



OPEN

Can information from citizen science data be used to predict biodiversity in stormwater ponds?

Frank Johansson¹✉, Jani Heino², Paul Coiffard¹, Richard Svanbäck¹, Jacob Wester¹ & Luis Mauricio Bini³

Citizen science data (CSD) have the potential to be a powerful scientific approach to assess, monitor and predict biodiversity. Here, we ask whether CSD could be used to predict biodiversity of recently constructed man-made habitats. Biodiversity data on adult dragonfly abundance from all kinds of aquatic habitats collected by citizen scientists (volunteers) were retrieved from the Swedish Species Observation System and were compared with dragonfly abundance in man-made stormwater ponds. The abundance data of dragonflies in the stormwater ponds were collected with a scientific, standardized design. Our results showed that the citizen science datasets differed significantly from datasets collected scientifically in stormwater ponds. Hence, we could not predict biodiversity in stormwater ponds from the data collected by citizen scientists. Using CSD from past versus recent years or from small versus large areas surrounding the stormwater ponds did not change the outcome of our tests. However, we found that biodiversity patterns obtained with CSD were similar to those from stormwater ponds when we restricted our analyses to rare species. We also found a higher beta diversity for the CSD compared to the stormwater dataset. Our results suggest that if CSD are to be used for estimating or predicting biodiversity, we need to develop methods that take into account or correct for the under-reporting of common species in CSD.

In the face of increasing human pressure on biodiversity¹, scientists need additional approaches to respond to the demand for information to guide environmental management, conservation planning and policymaking². One such approach consists of the involvement of citizens (volunteers) to gather data for scientific purposes (see reviews in Dickinson *et al.*^{3,4}). Citizen science data (CSD) are gathered by non-professionals, including biodiversity data across space and time⁵. Because many non-professional naturalists are involved in biodiversity data collection, one of the biggest advantages is that this approach generates a large amount of data³. These datasets have been successfully used in many studies, for example, in monitoring population trends of species, assessing the impacts of global warming on biodiversity, and forecasting species responses to forest management^{6,7}. Hence, CSD are important in providing large amounts of data for scientists with little financial resources devoted to large-scale biodiversity surveys.

However, CSD may also have several drawbacks, mostly associated with the lack of a proper sampling design and accuracy of records⁸. Snäll *et al.*⁹ listed several of these drawbacks, such as: (1) population records only include presence, but not absence, (2) sampling effort varies over space and time, (3) spatial coverage might vary, (4) methods of collection might vary, (5) records of rare and common species might be biased towards rare or common species, and (6) detectability of species varies among volunteers. Because of these drawbacks, more studies are needed to examine how well CSD can be used to predict, for example, colonization of new habitats. Such new habitats could be created by environmental disturbance or deliberately by humans.

One type of new habitats created by humans are stormwater ponds. These ponds are water bodies constructed for capturing and storing water from rainstorms; they can be temporary or permanent. In general, stormwater ponds have a high biodiversity, especially when considering invertebrates and amphibians^{10–14}. Because many stormwater ponds were recently constructed, and many more are being constructed^{14,15}, it is interesting to ask whether knowledge of the distribution and abundance of organisms in the surrounding areas can be used to

¹Department of Ecology and Genetics, Animal Ecology, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden. ²Finnish Environment Institute, Freshwater Centre, Paavo Havaksen Tie 3, FI-90570, Oulu, Finland.

³Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, 74690-900, GO, Brazil. ✉e-mail: frank.johansson@ebc.uu.se

predict biodiversity in the ponds. Here, we investigate whether CSD from records of aquatic insects from the Swedish Species Observation System (www.artportalen.se) can be used to predict the biodiversity of insects in recently constructed stormwater ponds. We used adult dragonflies (Odonata: Zygoptera and Anisoptera) as our focal group of insect biodiversity. Dragonflies are intermediate consumers both in the aquatic and the terrestrial life stage, i.e., they are predators of smaller invertebrate prey and are preyed upon by larger invertebrate and vertebrate predators¹⁶. Because they have an intermediate position in the food web, they should represent overall patterns in aquatic biodiversity relatively well¹⁶. In addition, Odonata species richness is positively correlated with species richness of many, but not all, invertebrate taxa and vegetation abundance^{17,18}.

To examine whether biodiversity patterns in stormwater ponds could be predicted from CSD, we used three approaches. First, we examined the similarity between biodiversity datasets (i.e., between CSD and those obtained in urban stormwater ponds) from past to recent years. Because Odonata have a short generation time (usually 1–2 years) and because they are good dispersers¹⁶, we expected that community patterns generated with data from recent years would be better predictors of community structure in stormwater ponds. Second, we examined how the inclusion of CSD covering different areas around the stormwater ponds predicted the Odonata diversity in the ponds. We expected that citizen science datasets covering an area with a larger diameter around our study area would be a better predictor since a large area covers more habitats. We also considered an alternative hypothesis, where predictability decreased with area, due to the distance decay of similarity in ecological communities, see e.g. Nekola & White¹⁹. Third, we examined whether including common or rare species affected the predictions. We expected that community patterns based on data from rare species would be more similar to the patterns based on our dataset (stormwater ponds) because these species are more actively sought by citizen scientists than common species^{3,20,21}.

