Received: 5 September 2023 | Revised: 21 March 2024 | Accepted: 2 June 2024

DOI: 10.1002/env.2870

RESEARCH ARTICLE

WILEY

Applying sequential adaptive strategies for sampling animal populations: An empirical study

Rosa M. Di Biase[1,2] | **Fulvia Mecatti**[1]

[1]Department of Sociology and Social Research, University of Milano-Bicocca, Milano, Italy

[2]NBFC, National Biodiversity Future Center, Palermo, Italy

Correspondence
Rosa M. Di Biase, Department of Sociology and Social Research, University of Milano-Bicocca, Milano, Italy.
Email: rosamaria.dibiase@unimib.it

Funding information
National Recovery and Resilience Plan (NRRP) European Union — NextGenerationEU, Grant/Award Number: CN00000033

Abstract

Traditional sampling methods may prove inadequate when dealing with spatially clustered populations or when studying rare events or traits that are not easily detectable across the target population. When both scenarios occur simultaneously, adaptive sampling strategies can represent a viable option to enhance the detectability of cases of interest. This paper delves into the application of a novel class of sequential adaptive sampling strategies to animal surveys. These strategies, originally proposed for human population tuberculosis prevalence surveys, allow oversampling of the rare interest variables while managing on-field constraints. This ensures that the unfixed sample size, typical of adaptive sampling, does not compromise overall cost-effectiveness. We explore a strategy within this class that integrates an adaptive component into a Poisson sequential selection. The aim is twofold: to intensify the detection of cases by exploiting the spatial clustering and to provide a flexible framework for managing logistics and budget constraints. To illustrate the strengths and weaknesses of this Poisson-based sequential adaptive sampling strategy compared to traditional sampling methods, a simulation study was conducted on a blue-winged teal population in Florida, USA. The results showcase the benefits of the proposed strategy and open avenues for future methodological and practical improvements.

KEYWORDS

adaptive rule, oversampling, pseudo Horvitz-Thompson estimator, waterfowl counts simulations

1 | INTRODUCTION

During the last century, there has been an increasing relevance in biodiversity conservation, driven by the need to counteract the rapid changes that human activities, such as habitat degradation, pollution, and climate change have caused to the ecosystems (Rands et al., 2010). Numerous programs and organizations have been established to address the pressing issue of biodiversity loss (Butchart et al., 2010) and, consequently, to promote ecological restoration (Wortley et al., 2013), with a special focus on habitat management. In particular, numerous habitat management programs have been funded to preserve wetland ecosystems through initiatives such as the International Waterbird Census, which started in 1967, and the North American Waterfowl Management Plan, signed in 1986. These initiatives play a crucial role in monitoring wetland habitats, enabling effective management and conservation of biodiversity, especially for wetland species, such as wintering waterfowl. Indeed, wintering waterfowl readily respond to climate changes (Nichols et al., 1983), altering their migratory patterns and seeking alternative habitats in different or uncommon regions. Waterfowl act as bioindicators, as their presence or absence may provide valuable insights into the overall health of wetland ecosystems (Green & Elmberg, 2013). Consequently, it is extremely important to manage and preserve the wetlands where they migrate to, also providing the adequate protection needed for their survival in an unfamiliar territory (Alagador et al., 2014). Typically, waterfowl counts are performed annually in mid-January to estimate their number and distribution in the wetlands (Gilissen et al., 2002), via land-based, plane, or helicopter surveys (Pihl et al., 1992). Aerial surveys are quick, allowing for a complete coverage of waterfowl assemblages, but the visibility is rarely perfect, and this compromises the detection of waterfowl. On the other hand, land-based surveys are time-consuming, but they allow for a more precise detection (Pihl et al., 1992). Usually, birdwatchers perform these counts by purposively (and subjectively) searching for waterfowl (Delany et al., 1999), making it difficult to check the reliability of the results. Indeed, the accuracy of the estimates can be evaluated from their sampling distributions only if probabilistic sampling schemes are used (Di Biase et al., 2024).

Based on the reasons above, in this paper, we propose to sample wintering waterfowl or, more generally, spatially clustered, and rare animal populations by means of a class of sequential adaptive sampling techniques originally proposed for adult population-based tuberculosis prevalence surveys (Mecatti et al., 2023), sponsored by the WHO and partners in several settings around the world (WHO, 2011). In particular, with sequential adaptive sampling techniques it is possible to improve the detection of the interest attribute (adaptive component) while creating a flexible framework to deal with budget and logistic constraints (sequential component).

The present paper explores the potential of applying such sampling strategies to animal surveys. This choice is motivated by the encouraging results of this pioneering work, where a major drawback of adaptive sampling, that is, the unfixed sample size (Hankin et al., 2019), can be balanced with practical advantages, such as oversampling the rare cases of interest and cost-effectiveness.

The paper is organized as follows. In Section 2, the general class of sequential adaptive sampling strategies is briefly introduced, with a focus on a special member of the class, based on Poisson sampling design, detailed as a viable, fully design-based alternative for sampling rare and clustered animal populations. Both the sample selection procedure and the associated design-unbiased estimation of the Horvitz-Thompson (HT) type are illustrated and discussed, along with estimation of design-based variance and inferential properties. Section 3 presents a simulation study on real data from a blue-winged teal population in Florida (USA) to illustrate the strengths and weaknesses of this Poisson-based sequential adaptive sampling strategy as compared to traditional sampling methods and to envision future methodological and practical improvements. Finally, Section 4 contains some conclusive remarks and research perspectives.

