

Adaptive sampling by citizen scientists improves species distribution model performance: A simulation study

Thomas Mondain-Monval¹  | Michael Pocock²  | Simon Rolph² | Tom August²  | Emma Wright³ | Susan Jarvis¹ 

¹UK Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster, UK

²UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

³JNCC, Quay House, Peterborough, UK

Correspondence

Thomas Mondain-Monval

Email: thoval@ceh.ac.uk

Funding information

Natural Environment Research Council, Grant/Award Number: NE/R016429/1 and NE/V003054/1

Handling Editor: Res Altwegg

Abstract

1. Volunteer recorders generate large amounts of biodiversity data through citizen science which is used in conservation planning and policy decision-making. Unstructured sampling, where the volunteer can record what they want, where they want, leads to spatial unevenness in these data. While there are many statistical techniques to account for the resulting biases, it may be possible to improve datasets by directing a subset of recorders to sample in the most informative locations, known as adaptive sampling. We investigated the potential for adaptive sampling to improve the performance of species distribution models built on citizen science data using simulated ecological communities.
2. We simulated ecological assemblages across Great Britain based on current butterfly data and modelled the distributions of each species. We then simulated the sampling of new data based on five adaptive sampling methods (one empirical method based simply on gap-filling, and four model-based methods using various measures from the model outputs) and one non-adaptive method (a method in which recording continued in the current pattern), and re-ran the species distribution models. In these, we also varied the rate of recording effort that was distributed according to adaptive sampling. The model predictions using the original and adaptively sampled data were compared to true species distributions to evaluate the performance of each method.
3. We found that all adaptive sampling approaches improved model performance, with greatest improvement for model-based approaches compared to the empirical sampling method (i.e. simple gap-filling). All four model-based adaptive sampling approaches provided similar benefits for model outputs. Improvements in model performance were greatest when the amount of adaptive sampling changed from no uptake to 1% uptake, indicating that only a small amount of change in recorder behaviour is needed to improve model performance.
4. Directing volunteer recorders to places where records are most needed, based on information from model outputs, can improve species distribution models built on citizen science data, even with minimal uptake of suggested locations. Our

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

results therefore suggest that adaptive sampling by recorders could be beneficial for real-world citizen science datasets.

KEYWORDS

adaptive sampling, butterfly, citizen science, simulations, species distribution model

1 | INTRODUCTION

The amount of biological data collected by citizen scientists has increased exponentially in recent decades, which is proving to be of great value for biodiversity monitoring (Feldman et al., 2021). One major benefit of citizen science recording is the extent and scale at which it operates, facilitating sampling from more locations and at a finer resolution than would be feasible through conventional surveys. This has allowed researchers and conservationists to monitor wildlife across spatial and temporal scales that were previously impossible (Weiskopf et al., 2022, eBird Status and Trends—<https://science.ebird.org/en/status-and-trends>) and are increasingly used to understand the drivers of biodiversity trends (Mancini et al., 2023; Woodcock et al., 2016). One of the major facilitators of this change in biological recording is that new technology has increased the volume of unstructured data, by allowing recorders to collect data more easily than ever before (August et al., 2015). This unstructured data collection has generated significant biases in such datasets. For example, observers tend to record near to their homes (which are not evenly distributed in space) and sometimes show preferences for particular taxa, habitats, locations or times of year (Bowler, Callaghan, et al., 2022; Isaac & Pocock, 2015). These recording patterns reflect the broad range of interests and motivations of recorders (August et al., 2020; Bowler, Bhandari, et al., 2022) and make analyses using these data particularly challenging (Johnston et al., 2022). Attempting to account for these biases when trying to understand patterns and changes in biodiversity has led to the development of increasingly complex statistical models (Johnston et al., 2022). However, rather than just statistically accounting for these biases post-hoc, it could also be beneficial to improve the data at the point of collection (Callaghan, Rowley, et al., 2019).

Adaptive sampling is an approach whereby the design of a survey or monitoring scheme changes as data are collected (Shanahan et al., 2021; Turk & Borkowski, 2005; Wikle & Royle, 2005). The rationale behind adaptive sampling is that by adjusting sampling effort based on existing knowledge, we can increase cost-effectiveness and reduce data redundancy (Lindenmayer & Likens, 2009). Adaptive sampling designs have been implemented in many different disciplines (e.g. sensor networks—Andersson et al., 2023; Jain & Chang, 2004, and the internet of things—Giouroukis et al., 2020), and are especially useful where there is change over time in the system of interest (Lermusiaux, 2007). In these, a traditional sampling scheme that is fixed from the outset may become suboptimal over time, as the system itself changes. In extreme cases, the original design can become unsuitable for answering the original question

of interest (Wikle & Royle, 2005), meaning that the ability to adjust our design to capture changes in the system is useful. This could become increasingly important given the current dramatic changes to biodiversity occurring globally. Even in cases where the system of interest does not change over time, adaptive designs may perform better than traditional methods because they can optimise data collection based on data already collected (Specht et al., 2017; Turk & Borkowski, 2005). Importantly, changes in the sampling design can be accounted for in our analyses. This allows us to draw robust conclusions despite changes in sampling design over time.

Adaptive sampling designs are not routinely used in ecology, where there is typically a strong focus on sampling design being established a priori. Where adaptive sampling designs have been used, they are usually in cases where the survey manager has full control over the data collection process, such as when contracted staff are sent to specific locations (Hooten et al., 2009; Pacifici et al., 2016; Wikle & Royle, 2005), or where the data are collected by sensor networks which can be managed remotely (Cardell-Oliver et al., 2005). However, adaptive sampling could be equally applicable where the data are collected by citizen or volunteer scientists (Callaghan, Poore, et al., 2019; Callaghan, Rowley, et al., 2019). This could be achieved by identifying regions in which sampling would have a large impact on the information content of the data or the performance of a statistical model (Callaghan, Rowley, et al., 2019). For example, Callaghan, Poore, et al. (2019) predicted the marginal benefit of sampling in all possible locations across a region for improving population trend estimates. From these, they created priority maps defining where future recording would be the most beneficial. If recorders could be influenced to record in such high priority regions, which evidence suggests is possible (Callaghan et al., 2021; Flint et al., 2023; Xue et al., 2016), then adaptive sampling could be used to enhance the impact of species recording in the field by volunteers.

The concept of redirecting volunteer effort in biological recording is not new. For example, comprehensively mapping species distributions through citizen science in atlasing projects requires recorders to be directed to unrecorded locations (Harris et al., 2021; Robertson et al., 2010). Recorders may also be asked to look for specific species (e.g. the Lost and Found Fungi Project—<https://www.kew.org/read-and-watch/lost-and-found-fungi>) or visit areas that have had few recent records or none at all (e.g. Targeting Revisits map: https://connect-apps.ceh.ac.uk/targeting_revisits_grasshoppers/ and the establishment of new Breeding Bird Survey routes: <https://www.pwrc.usgs.gov/BBS/RouteMap/Map.cfm>). Importantly, the objective of interest of any adaptive sampling design needs to be specifically defined a priori for its successful implementation. By doing this, we can

design a scheme which improves the information content of the data about the objective of interest (Callaghan, Poore, et al., 2019). The objective can be derived directly from the data, by filling gaps in the records of a species' distribution, which we term 'empirical adaptive sampling'. Alternatively, the objective can be derived from models of the data, i.e. improving model outputs in a pre-defined way, which we term 'model-based adaptive sampling'. In the latter approach, models are constructed from the data, and sampling locations are based on one or more of the model outputs. For example, model-based objectives could be to obtain better occupancy estimates of rare species (Pacifiçi et al., 2016; Specht et al., 2017) or to reduce uncertainty in estimates of species' current or future distributions (Reich et al., 2018; Williams et al., 2018). In the context of citizen science, there may also be secondary objectives, such as reduced spatial bias in observations, increased spatial coverage or increased participation (Xue et al., 2016).

Previous work has shown the potential for adaptive sampling to be beneficial for citizen science, but it is unclear which types of adaptive sampling (empirical or model-based) are most effective and how much survey effort needs to be redirected. To address these questions, we used a virtual ecologist approach (Zurell et al., 2010) and simulated the effects of adaptive sampling on species distribution modelling (SDMs) with citizen science data. We simulated assemblages of species, which allowed us to evaluate SDM performance against true species distributions, and applied six different scenarios of recording behaviour by volunteers: continued sampling with existing spatial biases, filling spatial gaps ('empirical adaptive sampling'), and four different model-based adaptive sampling methods. Within these scenarios, we varied the rate of uptake of adaptive sampling, that is the proportion of new locations that were sampled adaptively. The virtual ecologist approach enabled us to evaluate the impact of each sampling approach, at different uptake rates, on the performance of SDMs for the whole assemblage and for species individually.

