

Environmental and Ecological Statistics manuscript No. (will be inserted by the editor)

A framework for adapting survey design through time for wildlife population assessment

Fiona M. Underwood

Received: date / Accepted: date

Abstract Sampling strategies for monitoring the status and trends in wildlife populations are often determined before the first survey is undertaken. However, there may be little information about the distribution of the population and so the sample design may be inefficient. Through time, as data are collected, more information about the distribution of animals in the survey region is obtained but it can be difficult to incorporate this information in the survey design. This paper introduces a framework for monitoring motile wildlife populations within which the design of future surveys can be adapted using data from past surveys whilst ensuring consistency in design-based estimates of status and trends through time. In each survey, part of the sample is selected from the previous survey sample using simple random sampling. The rest is selected with inclusion probability proportional to predicted abundance. Abundance is predicted using a model constructed from previous survey data and covariates for the whole survey region. Unbiased design-based estimators of status and trends and their variances are derived from two-phase sampling theory. Simulations over the short and long-term indicate that in general more precise estimates of status and trends are obtained using this mixed strategy than a strategy in which all of the sample is retained or all selected with probability proportional to predicted abundance. Furthermore the mixed strategy is robust to poor predictions of abundance. Estimates of status are more precise than those obtained from a rotating panel design.

Keywords adaptive sampling · design-based estimation · monitoring strategies · wildlife population assessment

FM Underwood
Department of Mathematics and Statistics, University of Reading, Whiteknights, Reading
RG6 6BX UK
Tel.: +44 (0)118 378 7927
Fax: +44 (0)118 378 8032
E-mail: f.m.underwood@reading.ac.uk

1 Introduction

This paper is motivated by the requirement of many wildlife managers to obtain efficient estimates of the total number of individuals in a population they are managing and to monitor how this changes through time. In essence these questions can be answered by a series of surveys at times $t = 1, 2, \dots$, from which an estimate of $\tau^{(t)}$, the total number of individuals in the population at time t , can be obtained. Ideally these surveys are designed as part of a monitoring programme so that $\delta^{(t',t)} = \tau^{(t)} - \tau^{(t')}$ can also be efficiently estimated. This paper considers the question of which locations within a survey region should be sampled in any one survey so that efficient design-based estimates of $\tau^{(t)}$ and $\delta^{(t',t)}$ can be obtained.

In a design-based framework the most efficient estimate of $\tau^{(t)}$ would be obtained when the probability of inclusion for a sampling location is proportional to the response variable - in this case $y_i^{(t)}$, the number of individuals found at location i at time t . This is clearly never possible but if inclusion probabilities can be related to covariates that are correlated with $y_i^{(t)}$, the estimate of $\tau^{(t)}$ will be more efficient than if all units are selected with equal probability.

At the start of a monitoring programme, there may be little known about how $y_i^{(t)}$ varies over the survey region. Although there will often be auxiliary information about the survey region, such as habitat, distance from human disturbance, topography etc., the relationship between these variables and $y_i^{(t)}$ may not be clear so that this information cannot be used to determine inclusion probabilities. Alternatively the relationship may be known for some or all of the covariates individually but combined together it may be difficult to see how features of the overall survey design, such as stratification criteria or inclusion probabilities, should be determined.

Once a survey has been conducted, the survey data can be used to construct a model from which predictions of how species abundance varies through the survey region can be made. This paper proposes a design-based framework for incorporating this knowledge into the design of future surveys by using these model-based estimates to determine inclusion probabilities in future surveys. This is similar in spirit to the approach of Hansen et al (1983):

...design decisions may be guided and evaluated by models, but inferences concerning population characteristics should be made on the basis of induced randomization...