2 | MATERIALS AND METHODS

In this section, the key aspects of the methodology developed in Mecatti et al. (2023) are presented, focusing on a sampling procedure that appears to be promising for surveying rare and spatially clustered animal populations. Mecatti et al. (2023) propose a general class of sampling strategies, referred to as sequential adaptive, aiming to oversample the interest variable, that is rare and clustered into the target population, without compromising the overall cost-effectiveness due to the unfixed sample size that is typical of adaptive sampling. The class of sequential adaptive strategies also pursues practical advantages, by integrating an adaptive component (Thompson, 2006), which aims at improving the detection of the rare and clustered species of interest, into a sequential selection (see, among other, Tillé, 2006). The sequential feature aims at improving control on logistics and budget constraints in a flexible framework, both at the design stage of the survey and in real time during on-field operations. Such practical advantages come at the cost of increased theoretical complexity in the sampling design, which, in technical terms, becomes an *informative design* (details can be found in Cassel et al., 1977). This necessitated the development of the appropriate theory for sample selection and estimation (Mecatti et al., 2023), described in the following sub-sections.

2.1 | Preliminaries and notation

In their original setup, Mecatti et al. (2023) consider a unidimensional spatial setting. This implies that the target population is preordered according to a prechosen rule, in such a way that the population units are close or distant to each other based on such ordering.

Let consider a population U composed of N quadrats, or pixels, or cells, covering the area of interest of size A, denoted as $U = \{1, \dots, N\}$. In the following, the general term population *units* is also used. The population U is ordered, meaning that the units form a sequence from unit labeled 1 to unit N according to any natural or convenient rule. For example, when a grid of N quadrats is overlaid on a map of the survey area, a path across the N quadrats can be pre-designed by optimizing travel costs for on-field observations of the rare and clustered interest variable. Moreover, it could be convenient to tailor the ordering according to pre-existent natural features, such as following the banks of a river, or by exploiting man-made attributes, such as trails or paths in a woodland area.

Let Y denote the survey variable that takes value y_i on pixel $i \in U$, corresponding to the amount of interest variable in that pixel, and $T = \sum_{i=1}^{N} y_i$ is its total. For selecting a probabilistic sample of pixels, let S_i be a random variable indicating the sample membership of pixel i, that is, S_i takes value 1 if pixel i is included in the sample and 0 otherwise. Additionally, let D_i be a non-random variable equal to 1 if $y_i \geq 1$ and 0 otherwise. The final selected sample is then defined as a realization of the N-dimensional random variable $(S_1, \dots, S_i, \dots, S_N)$, which is an assorted sequence of 0s and 1s, indicating whether each pixel in the survey area is selected or not in the final sample of size $n = \sum_{i=1}^{N} S_i$. In practice, to select a random sample, each pixel $i \in U$ is visited step by step according to the pre-chosen order from 1 to N. At step i, a real-time decision is made whether pixel i is selected based on the result of a Bernoulli trial, for example, a 0/1 digital experiment. It is worth noting that in our notation, pixels and steps are identified by the same subscript because, in the sequential selection, pixels and steps are in fact interchangeable, in the sense that a specific pixel can only be visited at the same specific step of the procedure. At the end of the procedure, once all steps $i = 1, \dots, N$ have been completed, the final random sample is obtained based on the realized values of the sample membership indicators described above. The step-by-step selection starts with a chosen set of *initial* probabilities $\pi_1^{(0)}, \dots, \pi_i^{(0)}, \dots, \pi_N^{(0)}$. At the first step, the Bernoulli trial to select/not select pixel 1 is conducted with probability $\pi_1^{(0)}$, while in the subsequent steps $i = 2, \dots, N$, the Bernoulli trial to select/not select pixel j is performed with an *updated* probability according to the following *adaptive updating rule*.

$$\pi_j^{(i)} = \begin{cases} 1 & \text{if } j = i+1 \text{ and } S_i D_i = 1 \\ \pi_j^{(i-1)} - \left(S_i - \pi_i^{(i-1)}\right) w_{j-i}^{(i)} & \text{otherwise} \end{cases}. \tag{1}$$

The term $w_{j-i}^{(i)}$ in (1) defines the *updating weights*, subject to the constraint $0 < \pi_j^{(i)} \leq 1$. In practice, under the (general) updating rule (1), if at step i pixel $j = i$ is selected ($S_i = 1$) and the characteristic of interest is observed in the selected pixel ($D_i = 1$) suggesting that a cluster has been entered, then at the successive step $i + 1$ the subsequent pixel in the sequence ($j = i + 1$) is certainly included in the sample (first term in Equation 1). Otherwise, if pixel i is not selected ($S_i = 0$) or if it is selected but the characteristic of interest is not observed ($D_i = 0$), then probabilities of subsequent pixels are recomputed (updated) according to the second term in Equation (1).

Different choices for the updating weights can be considered, leading to various sampling schemes. In particular, a special choice of updating weights will be illustrated in the following sub-section. The interested readers should defer to the work of Mecatti et al. (2023) for further details and theoretical proofs, as they provide a comprehensive discussion on the subject.

2.2 | Conditional Poisson Sequential Adaptive (CPoSA) sampling

The simplest choice of updating weights $w_{j-i}^{(i)}$ in (1) is setting them equal to zero, determining the sampling strategy referred to as Poisson Sequential Adaptive (PoSA) sampling design (Mecatti et al., 2023). This straightforward updating rule only modifies the inclusion probability of the unit immediately following the selected unit and leaves unaltered at their initial value all the other units.

Despite its simplicity, PoSA does not guarantee a stable final sample size. In particular, it could be unnecessary large and inflated by non-cases, that is, pixels that do not contain the attribute of interest, or else too small to allow acceptable estimation precision.