2 | METHODS

2.1 | Replication statement

Scale of inference	Scale at which factor of interest is applied	Number of replicates at the appropriate scale
Assemblage	Assemblage	50 species × 50 assemblages = 2500

2.2 | Virtual species generation

All analyses were carried out in R v4.3.1 (R Core Team, 2023) and the code available online (Mondain-Monval et al., 2024a, 2024b). We used the *virtuallspecies* package (Leroy et al., 2016) to generate 50 virtual assemblages across GB, with each assemblage containing

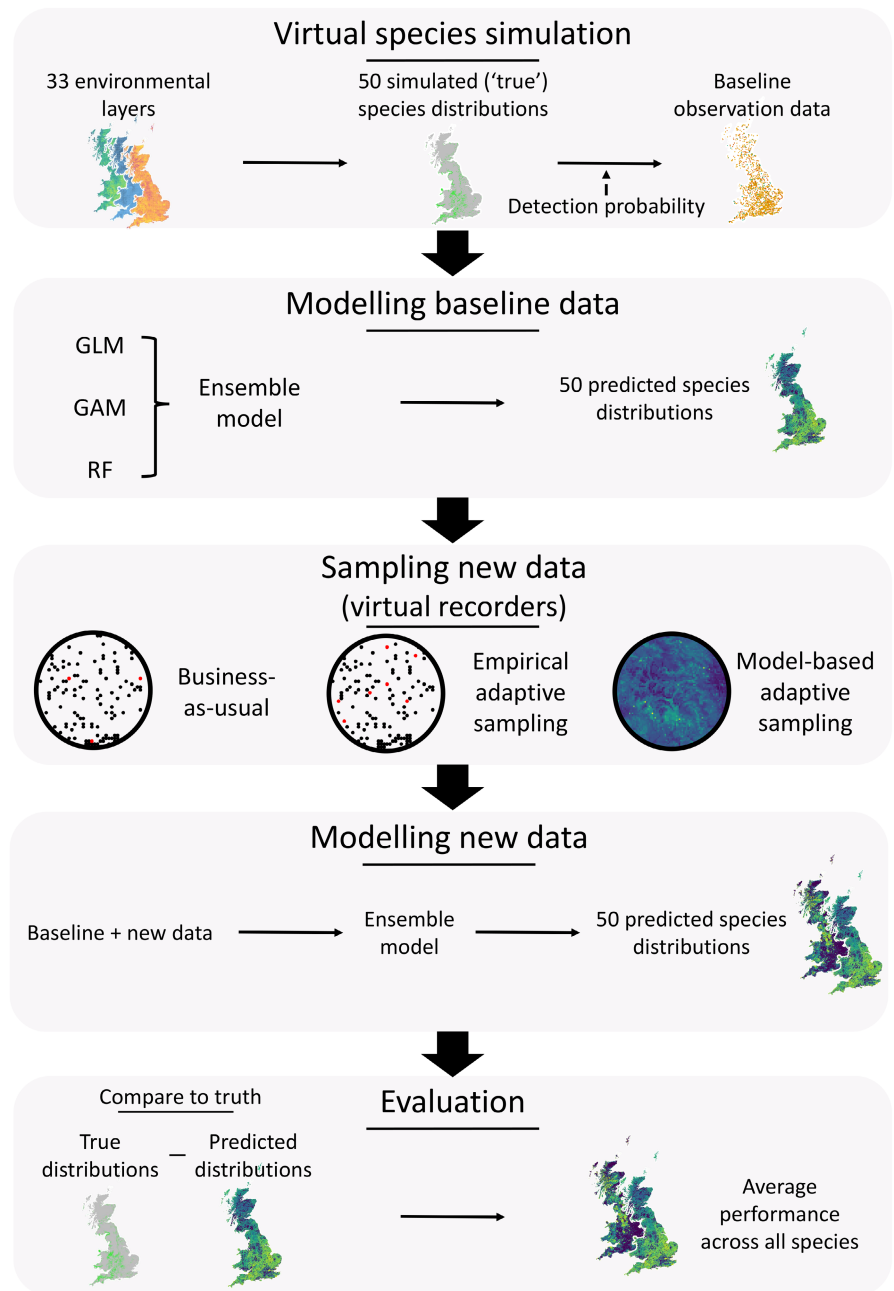
50 species (Figure 1). We chose 50 species because it approximates the number of butterfly species in GB, based on the Butterflies for the New Millennium (BNM) dataset, an unstructured recording scheme coordinated by Butterfly Conservation. Multiple virtual assemblages were simulated to account for natural variation in ecological communities; variation between assemblages was generated by stochasticity in the species simulations. We used 33 environmental input layers, describing climate (Hollis et al., 2018), elevation (Copernicus EU-Digital Elevation Model v1.1) and habitat variation (Rowland et al., 2017) across GB (Table S1). The habitat and elevation layers were provided at a 25 m resolution, these were aggregated to the 1 km resolution of the climatic variables (Table S1). We used a principal component analysis (PCA)-based approach to create species niches using differing combinations of the environmental layers (Leroy et al., 2016). Species distributions were simulated independently from one another to simplify our simulation. To ensure sufficient variation in species distributions, we did not include all 33 input layers in each species generation; instead, we randomly selected 10 layers from the 33 available. We generated a probability of occurrence ($p_{occ,i,j}$) for each virtual species j in each 1 km cell i in GB. This was used to generate a binary occurrence map for each species using the *convertToPA* function (Leroy et al., 2016). We did this using a probabilistic approach, assuming a logistic relationship between presence and environmental suitability (Leroy et al., 2016). This was defined as the true species distribution. We did not explicitly consider time, with only two 'time points': one for baseline sampling and a second time point for adaptive sampling (see below). For simplicity, we assumed that there was no change in species distributions between these two sampling points.

Each of the virtual species were simulated with 'narrow' species ranges using the *generateSpFromPCA* function (Leroy et al., 2016). Narrow ranges are calculated by limiting the standard deviations of the PCA axes to between 1% and 10%, essentially recreating species with low tolerance to variation in habitat and/or climate. We assessed how well the virtual assemblages matched real communities in terms of the relative distribution of common and rare species using rank abundance curves. For each of the 50 assemblages we generated, we verified that their rank abundance curves broadly matched the rank abundance curve generated for butterflies in GB in the BNM dataset (Figures S1 and S2), ensuring a realistic balance between common and rare species.

2.3 | Baseline sampling of virtual species

To include realistic existing patterns of sampling we used data from the BNM scheme to produce a layer of spatially varying effort over GB. We counted the number of unique days that each 1 km grid cell was visited (D_i) and at least one species was reported between 2001 and 2019, inclusive. We used this layer as weights to probabilistically sample our virtual species' distributions and generate baseline distribution data. To ensure that unvisited cells did not have zero probabilities of future visits, we added 1 to all cells ($D_{+1,i}$). The probability

FIGURE 1 Outline of the simulation and modelling process. We simulated 50 species distributions for each of the 50 assemblages, using 33 environmental variables. These species distributions were sampled according to the recording patterns in the Butterflies for the New Millennium dataset. This generated baseline species presence data which mimicked those collected by recorders in real-world recording schemes. We modelled these data using ensemble models to generate predicted distributions and uncertainty used for our model-based adaptive sampling methods. Ensemble models consisted of averaging predictions from generalised linear models (GLM), generalised additive models (GAM) and random forest (RF) models. Six different methods were used to sample new data: business-as-usual, empirical gap-filling adaptive sampling and four model-based adaptive sampling methods (see the main text for full details). These new data were combined with the baseline data and ensemble models were used to generate a new predicted distribution for each species and sampling method. The predicted distributions were evaluated against the true distributions of each species. We then compared the performance of models before and after sampling to determine the effectiveness of the six sampling methods.



of an unvisited grid cell being sampled was therefore half the probability of a grid cell which was visited once. This ensured that records could be made in areas with no visits in the real butterfly dataset, but that the broad pattern of intensity of simulated visits was strongly associated with the real pattern. The probability of sampling a cell i as part of the baseline sampling ($p_{\text{base},i}$) was

$$p_{\text{base},i} = \frac{(D_{+1,i} / \sum D_{+1,i})}{N_{\text{base}}}, \quad (1)$$

where N_{base} is the total number of cells sampled for baseline sampling.

To generate species observations we selected 20,000 cells (N_{base} , ~10% of the total $N=216,774$ 1 km cells) with the probability $p_{\text{base},i}$. We simulated occurrence of species j in each selected cell i by using the binary occurrence maps generated from the *virtuallspecies*

package to identify whether or not a species was present. To generate species observations, we applied a detection probability such that not all species present were observed. We assumed a constant detection probability of 0.2 for each species present (Isaac et al., 2011; Riva et al., 2020) for simplicity so we could understand the effect of adaptive sampling and uptake on model performance without additional noise from varying detection probabilities. This value was chosen based on studies of butterfly detectability in standardised Pollard Walk surveys (Isaac et al., 2011), and the effect of weather (Riva et al., 2020).

We did not simulate reporting probability, assuming that all species that were detected were reported (i.e. 'complete checklists'). Therefore, the data from each visited cell was a vector of detected species (non-detections were discarded to simulate a presence-only

dataset, as is the nature of many datasets collected by volunteer recorders).

2.4 | Modelling of baseline data

To inform our model-based adaptive sampling approaches, and to generate predictions which we could use to determine changes in model performance due to sampling, we needed an initial model of the species distributions from the simulated observations. To produce initial SDMs for each virtual species in each assemblage, we used an ensemble of three commonly applied models: logistic regression (GLM), generalised additive models (GAM) and random forests (RF) using code modified from the package *soaR* (<https://github.com/robboyd/soaR>; Boyd et al., 2023). Ensemble models are commonly used in analyses of species distributions, particularly when modelling large numbers of species for which it is infeasible to fine-tune the explanatory variables used for individual species (Hao et al., 2020). Each species was assumed to be independent from the other species in the assemblage and the distributions were modelled as a function of seven environmental variables randomly sampled from the 10 used to generate the species' distribution. This was done to mimic the imperfect knowledge of species distributions by the modeller. We selected pseudo-absences using the target background approach (Phillips et al., 2009). The estimate of occurrence probability was averaged across model types to give $\widehat{p}_{occ,i,j}$. Note that because detection was not explicitly modelled, we cannot separate occurrence and detection with our models, so $\widehat{p}_{occ,i,j}$ combines estimates of occurrence and detection.

