Adaptive sampling for ecological monitoring using biased data: A stratum-based approach

Abstract

Indicators of biodiversity change across large extents of geographic, temporal and taxonomic space are frequent products of various types of ecological monitoring and other data collection efforts. Unfortunately, many such indicators are based on data that are highly unlikely to be representative of the intended statistical populations. Where there is full control over sampling processes, individual spatial units within a geographical population have known inclusion probabilities, but these are unknown in the absence of any statistical design. This could be due to the voluntary nature of surveys and/or because of dataset aggregation. In these cases some degree of sampling bias is inevitable and, depending on error tolerance relative to some real-world goal, we may need to ameliorate it. One option is poststratification to adjust for uneven surveying of strata assumed to be important for unbiased estimation. We propose that a similar strategy can be used for the prioritisation of future data collection: that is, an adaptive sampling process focused on increasing representativeness defined in terms of inclusion probabilities. This is easily achieved by monitoring the proportional allocation of sampled units in strata relative to that expected under simple random sampling. The allocation of new units is thus that which reduces the departure from randomness (or, equivalently, that equalising unit inclusion probabilities), allowing an estimator to approach that level of error expected under random sampling. We describe the theory supporting this, and demonstrate its application using sample locations from the UK National Plant Monitoring Scheme, a citizen science monitoring programme with uneven uptake, and data on the true distribution of the plant *Calluna vulgaris*. This *in silico* example demonstrates how the successful application of the method depends on the extent to which proposed strata capture correlations between inclusion probabilities and the response of interest.

Keywords: survey error, survey quality, poststratification, weighting, response propensity, R-indicators, time-trends

Introduction

Ecologists are increasingly concerned with monitoring biodiversity change at a variety of spatial scales. Whilst this has long been an active area of research within conservation and related fields (e.g. Spellerberg [57]), in recent years

*Corresponding author

its importance has increased, with numerous species' time trends and associated multi-species indicators now based on a wide variety of data types [e.g. 40, 26, 14]. One consequence of this trend has been the increasing focus on the use of datasets for monitoring that lack any explicit survey design relative to the scientific question of interest. That is, the data used to estimate species' abundances or occupancies are frequently not a probability sample of the statistical target population. Unfortunately, inference using such nonprobability samples is considerably more difficult than has often been recognised in ecology [10]. The absence of sampling design typically means that model-based adjustments must be made to approach the answer that would have been obtained had sampling actually been probabilistic, and such adjustments can rarely, if ever, be shown to be absolutely reliable [33, 11, 65, 16, 1]. As a result, efforts to characterise biodiversity change from nonprobability samples have often received criticism for not being representative of their inferential target populations (e.g. Gonzalez et al. [20]), leading to a number of high-profile disagreements in the literature [10].

The technical elements of sampling design underlying these issues have been well-known in the statistical subdiscipline of survey sampling for decades [30, 2, 62, 33], yet many of these insights are frequently overlooked or misunderstood by ecologists (although by no means all, e.g. see many chapters within ref. [19]). One stumbling block may be the numerous definitions and types of "bias" available in the literature [19, 43, 49]; the lack of any well-known (to ecologists) unified mathematical definition of sampling bias may also have hindered communication and progress.

Within survey sampling focused on descriptive inference (i.e. characterising some directly measurable property of a population from a sample; [24]), statistical error has long been known to be driven in large part by correlations between the probability that any unit is in the sample π, the inclusion probability, and the property of interest y [e.g. 21, 4]. Note that in survey sampling π is also sometimes designated as the "response propensity", because there the key challenge is unknown probabilistic variation in subject responses to designed surveys, rather than the absence of design itself [30]. In ecology, this has also sometimes been discussed under the heading of preferential sampling [e.g. 1], although that label tends to imply a positive association, whereas the issue applies to correlations of either sign. Probability sampling ensures that this correlation is zero in expectation (i.e. across repeated, normally imaginary, realisations of the sampling mechanism; [33]). A conceptual complication here is that finite probability samples also have non-zero correlations between sample inclusion and the response variable, and that there is variation in the survey sampling literature relative to whether people refer to realised error in a sample as bias (when it may actually be a combination of sampling variance and a biased sampling mechanism), or whether the term sampling bias is reserved for situations where it is known (or strongly expected due to a lack of design) that $E[\rho(\pi, y)] \neq 0$; that is, that the sampling mechanism that produced the data had variable sampling unit inclusion probabilities, which, by definition, were not designed and so cannot be directly accounted for when estimating parameters like means from the data. This means that the expected value ($E[\cdot]$) of the correlation (ρ) between the sample inclusion probabilities π_i and the values of the response variable y_i is guaranteed to be non-zero, something that is only assured by probability sampling [33].

Regardless of these terminological issues, Meng [33] demonstrated how a standard formula for statistical error ($\overline{y}_n - \overline{y}_N$, the difference between the mean of the response variable in the sample and that of the variable in the full population) can be re-written as the product of three terms. One characterising the aforementioned correlation $\rho(\pi, y)$, given the name "data quality" by Meng, and two others representing the population fraction sampled ("data quantity") and the amount of variation in the response variable in the population ("problem difficulty"). (Note, however, that Meng approaches the correlation $\rho(\pi, y)$ from a finite population viewpoint, replacing the latent sampling unit inclusion probabilities π_i with the realised, binary sample inclusion indicators R_i.) The implications of this algebraic identity have been hailed in some areas as a "new paradigm" [3], and in our opinion the formula clarifies many issues that have previously sometimes only been intuitively understood in ecology [11, 7, 10].

The adjustment of nonprobability samples for approaching unbiased inference is one area that has been clarified by Meng's approach: in a subsequent paper, Meng [34] demonstrated how all such techniques (inverse probability weighting and poststratification, imputation or superpopulation modelling, and doubly-robust approaches) can be viewed as ways to minimise the correlation $\rho(\pi, y)$. This insight allows us to understand the assumptions of our methods, and therefore to justify our approaches and assess their limitations more clearly [8]. Here we apply these insights to the use of stratification in ecology, particularly its *post hoc* use to adjust unrepresentative sampling, demonstrating its use as an intelligent driver of adaptive sampling for many situations involving data that are biased for the estimation of some "estimand" (i.e. the real-world quantity of interest [32]).

A priori stratification is often used in survey design to achieve one or more of the following: good representation of a population relative to target variables of interest; to guarantee certain sample sizes within strata (which may be of intrinsic interest); for the convenience of survey administration, potentially including cost reduction via regional administration; and to increase the statistical efficiency of estimators [30, 62]. For the last point, error can be reduced by randomly sampling within strata of homogeneous units, i.e. those where subpopulation means and variances are expected to be similar [30].

Post hoc stratification, or, as it is more commonly known, "poststratification", can also be used to achieve this latter goal. That is, it can be used to increase the precision of estimators under known sampling schemes [56]. However, it can also be used as a way to remove potential biases arising from the use of nonprobability samples. In this sense it is part of the family of reweighting techniques intended to adjust a sample to better represent some population of interest [56, 65, 11].

