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

Oliver L. Pescott[1,*], Gary D. Powney[1], Rob J. Boyd[1]

[a] *UK Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, OX10 0PL, Oxfordshire, UK*

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: they are biased with respect to their estimands. Where there is full control over sampling processes, individual units within a population have known response propensities, but these are unknown in the absence of any statistical design. This could be due to the voluntary nature of surveys or because of data aggregation. In these cases some degree of sampling bias is inevitable and we must do something to ameliorate it. One such 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 actively increasing representativeness defined in terms of response propensities. 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 response propensities across population units), allowing an estimator to approach that level of error expected under random sampling. We describe the theory supporting this straightforward strategy, and demonstrate its application using the National Plant Monitoring Scheme, a UK-focused, structured citizen science monitoring programme with uneven uptake.

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 scales. Whilst this has long been an active area of research within conservation and related fields (e.g. Spellerberg [54]), in recent years its importance has increased, with numerous species' time trends, and associated

*Corresponding author
Email address: olipes@ceh.ac.uk (Oliver L. Pescott)

multi-species indicators, now based on a wide variety of data types [e.g. 37, 25, 13]. One consequence of this trend has been the increasing focus on the use of datasets for monitoring that lack any explicit 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 nonprobability samples is considerably more difficult than has often been recognised in ecology [9]. The absence of a statistical 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 [31, 10, 61, 15, 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. [19]), leading to a number of high-profile disagreements in the literature [9].

The technical elements of sampling design underlying these issues have been well-known in the statistical subdiscipline of survey sampling for decades [30, 2, 58, 31], yet many of these insights are frequently overlooked or misunderstood by ecologists (although by no means all, e.g. see many chapters within ref. [18]). One stumbling block may be the numerous definitions and types of "bias" available in the literature [18, 40, 46]; the lack of any well-known (to ecologists) unified mathematical definition of sampling bias may also have hindered progress and communication.

Within survey sampling focused on descriptive inference (i.e. characterising some directly measurable property of a population from a sample; [23]), 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 "response propensity") and the property of interest y [e.g. 20, 4]. 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; [31]). 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 $E[\rho(\pi, y)] \neq 0$ [e.g. 27].

Regardless of these terminological issues, Meng [31] demonstrated how the standard formula for statistical error (i.e. $\overline{y}_n - \overline{y}_N$, 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"). 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 [10, 6, 9].

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 [32] demonstrated how all such techniques (inverse probability weighting, imputation or "superpopulation" modelling, poststratification and doubly-robust approaches) ultimately boil down to minimising $\rho(\pi, y)$. This insight allows us to understand the assumptions of our methods better, and therefore to justify our approaches and assess their limitations more clearly [7]. 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.

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, 58]. 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 [53]. 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 [53, 61, 10].

The poststratification estimator [4], or "basic poststratification identity" [17], used to achieve this can be defined as:

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

where N is the population size, N_h is the overall size of stratum h, and H is the full set of strata into which the population is divided. The implication here is that within-stratum means substitute for individual unit values, and it is these which are averaged across the entire population once relative stratum sizes in the population have been accounted for (see [10] for a worked ecological example). This formulation implies that all i units within a given poststratum receive the same weight [4, 61], equal to

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

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

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

[61]. 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 [61]. 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 $\overline{y}_{ipw}$, is unbiased in expectation [53, 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 response propensities π and the outcome variable y [20, 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 almost guarantees that this correlation will be appreciably different from zero [9]. 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 [31, 10, 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:

1. The response of interest y_i is invariable within poststrata (i.e. $\sigma_{y(h)} = 0 \quad \forall\, h$).
2. The response propensities 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 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 [24]. In the second of these situations, the poststratification estimator reduces the bias, but not the variance [28, 26]. This is linked to the assertion of Gelman and Carlin [17] that poststratification is most important when correcting for differential nonresponse *between* poststrata. Assuming that response propensities are uniform within poststrata, but correlated with y within the overall population, then adjusting for poststratum membership renders $\rho(\pi, y)$ equal to zero [61, 32]. This puts such adjustments in the Missing At Random (MAR) category of Rubin's [45] missing data framework: π and y are independent conditional on some X, where here X is the vector of unit poststratum memberships [53].

Whilst poststratification and its variants [e.g. see 16] can be useful tools for adjusting existing samples [10], where monitoring is ongoing and survey organisers have some power to adapt data collection, adaptive sampling may be a more efficient way to reduce error compared to relying on poststratification of unrepresentative samples alone [51, 49]. Larger samples may be also desired for other reasons irrespective of the potential for using the poststratification estimator on a sample in hand. 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 appear to be increasingly important for ecological monitoring and conservation, especially where nonprobability samples are used [9]. Simple approaches to adaptive sampling, with few assumptions, are therefore likely to be of wide utility [21].