To improve control over the random sample size of PoSA, a conditional version, CPoSA for short, has been developed (Mecatti et al., 2023). In CPoSA, the appealing aspects of PoSA, namely its simplicity and the ability to oversample pixels that contain the attribute of interest, are retained, but with the introduction of a choice of the minimum sample size, to avoid the selection of undesirably small samples. In practice, to implement CPoSA the initial inclusion probabilities are set to sum up to the chosen minimum sample size: $n_{min} = \sum_{i=1}^{N} \pi_j^{(0)}$. Moreover, in the updating rule (1), the weights are replaced by $\frac{1}{N-i}$, resulting in the following CPoSA updating rule

$$\pi_j^{(i)} = \begin{cases} 1 & \text{if } j = i+1 \text{ and } S_i D_i = 1 \\ \max\left(0, \min\left(\pi_j^{(i-1)} - \left(S_i - \pi_i^{(i-1)}\right) \frac{1}{N-i}, 1\right)\right) & \text{otherwise} \end{cases}. \tag{2}$$

It is worth noting that by (2), the inclusion probabilities of all to-be-visited units $j \geq i+1$ are updated, in particular they are increased in the case of previous non-selection(s), to ensure that the stated minimum sample size is actually reached. On the other hand, inclusion probabilities become smaller whenever a unit is selected. Furthermore, in the case of consecutive selections of units with no interest variable once n_{min} has been reached, rule (2) ensures that the inclusion probabilities tend to 0, preventing the inflation of the final sample with non-cases, as it can occur under PoSA instead. Because of these modifications, the minimum sample size is ensured together with the oversampling of interest variable. Notice that, even if CPoSA still provides a random final sample size n, it simultaneously ensures $n \geq n_{min}$ and, at the same time, it allows control over excessively large n. This remark will be further discussed in Section 3 based on empirical evidence.

2.3 | Estimation under sequential adaptive sampling: The pseudo-HT estimator

In their original paper, Mecatti et al. (2023) introduced a pseudo Horvitz-Thompson (HT) estimator of the population mean, that is unbiased under a sequential adaptive sampling. Good inferential properties are also proved, particularly the consistency and the asymptotic Normal distribution of the pseudo-HT estimator. Exploiting these general results, we now illustrate pseudo-HT estimation on a sequential adaptive sample of pixels. The design-unbiased pseudo-HT estimators of total and, consequently, density (i.e., total per unit area) are given by:

$$\hat{T}_{PHT} = \sum_{i=1}^{N} \frac{S_i}{\pi_i^{(i-1)}} y_i, \quad \hat{\rho}_{PHT} = \frac{\hat{T}_{PHT}}{A}, \tag{3}$$

expressions close to Grafström (2012, Eq. 1). In practice, to compute either the total or density estimate, sampled pixel in (3) are identified by the index i such that $S_i = 1$, for which the observed value y_i is available and weighted by the inverse of the inclusion probability of pixel i updated at the previous step $(i-1)$ of the sequential selection.

For both estimators in (3), accuracy is assessed via an unbiased variance estimate, given by

$$\hat{V}_{PHT} = \sum_{i=1}^{N} \frac{S_i}{\pi_i^{(i-1)}} \left(\frac{1}{\pi_i^{(i-1)}} - 1 \right) y_i^2, \tag{4}$$

for the total estimator $\hat{T}_{PHT}$, and simply divided by A^2 for the density estimator $\hat{\rho}_{PHT}$.

Equations (3) and (4) are easily implemented and can be applied to all sampling designs in the sequential adaptive class, with probability $\pi_i^{(i-1)}$ computed according to the specific updating rule of the particular sequential adaptive sampling design used, as in Equation (2) for CPoSA. A simplified expression of (4) under CPoSA can be found in Mecatti et al. (2023).

3 | SIMULATION STUDY

The purpose of this section is to provide empirical evidence of the advantages and limitations of the presented sequential adaptive sampling strategy when dealing with rare and clustered animal populations. To achieve this goal, a simulation study has been conducted on a real population of blue-winged teal on a wildlife refuge in central Florida (Thompson, 2006). The main objective of the study is to explore whether Conditional Poisson Sequential Adaptive strategy (CPoSA) can offer benefits compared to traditional methods when applied to this case study. The simulation study has