The poststratification estimator $\overline{y}_{ps}$ [4], or "basic poststratification identity"

[18], used to achieve this can be defined as:

$$\overline{y}_{ps} = \frac{1}{N}\sum_{h=1}^{H} N_h \overline{y}_h \tag{1}$$

where N is the population size (here the total number of spatial units), H is the full set of strata into which the population is divided, N_h is the overall size of stratum h, and $\overline{y}_h$ is the mean within stratum h. The implication of (1) is that within-stratum means substitute for individual unit values, and it is these that are averaged across the entire population once relative stratum sizes in the population have been accounted for (see [11] for a worked ecological example). This formulation implies that all i units within a given poststratum receive the same weight [4, 65], equal to

$$w_{i(h)} = \frac{N_h/N}{n_h/n} \tag{2}$$

where n is the total sample size, and n_h is the size of the sample within stratum h. Equation (2) can be understood as upweighting units that are under-represented in the sample relative to the population and *vice versa*. These weights imply an individual unit inclusion probability of $\pi_{i(h)} = n_h/N_h$. And so it can be shown that

$$\bar{y}_{ps} = \bar{y}_{ipw} = \frac{1}{N}\sum_{h=1}^{H}\sum_{i\in n_h}\frac{y_{i(h)}}{\pi_{i(h)}} \tag{3}$$

[65]. Thus poststratification is a special case of inverse probability weighting (a.k.a. quasirandomization or propensity score weighting) where $\pi_{i(h)}$ is assumed to be constant within strata but to (potentially) vary between strata [65]. In the situation where a set of randomly sampled population units are surveyed with full response (i.e. no "loss" of design-based survey units), then this estimator, whether construed as $\overline{y}_{ps}$ or the inverse probability weighted estimator $\overline{y}_{ipw}$, is unbiased in expectation [56, 4]. However, as noted above, it is well known that in actual samples error will tend to increase as a function of the correlation between between inclusion probabilities π and the outcome variable y [21, 4].

In the case of uncontrolled (i.e. nonprobability) samples, whether based on a single survey such as a designed citizen science scheme with some nonresponse, or an aggregated sample such as one might retrieve from the Global Biodiversity Information Facility (GBIF) or other meta-database, the lack of statistical design control essentially guarantees that this correlation will be appreciably different from zero [10]. This will not merely be the bad luck of an unrepresentative random sample, but the expectation of a biased sampling mechanism; that is, $E[\rho(\pi, y)] \neq 0$. Here, increases in sample size will not help; in fact, they have been shown to make things worse in realistic scenarios, i.e. when $n << N$ and the standard deviation of y, σ_y, does not equal zero, as will generally be the case for most environmental monitoring at small scales [33, 11, 3].

With regards to poststratification, two situations will reduce this undesirable correlation [4]. These rely on the fact that if either of a pair of variables is fixed then they cannot be correlated. These are:

i. The response of interest y_i is invariable within poststrata (i.e. $\sigma_{y(h)} = 0 \quad \forall\, h$).

ii. The inclusion probabilities π_i are invariable within poststrata (i.e. $\pi_{i(h)} = \pi_h \quad \forall\, i \in h$), achieved by simple random sampling (SRS) within strata.

In the first of these situations, the poststratification estimator (1) will be more efficient (lower variance) than the arithmetic mean, and will reduce error wherever a random sampling design has yielded an unbalanced sample by chance [25]. In the second of these situations, the poststratification estimator reduces the bias, but not the variance [28, 27]. This is linked to the assertion of Gelman and Carlin [18] that poststratification is most important when correcting for differential nonresponse *between* poststrata. Assuming that inclusion probabilities are uniform within poststrata, but correlated with y within the overall population, then adjusting for poststratum membership renders $\rho(\pi, y)$ equal to zero [65, 34]: that is, π and y are independent conditional on some X, where here X is the vector of unit poststratum memberships [56].

Whilst poststratification and its variants [e.g. see 17] can be useful tools for adjusting existing samples [11], where monitoring is ongoing and survey organisers have some power to alter data collection, combining adaptive sampling with poststratification may be a more efficient way to reduce error compared to relying on poststratification of unrepresentative samples alone [54, 52]. Larger samples may be also desired for other reasons irrespective of the potential for using the poststratification estimator on a sample in hand (increases in power for example). The situations in which poststratification is likely to assist the sampler given above suggest a simple approach to adaptive sampling for researchers seeking to characterise a population parameter such as a mean. As noted above, such descriptive targets are increasingly important for ecological monitoring and conservation, especially where nonprobability samples are used [10]. Straightforward approaches to adaptive sampling, with few assumptions, are therefore likely to be of wide utility [22].

Here we propose an approach to the problem based on assessments of post-stratum sampling coverage. We show how this can be implemented easily with standard binomial formulae within an adaptive framework using data collected between 2015–2023 for the UK National Plant Monitoring Scheme, a designed citizen science programme with uneven site uptake [45]. Our approach has a direct link to the literature on the monitoring of survey quality via assessments of potential nonresponse bias [63, 38], and we use one such indicator (the R-indicator of Schouten, Shlomo and colleagues [50]) of variation in response propensities (or, as we have styled them here, inclusion probabilities) across strata to explore the potential improvements in survey representativeness (a measure of survey quality [51]) achievable using our approach. Finally, we investigate the performance of the approach through simulation and when confronted with a real dataset. Specifically, we examine how well stratum-based adaptive sampling performs in estimating the true 1 km^2 occupancy of the subshrub *Calluna vulgaris* (L.) Hull in Great Britain. We use both simulated locations and actual sampled sites from the NPMS to explore the potential strengths and weaknesses of the method relative to its key assumptions.

Methods

A stratum-based adaptive survey strategy

The approach proceeds as follows: for the population of interest (e.g. some geographic area over which the mean of some attribute of a population of units is desired), select a set of strata H considered to have some differential relationship with sample inclusion and/or the response variable(s) of interest. Each stratum need not be a single spatially contiguous unit, but each population unit should be assignable to a single stratum (geographical units may often require assigning to the stratum with the largest overlapping area). Many such strata will likely already exist, although the approach is not limited to existing strata, as any set of geographically indexed variables could be discretised and crossed to create strata [e.g. see 11]. For example, in the UK "land classes" have previously been erected based on covariation in numerous geographical and environmental variables [12] and then amalgamated into broader zones [61]; for Europe, biogeographic zones based on patterns of terrestrial and marine biodiversity exist [15]. Note that the strata do not have to be absolutely believed to have an invariable one-to-one relationship between stratum unit membership and inclusion probabilitiy, only that there is some nontrivial relationship, such that adjusting for its contribution to the correlation $\rho(\pi, y)$ will be better than assuming that the sample is equivalent to one selected at random [34].

For the existing sample for which future adaptive selections are required, compare the current distribution of units across strata to that expected for the same sample size under simple random sampling; this is known as proportional allocation in the survey sampling literature [62]. That is, a given set of strata H partitioning N will be sampled in proportion to n/N, such that, for stratum h, $n_h = (n/N) \cdot N_h$; if achieved, all response propensities would be equal, both within and between strata. The stratum for which the next unit should be collected will then be the one with the current largest negative departure from random expectation, quantified using z-statistics.

Monitoring representativeness

The link between inclusion probabilities and indicators of representativeness noted above was formalised by Schouten and colleagues [51]. They provide the following operational definition of "representative" in the survey sampling context:

$$\bar{\pi}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} \pi_{i(h)} = \pi \quad \forall\, h. \tag{4}$$

Equation (4) is a weaker version of statement (ii) given in the Introduction, as it does not state that all unit inclusion probabilities within a stratum are identical, only that the means across strata are equal. Based on this, the Schouten *et al.* R-indicator is $R(\pi) = 1 - 2s_{\overline{\pi}_h}$, where $s_{\overline{\pi}_h}$ is the weighted standard deviation of the mean inclusion probabilities across strata [51]. $R(\pi) = 1$ denotes maximum representativeness when the variance in inclusion probabilities across strata is zero.

Adaptive sampling algorithm

This proceeds as follows (also see the *R* code in Supplementary Material 1):

Step 1: Assign all population units N_i to a unique corresponding stratum h_i.

Step 2: Calculate each stratum's current z-statistic, z_h, by comparing the current empirical count ($\bar{x}_h = N_h \cdot (n_h/N_h) = n_h$, the current sample size) and binomial count standard deviation ($s_h = \sqrt{N_h{\cdot}n_h/N_h{\cdot}(1 - n_h/N_h)}$) to the expected count ($\hat{\mu}_h$) based on proportional allocation (i.e. $n/N{\cdot}N_h$). Then, $z_h = (\bar{x}_h - \hat{\mu}_h)/s_h$, the difference between the empirical and expected counts in standard deviation units.

Step 3: Across the H strata, select that h with the smallest z_h as the stratum most in need of additional sampling to reach the simple random sample benchmark. Call this the focal stratum h_f.