For each model fitted to each species within each assemblage, we used a k -fold cross-validation approach to evaluate model performance and generate comparable measures of uncertainty across

model types (see below). Each dataset is split into k folds and the models are run on $k-1$ folds while holding one out. Performance is then evaluated against this test fold and the process repeated k times. We chose $k=10$. We used these 10 cross-validated model runs to generate predicted probability of occurrence maps for GB. We calculated the standard deviation across these predictions for each 1 km square for each model type for each species. We then calculated the mean probability of presence and standard deviation across the 10 model runs, for each of the three model types (GLM, GAM, RF) separately. We chose this method rather than calculating standard deviation across all 30 iterations per species (3 models \times 10 runs) to allow each model to be investigated separately if required. We took the mean of these standard deviations for each model type to use as our metric of uncertainty in model estimates, $\widehat{unc}_{i,j}$.

2.5 | Sampling new data

To test the impact of different adaptive sampling methods, we simulated sampling of the virtual assemblages using six different sampling strategies (Figure 2), specifically: a single, non-adaptive, sampling method—'business-as-usual'; and five adaptive methods—'gap-filling', 'rare species', 'uncertainty only', 'uncertainty of rare species' and 'gap-filling with uncertainty' (see below for full descriptions). The 'business-as-usual' method was used as a comparison for the other methods, to show the change in model performance if no adaptive sampling was carried out. The five other methods were chosen because of their potential to increase the value of the new data in informing SDMs of the simulated species. In real-world adaptive sampling schemes, decisions are required about the number of times adaptive sampling will be done and the number of samples taken each time (dictated by, for

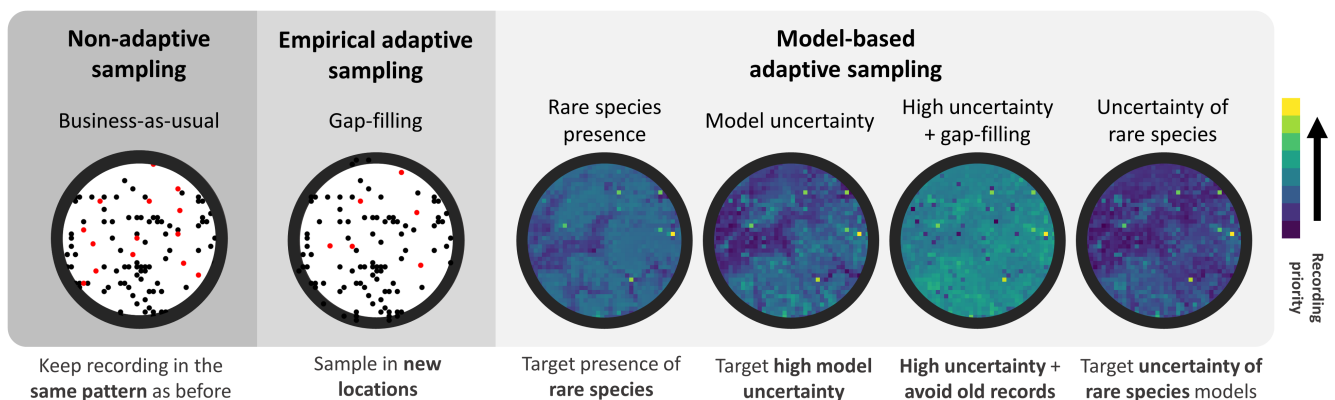


FIGURE 2 The empirical and model-based adaptive sampling methods evaluated in the simulation. From left to right, these are: Business-as-usual (non-adaptive), Gap-filling (empirical adaptive sampling), target areas where rare species are likely to occur, target areas of high model uncertainty, target gaps that have high model uncertainty across all species, and target areas with high model uncertainty of rare species (all of the last four—model-based). Filled circles in the left two images show old (black circles) and new (red circles) records. The four images on the right show the probability of each cell being sampled for each of the adaptive sampling metrics. For the adaptive sampling replicates, 2000 cells were sampled, with 1%, 10% or 50% according to each adaptive sampling method and the remainder as business-as-usual.

example, funding or species life-cycle characteristics). We simulated a single round of adaptive sampling for simplicity and due to computational limitations. For each of the sampling methods, we sampled 2000 cells in a single sampling effort. For the 'business-as-usual' method, we sampled 2000 cells as outlined in the 'baseline sampling of virtual species' section above. For each of the five adaptive sampling methods we sampled a set proportion of the 2000 cells (1% [20 cells], 10% [200 cells] or 50% [1000 cells] of the total) according to the adaptive sampling rules, and the remainder according to the baseline recording pattern (i.e. 'business-as-usual'). This was done to explore the importance of the uptake rate of adaptive sampling. The 2000 cells were selected independently from the previous 20,000 cells, so each new cell could have been in a location that was or was not previously sampled.

2.5.1 | Business-as-usual

This method samples additional locations probabilistically, according to the baseline recording pattern (i.e. samples are more likely to be made in areas with large numbers of butterfly records, see 'baseline sampling of virtual species' above). The probability of each cell being sampled was defined as p_{base} above, except that N_{base} was replaced with the number of newly sampled cells N_{adapt} (set to 2000). This method was used as a comparison ('control') for the other sampling methods described below.

2.5.2 | Gap-filling

This empirical adaptive method samples additional locations only in 1 km grid cells with no previous records, as is done in many existing adaptive strategies in citizen science. For each 1 km square we define Y_i as the number of observations in each cell obtained during baseline sampling and then sample N_{adapt} new locations with equal probability ($p_{\text{gap},i}$) where Y_i is zero.

$$p_{\text{gap},i} | Y_i > 0 = 0, \quad (2)$$

$$p_{\text{gap},i} | Y_i = 0 = \frac{(1/N_{Y_i=0})}{N_{\text{adapt}}}. \quad (3)$$

2.5.3 | Target rare species presence ('rare species')

In this model-based method, sampling locations are chosen based only on the predicted probability of occurrence of rare species, so that areas with high predicted probabilities are prioritised (Chiffard et al., 2020). While this approach might be useful for finding rare species (e.g. Pacifici et al., 2016), there is a danger that it might reinforce existing spatial biases in data by targeting effort only in well recorded locations. We defined prevalence for species j (P_j) as the proportion of all grid cells in which a species occurred (where Z_{ij} indicates the presence of species j in cell i), and defined rarer species

as those with lower prevalence. There were similar numbers of rare species in each assemblage (Figure S2).

$$P_j = \frac{\sum_i Z_{ij}}{N}. \quad (4)$$

We multiplied the predicted occurrence probability in each grid cell for each species $\widehat{p}_{\text{occ},i,j}$ by $1 - P_j$ to upweight the contribution of rarer species. To obtain an assemblage level adaptive sampling layer for each cell i we then calculated the mean across all species of the rarity-weighted occurrence probability layers so that each cell obtained a weight $W_{\text{rare},i}$ defined as below:

$$W_{\text{rare},i} = \frac{\sum_j (\widehat{p}_{\text{occ},i,j} * (1 - P_j))}{J}. \quad (5)$$

2.5.4 | Target uncertainty only ('uncertainty only')

In this model-based method, we chose new sampling locations based on the uncertainty derived from the SDMs, $\widehat{\text{unc}}_{i,j}$. This approach should improve model performance the most, as it directly considers where the model is most uncertain. However, it could lead to the selection of locations which are less appealing to recorders due to remoteness or low species richness (Mair & Ruete, 2016). We averaged uncertainty metrics across all 50 species to create a single adaptive sampling layer $W_{\text{unc},i}$. The resulting layer therefore represents locations with the highest average uncertainty across all species in the virtual assemblage

$$W_{\text{unc},i} = \frac{\sum_j (\widehat{\text{unc}}_{i,j})}{J}. \quad (6)$$

2.5.5 | Target high uncertainty while gap-filling ('gap-filling with uncertainty')

In this method, new cells are chosen based on having high model uncertainty and a low number of records from baseline sampling. This approach therefore combines both model-based and empirical adaptive methods. The problem with only targeting uncertainty is that future sampling locations could be chosen in areas that have previously been visited, which may make recorders feel like their previous records are not valued. We attempted to overcome this, so to down-weight grid cells according to the number of existing records, we multiplied uncertainty by $1/Y_i$, so that

$$W_{\text{ugap},i} | Y_i > 0 = \frac{\sum_j \widehat{\text{unc}}_{i,j}}{J} \cdot \frac{1}{Y_i}, \quad (7)$$

$$W_{\text{ugap},i} | Y_i = 0 = \frac{\sum_j \widehat{\text{unc}}_{i,j}}{J}. \quad (8)$$

2.5.6 | Target uncertainty of rare species models ('uncertainty of rare species')

This model-based method targets sampling where the uncertainty of SDMs of rare species models is highest. Optimising elements of both

uncertainty and prevalence of rare species might allow us to improve model performance while retaining user engagement by targeting locations that are desirable to visit, based on the chance of seeing rarer species. Combining elements of both uncertainty and prevalence may upweight areas containing rare species but could also reinforce existing biases. To calculate a score $W_{urare,i}$, we multiplied the rarity-weighted occurrence probability of each species by the uncertainty layer for each species. We then took the mean across all species in the assemblage

$$W_{urare,i} = \frac{\sum_j (\widehat{unc}_{i,j} * \widehat{p}_{occ,i,j} * (1 - P_j))}{J} \quad (9)$$

For each of the model-based adaptive sampling methods described above the derived weight layers were converted into probabilities of sampling based on N_{adapt} new locations:

$$p_{adapt,i} = \frac{(W_i / \sum W_i)}{N_{adapt}} \quad (10)$$

2.6 | Variation in uptake

We recognise that in citizen science, the level of uptake of adaptive sampling will vary. As stated above, we varied the proportion of the 2000 cells that were sampled according to the adaptive sampling rules. We defined a new variable called uptake (U) which is set between 0 and 1 and determines the relative influence of the adaptive sampling methods. Uptake is assumed to be constant across cells and across species. For the gap-filling method uptake is implemented by altering the proportion of N_{adapt} selected using gap-filling; when U is 0.1 then 10% of cells are selected using gap-filling and the remaining 90% are selected using business-as-usual.