Methods

Stormwater pond data. Biodiversity data of dragonflies in 18 stormwater ponds were obtained in the city of Uppsala, Sweden in 2018. These man-made ponds were recently constructed (i.e. between 2004 and 2014). The city of Uppsala has 150 000 inhabitants and covers an area of 26 km², and all stormwater ponds that are filled with water all year round were used for this study (Fig. 1; for more details on methods and pond description, see Johansson *et al.*²²). Dragonfly abundance was recorded every second week over a 10-week period by two trained researchers (P.C and J.W.), who walked one lap slowly around the ponds from 29 May until 5 August 2018. This period covers the emergence period of all species found at this latitude²³. Most species were identified visually. However, for some species, identifications were made after capture by a butterfly net. The speed of walking was adjusted with respect to vegetation of the pond and by the abundance of dragonflies, such that the speed was slow at ponds with a lot of vegetation and a high abundance of dragonflies. Total numbers of adults (including mating pairs and ovipositing individuals) were counted and used for the subsequent analyses. No counts were done during cloudy, windy (>30 km/h) or rainy days and, therefore, the biweekly counts were shifted 1–2 days on two of the sampling occasions. For the analyses, the week with the highest number of individuals was used for each species. These ponds and a modified data set was used in a recent study by Johansson *et al.*²² and therefore pond description and some of the methods overlap slightly with the information given in that study²².

Citizen science data (CSD). We used the Swedish Species Observation System to extract records of dragonfly abundance based on CSD. These observations were collected by citizens in a non-standardized way, meaning that data were gathered opportunistically without standardized methods and controlling for sampling effort²⁴. These observations included all water bodies in the study area, and thus CSD habitats might represent freshwater systems with a wider range of habitat characteristics. Unfortunately, the CSD did not have enough replicates of stormwater ponds, which would allow to control for other confounding factors (e.g., type of water body). However, we included only species that were recorded in our stormwater pond survey. Hence, species that did not occur in the stormwater ponds were excluded from the CSD set. When the number of individuals was available in the database, we used this number as our estimate, and when the record of a species only mentioned “observed”, we gave this record an abundance of 1. Two hundred forty-eight (248) localities were surveyed by citizens during 8 years. Furthermore, two species, *Calopteryx virgo* and *C. splendens*, are predominantly lotic specialists and typically do not occur in standing waters, such as stormwater ponds. They were therefore removed from the analyses. These species occurred in low numbers at the stormwater ponds (a total of 5 and 2 individuals for *C. virgo* and *C. splendens*, respectively).

To evaluate whether CSD could be used to predict pond biodiversity, we extracted data covering different diameters around the center of the stormwater ponds in the city (Fig. 2). The center of the study area (59°50′51.3″N; 17°39′3.4″E) was estimated approximately as the centroid of the coordinates of the stormwater ponds. Data records from the database were extracted on a yearly basis from 2010 until 2017 to examine whether more recent CSD performed better in predicting biodiversity in the stormwater ponds. In addition, we also examined whether citizen data records covering a larger area around the city center improved the prediction of stormwater pond biodiversity. We did this by extracting and comparing data covering a diameter of 10, 20 or 30 km around the center of the stormwater ponds in the city (Fig. 2). This analysis was performed for the year 2017.

Data analyses. We used a Canonical Analysis of Principal Coordinates (CAP; Anderson & Willis²⁵) to investigate the differences in community structure between the stormwater pond (sampled by us in 2018) and the citizen datasets in terms of dragonfly species abundance and composition. CAP is a constrained ordination method and, therefore, it uses an a priori hypothesis to produce an ordination plot. This hypothesis can then be tested using a generalized discriminant analysis based on distances²⁶. Our a priori hypothesis was represented by a categorical (explanatory) variable with levels representing the type of data (stormwater in 2018 and citizen datasets). CAP was based on the Bray-Curtis dissimilarity index on dragonfly abundance data. The resulting dissimilarity matrix was also used in a distance-based test for homogeneity of multivariate dispersions (PERMDISP;