In a monitoring programme it is not just $\tau^{(t)}$ that is required, but $\delta^{(t',t)}$ also. If $y_i^{(t)}$ the number of individuals observed at location i at time t is expected to be correlated with $y_i^{(t')}$ the number observed at that location in a previous survey, then a precise estimate of $\delta^{(t',t)}$ is obtained by sampling the same units in survey t as survey t' . In this paper the populations being considered are motile, that is individuals of the species can cover the whole of the survey region in a much smaller time than the time between surveys. Even

so, the distribution of the population over the survey region is expected to be relatively constant between surveys, unless there have been large changes in the environment, for example habitat or climate changes, or migrational shifts. That is, if $\mu_i^{(t)} = k\mu_i^{(t')}$ is the expected number of individuals at location i at time t then $y_i^{(t)}$ and $y_i^{(t')}$ will be correlated. A precise estimate of change in the population total, $\hat{\delta}^{(t',t)}$, could therefore be obtained by sampling the same units in each survey. However if the original sample design was not very efficient, then future survey design would also be inefficient. If however an entirely new design is used based on what has been learnt from the previous survey then again the estimate of $\hat{\delta}^{(t',t)}$ may be inefficient.

Rotating panel designs first proposed by Jessen (1942) and reviewed by Duncan and Kalton (1987) and Binder and Hidioglou (1988) are strategies where part of the sample is retained from one survey to another, to obtain precise estimates of $\delta^{(t',t)}$, and the rest of the sample is selected from those units not sampled previously. The idea is that by covering a greater part of the sample space a better estimate of $\tau^{(t)}$ can be obtained. For these designs the sampling strategy is determined at the start of the monitoring programme and remains fixed through time, even if the units in the sample vary. Hence for wildlife populations, if little is known at the start about the species distribution, the survey design will remain inefficient. Currently, these designs are rarely used in wildlife population assessment, although they were proposed by Skalski (1990). They are more commonly used in environmental monitoring - for example in the Environmental Monitoring Assessment Programme (EMAP) (Overton et al, 1990) - and in social surveys - for example the Canadian Labor Force Survey (Dufour et al, 1998)).

To deal with environmental change, Overton and Stehman (1996) propose a strategy for redefining strata to improve sample efficiency during the lifetime of a monitoring programme. They advise that this adjustment should only occur once or twice within the course of the monitoring programme. However, more frequent updating may be important when much is being learnt about the distribution of animals over the survey region in the first few surveys, or if it is varying through time. The methods developed in this paper are an initial step in addressing this problem.

This paper introduces a common notation before describing the general design framework and derivation of estimators. A simulated population is used to demonstrate the effectiveness of these methods when selecting a second survey after obtaining a model of predicted abundance from the first survey. Simulations are also used to investigate the robustness of the methodology if predicted abundance is a biased estimate of true abundance. Results are compared with standard rotating panel surveys. The paper also discusses and illustrates the application of the methods over a 10 year monitoring study in which a survey is taken each year.

2 Method

2.1 Notation

Consider a survey region U divided into N contiguous units $U = \{1, \dots, N\}$. There is information about the survey region in the form of a number p of covariates so that x_{i1}, \dots, x_{ip} are the value of the p covariates for unit i . These covariates might be physical characteristics of the survey region such as: elevation or aspect (eg Opsomer et al, 2003); habitat types based on some classification scheme (eg Buckland and Elston, 1993) who used those from the National Habitat Survey of Grampian Region; or human influences such as distance to transport networks (eg Khaemba and Stein, 2000). In addition, each unit is georeferenced providing two further covariates on location. At time t , the number of individuals in unit i is $y_i^{(t)}$, so the total number of individuals in the population is $\tau^{(t)} = \sum_{i=1}^N y_i^{(t)}$ which can be written $\sum_{i \in U} y_i^{(t)}$. The change in the population between time t and time t' is $\delta^{(t',t)} = \tau^{(t)} - \tau^{(t')}$ where $t' < t$.

In survey t a sample $s^{(t)} \subseteq U$ of n units is taken. Assuming that there is perfect detectability $y_i^{(t)}$ is recorded for these n units and a design-based estimate of $\tau^{(t)}$ is obtained using these data. In addition the data are used to predict $E[y_i^{(t)}] = \mu_i^{(t)}$ the expected number of animals in each unit in the survey region using a statistical model $\zeta^{(t)}$ fitted to data from the sampled units.