For the model-based methods U modifies the weight calculated for the corresponding method (W_i). To do so we calculate a modified weight $W_{uptake,i}$ so that as U increases to 1 the relative influence of adaptive sampling over business-as-usual increases:

$$W_{uptake,i} = W_i U + D_{+1,i} (1 - U) \quad (11)$$

$W_{uptake,i}$ then replaces W_i in Equation (10) to calculate $p_{adapt,i}$.

We consider three values of U ; 0.01, 0.1 and 0.5, which simulate a small (0.01 = 1%) to very strong (0.5 = 50%) influence of adaptive sampling (Xue et al., 2016). This means that with the amount of uptake at 50%, new locations are equally influenced by the adaptive sampling method and the existing pattern of recording. Our uptake parameter should be thought of as a measure of overall recorder behaviour, rather than simulating the actual number of visitors to suggested locations.

2.7 | Modelling new data

Once the 2000 new locations were selected, we used the binary occurrence maps for each species to identify all the species occurring in the new locations, as done for baseline sampling. A list of new observations was generated at each location by applying the

same detection probability of 0.2. We then ran SDM ensembles, as described above, of the combined existing and new data to update our predicted distributions of each species.

2.8 | Evaluation

We averaged the probability of presence predictions across each of the SDM types (GLM, GAM, RF) to get an average prediction for each species. Hereafter, we only present the predictions averaged across the three models and not each model type separately because they did not substantially differ from one another; we refer to them as 'model(s)' for simplicity. We evaluated the model predictions against the true distributions of each species using three traditional metrics of model performance; the AUC, the mean square error (MSE) and the correlation between true and predicted occurrence (correlation). We used AUC because it is a commonly reported metric in studies of species distributions and is a useful measure of the relative predictive accuracy of the model. MSE provides an absolute measure of each model's predictive performance against the true species distributions, and correlation provides a relative measure. In total, we ran three models per species, and there were 50 species in each of the 50 assemblages (3 models \times 50 species \times 50 assemblages = 7500 models). These models were run two separate times, once for the baseline modelling and once after sampling (2 \times 7500 models). To obtain the changes in model performance caused by additional sampling, we subtracted the values of the three evaluation metrics after additional sampling from those of the baseline models. We only present the changes in MSE (delta MSE) in the results because results were very similar across model performance metrics (see supplementary material for other metrics).

First, we investigated our adaptive sampling metrics at the assemblage level because sampling was conducted across all species in an assemblage (assemblage-level adaptive sampling). To calculate changes in assemblage-level model performance caused by additional sampling, we averaged the changes in the value of each evaluation metric across all species in an assemblage. This resulted in 50 values of change in model performance for each metric, one value for each assemblage averaged across all species in the given assemblage. We also reported the number of species models improving in MSE by at least 1%, to investigate the influence of adaptive sampling on individual species.

Second, we assessed the effect of the prevalence of individual species (P_j) on the benefit of adaptive sampling (i.e. not at the assemblage level). To do this, we split the continuous variable prevalence into deciles with approximately equal numbers of species in each (median = 237 species, 50 species \times 50 communities / 10 deciles = 250; models could not be run for the rarest species because of lack of data). Within each decile, we investigated the proportion of models whose MSE improved or deteriorated by 0%–5% and the proportion that improved or deteriorated by $\geq 5\%$.

Third, we were interested in understanding the mechanism through which adaptive sampling affected model performance. To

do this, for each of the sampling methods we extracted the number of observations that were made in new locations. We did this for each species within each assemblage, allowing us to investigate the effect of prevalence on the number of new observations of a species. We used this to calculate the proportion of each species' range that was sampled by visiting new locations for each of the sampling methods. This was done by dividing the number of new locations in which observations of a species were made (i.e. new grid cells visited and a species detected) by the total number of locations in which a species was present (i.e. the total number of grid cells in which a species was present across all of GB). Because multiple species could be observed at each sampling location, we also extracted the number of new locations sampled across all species in each assemblage. This provided information about the number of new locations in which observations were made across an assemblage.

3 | RESULTS

3.1 | Empirical versus model-based adaptive sampling

We found that adaptive sampling (of any kind) benefitted the performance of SDMs. Importantly, all of our model-based adaptive sampling had benefits even at low levels (1%) of uptake (Figure 3). Increasing the uptake of adaptively sampled locations benefitted model performance, but this effect was not linear. For example, for

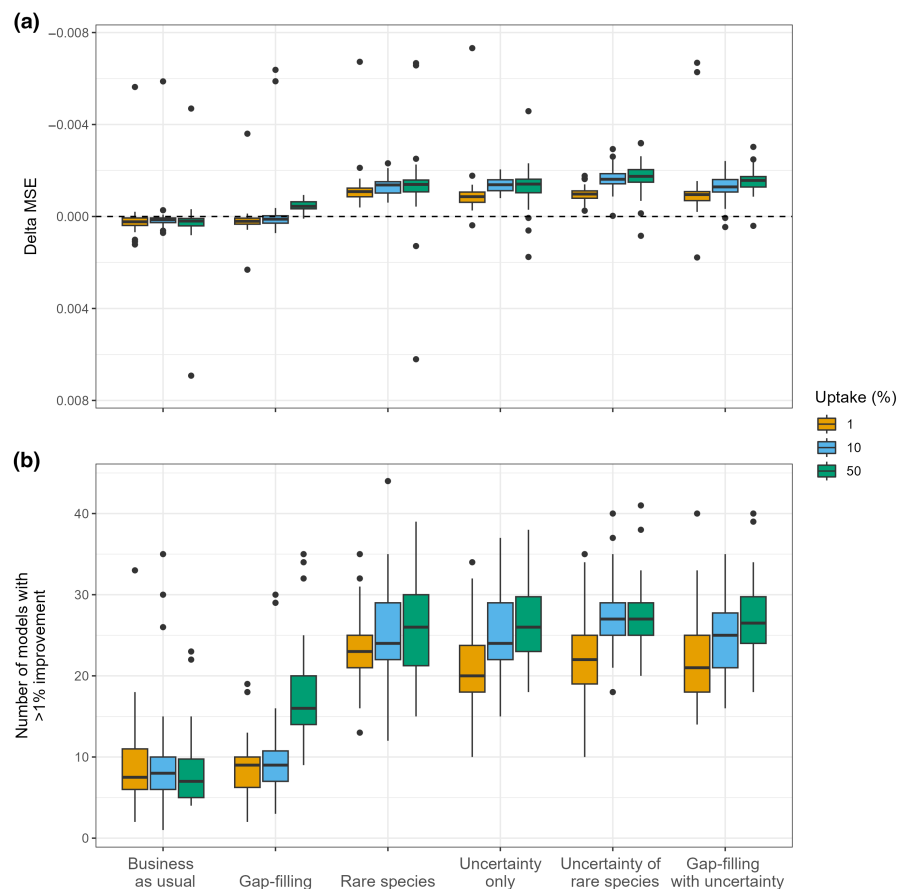
most of the model-based adaptive sampling methods, there was much more of a difference in model performance between no adaptive sampling and 1% uptake than the difference between 1% and 10% uptake (Figure 3a). The empirical sampling 'gap-filling' method (randomly selecting gaps) showed improvements in model performance only with large levels of uptake. Furthermore, even with the maximum uptake value we considered (50% uptake), model improvement of the empirical gap-filling was always smaller than any of the model-based adaptive sampling methods (Figure 3a). This general pattern was also observed when investigating the number of species models that improved by over 1% (Figure 3b).

While on average, models improved after sampling new data, we found that some models decreased in performance. This meant that extra sampling in these assemblages was detrimental to average model performance (Figure 3a). Interestingly, our results show that the business-as-usual method (i.e. continuing recording according to the current pattern) led to MSE becoming worse, on average (Figure 3a), despite increased sample size from additional sampling. However, this was not reflected in the other evaluation metrics, in which the business-as-usual method resulted in marginal improvements in model performance (Figures S4 and S5).

3.2 | The effect of species prevalence

We found that the models of prevalent species were more likely to improve than those of rare species under all forms of sampling

FIGURE 3 The effect of six sampling methods and the proportion of uptake on assemblage-level model improvements in MSE across 50 simulations of virtual recorders and virtual species assemblages. Plot (a) shows differences in MSE between model predictions before and after additional sampling and true species distributions, averaged across all species in an assemblage, for different levels of uptake. Note that the y-axis is reversed, so values above the dashed line, that is more negative, are better. Values falling more than two standard deviations away from the mean have been removed for clarity. Plot (b) shows the number of species in each assemblage whose models had a greater than 1% improvement in MSE for different levels of uptake. There are 50 data points in each box and whisker for both (a) and (b), one for each assemblage. For plots containing outliers and other evaluation metrics (AUC and correlation) see Figures S3–S5.