Step 4: Given the addition of a new site to stratum h_f, calculate the new values of $\overline{x}_h$ and s_h directly from the standard binomial formulae. The new target stratum site count expected under simple random sampling is also updated as $\hat{\mu} = (n + a)/N{\cdot}N_h$. In the following examples $a = 1$, but it could be any positive integer as there is no requirement to evaluate the switch after the addition of every single new sampling unit.

Step 5: After updating the current focal stratum h_f with the newly added site(s), recalculate the z-statistics for all strata, including h_f. Compare the updated $z_{h(f)}$ with the minimum z_h across all strata. If $z_{h(f)}$ is no longer the smallest, switch the focus to the stratum with the new smallest $z_{h(f)}$ denoted h_{f^*}. Begin sampling h_{f^*} if required, otherwise continue with h_f.

Step 6: Repeat Steps 2 to 5 K times until the desired new sample size allowed by current resourcing, $n + aK$, is reached, or until all strata are at their expected simple random sampling counts $(n + aK)/N{\cdot}N_h$.

We can monitor the progress of this algorithm by following the empirical stratum sampling proportions (n_h/N_h), and by calculating the corresponding R-indicator at each step.

Investigating performance

Empirical data and initial proof-of-concept

The UK National Plant Monitoring Scheme (NPMS) asks volunteers to record plant abundances in small plots located in particular habitats [64]. Plots are located within 1 km^2 squares (hereafter "sites") of the relevant country grid (the scheme currently covers Great Britain, Northern Ireland, the Isle of Man and the Channel Islands). The available sites within the scheme (see `https://www.npms.org.uk/square-near-me-public`) are originally a weighted-random selection, stratified by 100 × 100 km cells of the larger relevant grid; see [45] for more detail. Due to variable population density and other factors across the region, uptake of these sites is uneven, and some areas have far fewer survey returns than others [45]. A primary aim of the NPMS is the production of nationally representative indicators of habitat quality [41], and so, ideally, coverage of the area would be relatively even. We know that inclusion probability (i.e. site uptake) is related to such factors as human population density and correlated environmental variables such as altitude and land cover type, and that these variables are also correlated with the local abundances and occupancies of plant indicator species and their habitats [45]. North-west to south-east gradients of all these variables are well-known for Britain and Ireland [48, 58, 44, 23, 47]. We therefore assume that representation of broad environmental strata, in tandem with poststratification of results, is likely to be a positive step towards reducing potential bias in monitoring scheme outputs. One widely-used set of strata for Great Britain is the UK Countryside Survey (UKCS) Environmental Zones [61], based on a larger set of "land classes" created originally for the *a priori* stratification of national ecological and biogeographical surveys [12]. To these we add Northern Ireland as an additional stratum to better cover our area (Fig. 1). Surveyed NPMS sites [39] are overlaid on these zones in Figure 1 to show their current overall (2015–2023) coverage. We use these data to demonstrate an initial proof-of-concept, namely that the algorithm equalises stratum sampled proportions and maximises the R-indicator as proposed.

Reducing bias in a response variable of interest

Investigating the likely benefits of our strategy for a response variable of interest, such as a species' occupancy or average abundance, is more challenging, as it requires access to a species' true underlying state to evaluate (or a good estimate of this via a probability-based survey). Whilst pure simulation approaches could be used, we consider that these would be less illuminating than investigations more closely aligned to real-world datasets, because the theoretical principles underlying the approach are already well characterised. We use an approximation of the true 1 km^2 distribution (for 2000–2019) of the heathland subshrub *Calluna vulgaris* (L.) Hull ("Heather"), originally created for Boyd et al. [11]. This "true" distribution is based on the 2018 UKCEH Land Cover Map [37] (where "Heather" and "Heather grassland" are land covers derived from satellite images and other information) and occurrence data from the distribution mapping project *Plant Atlas 2020* [58]. See [11] for more information on the construction of the *Calluna* map.

Adaptive sampling based on simulated locations. First, we demonstrate the performance of the method when the key assumption regarding random sampling within strata is met. Here we only use empirical data from the NPMS [39] dataset to initialise stratum sample sizes for the adaptive algorithm. (Specifically we use data from 2019 for these investigations.) The intial samples themselves are new random selections within strata; the adaptive addition of sites uses our suggested algorithm. We refer to this approach as "Stratum SRS [Simple Random Sampling] + adaptive". The iterative estimates of the mean occupancy

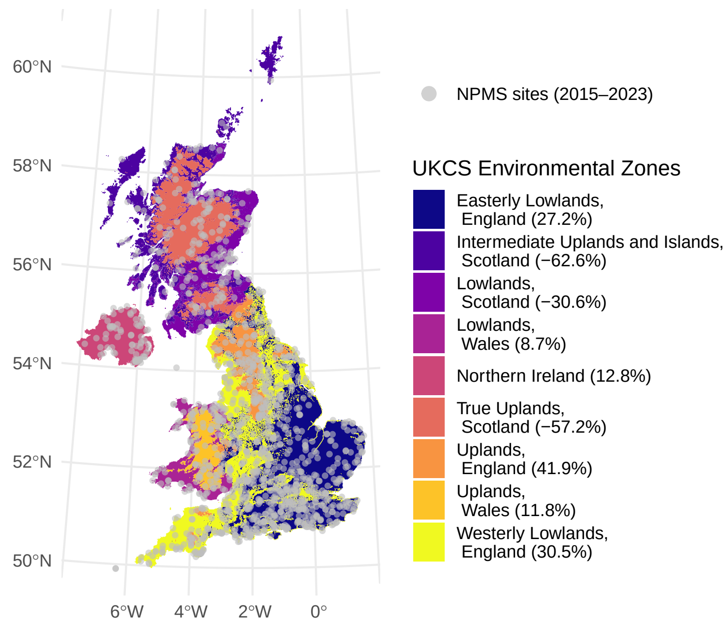

Figure 1: UK Countyside Survey (UKCS) Environmental Zones plus Northern Ireland. The numbers following the UKCS zone names give the difference between the empirical National Plant Monitoring Scheme (NPMS) square count and that expected under simple random sampling (SRS), expressed as a percentage difference (+/-) relative to the expected count. Percentages closer to zero therefore approach SRS counts. Grey circles are surveyed NPMS sites, 2015–2023.

of *Calluna* for this scenario use the poststratification estimator from the *R* package "survey" [31].

Adaptive sampling based on empirical locations. Second, we investigate the performance of the method using the actual sampled sites from 2019 in the NPMS [39] dataset. This approach provides insight into how the method might perform when the key assumption of random sampling within strata is unlikely to be fully met. We refer to this approach as "NPMS + adaptive". Again, iterative estimates of *Calluna* occupancy use the poststratification estimator from the *R* "survey" package [31]. We also include a scenario where our proposed strata are ignored, and new sites added to the existing NPMS 2019 sample are chosen randomly from the total site population of Great Britain. We call this approach "NPMS + SRS". *Calluna* occupancy estimates from this procedure are the simple (i.e. unadjusted) mean rather than the poststratified mean. For all three scenarios new sites added to the sample are labelled as unavailable for future iterations of the algorithm.

Results

Table 1 gives the current distribution of NPMS 1 km^2 sites by UKCS Environmental Zone stratum. These are given in order of their discrepancy from proportional allocation (i.e. simple random sampling) of the 2015–2023 sample of 1,204 sites that could be assigned to strata, from under- to over-sampled [39].

Figure 2 demonstrates the progress of the stratum-based adaptive sampling algorithm in terms of stratum sampled proportions and R-indicator. The example here uses 600 iterations (i.e. the final target sample size was $n+600 = 1804$). This amount of adaptive sampling may be unrealistic in most real world situations where there is existing nonresponse, but we use this number to demonstrate the point at which all strata become proportionally allocated, and to show the evolution of the R-indicator towards its maximum possible value of 1 (Fig. 2). As per Table 1, Figure 2a shows how, initially, only the Intermediate Uplands and Islands and True Uplands of Scotland are underallocated. These have the lowest stratum proportions sampled initially: up to around the 100th iteration it is only these sites that are being selected for new sampling locations. The other strata "flatline" up to this point, indicating that they are over-sampled relative to the number of samples they would expect if the total sample had actually been proportionally allocated. The most over-sampled stratum is the Uplands of England, as this does not see its sample size increased until around the 500th iteration. This is also the point at which the R-indicator (Fig. 2b) approaches its maximum value of 1 and itself flatlines; this indicates that all strata are now being sampled relative to the proportions expected under proportional allocation.