2.2 Design

At the start of the monitoring programme, $t = 1$, little is known about the distribution of animals over the survey region so the first sample $s^{(1)}$ is selected using simple random sampling without replacement (*srswor*). The data from the sample $s^{(1)}$ and the explanatory variables are used to obtain $\zeta^{(1)}$ from which $\mu_i^{(1)}$ is predicted for all units in the survey region.

In the second survey, at time $t = 2$, a sample $s^{(2)}$ is selected, also of size n . Part of the sample, say $s_1^{(2)}$ of size $n_1^{(2)}$, is selected from the first survey sample $s^{(1)}$ using *srswor*. The rest of the sample, $s_2^{(2)}$ of size $n_2^{(2)} = n - n_1^{(2)}$, is selected without replacement with inclusion probability proportional to $\hat{\mu}_i^{(1)}$ from $s_c^{(1)} = U - s^{(1)}$, those units that were not included in survey 1, so that $s_2^{(2)} \subseteq s_c^{(1)}$.

In future surveys, the strategy is similar. In survey t , $s_1^{(t)}$ can be selected from $s^{(1)}$ using *srswor*, or can be the same as $s_1^{(2)}$, depending on whether it is $\delta^{(1,t)}$ or $\delta^{(t',t)}$ where $t' > 1$ that is of most interest. Then $s_2^{(t)}$ will be selected with probability proportional to predicted abundance, $\hat{\mu}_i^{(t)}$, from $U - s^{(1)}$ or $U - s_1^{(2)}$ respectively.

2.3 Models of predicted abundance

After survey t data from all previous surveys, $t = 1, \dots, t$ can be used to construct a model $\zeta^{(t)}$ which describes the relationship between auxiliary data and species abundance. Because the $y_i^{(t)}$ are count data, the type of model used to predict the abundance of the species over the survey region will often be of the form: $y_i^{(t)} \sim \text{Poisson}(\mu_i^{(t)})$ although there will often be overdispersion. Generalised linear models McCullagh and Nelder (1989) or generalised additive models Wood (2006) can be used to model the relationship between the explanatory variables and the observed counts. Smooth terms for location, such as latitude and longitude, perhaps using a thin-plate spline, might be included in the model to describe spatial variability in the data. A very general population model would therefore be of the form:

$$\log(\mu_i^{(t)}) = \sum_{j=0}^P f_j^{(t)}(x_{ij}^{(t)}) + g(t)$$

where $f_j^{(t)}(x_{ij}^{(t)})$ is a linear or smooth function of the j^{th} auxiliary variable at time t and $g(t)$ is a function, linear or smooth, of time. Simpler versions of the model include those with a constant population size through time or with covariate effects remaining constant over time so that $\log(\mu_i^{(t)}) = \sum_{j=0}^P f_j(x_{ij}^{(t)}) + g(t)$. Then, data from all surveys would be equally important for modelling the relationship between habitat and abundance. By contrast, if $\zeta^{(t)}$ models a changing relationship between habitat and abundance, for example in the presence of climate change, then data from past surveys will contribute less to the construction of $\zeta^{(t)}$.

To design the survey at time $t + 1$, the model $\zeta^{(t)}$ is used to predict $\hat{\mu}_i^{(t+1)}$. This may be equivalent to $\hat{\mu}_i^{(t)}$ if there have been no habitat changes between the two surveys and the relationship between habitat and abundance remains constant over time. One alternative would be to use the same model $\zeta^{(t)}$ but recalculate $\hat{\mu}_i^{(t+1)}$ using updated values of the covariates.