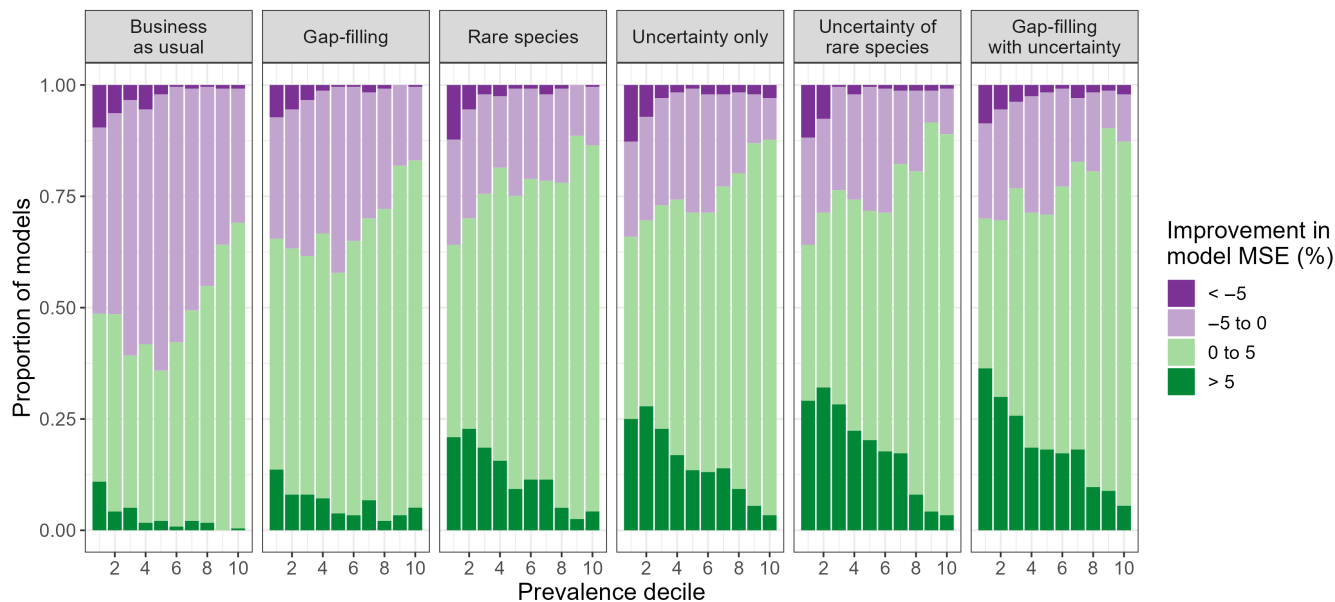


FIGURE 4 The proportion of models whose MSE improved or became worse by different percentages across prevalence deciles and sampling methods for the 50% uptake value. Prevalence deciles (going from 1: the rarest 10% of species to 10: the most prevalent 10% of species) were created to contain approximately the same number of species in each prevalence decile (median = 237 species) and prevalence ranges between 0.001 and 0.65. See [Figures S6–S8](#) for the changes in model performance as determined by AUC and correlation, and for all levels of uptake.

but particularly for the model-based adaptive sampling methods ([Figure 4](#)). However, the largest improvements in model performance were seen for those of rare species ([Figure 4](#); [Figures S6–S8](#)). These benefits were greatest when using the model-based adaptive sampling methods, with a much higher proportion of models improving by over 5% in these compared with either the business-as-usual or empirical gap-filling methods ([Figure 4](#)). Interestingly, the pattern seen in [Figure 3a](#), that overall model performance decreased in the business-as-usual method, appears mostly caused by the models of rarer species becoming worse ([Figure 4](#)). In fact, across all sampling methods and prevalence, some models became worse (i.e. improvement in MSE < 0%); approximately 50% of models in the business-as-usual method but closer to 25% in the model-based adaptive sampling methods. The proportion of models becoming worse (improvements in MSE < 0%) was lower for commoner species than rarer species, but the proportion of models showing substantial improvements (improvement in MSE > 5%) was greater for the rarer species. This suggests that, despite the variability, adaptive sampling is particularly beneficial for the models of rarer species ([Figure 4](#)).

3.3 | Exploring the mechanism influencing model performance

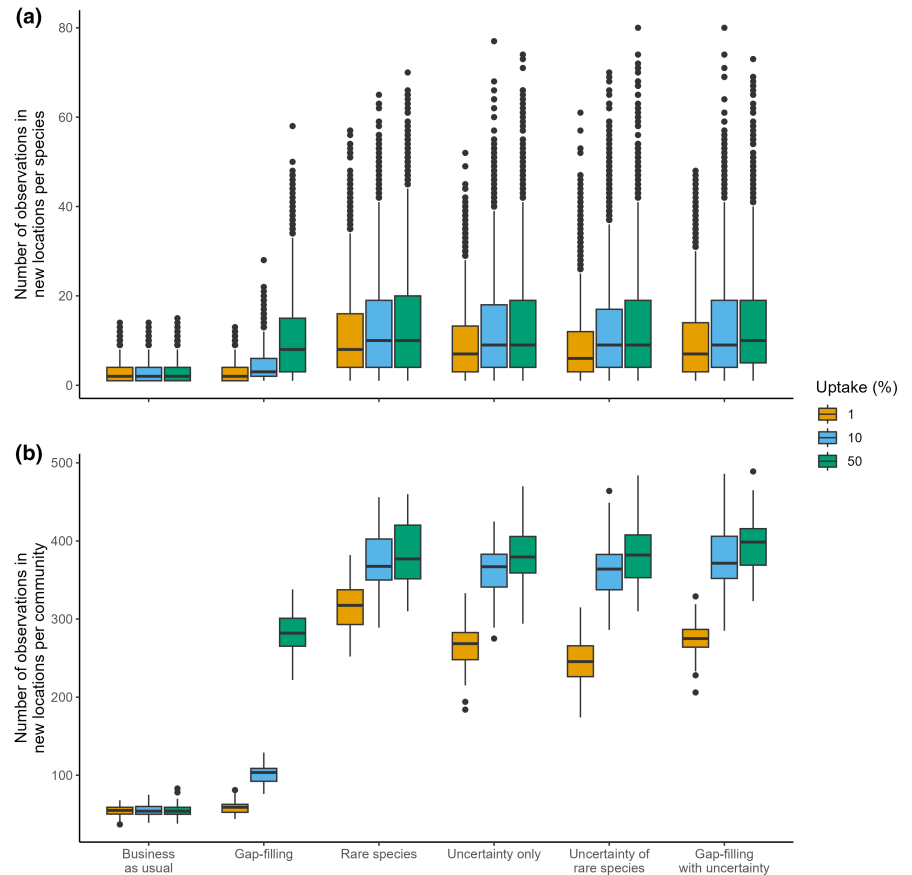
In order to investigate the mechanisms that might have driven the improvements in model performance due to adaptive sampling, we extracted the number and location of samples chosen by the various sampling methods. We found that each species was observed in

more new locations under the adaptive sampling methods than the business-as-usual method. For some individual species, this resulted in as many as 80 new records being made ([Figure 5a](#)). Furthermore, we found that in some assemblages almost 500 new locations were visited in total under adaptive sampling, compared with less than 100 new locations under the business-as-usual sampling method ([Figure 5b](#)). More observations from new locations were made after model-based adaptive sampling than empirical gap-filling. In the latter, only the highest value of uptake matched the model-based methods in terms of total numbers of observations. These records also increased the proportion of each species' range that had been sampled ([Figure S9](#)). This benefited the rarest species in each assemblage the most, with a greater proportion of their range being sampled by new the observations than those of more prevalent species ([Figure S10](#)).

4 | DISCUSSION

Our results show that adaptive sampling improved the performance of SDMs, even at low levels of uptake. This supports the proposition that optimising data collection by citizen scientists could be a powerful mechanism to maximise the potential of these datasets (Callaghan et al., 2023; Callaghan, Poore, et al., 2019; Callaghan, Rowley, et al., 2019; Kays et al., 2021; Xue et al., 2016). Furthermore, we showed that model-based adaptive sampling improved SDM performance more than a simple gap-filling (empirical) sampling approach. This demonstrates the value of optimising sampling based on models and not simply data gaps. Several other

FIGURE 5 The influence of the six sampling methods on (a) the number of times each species was seen at new locations and (b) the total number of observations from new locations across all species in each assemblage, for the three different uptake values. In (a), each data point is a single species (maximum number of data points per box and whisker, 50 species \times 50 assemblages = 2500 species). In (b), each data point represents an assemblage (50 assemblages per box and whisker).



studies have also found that model-based adaptive sampling designs can have important benefits for model performance (e.g. Camp et al., 2020; Flint et al., 2023; Shanahan et al., 2021, but see Bird et al., 2022). Interestingly, we found that the type of model-based method had little effect on model performance: provided some form of model-based sampling was used, the benefits to SDM performance were similar. It is very promising that our results suggest that even small levels of adaptive sampling (just 1% of new recording visits) can improve model performance, especially given the growing demand for citizen science derived modelling outputs to be used in policy-making decisions (Callaghan & Gawlik, 2015; Weiskopf et al., 2022).