Table 2 gives abridged output of the adaptive sampling algorithm underlying Figure 2. The top of the table shows how, initially, stratum number 5, the "Intermediate Uplands and Islands" zone of Scotland is targeted in isolation (as expected from its position at the top of Table 1). The bottom of Table 2 shows how, once all strata are undersampled relative to the addition of new

Table 1: The current distribution of NPMS sites by UKCS Environmental Zone strata, ordered from under- to over-sampled relative to simple random sampling (SRS). Exp. count is the expected number of squares under SRS. Pct sampled is the current percentage of the stratum area sampled; Count discrepancy is the difference between the actual square count and the expected count expressed as a percentage difference (+/-) relative to the expected count.

Stratum no.	Stratum	No. sites	Exp. count	Stratum area (km^2)	Pct sampled (%)	Count discrepancy (% of expected)
5	Intermediate Uplands and Islands, Scotland	53	142.2	29866	0.18	-62.7
6	True Uplands, Scotland	65	152.5	32034	0.20	-57.4
4	Lowlands, Scotland	76	109.9	23084	0.33	-30.9
7	Northern Ireland	73	67.4	14156	0.52	8.3
8	Lowlands, Wales	60	53.8	11309	0.53	11.4
9	Uplands, Wales	55	48.9	10272	0.54	12.5
1	Easterly Lowlands, England	395	311.6	65441	0.60	26.8
2	Westerly Lowlands, England	321	246.7	51815	0.62	30.1
3	Uplands, England	106	74.9	15739	0.67	41.5

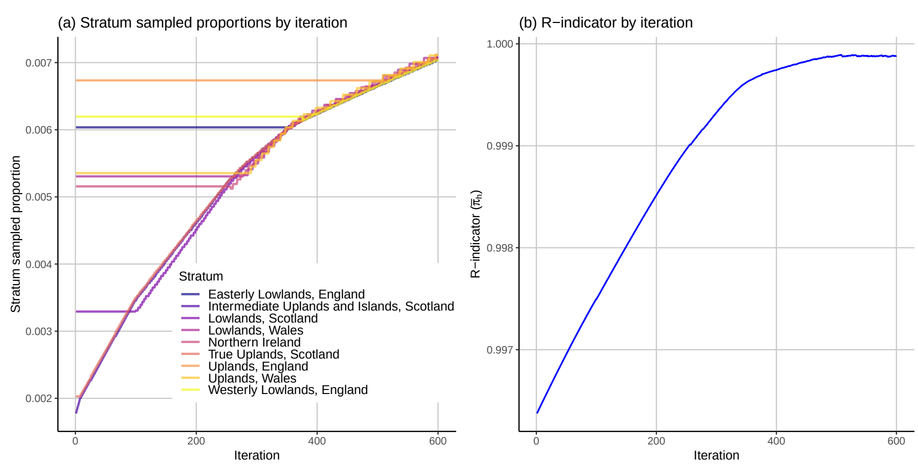

Figure 2: Evolution of UKCS Environmental Zone stratum mean sampled proportions and their R-indicator by iteration.

Table 2: Abridged adaptive sampling output for the first and last six added sites across 600 iterations. Stratum no. = stratum number of focal stratum (see Table 1 for stratum name); Mean prop. = sampled proportion for target stratum; SD = binomial st. dev. for site count within stratum.

Iteration	Stratum no.	z-value	Mean prop.	Site count	SD
1	5	-1.2e+01	1.8e-03	54	7.3
2	5	-1.2e+01	1.8e-03	55	7.4
3	5	-1.2e+01	1.8e-03	56	7.5
4	5	-1.2e+01	1.9e-03	57	7.5
5	5	-1.1e+01	1.9e-03	58	7.6
6	5	-1.1e+01	1.9e-03	59	7.7
...	...	...	...	...	...
595	3	1.1e-01	7.1e-03	111	10.5
596	1	1.0e-01	7.0e-03	460	21.4
597	5	9.9e-02	7.0e-03	214	14.6
598	4	9.1e-02	7.0e-03	164	12.8
599	2	8.7e-02	7.0e-03	368	19.1
600	7	8.5e-02	7.1e-03	102	10.1

sites, the target stratum switches with every iteration of the algorithm. The total population size of UK 1 km^2 sites assigned to UKCS Environmental Zone strata is 257,502; $1804/257502 = 0.0070$, hence the stratum sampled proportions achieved for the final six iterations at the bottom of Table 2 ("Mean prop." column).

Figure 3 shows the results of applying our algorithm to the case of estimating our "true" 1 km^2 occupancy of *Calluna vulgaris* (0.27). The simple (i.e. unadjusted) mean occupancy of the existing 2019 NPMS data for *Calluna* is 0.33. Taken together, the four elements of Figure 3 reveal both the potential strengths and weaknesses of the proposed method in improving on the unadjusted sample mean through the adaptive sampling algorithm. Figure 3a demonstrates how three different data/model scenarios can lead to better estimates of the true mean with increasing sample size. Given that both simple random sampling and stratified random sampling are standard methods in survey sampling, this is not surprising; it is the differences between the strategies investigated that provide useful insights into the likely performance of our approach when applied to real-world datasets. Figure 3a also shows that the initial poststratified estimate (iteration 1) of *Calluna* occupancy using the 2019 NPMS locations ("NPMS + adaptive"; see also Table 3) leads to the most biased estimate (0.41). In addition, the "NPMS + adaptive" estimates are worse than those estimated using the sample mean with new sites added through simple random sampling ("NPMS + SRS"). However, the "NPMS + adaptive" poststratified estimates approach the true value more quickly than "NPMS + SRS", presumably due to the important variation in *Calluna* occupancy across the strata used (Fig. 3d).

The third scenario, "Stratum SRS + adaptive", indicates the reason for the initially poor poststratified estimates under "NPMS + adaptive": the 2019 NPMS locations are biased towards the presence of *C. vulgaris* within all strata.

Table 3: Initial and final mean occupancies (with standard errors) of *Calluna vulgaris* for different adaptive sampling methods and starting data.

Iteration	Mean	SE	Method	Estimator
1	0.33	0.024	NPMS + SRS	Unadjusted
600	0.28	0.014	NPMS + SRS	Unadjusted
1	0.41	0.021	NPMS + adaptive	Poststratified
600	0.31	0.012	NPMS + adaptive	Poststratified
1	0.25	0.021	Stratum SRS + adaptive	Poststratified
600	0.27	0.011	Stratum SRS + adaptive	Poststratified

Evidence for this can be seen within Figure 3c; for example, the estimated occupancy within the Uplands of England is very strongly overestimated before it is incorporated into the adaptive sampling algorithm. Other strata show weaker patterns, but the pattern of initial overestimation is clear (Fig. 3d provides the “true” values for comparison). This means that some stratum occupancy estimates require the addition of many new sampling locations before their estimates increase in accuracy: before this point the poststratification estimator simply weights the biased stratum estimates according to their areas, resulting in important residual bias in the overall estimate. The “Stratum SRS + adaptive” scenario shows that rejecting the existing locations within the 2019 NPMS dataset and selecting new random sets of sites within strata results in more accurate poststratified estimates that rapidly improve (Fig. 3a, “Stratum SRS + adaptive”). This highlights that if the assumptions of the poststratification model are approximately correct (i.e. sampling is random conditional on the strata), then our approach can perform well: the estimates also show slightly decreased standard errors over the iterative series relative to the simple random sampling site-addition approach (Table 3). Finally, the R-indicators shown in Figure 3b demonstrate how these metrics are only as useful as the accuracy of the underlying assumptions [53]: the R-indicator for the “NPMS + adaptive” scenario shows the expected pattern of decreasing variation in the mean sample inclusion probability across strata (note that “Stratum SRS + adaptive” is not shown as it is identical to “NPMS + adaptive”), whereas the simple random sampling additions to the original NPMS sample do not aim to harmonise stratum sampling proportions on this basis.