2.4 Estimation

Here, a rationale and estimators for $\tau^{(2)}$ and $\delta^{(1,2)}$ given that $s_1^{(2)}$ is selected using *srswor* from $s^{(1)}$, and $s_2^{(2)}$ is selected with probability proportional to $\hat{\mu}_i^{(1)}$ from $s_c^{(1)}$ is given. The Appendix provides more details and general results for two surveys $s^{(t)}$ and $s^{(t)}$ and demonstrates that these estimators are unbiased.

Let $\pi_i^{(t)} = \text{Pr}(i \in s^{(t)})$ and $\pi_{ij}^{(t)} = \text{Pr}(i \& j \in s^{(t)})$. In survey $t = 1$ these are:

$$\pi_i^{(1)} = \sum_{s^{(1)} \ni i} p(s^{(1)}) = \frac{n}{N} \text{ and } \pi_{ij}^{(1)} = \frac{n(n-1)}{N(N-1)}$$

Standard estimators are derived using the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) and the Sen-Yates Grundy variance estimator (Yates and Grundy, 1953; Sen, 1973).

In the second survey, when some sample units in $s^{(2)}$ are selected from $s^{(1)}$ then

$$\pi_i^{(2)} = Pr(i \in s_1^{(2)} | i \in s^{(1)})Pr(i \in s^{(1)}) + Pr(i \in s_2^{(2)} | i \in s_c^{(1)})Pr(i \in s_c^{(1)})$$

The probability $Pr(i \in s_2^{(2)} | i \in s_c^{(1)}) = \sum_{s_c^{(1)} \ni i} \pi_{i_2 | s^{(1)}}^{(2)} p(s^{(1)})$ where $\pi_{i_2 | s^{(1)}}^{(2)}$ is the probability that unit i is included in $s_2^{(2)}$ given that the sample $s^{(1)}$ has been selected. This requires calculating the probability that unit i is included in $s_2^{(2)}$ for all possible samples $s^{(1)}$. This probability depends on $\hat{\mu}_i^{(1)}$ and the value of $\hat{\mu}_i^{(1)}$ for all other units in $s_c^{(1)}$. However $\hat{\mu}_i^{(1)}$ will depend on the sample $s^{(1)}$ that is selected. As only one sample is taken it is not possible to evaluate the second half of $\pi_i^{(2)}$ directly.

Instead, the principles of two-phase sampling and estimation can be used to develop estimators of $\tau^{(2)}$. Särndal et al (1992) describe unbiased design-based estimators for two phase sampling in which a sample s_2 is an unequal probability sample selected conditional on the selection of a sample s_1 . It is often used when s_2 is contained within s_1 , from which only auxiliary information is obtained. One use of these two-phase sampling schemes is for sampling through time, in which part of the sample is retained from one survey to another.

The estimate $\hat{\tau}^{(2)}$ is a weighted average, $\hat{\tau}^{(2)} = \omega \hat{\tau}_1^{(2)} + (1 - \omega) \hat{\tau}_2^{(2)}$, where $\hat{\tau}_k^{(2)}$ uses the data from $s_k^{(2)}$ only. Ideally, ω is selected to minimise the variance of $\hat{\tau}^{(2)}$ but in practice, given estimates of $var(\hat{\tau}_k^{(2)})$, $var(\hat{\tau}^{(2)})$ and $cov(\hat{\tau}_1^{(2)}, \hat{\tau}_2^{(2)})$, ω can be selected to minimise the estimated variance $\widehat{var}(\hat{\tau}^{(2)})$. Alternatively, weights can be based on the relative sample sizes so that $\omega = \frac{n_1}{n}$ as is implemented here.

The estimate $\hat{\tau}_1^{(2)}$ and its variance are obtained using the data from $s_1^{(2)}$. Because units are selected using simple random sampling without replacement, the inclusion probabilities $\pi_{i_1}^{(2)} = Pr(i \in s_1^{(2)} | i \in s^{(1)})Pr(i \in s^{(1)})$ are $\pi_{i_1}^{(2)} = \frac{n_1}{n} \frac{n}{N} = \frac{n_1}{N}$. Similarly the joint inclusion probability $\pi_{(ij)_1}^{(2)} = \frac{n_1(n_1-1)}{N(N-1)}$ and so the Horvitz-Thompson estimator and the Sen-Yates-Grundy estimator of the variance can be used to obtain $\hat{\tau}_1^{(2)}$ and $\widehat{var}(\hat{\tau}_1^{(2)})$.