We outline an approach to the problem based on assessments of poststratum sampling coverage. We show how this can be easily implemented using standard binomial formulae within an adaptive framework using an empirical example: data collected between 2015–2023 for the UK National Plant Monitoring Scheme, a designed citizen science programme with uneven site uptake to date [42]. Our approach has a direct link to the literature on the monitoring of survey quality via assessments of potential nonresponse bias [59, 35], and we use one such indicator (the R-indicator of Schouten, Shlomo and colleagues [47]) of variation in response propensities across strata to demonstrate the potential improvements in survey representativeness, a measure of survey quality [48], achievable using our approach.

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 sampling response 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 10]. For example, in the UK "land classes" have previously been erected based on covariation in numerous geographical and environmental variables [11] and then amalgamated into broader zones [57]; for Europe, biogeographic zones based on patterns of terrestrial and marine biodiversity exist [14]. Note that the strata do not have to be absolutely believed to have an invariable one-to-one relationship between stratum unit membership and response propensity, 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 [32].

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 SRS; this is known as proportional allocation in the survey sampling literature [58]. 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 response propensities and indicators of representativeness noted above was formalised by Schouten and colleagues [48]. They provide the following operational definition of "representative" in the survey sampling context:

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

Note that this is a weaker version of (2) given in the Introduction above, as it does not state that all response propensities within a stratum are identical, only that the means across strata are equal. Based on this, the Schouten *et al.* R-indicator is $R(\pi) = 1 - 2\sigma^2_{\bar{\pi}_h}$, where $\sigma^2_{\bar{\pi}_h}$ is the variance of the mean response propensities across strata. $R(\pi) = 1$ denotes maximum representativeness (equivalent to SRS), when the variance in response propensities across strata is zero.

Adaptive sampling algorithm

This proceeds as follows (see also 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 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 SRS 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 $\bar{x}_h$ and s_h directly from the standard binomial formulae. The new target stratum site count expected under SRS is also updated as $\hat{\mu} = (n + a)/N{\cdot}N_h$. In the following example $a = 1$, 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; the supporting code allows for this parameter to be varied.

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 SRS 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.

An empirical example: the UK National Plant Monitoring Scheme

The UK National Plant Monitoring Scheme (NPMS) asks volunteers to record plant abundances in small plots located in particular habitats [60]. Plots are located within 1 km^2 squares (hereafter "sites") of the relevant country grid (the scheme 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 [42] 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 surveyed than others [42]. A primary aim of the NPMS is the production of nationally representative indicators of habitat quality [38], and so, ideally, coverage of the area would be relatively even. We know that response propensity (i.e. site uptake) is related to such factors as 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 [42]. North-west to south-east gradients of all these variables are well-known in the British Isles [44, 55, 41, 22, 43]. 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 the monitoring scheme's estimands. One widely-used set of strata for Great Britain is the UK Countryside Survey (UKCS) Environmental Zones [57], based on a larger set of "land classes" created originally for the *a priori* stratification of national ecological and biogeographical surveys [11]. To these we add Northern Ireland as an additional stratum to better cover our area (Fig. 1). Surveyed NPMS sites [36] are overlaid on these zones in Figure 1 to show their current (2015–2023) coverage.