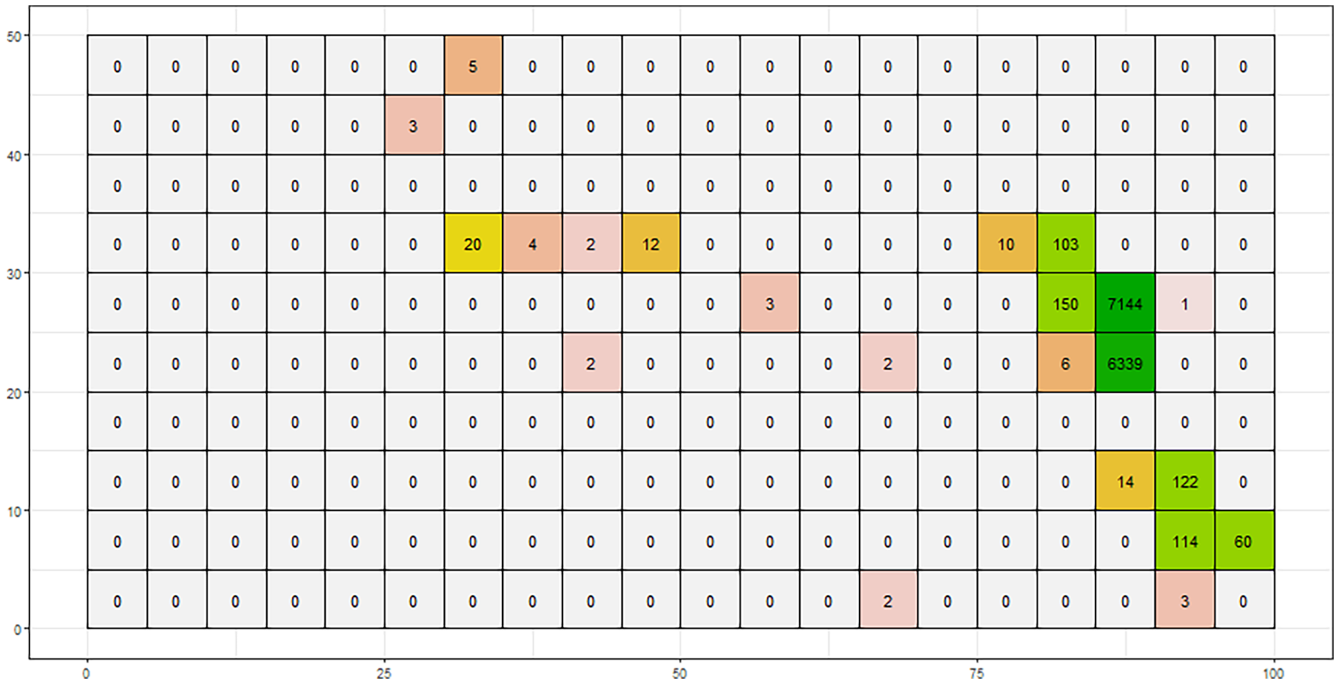

FIGURE 1 Real population of blue-winged teal distributed within a 200-unit surface (Smith et al., 1995), referred to as Pop B.

the goal of comparing the performance of CPoSA with respect to Adaptive Cluster (AC) sampling (Thompson, 2012) and Simple Random Sampling without replacement (SRSWOR), viewed as traditional sampling choices with an unfixed and fixed, respectively, sample size.

The real study region is a 100×50 km^2 rectangle with the southwest corner at 0438000, 3056000 (UTM coordinates, zone 17). Blue-winged teal, the species of interest, were counted by two biologists who flew over the entire region in two different helicopters (Smith et al., 1995), obtaining $T = 14{,}121$ and $\rho = 2.82$ km^{-2}, which serve as the true total number of teal and true density for the simulation. The study area is partitioned into $N = 200$ units (cells or pixels), with the number of teal registered in each cell as displayed in Figure 1. Sampling concerns pixels, and all the teal present in the sampled pixels are counted in the simulation.

Under SRSWOR, $n = 24$ pixels were randomly selected, with equal first-order inclusion probabilities $\pi_j = \frac{n}{N} = \frac{24}{200} = 0.12$. The customary Horvitz-Thompson estimator is used for the estimation of the total.

To simulate CPoSA, all pixels are supposed to have equal initial first order inclusion probabilities $\pi_j^{(0)} = \frac{n_{min}}{N}$. Clearly, n_{min} is a relevant simulation choice, hence three different values have been considered: smaller than the fixed sample size n under SRSWOR, slightly less than n, and equal to n (similarly to Mecatti et al., 2023). In particular, n_{min} was set equal to 17, 19, and 24, resulting in initial first-order inclusion probabilities equal to 0.085, 0.095, and 0.12, respectively. For CPoSA, the chosen spatial pre-ordering is a rectangular-spiral path across all 200 pixels, starting from the bottom-left corner toward the center. This choice is motivated by our study population itself. Indeed, in this case, the spiral path appears as the most feasible choice to simulate controlled scenarios of spatial clustering. Obviously, several other choices are possible, including going up-and-down or left-to-right. For both simplicity and computational burden, other possibly more complex orderings (see Dickson & Tillé, 2016; Stevens & Olsen, 2004) are not considered in this paper simulations. In practice, it is worth noting that CPoSA allows the choice of the route/ordering that better accommodate logistics and budget constraints, including according to pre-existent features of the survey area (e.g., roads, trails, and riverbanks) or prior information about the location of the species of interest (e.g., preliminary investigations with photo traps).

Following the procedure described in Thompson (2012), the initial AC sample is selected by SRSWOR. Then, whenever one or more teal are observed in a selected pixel, all the adjacent (top, bottom, left, and right) pixels are added to the sample and so on. The initial AC sample size n_0 is set equal to n_{min} in CPoSA, in such a way that if no pixels containing teal are sampled, then the final samples for both sampling procedures have the same number of selected pixels, namely n_0 for AC and n_{min} for CPoSA. For the sake of simplicity, in the sequel the symbol n^* will indicate the simulated value for either the