To obtain $\hat{\tau}_2^{(2)}$ a similar argument to Särndal et al (1992) can be followed. The sample $s_2^{(2)}$ is a probability sample of $s_c^{(1)}$ which itself is a probability sample of U . So $\pi_{i_c}^{(1)} = Pr(i \in s_c^{(1)}) = \frac{N-n}{N}$ and an unbiased estimate of $\tau^{(2)}$ would be $\sum_{i \in s_c^{(1)}} \frac{y_i^{(2)}}{\pi_{i_c}^{(1)}} = \frac{N-n}{n} \sum_{s_c^{(1)}} y_i^{(2)}$. The $y_i^{(2)}$ are only known for $i \in s_2^{(2)}$

rather than for all $i \in s_c^{(1)}$. Because $s_2^{(2)}$ is a probability sample from $s_c^{(1)}$ where unit i is selected with the inclusion probability $\pi_{2 | s^{(1)}}^{(2)}$ proportional to $\hat{\mu}_i^{(1)}$ an unbiased estimator of $\sum_{i \in s_c^{(1)}} y_i^{(2)}$ is $\sum_{i \in s_2^{(2)}} \frac{y_i^{(2)}}{\pi_{2 | s^{(1)}}^{(2)}}$.

Hence $\tau_2^{(2)}$ is estimated as:

$$\hat{\tau}_2^{(2)} = \sum_{s_2^{(2)}} \frac{y_i^{(2)}}{\pi_{i|s_c^{(1)}}^{(2)} \pi_{i_c}^{(1)}} = \sum_{s_2^{(2)}} \frac{y_i^{(2)}}{\pi_i^{(2)\dagger}}$$

and the probability $\pi_k^{(2)\dagger} = \pi_{k|s_c^{(1)}}^{(2)} \pi_{k_c}^{(1)}$ where $k = i$ or $k = ij$ for joint inclusion probabilities.

The variance of this estimator has two components that express the variability in estimating $\hat{\tau}_2^{(2)}$ using the units from $s_c^{(1)}$ and the expected variance given that the sample $s_c^{(1)}$ has been selected. Its estimator can be calculated explicitly as:

$$\widehat{var}(\hat{\tau}_2^{(2)}) = \sum_{s_2^{(2)}} \sum_{s_2^{(2)}} \frac{\Delta_{(ij)_c}^{(1)} y_i^{(2)} y_j^{(2)}}{\pi_{ij}^{(2)\dagger} \pi_{i_c}^{(1)} \pi_{j_c}^{(1)}} + \sum_{s_2^{(2)}} \sum_{s_2^{(2)}} \frac{\Delta_{(ij)2|s_c^{(1)}} y_i^{(2)} y_j^{(2)}}{\pi_{(ij)2|s_c^{(1)}}^{(2)} \pi_i^{(2)\dagger} \pi_j^{(2)\dagger}}$$

where $\Delta_{(ij)k} = \pi_{(ij)_k} - \pi_{(i)_k} \pi_{(j)_k}$

The covariance $cov(\hat{\tau}_1^{(2)}, \hat{\tau}_2^{(2)})$ is the negative of the variance of the data in U . If all N units had the same value so that $y_i^{(2)} = y^{(2)}$ then the covariance would be zero as the totals $\hat{\tau}_1^{(2)}$ and $\hat{\tau}_2^{(2)}$ would remain the same for all possible $s_1^{(2)}$ and $s_2^{(2)}$. Its estimator only uses the data from $s_1^{(2)}$. This is reasonable because $s_1^{(2)}$ is a probability sample from U but it does mean that if n_1 is small then the estimator will not be very precise.