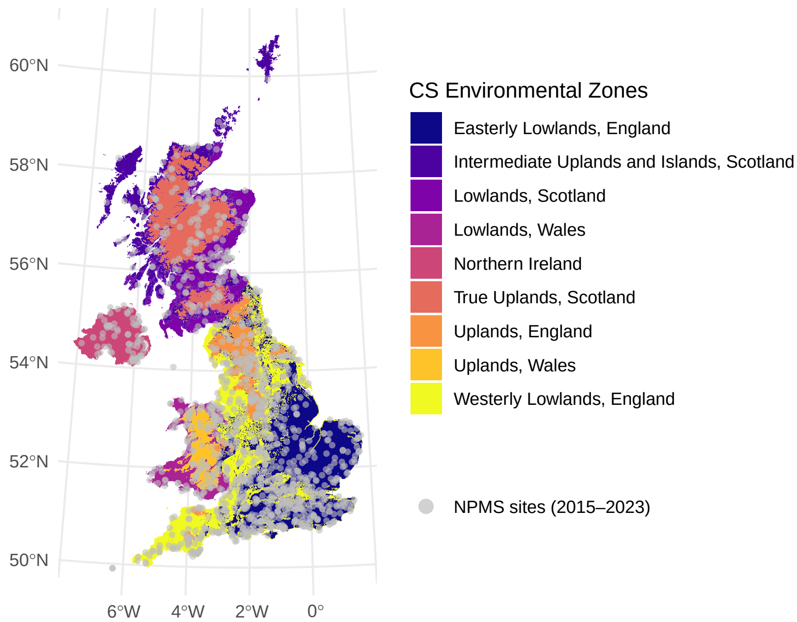

Figure 1: UK Countyside Survey Environmental Zones plus Northern Ireland. Grey circles are the 2015–2023 NPMS sites with survey data.

Results

Table 1 gives the current distribution of NPMS 1 km^2 sites by UKCS stratum. These are given in order of their discrepancy from proportional allocation (i.e. SRS) of the current sample of 1,204 sites that could be assigned to strata, from under- to over-sampled [36].

Table 1: The current distribution of NPMS sites by UKCS Environmental Zone strata, ordered from under- to over-sampled relative to SRS. Mean prop. is the current proportion of the stratum sampled; Discrepancy is the difference between Mean prop. and that expected under SRS.

Stratum no.	Stratum	No. sites	Stratum area (km^2)	Mean prop.	Discrepancy
5	Intermediate Uplands and Islands, Scotland	53	29866	1.8e-03	-2.9e-03
6	True Uplands, Scotland	65	32034	2.0e-03	-2.7e-03
4	Lowlands, Scotland	76	23084	3.3e-03	-1.4e-03
7	Northern Ireland	73	14156	5.2e-03	4.7e-04
8	Lowlands, Wales	60	11309	5.3e-03	6.1e-04
9	Uplands, Wales	55	10272	5.4e-03	6.6e-04
1	Easterly Lowlands, England	395	65441	6.0e-03	1.3e-03
2	Westerly Lowlands, England	321	51815	6.2e-03	1.5e-03
3	Uplands, England	106	15739	6.7e-03	2.0e-03

Figure 2 demonstrates the progress of the stratum-based adaptive sampling algorithm in terms of stratum means 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 points 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).

Table 2 gives abridged output of the adaptive sampling algorithm. 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 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 acheived for the final six iterations at the bottom of Table 2.

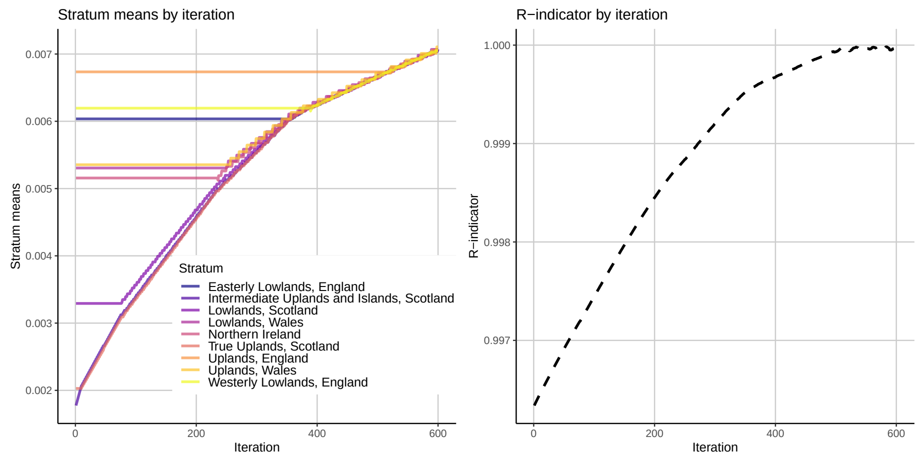

Figure 2: Evolution of UKCS Environmental Zone stratum mean sampled proportions and R-indicator by iteration.

Discussion

Nonprobability samples of different types are now routinely used within ecology and conservation for various monitoring aims [9]. Not infrequently these relate to the desire to produce large-scale indicators of biodiversity change, with representativeness of large geographical areas often 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 [10, 32], targeting new effort in order to reduce such biases is likely to be a useful complementary strategy [49, 51]. We suggest that the use of poststrata, considered to capture important relationships between response propensities and the variable(s) of interest, is a straightfoward starting point for adaptive sampling for projects with descriptive goals (i.e. those where the aim is to estimate some directly measurable property of a population from a sample; [23]).