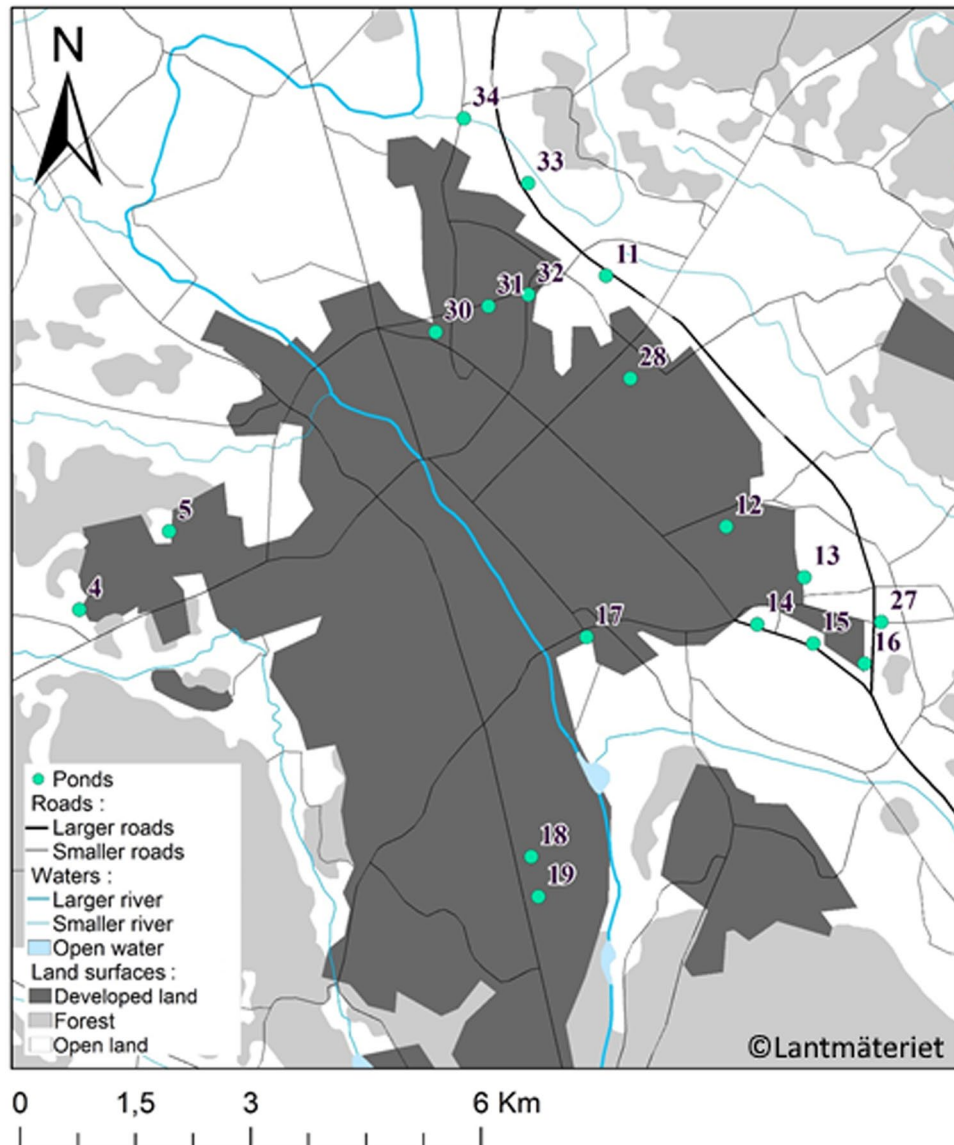


Figure 1. Map of the City of Uppsala, Sweden showing the location of the 18 stormwater ponds. The map was created in the software ArcGIS 10.0, ESRI Inc., Redlands CA (<https://desktop.arcgis.com/en/>) with permission from Lantmäteriet⁴³ through Uppsala University (license number: I2018/00145).

see Anderson,²⁷; Anderson *et al.*²⁸). This was done to test whether beta diversity values differed according to the type of the data. Since our main interest was focused on comparing the data obtained in the 2018 stormwater ponds with the data obtained by citizens, we carried out a set of planned comparisons. First, we repeated CAP to compare the 2018 dataset to each year separately (from 2010 to 2017) considering sites (surveyed by citizens) that were within a diameter of 10 km around the center of the stormwater ponds in the city. Second, we used CAP to compare the 2018 dataset with sites within diameters of 10 km, 20 km and 30 km around the center of the stormwater ponds. This analysis was restricted to the year 2017. Third, to evaluate the likely effect of a biased search for rarer species in the CSD, we also divided the dataset into two parts. The first part included the 16 most abundant species in our stormwater dataset (see Fig. S1), whereas the second part included the 11 rarer species. This splitting of the dataset was based on the location of an inflection point exhibited in a Whittaker plot (rank-abundance curve) and on an attempt to balance the number of rare and common species in the analyses. Thereafter, we ran independent analyses using these datasets. All analyses were carried out using functions (*vegdist*, *capscale* and *betadisper*) available in the vegan package²⁹. Significance tests were based on 999 permutations.

Results

Twenty-nine species of Odonata were found in the 18 studied stormwater ponds (Table 1), but only 27 were included in the main analyses since the two lotic *Calopteryx* species were excluded. The total number of species represents 61% of the Odonata species recorded in the province of Uppland in Sweden. The average species richness in the stormwater ponds was 10, with a range between 3 to 19. The most common species were *Libellula*