Discussion

Nonprobability samples of different types are now routinely used within ecology and conservation for various monitoring aims, often with minimal critical assessment [8, 10]. Not infrequently such projects relate to the desire to produce large-scale indicators of biodiversity change, with representativeness of large geographical areas implied as a consequence. Whilst estimates based on such data can potentially be partially adjusted for sampling bias using a family of reweighting techniques including poststratification [11, 34], targeting new effort

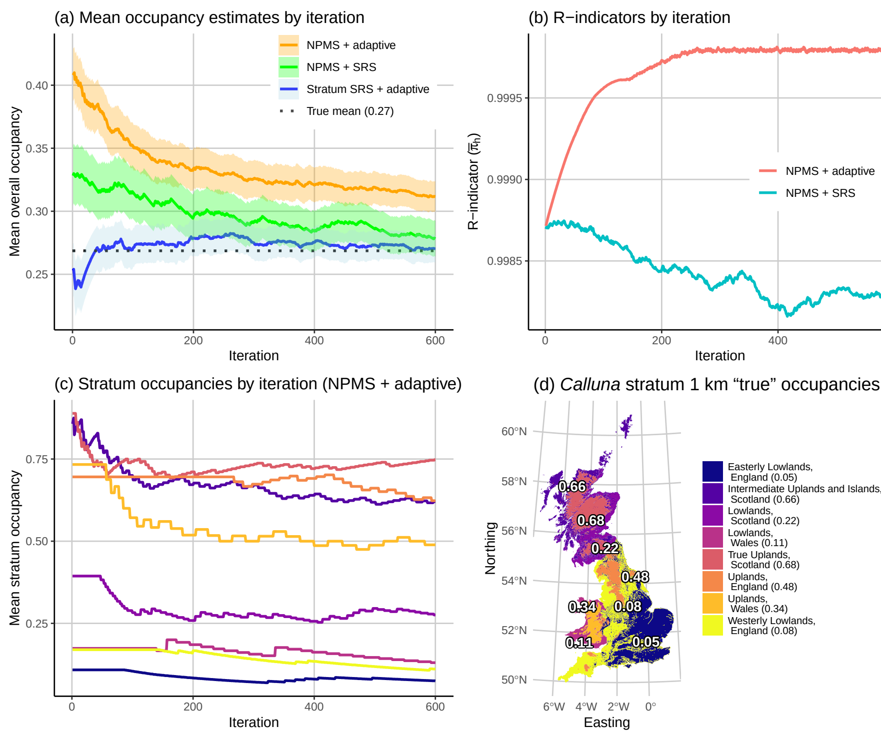

Figure 3: Adaptive sampling of *Calluna vulgaris* occupancy within the National Plant Monitoring Scheme, 2019. (a) Overall Great Britain occupancy estimates for different adaptive sampling scenarios, with binomial proportion standard error ribbons, compared to the "true" mean (0.27). The simple (i.e. unadjusted) mean of the initial 2019 NPMS data is 0.33 (the starting point of the NPMS + SRS curve plotted in green). The estimates shown by the orange (NPMS + adaptive) and blue (Stratum SRS + adaptive) curves use our adaptive algorithm coupled with poststratified estimates of the mean; (b) R-indicators for stratum unit inclusion probabilities, NPMS + adaptive and + SRS scenarios; (c) Stratum occupancies for the NPMS + adaptive method by iteration, colour-coding follows (d); (d) Estimated "true" mean occupancies of *Calluna* by UKCS Environmental Zone stratum (these are displayed at, or near, stratum centroids on the map).

in order to reduce such biases is likely to be a useful complementary strategy [52, 54]. We suggest that the use of strata, hypothesized to capture important relationships between inclusion probabilities and the response variable(s) of interest, is a useful and simple theoretical starting point for adaptive sampling for projects with descriptive goals (i.e. where the aim is to estimate some directly measurable property of a population from a sample; [24]).

If the strata are well-chosen relative to their potential to reduce correlations representing sampling bias, our adaptive approach aimed at a random sample stratified using proportional allocation can improve matters. An example would be where a common plant has near 100% occupancy at some broad scale (e.g. a 10 x 10 km grid), but its average local cover (e.g. at the square-metre scale) varies with an environmental gradient. If sampling co-varies along the same gradient (e.g. due to population density, as in the UK National Plant Monitoring Scheme; [45]) then estimates of average abundance are likely to exhibit important bias. However, if some set of strata partition the environment into areas where sampling is close to random with respect to regional variation in the species' abundance, then this bias will be significantly reduced: the national correlation is removed by estimating means within smaller areas and then combining these in relation to their expected national proportions to better represent the total population [11]. Whilst it is true that in such a case the poststratification estimator will theoretically reduce bias anyway [4, 18, 13, 53], the combination of adaptive sampling and reweighting has been shown to be superior to relying on reweighting alone, both in theory and in empirical investigations in the survey sampling literature [52, 53]. Adding new sites to the sample in this way can reduce variance, as well as keeping bias low [66]. Regardless of this, monitoring programs will often have a focus on increasing uptake for other reasons (e.g. engagement, increasing power; [22]), and so targeted approaches to selecting new sites are likely to be required irrespective of existing analytical options for potential bias reduction of the sample in hand [11]. In theory, such approaches could also be applied to sampling in other dimensions, e.g. to prioritise the digitisation of literature or museum records to improve spatial and/or temporal representativeness in historic time periods.

Researcher domain knowledge is crucial to the successful application of the strategy explored here and elsewhere [53]. Reweighting nonprobability samples via any analytical technique requires a substantive understanding of plausible relationships between variables driving the sampling process and those driving the response [11, 13, 35]. If strata are in fact random with respect to both y and π, that is they have no relationship with the correlation between sample inclusion and variable of interest, then new locations based on them should not contribute to estimator bias, although variance may be increased. It is also possible that selected strata increase bias. As our *Calluna* example demonstrates, this may be due to the poststratification step amplifying poor within-stratum estimates (i.e. those with substantial remaining biases). Theoretically the adaptive sampling step itself should not increase bias if it is a probability-based selection. In reality, constraints on the sampling of new locations within strata could increase or maintain bias for the same reasons that the sample in-hand was initially biased, for example due to land access issues.

A similar situation might occur if an adaptive sampling strategy was applied to a finite pool of interested surveyors, and the strategy ended up merely shifting attention from one area to another, introducing a bias that might change over time if left unadjusted. Whilst poststratification could continue to reduce such biases if the underlying strata were effective, survey organisers would presumably want to monitor such situations given that they may represent no net gain in accuracy. There would be little point in attempting to manipulate data collection if it merely led to a new sample configuration with biases of a similar size unless other inferential aims were in play: the desire to cover some environmental gradient to better estimate predictive or causal regression coefficients for use in species distribution modelling or similar across broader time-slices, for example [36]). A related issue is that our algorithm only considers the addition of new sampling units, not their removal. In theory, removing existing sites could also reduce bias: for example, in our *Calluna* example, even if we did not have access to the "true" distribution, a coarser map of habitat types might clearly indicate oversampling of heathland and other relevant habitats within strata. Whether or not reducing survey effort in this way is a sensible option will of course be survey specific.

Other practical issues also need considering. Spatial bias in citizen science surveys is not unexpected given the vounteer effort underpinning them [46], and so it may not be realistic to recruit surveyors for locations selected according to theories of statistical optimisation. Some schemes may be able to avoid this issue through the combination of volunteer and professional effort; for example, the UK Pollinator Monitoring Scheme currently relies on both [60]. In other cases low uptake in some areas can be very challenging, and substantial effort may be required to understand the reasons for nonresponse. An example is the "Upland Rovers" scheme of the UK Breeding Bird Survey, where substantial effort has gone into trialling different approaches to increasing surveyor uptake of upland squares [5].