If the strata are well-chosen relative to their potential to reduce correlations driving sampling bias, our adaptive approach aimed at a random sample stratified using proportional allocation will 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 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; [42]) 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

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

their expected national proportions to better represent the total population [10].

Whilst it is true that in such a case the poststratification estimator will reduce bias anyway [4, 17, 12, 50], 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 [49, 50]. This is because a MAR assumption underpinning standard poststratification may well be incorrect if it has been based on the available sample; however, adding new sites to the sample always admits of the possibility that new elements of the relationship between response propensity and a target variable will be uncovered. Regardless of this, monitoring programs will often have a focus on increasing uptake for other reasons (e.g. engagement, increasing power; [21]), 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 [10].

If the strata are in fact random with respect to both y and π (i.e. they explain nothing), then new locations based on them should not contribute to estimator bias, although variance may be increased. However, if the analyst is unlucky, it is theoretically possible that the “true” poststrata that would have reduced bias are totally uncaptured by the prioritised selection. The worse case might be that they constitute a set nested within some other stratum that appears to be well-sampled relative to the proportional allocation implied by random sampling. The use of researcher domain knowledge should help to avoid this situation [50], just as it has been repeatedly flagged as essential for the choice of adjustment variables in the first place [10, 12, 33]. 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. This situation could in theory be remedied by applying the poststratification estimator within time-slices, although no doubt survey organisers and metadatabase curators would also want to monitor such situations. Adjusting sampling behaviour is obviously not cost-free, and 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 (although not necessarily direction), unless other inferential targets were in play (e.g. 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; [34]). Ultimately, 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 confidence is actually sampled representatively. That is, do not make inferential claims that are significantly larger than the evidence [7]. An example would be claiming that the time series of a butterflies' local abundance in England was actually indicative of that average across the whole of Great Britain when there is clear evidence for temporal shifts in sample coverage of the statistical population [5].

Adaptive sampling in environmental monitoring is not new [e.g. 52], 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" [56]. 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 [21]. Whilst there is considerable mathematical overlap between these existing approaches to adaptive sampling [56] and that considered here, those approaches have tended to use the response values of interest to guide the selection of new sampling locations [56], whereas here we follow the recently developed survey sampling approach of focusing on how to equilibriate response propensities across units to reduce correlations between these and the variable(s) of interest [50]. Such approaches fall within the second category of Wagner's typology of nonresponse bias indicators [59], 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 propensity weighting in the context of reducing bias (or, equivalently, improving representation) for descriptive inference. Following Meng [31] and others [4, 61], we have characterised this bias as a non-zero correlation between response propensities and the variable(s) of interest and clarified the assumptions required to justify this approach. A recent review of adaptive sampling in ecology [21] suggested that the complexity of some techniques in the literature likely constituted an important barrier to uptake, and our simple approach may help overcome this problem. The approach proposed here relies on MAR 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, 10], and the ongoing development of R-indicators and related tools points to numerous opportunities for ecologists in these areas [e.g. 49, 35, 50]. 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 [58]. It may be that modelling response propensities using multivariable approaches, and using "partial" R-indicators based on these, will allow finer-grained exploration and control of adaptive sampling strategies relative to response propensity variance in the future [51].

We reiterate that our approach is not a panacea. In general, if the missing data mechanism is still Missing Not At Random [29] even 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 MCAR without critical inspection). Best practice is likely to involve sensitivity analyses [29, 39], and both quantitative [8] and qualitative assessments of the potential for bias relative to key research goals [7, 40].

Acknowledgements

OP and RB acknowledge NERC award number NE/R016429/1 as part of the UK Status, Change and Projections of the Environment (UK-SCAPE) program delivering National Capability and Joint Nature Conservation Committee funding to the NPMS. GP was supported through NERC award number NE/V006878/1 as part of the DRUID (Drivers and Repercussions of UK Insect Declines) project.

Supplementary Material 1

`https://doi.org/10.5281/zenodo.13736327`

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] 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 .

[6] 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`.

[7] 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`.

[8] 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`.

[9] 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`.

[10] 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`.

[11] 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`.

[12] 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`.

[13] 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.

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

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

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

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

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

[27] Little, R.J., 2023. The "Law of Large Populations" Does Not Herald a Paradigm Shift in Survey Sampling. Harvard Data Science Review 5. doi:10.1162/99608f92.6b049957.

[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] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

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

[45] Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592. doi:10.1093/biomet/63.3.581.

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

[47] 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.

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

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

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

[51] 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.

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

[53] 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.

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

[55] 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.

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

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

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

[59] 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.

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

[61] 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.