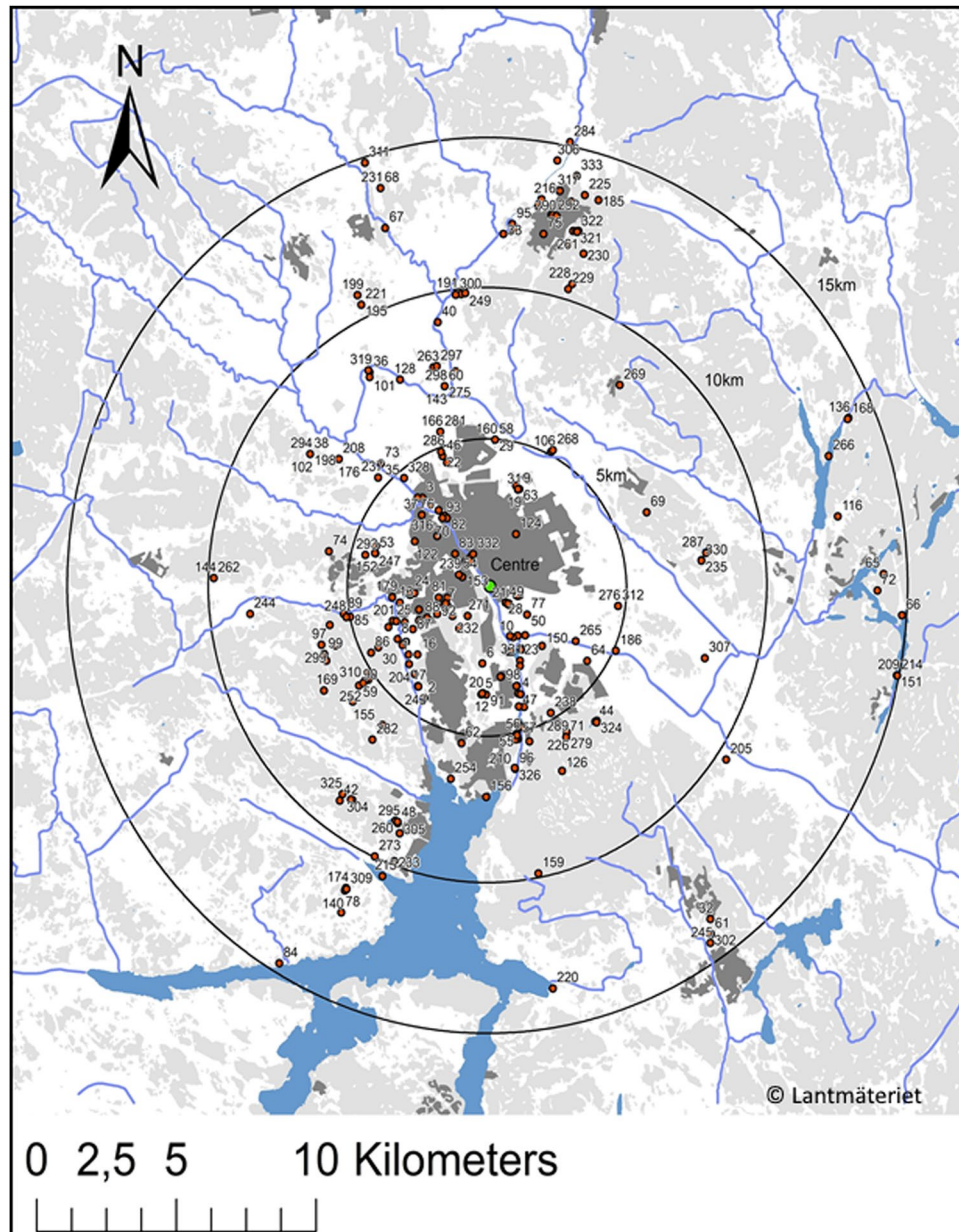


Figure 2. Distribution of citizen science data records (red dots). The green dot is the center of the study area and the circles show a 10, 20 and 30 km diameter from the center. Note that the dots for the 20 and 30 km only include the year 2017, while the 10 km radius include dots from 2010–2017. The map was created in the software ArcGIS 10.0, ESRI Inc., Redlands CA (<https://desktop.arcgis.com/en/>) with permission from Lantmäteriet⁴³ through Uppsala University (license number: I2018/00145).

quadrimaculata, *Sympetrum vulgatum*, *Lestes sponsa* and *Coenagrion puella*. Five species were only recorded in one pond (*Platycnemis pennipes*, *Aeshna mixta*, *Orthetrum coerulescens*, *Leucorrhinia pectoralis* and *Leucorrhinia rubicunda*; Table 1). Considering the CSD-set, the most common species were *Aeshna cyanea* (2010), *Sympetrum vulgatum* (2011 and 2014), *Aeshna grandis* (2012), *Somatochlora metallica* (2012), *Calopteryx virgo* (2013 and 2015), *Coenagrion hastulatum* (2016 and 2017; Table 1). We found a strong negative relationship between the difference in occupancy given by the datasets (average of the species frequencies of occurrence over the years in the CSD - frequency of occurrence in the stormwater ponds) and the mean abundance in the stormwater ponds (Fig. 3; $r = -0.88$; $P < 0.0001$ with all species and $r = -0.87$; $P < 0.0001$ after removing *C. virgo* and *C. splendens*). This result indicates a bias against abundant species in the CSD.

We found a significant difference between the citizen dataset (considering different years, 2010–2017) and the stormwater pond dataset from 2018 (Fig. 4; $F = 1.97$; $P < 0.001$), and the community structure from the more recent years of CSD-set were not more similar to that of the stormwater ponds community structure (Fig. 4). In addition, variations in community structure, i.e. beta diversity, as given by the citizen science datasets were much higher than that given by the 2018 stormwater pond dataset (Fig. 5; $F = 6.77$; $P < 0.001$). We also found that the

Species/Years	2010	2011	2012	2013	2014	2015	2016	2017	2018
<i>Aeshna cyanea</i>	25.0	15.0	30.8	15.2	21.4	11.1	7.7	17.6	11.1
<i>Aeshna grandis</i>	16.7	25.0	46.2	27.3	7.1	0.0	15.4	17.6	61.1
<i>Aeshna juncea</i>	8.3	15.0	15.4	12.1	7.1	0.0	7.7	0.0	11.1
<i>Aeshna mixta</i>	0.0	0.0	7.7	12.1	0.0	0.0	3.8	0.0	5.6
<i>Aeshna serrata</i>	0.0	0.0	7.7	6.1	0.0	0.0	3.8	0.0	22.2
<i>Brachytron pratense</i>	0.0	0.0	0.0	6.1	7.1	0.0	0.0	11.8	22.2
<i>Cordulia aenea</i>	0.0	5.0	7.7	6.1	0.0	3.7	3.8	5.9	16.7
<i>Coenagrion hastulatum</i>	0.0	10.0	30.8	18.2	14.3	11.1	26.9	29.4	61.1
<i>Coenagrion lunulatum</i>	0.0	0.0	0.0	0.0	0.0	0.0	15.4	0.0	16.7
<i>Coenagrion puella</i>	0.0	0.0	23.1	9.1	21.4	14.8	23.1	11.8	77.8
<i>Coenagrion pulchellum</i>	8.3	5.0	23.1	18.2	14.3	22.2	19.2	17.6	61.1
<i>Calopteryx splendens</i>	8.3	15.0	23.1	12.1	14.3	11.1	23.1	23.5	11.1
<i>Calopteryx virgo</i>	8.3	25.0	7.7	45.5	21.4	29.6	19.2	17.6	27.8
<i>Enallagma cyathigerum</i>	0.0	5.0	15.4	9.1	7.1	7.4	23.1	17.6	66.7
<i>Erythromma najas</i>	8.3	5.0	23.1	9.1	7.1	7.4	19.2	5.9	27.8
<i>Ischnura elegans</i>	0.0	0.0	7.7	3.0	14.3	7.4	7.7	11.8	50.0
<i>Libellula depressa</i>	0.0	10.0	7.7	0.0	14.3	3.7	7.7	11.8	33.3
<i>Leucorrhinia pectoralis</i>	0.0	0.0	7.7	3.0	0.0	0.0	3.8	0.0	5.6
<i>Libellula quadrimaculata</i>	16.7	10.0	23.1	9.1	0.0	3.7	7.7	0.0	94.4
<i>Leucorrhinia rubicunda</i>	8.3	0.0	0.0	6.1	0.0	0.0	3.8	0.0	5.6
<i>Lestes sponsa</i>	0.0	15.0	15.4	24.2	14.3	11.1	11.5	11.8	83.3
<i>Orthetrum coerulescens</i>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6
<i>Platycnemis pennipes</i>	0.0	5.0	23.1	9.1	7.1	22.2	15.4	17.6	5.6
<i>Sympetrum danae</i>	0.0	5.0	23.1	12.1	7.1	11.1	11.5	11.8	16.7
<i>Sympetrum flaveolum</i>	0.0	5.0	23.1	6.1	7.1	3.7	3.8	0.0	38.9
<i>Sympecma fusca</i>	8.3	5.0	15.4	3.0	7.1	7.4	19.2	23.5	16.7
<i>Somatochlora metallica</i>	0.0	25.0	46.2	18.2	7.1	7.4	0.0	17.6	22.2
<i>Sympetrum sanguineum</i>	16.7	20.0	30.8	21.2	21.4	14.8	7.7	17.6	44.4
<i>Sympetrum vulgatum</i>	0.0	30.0	38.5	33.3	28.6	11.1	7.7	23.5	94.4