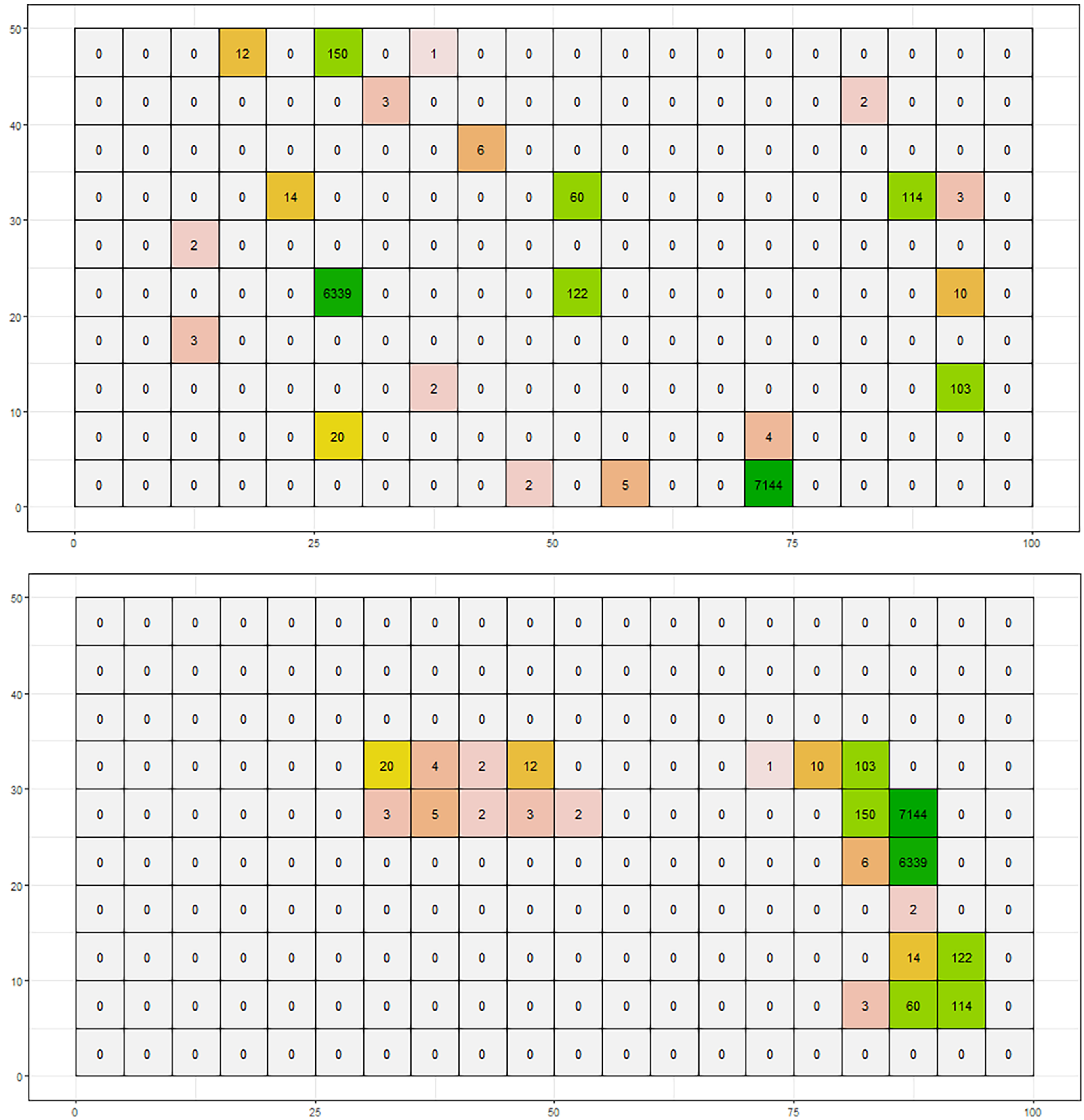

FIGURE 2 Top: scenario of a less clustered population with randomly distributed pixels containing teal (Pop A). Bottom: scenario of a more clustered population of pixels containing teal (Pop C).

AC initial or the CPoSA minimum sample size. For the estimation of the total, we use the modified Horvitz-Thompson estimator (Thompson, 1990) because, among the other estimators proposed by the same author, it is known to be a good compromise between low variance and computational effort (Smith et al., 1995).

Starting from the real population of blue-winged teal (henceforth Pop B), two further scenarios have been simulated. To assess the performance of the considered sampling strategies under different levels of spatial clustering, we generate a less clustered scenario, referred to as Pop A, by randomly distributing the pixels within the area (top panel of Figure 2). We also create a more clustered scenario, named Pop C, where the pixels containing teal are closer to each other (bottom panel of Figure 2). The level of spatial clustering, as measured by the mean length of the networks created by adjacent pixels containing teal, inspired by network sampling (Birnbaum & Sirken, 1965), is equal to 1.6 for Pop B, to 1 for Pop A, and 4.4 for Pop C. For each population, the simulation is based on 10,000 Monte Carlo runs.

3.1 | Simulation results

For the sake of simplicity, empirical results are reported only for the total number of teal as the population parameter to be estimated, from which density results may be easily derived.

With respect to AC and SRSWOR, the performance of CPoSA is evaluated considering: (1) the final sample size, that is, the number of pixels selected in each simulation run; (2) the number of observed teal in each simulated sample, which measures the ability of the sampling scheme to effectively detect the target species; (3) the accuracy of the final estimate; (4) the total costs.

As depicted in Figure 3, CPoSA effectively controls the final sample size. With respect to SRSWOR, if the minimum sample size is chosen slightly smaller than n, the final sample size of CPoSA is also smaller than the fixed size n under SRSWOR. Simulation results suggest that the minimum sample size under CPoSA should be set around 0.8 times smaller than the fixed size of a non-sequential adaptive traditional design. This choice of n^* determines a final sample size approaching n for the population with the highest level of spatial clustering.

Compared to AC, CPoSA uniformly and significantly reduces the final sample size as well as its variability, thus avoiding the selection of excessively small or excessively large samples.

Moreover, Figure 3 shows that the final sample size under AC tends to increase faster than under CPoSa as the spatial clustering increases. This notable difference can hamper the fair comparison between CPoSA and AC with respect to both their ability to detect teal and the accuracy of the final estimates, aspects that are strictly linked to the actual final sample size.

Figure 4 shows simulation results concerning the total number of teal. In particular, the left panel in Figure 4 highlights CPoSA potential to oversample blue-winged teal, rapidly outperforming SRSWOR, as the spatial clustering increases. The number of teal detected under CPoSA can be 2.3–3 times larger than the traditional SRSWOR design, as the minimum sample size increases up to the same size under SRSWOR. Figure 4 also shows that larger oversampling capacity of AC over CPoSA, although such superiority comes to the cost of much larger final sample sizes (Figure 3). Moreover, by considering the ratio of the number of observed teal over the final sample size (right panel of Figure 4), it is apparent that both sampling procedures ensure the oversampling capacity with respect to SRSWOR. Moreover, CPoSA and AC guarantee similar results for Pop A and Pop C, while in Pop B AC doubles the number of teal per pixel selected under CPoSA.