Some previous studies have suggested that adaptive sampling is only beneficial in specific scenarios. For example, Pacifici et al. (2016) found adaptive cluster sampling to be an improvement over random sampling only when detection probability was low (Bird et al., 2022; Camp et al., 2020). The contrast with our results, where adaptive sampling was generally beneficial, might be explained by the non-random and spatially biased nature of the citizen science datasets on which our simulations were based. The gains from adaptive sampling may be smaller when initial data are randomly sampled or when compared against the data obtained by structured or random sampling designs. However, it is likely that newly implemented adaptive sampling schemes will use existing data, which could be biased in various ways (e.g. towards certain species or locations). Continuing to record wildlife without correcting for such biases could reduce the

suitability of data for answering ecological questions. Our simulation used real patterns of butterfly recording in the UK to generate initial data, which had a strong spatial bias towards well-recorded locations as do many ecological datasets (Isaac & Pocock, 2015). Our results suggest that adaptive sampling may be particularly beneficial in citizen science, where redirecting effort to new locations could help to optimally address the biases generated by non-random sampling.

Our results support the idea that redistributing recording activity into new areas could benefit SDM performance, particularly for rare species (Johnston et al., 2022). Our simulated adaptive sampling resulted in more observations from new locations, which increased the proportion of each species' range that was sampled. This proportion increase in range coverage was greater for rare species than common species, likely driving greater model improvements in the former. However, current recording patterns may be biased towards places with high species diversity and rare species (August et al., 2020; Isaac & Pocock, 2015). This means that datasets in well recorded countries or regions may already sample most of the true range of even the rarest species. In this case, the benefits from adaptive sampling that we found may be exaggerated compared with their real-world benefit. Additionally, we found that rarer species showed greater variability, with a higher proportion of the models of rare species becoming worse (as well as a higher proportion also becoming better) through sampling compared to those of common species. Indeed, it is unlikely that a single approach to adaptive sampling is suitable for all species (Specht et al., 2017; Turk & Borkowski, 2005).

Adaptive schemes therefore need to be specifically designed for the questions of interest. For example, assemblage-level adaptive sampling, as we implemented, may not effectively capture all species within large assemblages. In these, different adaptive sampling methods could be considered, perhaps focussing on specific subsets of species. While our findings do suggest that model-based adaptive sampling improves model-based inference from citizen science data, more work is needed to determine its influence in a range of recording scenarios, for assemblages of different sizes and for species with differing distributions and traits. Our study provides a useful framework for testing and implementing such designs.

The engagement of citizen science recorders in adaptive sampling schemes is likely to be an important determinant of its effectiveness (Callaghan et al., 2023) and will impact spatial coverage and representativeness (Pocock et al., 2023). Fortunately, studies have found that recorder motivations often align with the goal of sampling schemes (Thompson et al., 2023), suggesting that citizen scientist-based adaptive sampling could be a very powerful tool (Callaghan et al., 2023; Thompson et al., 2023). In our simulations, SDMs using citizen science datasets would benefit even if only 1% of data were sampled adaptively. The improvements associated with increased engagement were also nonlinear; the incremental benefit to models decreased as uptake increased. Xue et al. (2016) and Kays et al. (2021) also found that models improved with only moderate amounts of engagement. The amount of engagement in citizen science has been shown to vary according to skill, experience and demographic characteristics (Isaac & Pocock, 2015; Johnston et al., 2022; Rotman et al., 2012; West et al., 2021). An adaptive scheme that does not consider recorder preferences could reduce recorder engagement and risk decreasing the number or information value of observations. One solution could be to implement mixed adaptive designs, which combine elements of both empirical (i.e. data-based) and model-based methods. These could be relatively simple, such as our uncertainty with gap-filling method. More complex methods could tailor recommended sampling locations to the preferences of individual recorders. Alternatively, professional contractors could be used to supplement data in unappealing and data-deficient locations. Developing adaptive sampling methods that appeal to recorders with a diverse range of motivations and skills is an exciting opportunity for further investigation. Regardless, we found that if adaptive sampling is intelligently applied (i.e. model-based) then even only small levels of uptake have positive impact, thus supporting its use in ecological citizen science.

4.1 | Future research directions

Our probabilistic approach to adaptive sampling, i.e. locations for sampling are drawn with a probability proportional to the predicted utility of each location, meant that slight shifts in the underlying probability layer could cause relatively large changes in sampling distribution. In reality, citizen science datasets are biased, with records close to the homes of recorders, from a few particular 'hotspots'

and are affected by land access rights and infrastructure (Bowler, Callaghan, et al., 2022; Isaac et al., 2014; Mair & Ruete, 2016). These are likely to limit the benefit of adaptive designs as they restrict the potential movements of recorders. Further simulation work could focus on the effect of these limitations on the benefit afforded by adaptive sampling schemes compared to traditional survey methods. Our study also considered changes in overall recorder behaviour, but uptake could also be influenced by recorder identity and their capacity and willingness to change behaviour. This should be investigated further in both empirical studies of volunteer recorders and simulation studies.

Our work highlights the need to understand the influence of spatial structure in citizen science data and modelling frameworks. Spatially explicit SDMs are not often employed in ecological literature despite some evidence that these may improve model performance (Domisch et al., 2019; Hao et al., 2019). Without spatially explicit model terms, adaptive sampling methods target gaps in environmental, rather than geographic, space. Including spatial terms in models would likely change sampling priority maps, as they would account for the current distribution of records. While we did not investigate the effect of spatially explicit models on the adaptive sampling layers, we did find that model-based adaptive sampling resulted in more new locations being visited than with empirical methods. In general, uncertainty is likely to increase as distance from current sampling locations also increases. Therefore, including spatial components in models could result in regions of high priority being identified even further from current recording locations than we found with our current models. This is interesting considering that the current true recording pattern across GB is correlated with species diversity and rarity which means that diversifying recording locations might not lead to improved sampling of species' ranges (Isaac & Pocock, 2015). It might therefore be helpful to run simulations which account for the current distributions of species across the area of interest when considering adaptive sampling schemes.

While variations of adaptive sampling designs have been implemented in ecology for decades, such as recording in unvisited grid squares for atlas mapping, there are still significant barriers to their mainstream use. For example, many schemes try to maintain the same sampling design because they are interested in trends through time, as well as increasing spatial coverage of their data. We, and others, have shown that adaptive sampling can work for improving estimates of species' distributions (Callaghan et al., 2023; Camp et al., 2020; Kays et al., 2021; Shanahan et al., 2021) and trend detection (Callaghan, Poore, et al., 2019; Lindenmayer & Likens, 2009). However, further work needs to consider trade-offs between the intended and unintended outcomes of adaptive designs. For example, intentionally changing the distribution of site coverage could impact occupancy estimates but not trend estimates (Pocock et al., 2023). Determining the influence of adaptive sampling for estimating species' distributions on the ability to detect trends, and vice versa, is likely to be key for its mainstream use.

More work is needed to test adaptive sampling in real world contexts. For example, the magnitude of improvements to models needed to make real world changes will depend on the objectives of interest. Our simulations assumed a constant value of detectability (0.2). However, in reality, detectability is likely to be species-specific and to change spatially and throughout the year, particularly as volunteers differ in their recording expertise (August et al., 2020; Isaac & Pocock, 2015). Adaptive sampling may be more effective for more easily detectable species than cryptic species; the latter could be missed in areas which were identified as important by the sampling method. At the assemblage level, variation in detectability may preferentially benefit highly detectable species, which may not be the goal of the sampling method. This might mean that variability in detection would lead to smaller gains in performance across the whole assemblage, suggesting further work is required to understand its impact. Sampling methods could also be informed by the detectability of the species of interest if this can be estimated. For example, adaptive sampling could be targeted to areas that have a high probability of containing species with low detectability.

5 | CONCLUSIONS

Our study highlights the large potential benefit of adaptive sampling for improving unstructured citizen science datasets. We showed that even small amounts of uptake using model-derived adaptive sampling metrics have the potential to dramatically improve model performance in simulated assemblages. Given the increased use of citizen science datasets for SDM (Feldman et al., 2021), there is ample opportunity to develop metrics to improve the quality of data being collected. Jansen et al. (2022) highlight the importance of presenting uncertainty maps as a key part of SDM outputs. We suggest that these maps could also be used to identify optimal sampling locations. More work is needed to determine the barriers to the implementation of adaptive sampling. To do this, it is key that future simulation work engages with recorders, to implement and assess real-world adaptive recording activities. Such projects could be very successful given the desire of recorders to help with conservation focused wildlife recording (Callaghan et al., 2023; Thompson et al., 2023). Adaptive sampling clearly has a large amount of potential for improving citizen science datasets (Callaghan et al., 2023; Callaghan, Rowley, et al., 2019; Kays et al., 2021) and more work is needed to determine whether and how it can be exploited to address existing biases (Callaghan, Poore, et al., 2019; Isaac & Pocock, 2015).