Table 1. List of dragonfly species recorded by citizen scientists (from 2010 up to 2017) and in 18 stormwater ponds in the city of Uppsala, Central Sweden (2018). Also shown is the frequency of occurrence of each species (number of occurrences/number of sites). Note that the lotic species, *Calopteryx splendens* and *C. virgo*, were excluded from the statistical analyses.

differences among the datasets remained independently of the diameter (10, 20 and 30 km) used to extract the CSD ($F = 2.61$; $P = 0.001$). Running these analyses after including data on lotic species did not change the results qualitatively (results not shown).

After splitting the dataset according to species abundance (see Fig. S1), we found that the differences between citizen science datasets and the 2018 dataset were remained only when common species were considered in the analyses (Table 2; Fig. 6). In contrast, we did not find any significant difference in community structure (between CSD and 2018 dataset) when rare species were considered (Table 2; Fig. 6). The same pattern was recorded when the comparisons were based on the years 2018 and 2017 for different diameters: the differences among the datasets were significant when the analysis were done using all species ($F = 2.61$; $P = 0.001$) or common species ($F = 2.75$; $P = 0.001$), but not when the analysis was based on rare species data ($F = 0.84$; $P = 0.77$; Fig. S2).

Discussion

Colonization of new man-made habitats is an important process that may counteract biodiversity loss³⁰. Stormwater ponds is one category of such new man-made habitats and, therefore, it is important to examine whether knowledge on the biodiversity in the area surrounding the ponds based on citizen science data (CSD) can be used to predict the biodiversity in these ponds. Our results suggest that the diversity of rare species in these new man-made habitats could be predicted from the CSD. However, the results also suggest that CSD cannot be used to portray the overall biodiversity of dragonflies in stormwater ponds. There could be several reasons underlying this result.

First, there is a tendency that opportunistic CSD are biased towards over-reporting rare species and under-reporting more common species^{20,21}. We found support for this bias because ordination scores obtained with our dataset and with CSD overlapped mainly when rare species were used in the analyses. Similarly, Snäll *et al.*²⁰ found that common bird species were not regularly reported in the Swedish Species Observation System, and we suggest that the same holds for dragonflies. For example, *L. quadrimaculata* and *S. vulgatum* were found in 94.4% of the ponds in 2018, but in the CSD set (radius 10 km) these two species were reported, in average, with much lower frequencies (8.8 % and 21.6 %, respectively). In contrast, species less abundant, in our dataset of stormwater ponds, were reported with a similar frequency in the CSD-set. Currently, methods are being

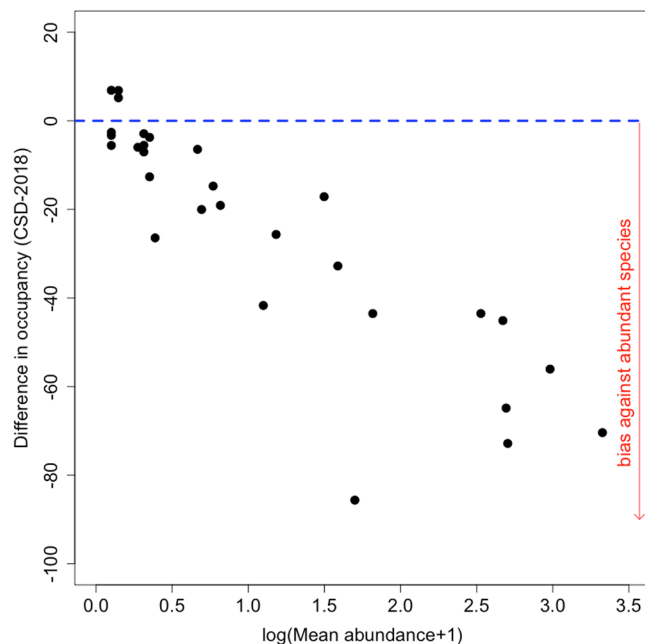


Figure 3. Relationship between the difference in occupancy given by the datasets [average of the species frequencies of occurrence over the years in the CSD - frequency of occurrence in the stormwater ponds (2018)] and the mean abundance in the stormwater ponds. Each point represents a species. The dashed line represents no difference and negative values indicate that occupancies as given in CSD were lower than those in the stormwater ponds.