Even if practical implementation is difficult, our approach can have value as a conceptual tool for the investigation of existing biases via simulation exercises in a similar way to the *Calluna* example given here. Discretised species distribution models, or simply habitat or land cover maps, could still provide insight into likely biases affecting the sampling of a species' abundance or occupancy, and this type of information could be used to better construct adjustment poststrata and/or adjust uncertainty intervals for estimates [11, 42]. If large biases are suspected to remain, even after the exploration of adaptive sampling or poststratification, then other bias reducton strategies should be explored, the simplest being to adjust the estimand to a population that one has more confidence of being sampled representatively. That is, do not make inferential claims that are significantly larger than the evidence [8]. An example would be claiming that a time series of a butterfly's local abundance was actually indicative of that across the whole of a country in the face of strong evidence for geographic bias and temporal shifts in such over time [cf. 6].

Adaptive sampling in environmental monitoring is not new [e.g. 55], however, a majority of previous investigations in this area have primarily aimed at taking "advantage of population characteristics to obtain more precise estimates of population abundance or density, for a given size or cost, than is possible with conventional designs" [59]. Indeed, work in this area of ecology has tended to focus on the reduction of variance conditional on controlled design, and seems rarely to have asked the question from the point of view of adding units to reduce estimator bias relative to a baseline of unrepresentative sampling for descriptive inference [22]. Whilst there is considerable mathematical overlap between these existing approaches to adaptive sampling [59] and that considered here, those approaches have tended to use the response values of interest to guide the selection of new sampling locations [59], whereas here we follow the recently developed survey sampling approach of focusing on how to equilibriate inclusion probabilities across units to reduce correlations between these and the response variable(s) of interest [53]. Such approaches fall within the second category of Wagner's typology of nonresponse bias indicators [63], as they require data on survey response and sampling frame information at the population level (here stratum membership), but not on the survey outcome variables themselves.

Conclusion

We have laid out the relationship between poststratum-based adjustment strategies and inverse probability weighting in the context of reducing bias (or, equivalently, improving representation) for descriptive inference. Following Meng [33] and others [4, 65], we have characterised this bias as a non-zero correlation between inclusion probabilities and the variable(s) of interest and clarified the assumptions required to justify this approach. A recent review of adaptive sampling in ecology [22] suggested that the complexity of some techniques in the literature likely constituted an important barrier to uptake, and our simple approach may help to overcome this problem. The approach proposed here relies on assumptions that are typically impossible to verify without separate survey efforts, but this is no different to the assumptions required to reweight existing samples to improve representativeness [3, 2, 11], and the ongoing development of R-indicators and related tools points to numerous opportunities for ecologists in these areas [e.g. 52, 38, 53]. We have focused on a single categorical driver of sampling bias to target adaptive sampling, but, in principle, one could cross-tabulate many categorical variables and/or discretise continuous ones for crossing [62]. It may be that modelling inclusion probabilities using multivariable approaches, and using "partial" R-indicators based on these, will allow finer-grained exploration and control of adaptive sampling strategies relative to inclusion probability variance in the future [54].

We reiterate that our approach is not a panacea. In general, if sample inclusion probabilites and the response variable are still correlated after poststratification (i.e. $|\rho(\pi_{i(h)}, y_{i(h)})| >> 0$), then calculated statistics may still contain important bias relative to any given research question. However, this applies to all such strategies based on weighting adjustments, and certainly applies to ignoring the problem altogether (i.e. assuming that the sampling mechanism is already equivalent to a probability sample without critical inspection). Best practice is likely to involve sensitivity analyses [29, 42], and both quantitative [9] and qualitative assessments of the potential for bias relative to key research goals [8, 43].

Acknowledgements

Redacted for double-blind review.

Supplementary Material 1

https://doi.org/10.5281/zenodo.13736326

References

[1] Aubry, P., Francesiaz, C., Guillemain, M., 2024. On the impact of preferential sampling on ecological status and trend assessment. Ecological Modelling 492, 110707. doi:10.1016/j.ecolmodel.2024.110707.

[2] Bailey, M., 2023a. Polling at a Crossroads: Rethinking Modern Survey Research. Cambridge University Press, New York, NY.

[3] Bailey, M.A., 2023b. A New Paradigm for Polling. Harvard Data Science Review 5. doi:10.1162/99608f92.9898eede.

[4] Bethlehem, J., 2002. Weighting nonresponse adjustments based on auxiliary information, in: Groves, R., Dillman, D., Eltinge, J., Little, R. (Eds.), Survey Nonresponse. John Wiley & Sons, Inc., New York, pp. 275–288.

[5] Border, J., Gillings, S., Newson, S., Logie, M., August, T., Robinson, R., Pocock, M., 2019. The JNCC Terrestrial Biodiversity Surveillance Schemes: An Assessment of Coverage. JNCC Report 646. JNCC. Peterborough.

[6] Boyd, R.J., Botham, M., Dennis, E., Fox, R., Harrower, C., Middlebrook, I., Roy, D., Pescott, O., 2024a. Using causal diagrams and superpopulation models to correct geographic biases in biodiversity monitoring data. ecoEvoRxiv .

[7] Boyd, R.J., Bowler, D.E., Isaac, N.J.B., Pescott, O.L., 2024b. On the trade-off between accuracy and spatial resolution when estimating species occupancy from geographically biased samples. Ecological Modelling 493, 110739. doi:10.1016/j.ecolmodel.2024.110739.

[8] Boyd, R.J., Powney, G.D., Burns, F., Danet, A., Duchenne, F., Grainger, M.J., Jarvis, S.G., Martin, G., Nilsen, E.B., Porcher, E., Stewart, G.B., Wilson, O.J., Pescott, O.L., 2022. ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. Methods in Ecology and Evolution 13, 1497–1507. doi:10.1111/2041-210X.13857.

[9] Boyd, R.J., Powney, G.D., Carvell, C., Pescott, O.L., 2021. occAssess: An R package for assessing potential biases in species occurrence data. Ecology and Evolution 11, 16177–16187. doi:10.1002/ece3.8299.

[10] Boyd, R.J., Powney, G.D., Pescott, O.L., 2023. We need to talk about nonprobability samples. Trends in Ecology & Evolution doi:10.1016/j.tree.2023.01.001.

[11] Boyd, R.J., Stewart, G.B., Pescott, O.L., 2024c. Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists. Ecology 105, e4214. doi:10.1002/ecy.4214.

[12] Bunce, R.G., Barr, C.J., Gillespie, M.K., Howard, D.C., 1996. The ITE Land classification: Providing an environmental stratification of Great Britain. Environmental Monitoring and Assessment 39, 39–46. doi:10.1007/BF00396134.

[13] Caughey, D., Berinsky, A.J., Chatfield, S., Hartman, E., Schickler, E., Sekhon, J.S., 2020. Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias. Elements in Quantitative and Computational Methods for the Social Sciences doi:10.1017/9781108879217.