The estimated variance of $\hat{\delta}^{(1,2)} = \hat{\tau}_2^{(2)} - \hat{\tau}_1^{(1)}$ requires the covariance term $cov(\hat{\tau}_2^{(2)}, \hat{\tau}_1^{(1)})$, which is a weighted average of $cov(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)})$ and $cov(\hat{\tau}_1^{(1)}, \hat{\tau}_2^{(2)})$. These covariances require information from units that were sampled in both time periods so only those data from $s_1^{(2)}$ are used. These covariances will be poorly estimated when $s_1^{(2)}$ is small. Further work to develop estimators that use all the data would be useful.

3 Simulation

A population was generated for a square area A . Four covariates were generated and combined to produce a density surface $\lambda_{\mathcal{L}}$. An inhomogeneous Poisson process with intensity $\lambda_{\mathcal{L}}$ was used to generate populations for the two surveys. No trend in the population or in the density surface was incorporated. Further details of how $\lambda_{\mathcal{L}}$ and the four covariates were generated are provided in the on-line information.

The area A was divided into $N = 1296$ units in a 36 x 36 grid. The value of each covariate was recorded for the central point in each unit, and the number of individuals in unit i at time t recorded as $y_i^{(t)}$.

3.1 Survey 1

In survey 1 a without replacement simple random sample $s^{(1)}$ of $n = 100$ units was taken and $y_i^{(1)}$ recorded for $i \in s^{(1)}$ and an estimate of $\tau^{(1)}$ obtained. These data and covariate information for the n units in $s^{(1)}$ were also used to construct a model of the form: $y_i^{(1)} \sim \text{Poisson}(\mu_i^{(1)})$ where $\log(\mu_i^{(1)}) = \beta_0 + \sum_{k=1}^4 \beta_k x_{ik}$. A step-wise AIC procedure Venables and Ripley (2002) was then used to select covariates for the final model. This model was then used to predict $\hat{\mu}_i^{(1)}$ for all units in the survey region.

3.2 Survey 2

In survey 2 a sample $s_1^{(2)}$ of n_1 units was selected from $s^{(1)}$ and a sample $s_2^{(2)}$ of n_2 units was selected with probability proportional to $\hat{\mu}_i^{(1)}$ from $s_c^{(1)}$. Five different values of $n_1 = 100, 75, 50, 25, 0$ were selected. When $n_1 = 100$ all units from the first survey, $s^{(1)}$, are retained and when $n_1 = 0$, all units are selected with probability proportional to $\hat{\mu}_i^{(1)}$ from $s_c^{(1)}$. Figure 1 shows an example for one set of $\hat{\mu}_i^{(1)}$ from a model $\zeta^{(1)}$ where different proportions of $s^{(2)}$ were sampled with inclusion probability proportional to $\hat{\mu}_i^{(1)}$. As n_2 increases the number of samples seen in the high density areas, the light coloured areas, increases. This whole procedure was repeated $B = 1000$ times so that 1000 samples for survey 1 were taken and a sample from each of these taken for the five values of n_1 . In addition, a second set of surveys was run to compare with results with a repeating panel survey. In this case all of the second survey sample $s^{(2)}$ was selected using *srswor* where $s_1^{(2)}$ of size n_1 was selected from $s^{(1)}$ and $s_2^{(2)}$ of size n_2 from $s_c^{(1)}$.

Sunter's method (Sunter, 1977a,b) was used to select units with inclusion probability proportional to $\hat{\mu}_i^{(1)}$. Using this strategy, units are ordered from those with largest $\hat{\mu}_i^{(1)}$ to smallest $\hat{\mu}_i^{(1)}$ and units with small values of $\hat{\mu}_i^{(1)}$, below a certain threshold, are selected with simple random sampling. This could be seen as smoothing $\hat{\mu}_i^{(1)}$ when $\hat{\mu}_i^{(1)}$ is small. Sunter states that it is often desirable that the smallest valued units are sampled with equal probability as the correlation between the size measure and the variable of interest can become unstable for these units. This seems appropriate here as interest is in the units with high values of $\hat{\mu}_i^{(1)}$. Alternative methods that could be used are Chao (Chao, 1982) or those described by Brewer and Hanif (1982) or Tillé (2006).