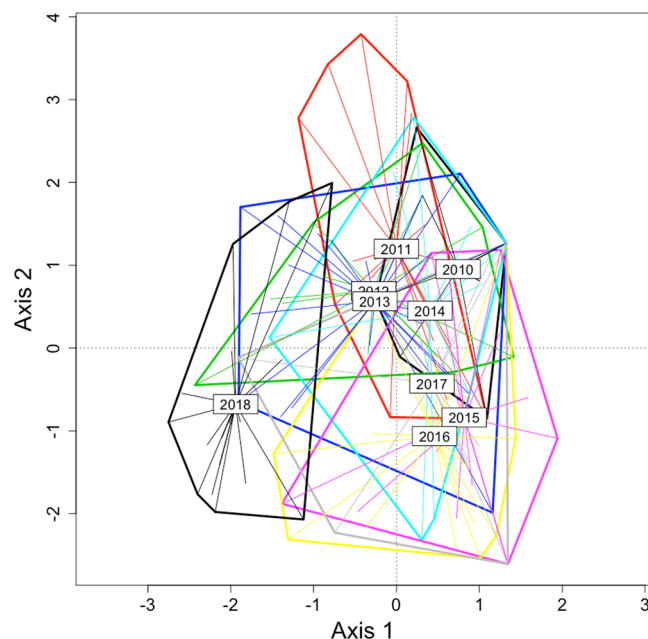


Figure 4. Ordination of sites (Canonical Analysis of Principal Coordinates) based on dragonfly species recorded in stormwater ponds (2018) and by citizen scientists (from 2010 up to 2017).

developed to correct for these biases³¹, and they should be used in future studies when predicting colonization of new man-made habitats. However, it is worth noting that citizen scientists may be especially interested in finding rare species with conservation interest³². Thus, we also found some support for this view, which suggests that even biased data may be beneficial for conservation purposes²⁴.

Second, the stormwater ponds may not be comparable to the more natural habitats sampled in the CSD. For example, since the stormwater ponds are more recent habitats, they might be at early successional stages preferred by only certain dragonfly species. In contrast, CSD habitats might represent freshwater systems with a wider range

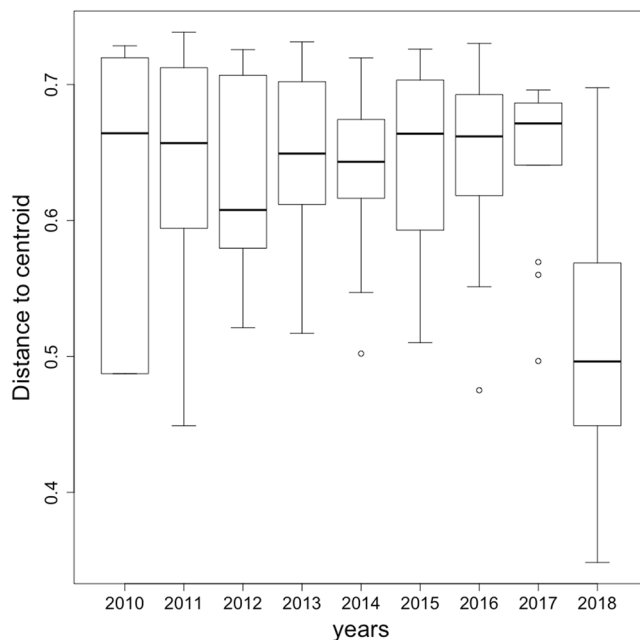


Figure 5. Distance to centroid within each year for citizen datasets (2010 to 2017) and for the stormwater ponds dataset (2018).

	Year	df	F	P
Whole	2010	26	5.51	0.001
	2011	33	4.42	0.001
	2012	28	3.13	0.001
	2013	41	3.57	0.001
	2014	28	4.16	0.001
	2015	37	6.54	0.001
	2016	37	5.47	0.001
	2017	29	4.15	0.001
Common	2010	22	4.17	0.001
	2011	29	4.00	0.001
	2012	27	2.86	0.003
	2013	38	3.13	0.001
	2014	26	4.02	0.001
	2015	29	5.50	0.001
	2016	33	5.05	0.001
	2017	26	3.81	0.001
Rare	2010	14	2.54	0.019
	2011	22	0.95	0.489
	2012	19	1.21	0.303
	2013	24	1.89	0.034
	2014	16	0.82	0.565
	2015	22	2.32	0.020
	2016	18	2.43	0.019
	2017	17	1.03	0.443

Table 2. Results from generalized discriminant analysis based on distances testing differences between the data obtained in 2018 and the citizen science data gathered in different years (from 2010 to 2017). For each set of results, F-values were considered significant at $P < 0.05/8$ according to the Bonferroni method. df = residual degrees of freedom. Significant F- values are in bold.

of habitat characteristics or at later successional stages, which are preferred by a different set of dragonfly species. However, studies have found that aquatic insect communities in newly-created ponds may reach community structure similar to that in the natural ponds within one or two years^{33–35}. Thus, since the youngest stormwater