[14] Dornelas, M., Antão, L.H., Moyes, F., Bates, A.E., Magurran, A.E., Adam, D., Akhmetzhanova, A.A., Appeltans, W., Arcos, J.M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A.H., Barbosa, M., Barreto, T.E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., Bett, B.J., Bjorkman, A.D., Błażewicz, M., Blowes, S.A., Bloch, C.P., Bonebrake, T.C., Boyd, S., Bradford, M., Brooks, A.J., Brown, J.H., Bruelheide, H., Budy, P., Carvalho, F., Castañeda-Moya, E., Chen, C.A., Chamblee, J.F., Chase, T.J., Siegwart Collier, L., Collinge, S.K., Condit, R., Cooper, E.J., Cornelissen, J.H.C., Cotano, U., Kyle Crow, S., Damasceno, G., Davies, C.H., Davis, R.A., Day, F.P., Degraer, S., Doherty, T.S., Dunn, T.E., Durigan, G., Duffy, J.E., Edelist, D., Edgar, G.J., Elahi, R., Elmendorf, S.C., Enemar, A., Ernest, S.K.M., Escribano, R., Estiarte, M., Evans, B.S., Fan, T.Y., Turini Farah, F., Loureiro Fernandes, L., Farneda, F.Z., Fidelis, A., Fitt, R., Fosaa, A.M., Daher Correa Franco, G.A., Frank, G.E., Fraser, W.R., García, H., Cazzolla Gatti, R., Givan, O., Gorgone-Barbosa, E., Gould, W.A., Gries, C., Grossman, G.D., Gutierréz, J.R., Hale, S., Harmon, M.E., Harte, J., Haskins, G., Henshaw, D.L., Hermanutz, L., Hidalgo, P., Higuchi, P., Hoey, A., Van Hoey, G., Hofgaard, A., Holeck, K., Hollister, R.D., Holmes, R., Hoogenboom, M., Hsieh, C.h., Hubbell, S.P., Huettmann, F., Huffard, C.L., Hurlbert, A.H., Macedo Ivanauskas, N., Janík, D., Jandt, U., Jażdżewska, A., Johannessen, T., Johnstone, J., Jones, J., Jones, F.A.M., Kang, J., Kartawijaya, T., Keeley, E.C., Kelt, D.A., Kinnear, R., Klanderud, K., Knutsen, H., Koenig, C.C., Kortz, A.R., Král, K., Kuhnz, L.A., Kuo, C.Y., Kushner, D.J., Laguionie-Marchais, C., Lancaster, L.T., Min Lee, C., Lefcheck, J.S., Lévesque, E., Lightfoot, D., Lloret, F., Lloyd, J.D., López-Baucells, A., Louzao, M., Madin, J.S., Magnússon, B., Malamud,

S., Matthews, I., McFarland, K.P., McGill, B., McKnight, D., McLarney, W.O., Meador, J., Meserve, P.L., Metcalfe, D.J., Meyer, C.F.J., Michelsen, A., Milchakova, N., Moens, T., Moland, E., Moore, J., Mathias Moreira, C., Müller, J., Murphy, G., Myers-Smith, I.H., Myster, R.W., Naumov, A., Neat, F., Nelson, J.A., Paul Nelson, M., Newton, S.F., Norden, N., Oliver, J.C., Olsen, E.M., Onipchenko, V.G., Pabis, K., Pabst, R.J., Paquette, A., Pardede, S., Paterson, D.M., Pélissier, R., Peñuelas, J., Pérez-Matus, A., Pizarro, O., Pomati, F., Post, E., Prins, H.H.T., Priscu, J.C., Provoost, P., Prudic, K.L., Pulliainen, E., Ramesh, B.R., Mendivil Ramos, O., Rassweiler, A., Rebelo, J.E., Reed, D.C., Reich, P.B., Remillard, S.M., Richardson, A.J., Richardson, J.P., van Rijn, I., Rocha, R., Rivera-Monroy, V.H., Rixen, C., Robinson, K.P., Ribeiro Rodrigues, R., de Cerqueira Rossa-Feres, D., Rudstam, L., Ruhl, H., Ruz, C.S., Sampaio, E.M., Rybicki, N., Rypel, A., Sal, S., Salgado, B., Santos, F.A.M., Savassi-Coutinho, A.P., Scanga, S., Schmidt, J., Schooley, R., Setiawan, F., Shao, K.T., Shaver, G.R., Sherman, S., Sherry, T.W., Siciński, J., Sievers, C., da Silva, A.C., Rodrigues da Silva, F., Silveira, F.L., Slingsby, J., Smart, T., Snell, S.J., Soudzilovskaia, N.A., Souza, G.B.G., Maluf Souza, F., Castro Souza, V., Stallings, C.D., Stanforth, R., Stanley, E.H., Mauro Sterza, J., Stevens, M., Stuart-Smith, R., Rondon Suarez, Y., Supp, S., Yoshio Tamashiro, J., Tarigan, S., Thiede, G.P., Thorn, S., Tolvanen, A., Teresa Zugliani Toniato, M., Totland, Ø., Twilley, R.R., Vaitkus, G., Valdivia, N., Vallejo, M.I., Valone, T.J., Van Colen, C., Vanaverbeke, J., Venturoli, F., Verheye, H.M., Vianna, M., Vieira, R.P., Vrška, T., Quang Vu, C., Van Vu, L., Waide, R.B., Waldock, C., Watts, D., Webb, S., Wesołowski, T., White, E.P., Widdicombe, C.E., Wilgers, D., Williams, R., Williams, S.B., Williamson, M., Willig, M.R., Willis, T.J., Wipf, S., Woods, K.D., Woehler, E.J., Zawada, K., Zettler, M.L., 2018. BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography 27, 760–786. doi:`10.1111/geb.12729`.

[15] EEA, 2002. Europe's Biodiversity – Biogeographical Regions and Seas. Biogeographical Regions in Europe. Technical Report.

[16] Elliott, M.R., Valliant, R., 2017. Inference for Nonprobability Samples. Statistical Science 32, 249–264. doi:`10.1214/16-STS598`.

[17] Gelman, A., 2007. Struggles with Survey Weighting and Regression Modeling. Statistical Science 22, 153–164. `arXiv:27645813`.

[18] Gelman, A., Carlin, J.B., 2002. Poststratification and weighting adjustments, in: Groves, R., Dillman, D., Eltinge, J., Little, R. (Eds.), Survey Nonresponse. John Wiley & Sons, Inc., New York, pp. 289–302.

[19] Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), 2012. Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.

[20] Gonzalez, A., Cardinale, B.J., Allington, G.R.H., Byrnes, J., Arthur Endsley, K., Brown, D.G., Hooper, D.U., Isbell, F., O'Connor, M.I., Loreau, M., 2016. Estimating local biodiversity change: A critique of papers claiming no net loss of local diversity. Ecology 97, 1949–1960. doi:10.1890/15-1759.1.

[21] Groves, R.M., 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly 70, 646–675. doi:10.1093/poq/nfl033.

[22] Henrys, P.A., Mondain-Monval, T.O., Jarvis, S.G., 2024. Adaptive sampling in ecology: Key challenges and future opportunities. Methods in Ecology and Evolution n/a. doi:10.1111/2041-210X.14393.

[23] Hill, M.O., 1991. Patterns of Species Distribution in Britain Elucidated by Canonical Correspondence Analysis. Journal of Biogeography 18, 247–255. doi:10.2307/2845395, arXiv:2845395.

[24] Hodges, J.S., 1996. Statistical Practice as Argumentation: A Sketch of a Theory of Applied Statistics, in: Lee, J.C., Johnson, W.O., Zellner, A. (Eds.), Modelling and Prediction Honoring Seymour Geisser. Springer, New York, NY, pp. 19–45. doi:10.1007/978-1-4612-2414-3_2.

[25] Holt, D., Smith, T.M.F., 1979. Post Stratification. Royal Statistical Society. Journal. Series A: General 142, 33–46. doi:10.2307/2344652.

[26] Ledger, S.E.H., Loh, J., Almond, R., Böhm, M., Clements, C.F., Currie, J., Deinet, S., Galewski, T., Grooten, M., Jenkins, M., Marconi, V., Painter, B., Scott-Gatty, K., Young, L., Hoffmann, M., Freeman, R., McRae, L., 2023. Past, present, and future of the Living Planet Index. npj Biodiversity 2, 1–13. doi:10.1038/s44185-023-00017-3.

[27] Little, R., 2009. Weighting and Prediction in Sample Surveys. Working Paper 81. University of Michigan. University of Michigan School of Public Health.

[28] Little, R.J.A., 1986. Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review / Revue Internationale de Statistique 54, 139–157. doi:10.2307/1403140, arXiv:1403140.