AUTHOR CONTRIBUTIONS

Thomas Mondain-Monval, Tom August, Michael Pocock, Emma Wright and Susan Jarvis conceived the ideas for the paper. Thomas Mondain-Monval and Susan Jarvis designed methodology, ran the simulations and analysed the data. Thomas Mondain-Monval and Susan Jarvis led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank all the volunteer recorders who have contributed data to the Butterflies for the New Millennium database, and to Richard Fox and Butterfly Conservation for providing the data which were used to inform the simulated assemblages and sampling. We would like to thank the entire DECIDE team for the helpful discussions surrounding adaptive sampling and citizen science. Thanks to Diana Bowler who commented on an earlier version of the manuscript. This work was part of the DECIDE project (Delivering Enhanced Biodiversity Information with Adaptive Citizen Science and Intelligent Digital Engagements) funded by Natural Environment Research Council (NE/V003054/1) and supported by the Natural Environment Research Council award np. NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14355>.

DATA AVAILABILITY STATEMENT

All code is available on GitHub: https://github.com/BiologicalRecordsCentre/DECIDEWP1_simulation.git, archived here: <https://zenodo.org/doi/10.5281/zenodo.11082192> (Mondain-Monval et al., 2024a). The environmental layers used to simulate the communities are freely available online. The butterfly sampling effort layer contains sensitive distribution data and as such is withheld but species record data can be requested from Butterfly Conservation. Any 1 km resolution raster layer will work as a replacement; we suggest the 'urban' and/or 'suburban' layers in the Land Cover Map 2015. All evaluation files are available here <https://zenodo.org/records/10609592> (Mondain-Monval et al., 2024b). For more information please contact the corresponding author.

ORCID

Thomas Mondain-Monval  <https://orcid.org/0000-0002-6770-2002>

Michael Pocock  <https://orcid.org/0000-0003-4375-0445>

Tom August  <https://orcid.org/0000-0003-1116-3385>

Susan Jarvis  <https://orcid.org/0000-0001-5382-5135>

REFERENCES

- Andersson, T., Bruinsma, W., Markou, S., Requeima, J., Coca-Castro, A., Vaughan, A., Ellis, A. L., Lazzara, M. A., Jones, D., Hosking, S., & Turner, R. (2023). Environmental sensor placement with convolutional Gaussian neural processes. *Environmental Data Science*, 2, E32. <https://doi.org/10.1017/eds.2023.22>
- August, T., Fox, R., Roy, D. B., & Pocock, M. J. O. (2020). Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Scientific Reports*, 10, 11009. <https://doi.org/10.1038/s41598-020-67658-3>

- August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T., & Jepson, P. (2015). Emerging technologies for biological recording. *Biological Journal of the Linnean Society*, 115(3), 731–749. <https://doi.org/10.1111/bij.12534>
- Bird, J. P., Terauds, A., Fuller, R. A., Pascoe, P. P., Travers, T. D., McInnes, J. C., Alderman, R., & Shaw, J. D. (2022). Generating unbiased estimates of burrowing seabird populations. *Ecography*, 2022, e06204. <https://doi.org/10.1111/ecog.06204>
- Bowler, D. E., Bhandari, N., Repke, L., Beuthner, C., Callaghan, C. T., Eichenberg, D., Henle, K., Klenke, R., Richter, A., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Decision-making of citizen scientists when recording species observations. *Scientific Reports*, 12, 11069. <https://doi.org/10.1038/s41598-022-15218-2>
- Bowler, D. E., Callaghan, C. T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Temporal trends in the spatial bias of species occurrence records. *Ecography*, 2022, e06219. <https://doi.org/10.1111/ecog.06219>
- Boyd, R. J., Harvey, M., Roy, D. B., Barber, T., Haysom, K. A., Macadam, C. R., Morris, R. K. A., Palmer, C., Palmer, S., Preston, C. D., Taylor, P., Ward, R., Ball, S. G., & Pescott, O. L. (2023). Causal inference and large-scale expert validation shed light on the drivers of SDM accuracy and variance. *Diversity and Distributions*, 29, 774–784. <https://doi.org/10.1111/ddi.13698>
- Callaghan, C. T., & Gawlik, D. E. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *Journal of Field Ornithology*, 86, 298–304. <https://doi.org/10.1111/jof.12121>
- Callaghan, C. T., Poore, A. G. B., Major, R. E., Rowley, J. J. L., & Cornwell, W. K. (2019). Optimizing future biodiversity sampling by citizen scientists. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20191487.
- Callaghan, C. T., Poore, A. G. B., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J. L., Vergés, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, 71(1), 55–63. <https://doi.org/10.1093/biosci/biaa131>
- Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. *PLoS Biology*, 17(6), e3000357. <https://doi.org/10.1371/journal.pbio.3000357>
- Callaghan, C. T., Thompson, M., Woods, A., Poore, A. G. B., Bowler, D. E., Samonte, F., Rowley, J. J. L., Roslan, N., Kingsford, R. T., Cornwell, W. K., & Major, R. E. (2023). Experimental evidence that behavioral nudges in citizen science projects can improve biodiversity data. *BioScience*, 73(4), 302–313. <https://doi.org/10.1093/biosci/biad012>
- Camp, R. J., Miller, D. L., Thomas, L., Buckland, S. T., & Kendall, S. J. (2020). Using density surface models to estimate spatio-temporal changes in population densities and trend. *Ecography*, 43, 1079–1089. <https://doi.org/10.1111/ecog.04859>
- Cardell-Oliver, R., Kranz, M., Smettem, K., & Mayer, K. (2005). A reactive soil moisture sensor network: Design and field evaluation. *International Journal of Distributed Sensor Networks*, 1, 149–162. <https://doi.org/10.1080/15501320590966422>
- Chiffard, J., Marciau, C., Yoccoz, N. G., Mouillot, F., Duchateau, S., Nadeau, I., Fontanilles, P., & Besnard, A. (2020). Adaptive niche-based sampling to improve ability to find rare and elusive species: Simulations and field tests. *Methods of Ecology and Evolution*, 11, 899–909. <https://doi.org/10.1111/2041-210X.13399>
- Domisch, S., Friedrichs, M., Hein, T., Borgwardt, F., Wetzig, A., Jähnig, S. C., & Langhans, S. D. (2019). Spatially explicit species distribution models: A missed opportunity in conservation planning? *Diversity and Distributions*, 25, 758–769. <https://doi.org/10.1111/ddi.12891>
- Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS One*, 16(3), e0234587. <https://doi.org/10.1371/journal.pone.0234587>
- Flint, I., Wu, C.-H., Valavi, R., Chen, W.-J., & Lin, T.-E. (2023). Maximising the informativeness of new records in spatial sampling design. *Methods in Ecology and Evolution*, 15, 178–190. <https://doi.org/10.1111/2041-210X.14260>
- Giouroukis, D., Dadiani, A., Traub, J., Zeuch, S., & Markl, V. (2020). A survey of adaptive sampling and filtering algorithms for the internet of things. In *Proceedings of the 14th ACM International Conference on Distributed and Event-based Systems* (pp. 27–38). Association for Computing Machinery. <https://doi.org/10.1145/3401025.3403777>
- Hao, T., Elith, J., Guillera-Aroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25, 839–852. <https://doi.org/10.1111/ddi.12892>
- Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43, 549–558. <https://doi.org/10.1111/ecog.04890>
- Harris, S. J., Massimino, D., Balmer, D. E., Eaton, M. A., Noble, D. G., Pearce-Higgins, J. W., Woodcock, P., & Gillings, S. (2021). The breeding bird survey 2020. *BTO research report*, 736.
- Hollis, D., McCarthy, M., Kendon, M., Legg, T., & Simpson, I. (2018). *HadUK-grid gridded and regional average climate observations for the UK*. Centre for Environmental Data Analysis. <http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb>
- Hooten, M. B., Wikle, C. K., Sheriff, S. L., & Rushin, J. W. (2009). Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, 20, 639–649. <https://doi.org/10.1111/j.1654-1103.2009.01040.x>
- Isaac, N. J. B., Cruickshanks, K. L., Weddle, A. M., Marcus Rowcliffe, J., Brereton, T. M., Dennis, R. L. H., Shuker, D. M., & Thomas, C. D. (2011). Distance sampling and the challenge of monitoring butterfly populations. *Methods in Ecology and Evolution*, 2, 585–594. <https://doi.org/10.1111/j.2041-210X.2011.00109.x>
- Isaac, N. J. B., & Pocock, M. J. O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, 115(3), 522–531. <https://doi.org/10.1111/bij.12532>
- Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5, 1052–1060. <https://doi.org/10.1111/2041-210X.12254>
- Jain, A., & Chang, E. Y. (2004). Adaptive sampling for sensor networks. In *Proceedings of the 1st international workshop on data management for sensor networks: In conjunction with VLDB 2004 (DMSN '04)* (pp. 10–16). Association for Computing Machinery. <https://doi.org/10.1145/1052199.1052202>
- Jansen, J., Woolley, S. N. C., Dunstan, P. K., Foster, S. D., Hill, N. A., Haward, M., & Johnson, C. R. (2022). Stop ignoring map uncertainty in biodiversity science and conservation policy. *Nature Ecology & Evolution*, 6, 828–829. <https://doi.org/10.1038/s41598-022-01778-z>
- Johnston, A., Matechou, E., & Dennis, E. B. (2022). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, 00, 1–14. <https://doi.org/10.1111/2041-210X.13834>
- Kays, R., Lasky, M., Parson, A. W., Pease, B., & Pacifici, K. (2021). Evaluation of the spatial biases and sample size of a statewide citizen science project. *Citizen Science: Theory and Practice*, 6(1), 34. <https://doi.org/10.5334/cstp.344>
- Lermusiaux, P. F. J. (2007). Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D: Nonlinear Phenomena*, 230(1–2), 172–196.
- Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). virtual-species: An R package to generate virtual species distributions.