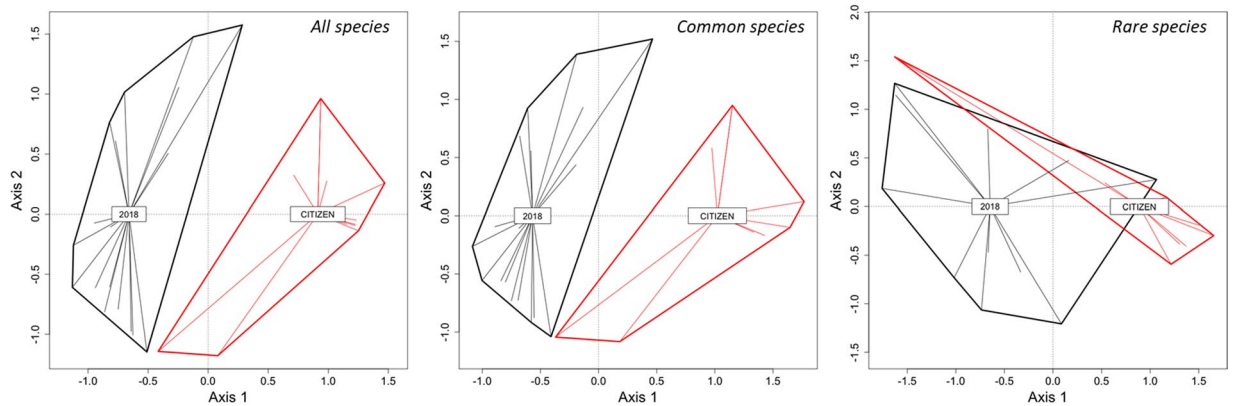


Figure 6. Ordination of sites (Canonical Analysis of Principal Coordinates) based on dragonfly species recorded in stormwater ponds (2018) and by citizen scientists (CITIZEN). Shown are the results obtained with all species (left), common species (center) and rare species (right) for the year 2017 (as an example).

ponds were over four years old, we suggest that differences in successional stages were not the reason why we could not predict the biodiversity of stormwater ponds based on the CSD. In addition, studies using systematic protocols that were designed to compare the biodiversity of aquatic insects (including dragonflies) in stormwater ponds and more natural ponds have found similar results¹¹.

Third, the CSD-set was collected from all kinds of aquatic habitats while the stormwater ponds probably represent a more similar range of aquatic habitats. If we had restricted our CSD-set to stormwater ponds, we might have found that CSD could predict the community composition in the stormwater ponds. Unfortunately, the CSD did not have enough replicates of stormwater ponds for such a comparison. Hence, we emphasize that our goal was to ask whether CSD collected from aquatic habitats could predict the community of Odonata in stormwater ponds.

We expected that the use of CSD collected from a larger diameter around our study area would make a better prediction, since a large area covers more habitats compared with more limited areas^{36,37}. Conversely, one could expect a decrease in predictability with area, due to the distance decay of similarity in ecological communities^{19,38}. However, we did not find support for any of these expectations because, independently of the diameter by which we based our comparison on, there was a significant difference in species composition between our stormwater pond data and the CSD. Thus, we did not observe that Odonata communities from aquatic habitats surveyed by volunteers, which were geographically closer to our study area, were more similar to the stormwater pond data. This finding suggests that under-reporting common species and over-reporting rare species still overrides the effect of area increase in our dataset.

We also found that variation in community structure (i.e., beta diversity) was significantly higher in the CSD compared to our stormwater pond data. We suggest that the main reason for this pattern is that stormwater ponds are more similar to each other than the water bodies from the CSD-set. The water bodies in the CSD included all kinds of freshwaters, from lentic to lotic and from temporary to large permanent lakes. Differences in biodiversity among different landscape or habitat types have also been found in previous studies on freshwater invertebrates. For example, urban and rural ponds may support different invertebrate communities (e.g.³⁹), forest lakes and bog lakes typically harbor distinct invertebrate communities (e.g.⁴⁰), and lotic and lentic ecosystems generally show contrasting invertebrate communities in terms of alpha and beta diversity (e.g.⁴¹). In this sense, our CSD-set should show overall higher levels of biodiversity than our stormwater pond dataset. An alternative, but not mutually exclusive, explanation for the difference in beta diversity could be the under-representation of common species in the CSD, which thereby could inflate the estimates of beta-diversity in the CSD.

In summary, we were unable to predict patterns in dragonfly biodiversity in stormwater ponds based on data collected by citizen scientists. We suggest that the main reason for this result is that common species are under-reported and rare species are over-reported by citizen scientists. Similar problems with the under-reporting of common species have been found in studies estimating annual variation in birds, species richness in beetles, and spatio-temporal variation in beetle abundance^{9,19,20,42}. There is thus a need for predictive models that take into account or correct for the under-reporting of common species in CSD, and such models should provide better predictions of population trends and colonization of man-made habitats by species.

Received: 19 February 2020; Accepted: 19 May 2020;

Published online: 10 June 2020

References

- Venter, O. *et al.* Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications* **7**, <https://doi.org/10.1038/ncomms12558>. (2016).
- Troudet, J. *et al.* Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* **7**, 9132 (2017).
- Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics* **41**, 149–172 (2010).