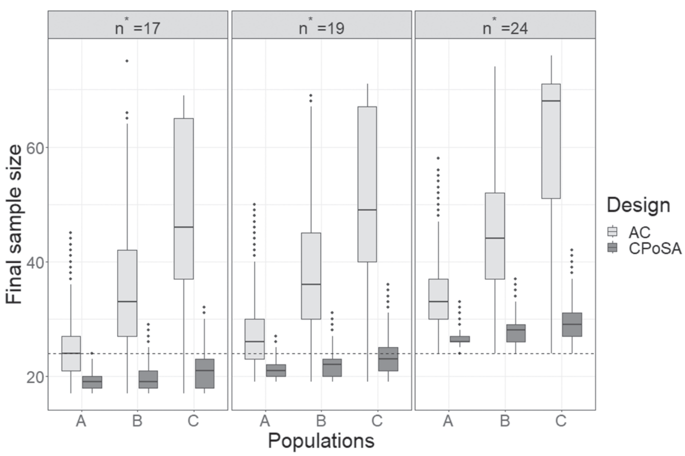

FIGURE 3 Boxplots of the Monte Carlo distributions of the final sample size under CPoSA (dark gray) and AC (light gray). Dashed line refers to the fixed sample size under SRSWOR.

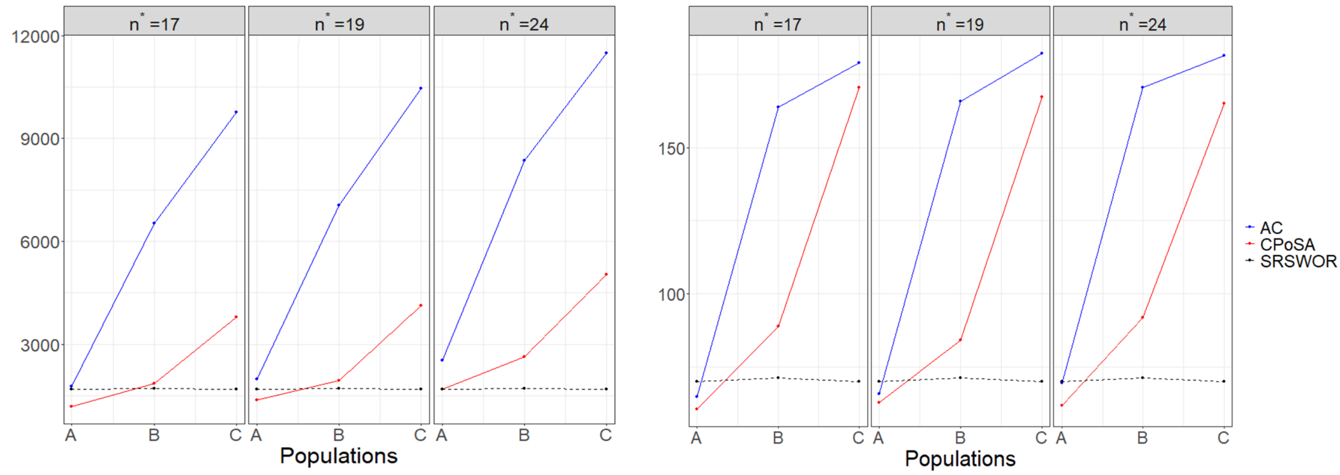

FIGURE 4 Left: number of observed teal in the three populations. Right: number of observed teal per selected pixels in the three populations.

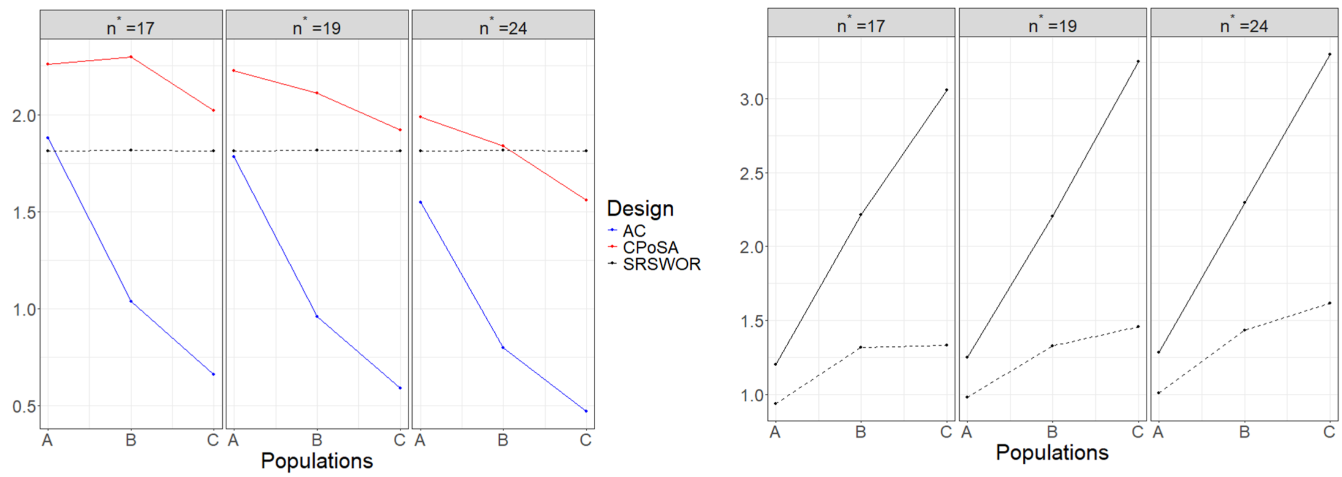

FIGURE 5 Left: accuracy of the final estimates in the three populations. Right: ratios of CPoSA RRMSE over AC RRMSE (solid line) and ratios of CPoSA weighted RRMSE over AC weighted RRMSE (dashed line).

