackburn, T.M.** (1993) Abundance, body size and biomass of arthropods in tropical forest. *Oikos*, 67(3), 483–489. DOI: https://doi.org/10.2307/3545360

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., and **Kelling, S.** (2009) eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282–2292. DOI: https://doi.org/10.1016/j.biocon.2009.05.006

Swanson, A., Kosmala, M., Lintott, C., and **Packer, C.** (2016) A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, 30(3), 520–531. DOI: https://doi.org/10.1111/cobi.12695

Versluijs, T.S.L., Zhemchuzhnikov, M.K., Kutcherov, D., Roslin, T., Schmidt, N.M., Gils, J.A. van, and **Reneerkens, J.** (2023) Site-specific length-biomass relationships of arctic arthropod families are critical for accurate ecological inferences. *PeerJ*, 11, e15943. DOI: https://doi.org/10.7717/peerj.15943

von Gönner, J., Bowler, D.E., Gröning, J., Klauer, A.-K., Liess, M., Neuer, L., and **Bonn, A.** (2023) Citizen science for assessing pesticide impacts in agricultural streams. *Science of The Total Environment*, 857, 159607. DOI: https://doi.org/10.1016/j.scitotenv.2022.159607

Whitener, C. and **Hurlbert, A.H.** (2026) Arthropods on native versus alien woody plants: Understanding variation across host plant taxonomy and geography. *Ecosphere*, 17(5), e70615. DOI: https://doi.org/10.1002/ecs2.70615

Wood, S.A., Robinson, P.W., Costa, D.P., and **Beltran, R.S.** (2021) Accuracy and precision of citizen scientist animal counts from drone imagery. *PLOS ONE*, 16(2), e0244040. DOI: https://doi.org/10.1371/journal.pone.0244040

TO CITE THIS ARTICLE:
Hurlbert, A.H. and Weaver, E. 2026. Are the Data Buggy? Sources and Magnitudes of Error in an Arthropod Citizen Science Project. *Citizen Science: Theory and Practice*, 11(1): 33, pp. 1–14. DOI: https://doi.org/10.5334/cstp.1012

Submitted: 24 February 2026 **Accepted:** 10 September 2026 **Published:** 19 October 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.