4. Dickinson, J. L. *et al.* The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment* **10**, 291–297 (2012).
5. Silvertown, J. *Trends in Ecology and Evolution* **24**, 467–471 (2009).
6. Devictor, V. *et al.* Citizen science programmes as useful tools for conservation biogeography. *Diversity and Distribution* **16**, 354–362 (2010).
7. Mair, L. *et al.* Evaluating citizen science data for forecasting species responses to national forest management. *Ecology and Evolution* **7**, 368–378 (2016).
8. Aceves-Bueno *et al.* The Accuracy of Citizen Science Data: A Quantitative Review. *Bulletin of the Ecological Society of America* **98**, 278–290 (2017).
9. Snäll, T. *et al.* Evaluating temporal variation in citizen science data against temporal variation in the environment. *Ecography* **37**, 293–300 (2014).
10. Scher, O. & Thiery, A. Odonata, Amphibia and environmental characteristics in motorway stormwater retention ponds (Southern France). *Hydrobiologia* **551**, 237–251 (2005).
11. Le Viol, I., Mocq, J., Julliard, R. & Kerbiriou, C. The contribution of motorway stormwater retention ponds to the biodiversity of aquatic macroinvertebrates. *Biological Conservation* **142**, 3163–3171 (2009).
12. Hassall, C. & Anderson, S. Stormwater ponds can contain comparable biodiversity to unmanaged wetlands in urban areas. *Hydrobiologia* **745**, 137–149 (2015).
13. Holtmann, L. *et al.* Stormwater ponds promote dragonfly (Odonata) species richness and density in urban areas. *Ecological Engineering* **118**, 1–11 (2018).
14. CEDR. Management of contaminated runoff water: current practice and future research needs. SBN: 979-10-93321-18-9 (2016).
15. European Commission, The EU Floods Directive. [online] Available at, http://ec.europa.eu/environment/water/flood_risk/, [Accessed 16 March 2019] (2012).
16. Corbet, P. S. Dragonflies: Behaviour and Ecology of Odonata. — Harley Books, Martins. (1999).
17. Oertli, B. *et al.* Does size matter? The relationship between pond area and biodiversity. *Biological Conservation* **104**, 59–70 (2002).
18. Hassall, C., Hollingshead, J. & Hull, A. Environmental correlates of plant and invertebrate species richness in ponds. *Biodiversity and Conservation* **20**, 3189–3222 (2011).
19. Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* **26**, 867–878 (1999).
20. Snäll, T. *et al.* Evaluating citizen-based presence data for bird monitoring. *Biological Conservation* **144**, 804–810 (2011).
21. Gardiner, M. M. *et al.* Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Frontiers in Ecology and the Environment* **10**, 471–76 (2012).
22. Johansson, F. *et al.* Environmental variables drive differences in the beta diversity of dragonfly assemblages among urban stormwater ponds. *Ecological Indicators* **106**, 105529 (2019).
23. Sahlén, G. Sveriges Trollsländor. Fältbiologerna, Stockholm (1996).
24. Tiago, P., Pereira, H. M. & Capinha, C. Using citizen science data to estimate climatic niches and species distributions. *Basic and Applied Ecology* **20**, 75–85 (2017).
25. Anderson, M. J. & Willis, T. J. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84**, 511–525 (2003).
26. Anderson, M. J. & Robinson, J. Generalized discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics* **45**, 301–318 (2003).
27. Anderson, M. J. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245–253 (2006).
28. Anderson, M. J., Ellingsen, K. E. & McArdle, B. H. Multivariate dispersion as a measure of beta diversity. *Ecology Letters* **9**, 683–693 (2006).
29. Oksanen, J. *et al.* Vegan: Community Ecology Package. *R package version 2*, 3–0 (2017).
30. Clobert, J., Baguette, M., Benton, T.G., & Bullock, J.M. Dispersal ecology and evolution. Oxford: Oxford Univ. Press. (2012).
31. Bird, T. J. *et al.* Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation* **173**, 144–154 (2014).
32. Losey, J. E., Perlman, J. E. & Hoebeke, E. R. Citizen scientist rediscovers rare nine-spotted lady beetle, *Coccinella novemnotata*, in eastern North America. *Journal of Insect Conservation* **11**, 415–417 (2007).
33. Street, M. & Titmus, G. The colonisation of experimental ponds by Chironomidae Diptera. *Aquatic Insects* **4**, 233–244 (1979).
34. Gee, J. H. R., Smith, B. D., Lee, K. M. & Griffiths, S. W. The ecological basis of freshwater pond management for biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems* **7**, 91–104 (1997).
35. Bloechl *et al.* Abundance, diversity and succession of aquatic Coleoptera and Heteroptera in a cluster of artificial ponds in the North German Lowlands. *Limnologia* **40**, 215–225 (2010).
36. MacArthur R. H., & Wilson E. O. The theory of island biogeography. Princeton, NJ: Princeton University Press (1967).
37. Kohn, D. D. & Walsh, D. M. Plant species richness – the effect of island and habitat diversity. *Journal of Ecology* **82**, 367–377 (1994).
38. Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological communities. *Ecography* **30**, 3–12 (2007).
39. Hill, M. J., Heino, J., Thornhill, L., Ryves, D. B. & Wood, P. J. Effects of dispersal mode on the environmental and spatial correlates of nestedness and species turnover in pond communities. *Oikos* **126**, 1575–1585 (2017).
40. Heino, J. Does dispersal ability affect the relative importance of environmental control and spatial structuring of littoral macroinvertebrate communities? *Oecologia* **171**, 971–980 (2013).
41. Williams, P. *et al.* Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. *Biological Conservation* **115**, 329–341 (2004).
42. Jeppson, T. *et al.* The use of historical collections to estimate population trends: A case study using Swedish longhorn beetles (Coleoptera: Cerambycidae). *Biological Conservation* **143**, 1940–1950 (2010).
43. Lantmateriet, Geodataportalen. URL, <https://www.geodata.se> (2020).

Acknowledgements

Thanks to Johanna Liljestränd Rönn for help with extracting the SCD from the Swedish Species Observation System, and to Jelle Brealey comments on a previous version of this article. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Work by LMB was done in the context of the National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, supported by MCTIC/CNPq (proc. 465610/2014-5) and FAPEG. Open access funding provided by Uppsala University.

Author contributions

F.J. conceived the study and wrote the first draft. L.M.B. designed and performed all the statistics with communication from J.H. P.C and J.W. conducted the field work. F.J., J.H., P.C., R.S., J.W. and L.M.B. contributed to writing and reviewing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66306-0>.

Correspondence and requests for materials should be addressed to F.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020