The accuracy of the final estimate has been measured by the Monte Carlo Relative Root Mean Squared Error of the total estimator

$$\mathrm{RRMSE} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{T}_r - T\right)^2}/T,$$

where, R refers to the number of Monte Carlo runs in the simulation study, T is the true total number of teal and $\hat{T}_r$ is the number of teal estimated in the rth simulation run. In particular, $\hat{T}_r$ is the Horvitz-Thompson estimator for SRSWOR, the pseudo Horvitz-Thompson estimator for CPoSA and the modified Horvitz-Thompson estimator for AC. In the left panel of Figure 5, RRMSEs under CPoSA and those achieved under SRSWOR show similar performances for all three scenarios, with CPoSA values that quickly approach and then improve those of SRSWOR. On the other hand, AC performs better, but, once again, this could be due to the extremely different final sample sizes. Indeed, if we "penalize" the RRMSE multiplying it by the mean number of selected pixels, obtaining a weighted RRMSE, then the performances of CPoSA and AC are more similar (right panel of Figure 5). In particular, RRMSEs under CPoSA are 3.3 times higher than those under AC, whereas if we consider the weighted RRMSEs, the ratios are smaller than 1.6.

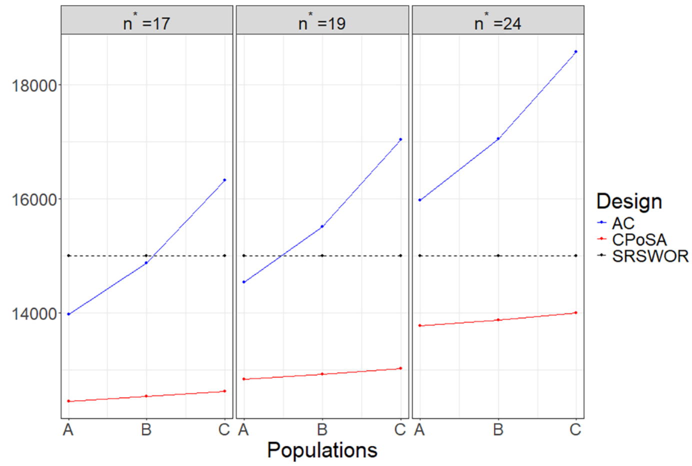

FIGURE 6 Average costs in the three populations under the three considered sampling schemes.

Regarding the final cost of the sampling procedures, a linear cost function has been determined following the one proposed by Thompson (2012)

$$C = c_0 + c_1 n' + c_2 n'',$$

where, c_0 represents the fixed costs, for example due to survey planning, on-field preparation, and team training; c_1 is the cost per visited pixel associated with the n' pixels selected: (1) as the initial sample size under AC, (2) because of the Bernoulli trial under CPoSA, and (3) as the fixed sample size under SRSWOR. Consequently, c_1 covers the costs of the actual teal searching procedure and those of traveling between pixels, which can be quite distant from each other. However, considering that CPoSA can control logistics, such as choosing the initial order of pixels by minimizing travel costs, a discount of 20% has been applied to c_1. A second marginal cost c_2 is assigned to the remaining n'' pixels, specifically the subsequent pixels selected under AC and the pixels selected after the previous selection of a non-empty pixel under CPoSA. Obviously, n'' is equal to zero under SRSWOR. It is worth noting that c_2 is chosen smaller than c_1 to reflect the reduced travel costs between neighboring pixels but maintaining the cost component due to the searching procedure within each pixel. Results in Figure 6 confirm the capacity of CPoSA of controlling the survey costs with respect to the other considered sampling designs.

It is worth noting that the cost-effectiveness is a key aspect to consider when planning a survey. Indeed, adaptive sampling techniques are often discarded because of the impossibility of controlling in advance the final sampling effort, and hence the total costs (Yang et al., 2011), as required by many ecological and environmental surveys (Turk & Borkowski, 2005). Moreover, if clusters are very large, the total costs, owing to the excessively large final sampling size, may be prohibitive (Gattone & Di Battista, 2011).

4 | DISCUSSION AND CONCLUDING REMARKS

Wintering waterfowl play a crucial role in biodiversity conservation and wetland habitat restoration. Gaining insight into their presence or absence in wetland areas is essential. However, accurate assessing of their total number or density can be challenging, as they tend to be highly spatially clustered, making them difficult to detect using traditional sampling strategies. In this paper, we explore the application of a sequential adaptive sampling strategy to rare and spatially clustered animal populations. The proposed strategy, namely Conditional Poisson Sequential Adaptive (CPoSA), aims to oversample the rare and clustered interest variable while managing logistics and budget constraints. Through a simulation study, we show that CPoSA outperforms the traditional SRSWOR with respect to all the considered aspects. Furthermore, adding a substantial variation from the work of Mecatti et al. (2023), in the simulation study we have also compared the performance of CPoSA with respect to Adaptive Cluster (AC) sampling (Thompson, 2012) and obtained original results with a clear practical advantage. Indeed, CPoSA proves to be able to improve the tendency of AC to provide unpredictable large samples, which is a well-known drawback of adaptive sampling that can limit applicability. In our simulation CPoSA provides sample sizes significantly smaller and more stable than AC. At the same time, CPoSA shows an effective oversampling capacity while maintaining cost-effectiveness, which increases with a careful planning of the minimum sample size.