- Ecography*, 39, 599–607. <https://doi.org/10.1111/ecog.01388/abstract>
- Lindenmayer, D. B., & Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in ecology & evolution*, 24(9), 482–486.
- Mair, L., & Ruete, A. (2016). Explaining spatial variation in the recording effort of citizen science data across multiple taxa. *PLoS One*, 11(1), e0147796. <https://doi.org/10.1371/journal.pone.0147796>
- Mancini, F., Cooke, R., Woodcock, B. A., Greenop, A., Johnson, A. C., & Isaac, N. J. B. (2023). Invertebrate biodiversity continues to decline in cropland. *Proceedings of the Royal Society B: Biological Sciences*, 290, 20230897. <https://doi.org/10.1098/rspb.2023.0897>
- Mondain-Monval, T., Pocock, M., Rolph, S., August, T., Wright, E., & Jarvis, S. (2024a). v1 adaptive sampling by citizen scientists improves species distribution model performance: A simulation study; code archive. <https://doi.org/10.5281/zenodo.11082192>
- Mondain-Monval, T., Pocock, M., Rolph, S., August, T., Wright, E., & Jarvis, S. (2024b). Evaluation data for: Adaptive sampling by citizen scientists improves species distribution model performance: A simulation study. <https://zenodo.org/records/10609592>
- Pacifici, K., Reich, B. J., Dorazio, R. M., & Conroy, M. J. (2016). Occupancy estimation for rare species using a spatially-adaptive sampling design. *Methods in Ecology and Evolution*, 7, 285–293. <https://doi.org/10.1111/2041-210X.12499>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications*, 19(1), 181–197.
- Pocock, M. J. O., Logie, M., Isaac, N. J. B., Fox, R., & August, T. (2023). The recording behaviour of field-based citizen scientists and its impact on biodiversity trend analysis. *Ecological Indicators*, 151, 110276.
- R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for statistical computing. <https://www.R-project.org>
- Reich, B. J., Pacifici, K., & Stallings, J. W. (2018). Integrating auxiliary data in optimal spatial design for species distribution modeling. *Methods in Ecology and Evolution*, 9, 1626–1637. <https://doi.org/10.1111/2041-210X.13002>
- Riva, F., Gentile, G., Bonelli, S., Acorn, J. H., Denes, F. V., Crosby, A. D., & Nielsen, S. E. (2020). Of detectability and camouflage: Evaluating pollard walk rules using a common, cryptic butterfly. *Ecosphere*, 11(4), e03101. <https://doi.org/10.1002/ecs2.3101>
- Robertson, M. P., Cumming, G. S., & Erasmus, B. F. N. (2010). Getting the most out of atlas data. *Diversity and Distributions*, 16, 363–375. <https://doi.org/10.1111/j.1472-4642.2010.00639.x>
- Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., & Jacobs, D. (2012). Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)* (pp. 217–226). Association for Computing Machinery. <https://doi.org/10.1145/2145204.2145238>
- Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., O'Neil, A. W., & Wood, C. M. (2017). *Land cover map 2015 (25m raster, GB)*. NERC Environmental Information Data Centre. <https://doi.org/10.5285/bb15e200-9349-403c-bda9-b430093807c7>
- Shanahan, E., Wright, W. J., & Irvine, K. M. (2021). Adaptive monitoring in action: Reconsidering design-based estimators reveals underestimation of whitebark pine disease prevalence in the Greater Yellowstone Ecosystem. *Journal of Applied Ecology*, 58, 1079–1089. <https://doi.org/10.1111/1365-2664.13837>
- Specht, H. M., Reich, H. T., Iannarilli, F., Edwards, M. R., Stapleton, S. P., Weegman, M. D., Johnson, M. K., Yohannes, B. J., & Arnold, T. W. (2017). Occupancy surveys with conditional replicates: An alternative sampling design for rare species. *Methods in Ecology and Evolution*, 8, 1725–1734. <https://doi.org/10.1111/2041-210X.12842>
- Thompson, M. M., Moon, K., Woods, A., Rowley, J. J., Poore, A. G., Kingsford, R. T., & Callaghan, C. T. (2023). Citizen science participant motivations and behaviour: Implications for biodiversity data coverage. *Biological Conservation*, 282, 110079.
- Turk, P., & Borkowski, J. J. (2005). A review of adaptive cluster sampling: 1990–2003. *Environmental and Ecological Statistics*, 12, 55–94. <https://doi.org/10.1007/s10651-005-6818-0>
- Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller, B. W., Myers, B. J. E., Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., Fulton, E. A., Harfoot, M., Isbell, F., Johnson, J. A., Mori, A. S., Weng, E., & Rosa, I. (2022). Increasing the uptake of ecological model results in policy decisions to improve biodiversity outcomes. *Environmental Modelling & Software*, 149, 105318. <https://doi.org/10.1016/j.envsoft.2022.105318>
- West, S., Dyke, A., & Pateman, R. (2021). Variations in the motivations of environmental citizen scientists. *Citizen Science: Theory and Practice*, 6(1), 14. <https://doi.org/10.5334/cstp.370>
- Wikle, C. K., & Royle, J. A. (2005). Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics*, 16, 507–522. <https://doi.org/10.1002/env.718>
- Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., & Bower, M. R. (2018). Monitoring dynamic spatio-temporal ecological processes optimally. *Ecology*, 99, 524–535. <https://doi.org/10.1002/ecs2.2120>
- Woodcock, B., Isaac, N., Bullock, J., Roy, D. B., Garthwaite, D. G., Crowe, A., & Pywell, R. F. (2016). Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nature Communications*, 7, 12459. <https://doi.org/10.1038/ncomm12459>
- Xue, Y., Davies, I., Fink, D., Wood, C., & Gomes, C. P. (2016). Avicaching: A two stage game for bias reduction in citizen science. In *Proceedings of the 2016 international conference on autonomous agents & multiagent systems* (pp. 776–785). International Foundation for Autonomous Agents and Multiagent Systems.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119, 622–635. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>

SUPPORTING INFORMATION

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

Figure S1: Rank-abundance curve for butterfly species at a 1 km resolution in Great Britain using the Butterflies for the New Millennium dataset.

Figure S2: Rank-abundance curves for all 50 simulated assemblages at a 1 km resolution.

Figure S3: The effect of six sampling methods and the proportion of uptake on assemblage-level model improvements in MSE including all outliers. The maximum number of species in each assemblage was 50. Note that the y-axis is reversed, so values above the dashed line, that is more negative, are better. There are 50 data points in each box and whisker.

Figure S4: The effect of six sampling methods and the proportion of uptake on assemblage-level model improvements in AUC. Plot (a) shows differences in AUC between model predictions after additional sampling and before additional sampling, averaged across all species in an assemblage, for different levels of uptake. Plot (b)

shows the number of species in each assemblage whose models (out of a maximum of 50 models) had a greater than 1% improvement in AUC for different levels of uptake. There are 50 data points in each box and whisker. Data points above the dashed line for plot (a) are assemblages for which the models improved.

Figure S5: The effect of six sampling methods and the proportion of uptake on assemblage-level model improvements in correlation. Plot (a) shows differences in correlation between model predictions after additional sampling and true species distributions, averaged across all species in an assemblage, for different levels of uptake. Plot (b) shows the number of species in each assemblage whose models (out of a maximum of 50 models) had a greater than 1% improvement in correlation for different levels of uptake. There are 50 data points in each box and whisker. Data points above the dashed line for plot (a) are assemblages for which the models improved.

Figure S6: The proportion of models whose MSE improved or became worse by different percentages across prevalence deciles and sampling methods for the three uptake values. Prevalence deciles (going from 1: the rarest 10% of species to 10: the most prevalent 10% of species) were created to contain approximately the same number of species in each prevalence decile (median = 237 species).

Figure S7: The proportion of models whose AUC improved or became worse by different percentages across prevalence deciles and sampling methods for the three uptake values. Prevalence deciles (going from 1: the rarest 10% of species to 10: the most prevalent 10% of species) were created to contain approximately the same number of species in each prevalence decile (median = 237 species).

Figure S8: The proportion of models whose correlation improved or became worse by different percentages across prevalence deciles and sampling methods for the three uptake values. Prevalence

deciles (going from 1: the rarest 10% of species to 10: the most prevalent 10% of species) were created to contain approximately the same number of species in each prevalence decile (median = 237 species).

Figure S9: Proportion increase in the true range of a species, as defined by the number of grid cells in which a species is present, that is covered by observations for each sampling method and uptake value. In (a) all data are shown, in (b) values over three standard deviations away from the mean are excluded. In (a), each data point is a single species (maximum number of data points per box and whisker, 50 species × 50 assemblages = 2500 species). In (b), each data point represents an assemblage (50 assemblages per box and whisker).

Figure 10: The proportion of the true range of a species (in percentage) sampled by new observations for the different sampling methods and for all three of the uptake values. Prevalence deciles (going from 1: the rarest 10% of species to 10: the most prevalent 10% of species) were created to contain approximately the same number of species in each prevalence decile (median = 237 species).

Table S1: The environmental layers used in the simulation and modelling, their sources and the aggregation methods used to combine them.

How to cite this article: Mondain-Monval, T., Pocock, M., Rolph, S., August, T., Wright, E., & Jarvis, S. (2024). Adaptive sampling by citizen scientists improves species distribution model performance: A simulation study. *Methods in Ecology and Evolution*, 15, 1206–1220. <https://doi.org/10.1111/2041-210X.14355>