[29] Little, R.J.A., Rubin, D.B., 2020. Statistical Analysis with Missing Data. 3rd ed., Wiley, Hoboken, N.J.

[30] Lohr, S., 2019. Sampling: Design and Analysis. 3rd ed., CRC Press, Boca Raton, FL.

[31] Lumley, T., 2010. Complex Surveys: A Guide to Analysis Using R. John Wiley and Sons.

[32] Lundberg, I., Johnson, R., Stewart, B.M., 2021. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. American Sociological Review 86, 532–565. doi:10.1177/00031224211004187.

[33] Meng, X.L., 2018. Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. The Annals of Applied Statistics 12, 685–726. doi:10.1214/18-AOAS1161SF.

[34] Meng, X.L., 2022. Comments on "Statistical inference with non-probability survey samples" – Miniaturizing data defect correlation: A versatile strategy for handling non-probability samples. Survey Methodology 48, Paper available at http://www.statcan.gc.ca/pub/12–001–x/2022002/article/00006–eng.htm.

[35] Mercer, A.W., Kreuter, F., Keeter, S., Stuart, E.A., 2017. Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference. Public Opinion Quarterly 81, 250–271. doi:10.1093/poq/nfw060.

[36] 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. doi:10.1111/2041-210X.14355.

[37] Morton, R., Marston, C., O'Neil, A., Rowland, C., 2022. Land Cover Map 2018 (1 km Summary Rasters, GB and N. Ireland). doi:10.5285/9b68ee52-8a95-41eb-8ef1-8d29e2570b00.

[38] Nishimura, R., Wagner, J., Elliott, M.R., 2016. Alternative indicators for the risk of non-response bias: A simulation study. International statistical review = Revue internationale de statistique 84, 43–62. doi:10.1111/insr.12100.

[39] NPMS, 2024. National Plant Monitoring Scheme survey data (2015-2023). doi:10.5285/eb135726-9039-441c-8335-1aab5f6dda21.

[40] Outhwaite, C.L., Powney, G.D., August, T.A., Chandler, R.E., Rorke, S., Pescott, O.L., Harvey, M., Roy, H.E., Fox, R., Roy, D.B., Alexander, K., Ball, S., Bantock, T., Barber, T., Beckmann, B.C., Cook, T., Flanagan, J., Fowles, A., Hammond, P., Harvey, P., Hepper, D., Hubble, D., Kramer, J., Lee, P., MacAdam, C., Morris, R., Norris, A., Palmer, S., Plant, C.W., Simkin, J., Stubbs, A., Sutton, P., Telfer, M., Wallace, I., Isaac, N.J.B., 2019. Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK, 1970–2015. Scientific Data 6, 1–12. doi:10.1038/s41597-019-0269-1.

[41] Pescott, O., Walker, K.J., Powney, G., 2019a. Developing a Bayesian Species Occupancy/Abundance Indicator for the UK National Plant Monitoring Scheme. Unpublished Report to JNCC/Defra.. NERC/Centre for Ecology & Hydrology and BSBI. Wallingford, UK.

[42] Pescott, O.L., 2023. Seek a Paradigm and Distrust It? Statistical Arguments and the Representation of Uncertainty. Harvard Data Science Review 5. doi:10.1162/99608f92.a02188d0.

[43] Pescott, O.L., Boyd, R.J., Powney, G.D., Stewart, G.B., 2023. Towards a unified approach to formal risk of bias assessments for causal and descriptive inference. doi:10.48550/arXiv.2308.11458, arXiv:2308.11458.

[44] Pescott, O.L., Preston, C., 2014. Some environmental factors influencing the distribution of bryophytes in Britain and Ireland, in: Blockeel, T., Bosanquet, S., Hill, M., Preston, C. (Eds.), Atlas of British and Irish Bryophytes. Pisces Publications, Newbury, UK. volume 1.

[45] Pescott, O.L., Walker, K.J., Harris, F., New, H., Cheffings, C.M., Newton, N., Jitlal, M., Redhead, J., Smart, S.M., Roy, D.B., 2019b. The design, launch and assessment of a new volunteer-based plant monitoring scheme for the United Kingdom. PLoS ONE 14, e0215891. doi:10.1371/journal.pone.0215891.

[46] Pescott, O.L., Walker, K.J., Pocock, M.J.O., Jitlal, M., Outhwaite, C.L., Cheffings, C.M., Harris, F., Roy, D.B., 2015. Ecological monitoring with citizen science: The design and implementation of schemes for recording plants in Britain and Ireland. Biological Journal of the Linnean Society 115, 505–521. doi:10.1111/bij.12581.

[47] Preston, C.D., Hill, M.O., Harrower, C.A., Dines, T.D., 2013. Biogeographical patterns in the British and Irish flora. New Journal of Botany 3, 96–117.

[48] Preston, C.D., Pearman, D.A., Dines, T.D. (Eds.), 2002. New Atlas of the British and Irish Flora. Oxford University Press, Oxford, England.

[49] Sackett, D., 1979. Bias in analytic research. Journal of Chronic Diseases 32, 51–63.

[50] Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., Skinner, C., 2012. Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators. International Statistical Review 80, 382–399. doi:10.1111/j.1751-5823.2012.00189.x.

[51] Schouten, B., Cobben, F., Bethlehem, J., 2009. Indicators for the representativeness of survey response. Survey Methodology 35, 101–113.

[52] Schouten, B., Cobben, F., Lundquist, P., Wagner, J., 2014. Theoretical and Empirical Support for Adjustment of Nonresponse by Design. Discussion Paper 2014-15.

[53] Schouten, B., Peytchev, A., Wagner, J., 2017. Adaptive Survey Design. 1st ed., Chapman and Hall/CRC, New York.

[54] Schouten, B., Shlomo, N., 2017. Selecting Adaptive Survey Design Strata with Partial R-indicators. International Statistical Review 85, 143–163. doi:10.1111/insr.12159.

[55] Seber, G., Thompson, S., 1994. Environmental Adaptive Sampling, Elsevier Science B.V.. number 12 in Handbook of Statistics.

[56] Smith, T.M.F., 1991. Post-Stratification. Journal of the Royal Statistical Society. Series D (The Statistician) 40, 315–323. doi:10.2307/2348284, arXiv:2348284.

[57] Spellerberg, I.F., 2005. Monitoring Ecological Change. Cambridge University Press, Cambridge, UK.

[58] Stroh, P., Walker, K., Humphrey, T., Pescott, O., Burkmar, R. (Eds.), 2023. Plant Atlas 2020. Mapping Changes in the Distribution of the British and Irish Flora. Botanical Society of Britain and Ireland & Princeton University Press, Princeton.

[59] Thompson, S.K., 2012. Sampling. Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, N.J.

[60] UK Pollinator Monitoring Scheme, 2024. The UK PoMS Annual Report 2023. Technical Report. UK Centre for Ecology & Hydrology and Joint Nature Conservation Committee.

[61] UKCEH Countryside Survey, 2013. Countryside Survey Environmental Zones. doi:10.5285/0cfd454a-d035-416c-80dc-803c65470ea2.

[62] Valliant, R., Dever, J., Kreuter, F., 2018. Practical Tools for Designing and Weighting Survey Samples. 2nd ed., Springer International Publishing, Cham, Switzerland.

[63] Wagner, J., 2012. A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. Public Opinion Quarterly 76, 555–575. doi:10.1093/poq/nfs032.

[64] Walker, K., Pescott, O., Harris, F., Cheffings, C., New, H., Bunch, N., Roy, D., 2015. Making plants count. British Wildlife 26, 243–250.

[65] Wu, C., 2022. Statistical inference with non-probability survey samples. Survey Methodology 48, Paper available at http://www.statcan.gc.ca/pub/12–001–x/2022002/article/00002–eng.htm.

[66] Zhang, S., Wagner, J., 2024. The Additional Effects of Adaptive Survey Design Beyond Post-Survey Adjustment: An Experimental Evaluation. Sociological Methods & Research 53, 1350–1383. doi:10.1177/00491241221099550.