Empirical findings highlight the potential of sequential adaptive sampling strategies when applied to animal populations, but further research is warranted to explore their full capacity. In particular, other choices of the updating rule, better tailored to animal populations, can yield further improvements compared to traditional non-sequential adaptive random sampling. A simple modification involves tightening the condition required for indicator D_i to be equal to 1, in contrast with the current broader condition "the characteristic of interest is detected at least once in the selected pixel." This might entail establishing a "threshold" specifying the minimum quantity of characteristic of interest to be observed in order to trigger the updating rule. By implementing this modification in the simulation study, for example specifying that at least two teal need to be detected in the selected pixel, CPoSA would avoid selecting teal-less pixels located after a pixel with only two teal and no cluster. This would help to further reduce the final sample size for the same total number of observed teal, leading to reduced costs.

Future research will consider exploiting available prior knowledge and auxiliary information to improve efficiency. For instance, auxiliary variables of the survey area, such as the location of shelters or areas with increased availability of food, can be employed to assign unequal initial inclusion probabilities to pixels, that is, higher probability to pixels where shelters/more food are available and animals are more likely to be counted, and reduced inclusion probability otherwise. Finally, spatial coordinates of pixels can be exploited as auxiliary variables to avoid the unidimensional ordering by defining a distance function between pixels within a two-dimensional space (see for instance spatially correlated Poisson sampling in Grafström, 2012).

ACKNOWLEDGMENTS

Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP B63C22000650007, Project title "National Biodiversity Future Center - NBFC". Open access publishing facilitated by Universita degli Studi di Milano-Bicocca, as part of the Wiley - CRUI-CARE agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Fulvia Mecatti https://orcid.org/0000-0002-0403-9231

REFERENCES

Alagador, D., Cerdeira, J. O., & Araujo, M. B. (2014). Shifting protected areas: Scheduling spatial priorities under climate change. *Journal of Applied Ecology*, *51*, 703–713.

Birnbaum, Z. W., & Sirken, M. G. (1965). Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates. *Public Health Service*, 1000.

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*, 1164–1168.

Cassel, C., Särndal, C., & Wretman, J. H. (1977). *Foundations of inference in survey sampling*. Wiley.

Delany, S., Reyes, C., Hubert, E., Pihl, S., Rees, E., Haanstra, L., & Van Strien, A. (1999). *Results from the International Waterbird Census in the Western Palearctic and Southwest Asia 1995 and 1996* (p. 54). Wetlands International Publication.

Di Biase, R. M., Fattorini, L., & Marcelli, M. (2024). *A design-based view of species richness estimation in environmental surveys*.

Dickson, M. M., & Tillé, Y. (2016). Ordered spatial sampling by means of the traveling salesman problem. *Computational Statistics*, *31*, 1359–1372.

Gattone, S. A., & Di Battista, T. (2011). Adaptive cluster sampling with a data driven stopping rule. *Statistical Methods & Applications*, *20*, 1–21.

Gilissen, G., Haanstra, L., Delany, S., Boere, G., & Hagemeijer, W. (2002). *Numbers and distribution of wintering waterbirds in the Western Palearctic and Southwest Asia in 1997, 1998 and 1999. Results from the International Waterbird Census*. Wetlands International Global Series 11.

Grafström, A. (2012). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference*, *142*, 139–147.

Green, A. J., & Elmberg, J. (2013). Ecosystem services provided by waterbirds. *Biological Reviews*, *89*, 105–122.

Hankin, D. G., Mohr, M. S., & Newman, K. N. (2019). *Sampling theory: For the ecological and natural resource sciences*. Oxford University Press.

Mecatti, F., Sismanidis, C., Furfaro, E., & Conti, P. L. (2023). Sequential adaptive strategies for sampling rare clustered populations. *Statistical Methods & Applications*, *32*, 1659–1693.

Nichols, J. D., Reinecke, K. J., & Hines, J. E. (1983). Factors affecting the distribution of mallards wintering in the Mississippi Alluvial Valley. *The Auk*, *100*, 932–946.

Pihl, S., Webb, A., Frikke, J., & Durinck, J. (1992). *The importance of monitoring waterfowl and seabirds*. In J. Komdeur, J. Bertelsen, & G. Cracknell (Eds.), *Manual for aeroplane and ship surveys of waterfowl and seabirds*. IWRB Special Publication.

Rands, M. R. W., Adams, W. M., Bennun, L., Butchart, S. H. M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J. P., Sutherland, W. J., & Vira, B. (2010). Biodiversity conservation: Challenges beyond 2010. *Science*, *329*, 1298–1303.

Smith, D. R., Conroy, M. J., & Brakhage, D. H. (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, *51*, 777–788.

Stevens, D. L., Jr., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, *99*, 262–278.

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, *85*, 1050–1059.

Thompson, S. K. (2006). Adaptive web sampling. *Biometrics*, *62*, 1224–1234.

Thompson, S. K. (2012). *Sampling*. Wiley.

Tillé, Y. (2006). *Sampling algorithms*. Springer.

Turk, P., & Borkowski, J. J. (2005). A review of adaptive cluster sampling: 1990–2003. *Environmental and Ecological Statistics*, *12*, 55–94.

WHO. (2011). *Tuberculosis prevalence surveys: A handbook (the lime book)*. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/44481/9789241548168_eng.pdf

Wortley, L., Hero, J. M., & Howes, M. (2013). Evaluating ecological restoration success: A review of the literature. *Restoration Ecology*, *21*, 537–543.

Yang, H., Kleinn, C., Fehrmann, L., Tang, S., & Magnussen, S. (2011). A new design for sampling with adaptive sample plots. *Environmental and Ecological Statistics*, *18*, 223–237.

How to cite this article: Di Biase, R. M., & Mecatti, F. (2024). Applying sequential adaptive strategies for sampling animal populations: An empirical study. *Environmetrics*, e2870. https://doi.org/10.1002/env.2870