Results are given in table 1. Unbiased estimates of $\tau^{(2)}$, and $\hat{\delta}^{(1,2)}$ were obtained using the new method. The estimated variance $\widehat{\text{var}}(\hat{\tau}^{(2)})$ decreases as the proportion of the sample selected with probability proportional to $\hat{\mu}_i^{(2)}$ increases. In comparison, when $s_2^{(2)}$ is selected using *srswor* the estimated variance $\sqrt{\widehat{\text{var}}(\hat{\tau}^{(2)})}$ remains constant as n_2 increases. Retaining units from $s^{(1)}$ to $s^{(2)}$ leads to a more precise estimate of $\hat{\delta}^{(t',t)}$ because $\text{cor}(y_i^{(1)}, y_i^{(2)}) > 0$

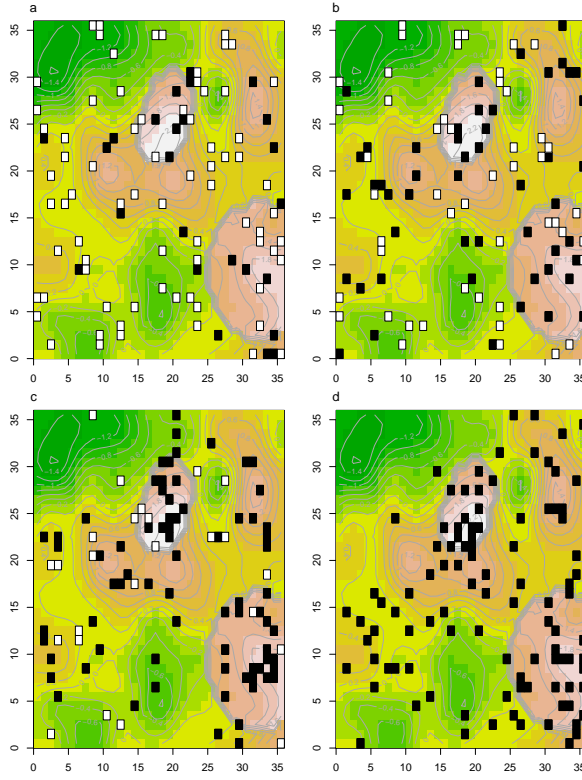


Fig. 1 Illustrations of different designs for predicted abundance. Lighter areas in the survey region indicate higher abundance. White squares indicate $s_1^{(2)}$, those retained from $s_1^{(1)}$, and black squares indicate those sampled with probability proportional to $\hat{\mu}_i^{(t)}$

Table 1 Results from survey 2 where from a total sample size of $n = 100$ units, $s_2^{(2)}$ of size n_2 is selected with inclusion probability proportional to $\hat{\mu}_i^{(1)}$ and $s_1^{(2)}$ of size $n_1 = n - n_2$ is selected from $s_1^{(1)}$ using simple random sampling without replacement. Results show mean (and standard deviation) from the 1,000 simulations of: $\hat{\tau}^{(2)}$ and $\sqrt{\widehat{\text{var}}(\hat{\tau}^{(2)})}$ when $s_2^{(2)}$ of size $n_2^{(2)}$ is selected with inclusion probability proportional to $\hat{\mu}_i^{(1)}$ from $s_c^{(1)}$; $\sqrt{\widehat{\text{var}}(\hat{\tau}^{(2)})}$ srs when $s_2^{(2)}$ of size $n_2^{(2)}$ is selected using simple random sampling from $s_c^{(1)}$, and $\sqrt{\widehat{\text{var}}(\hat{\delta}^{(1,2)})}$ under both sampling strategies just described.

n_2	$\hat{\tau}^{(2)}$	$\sqrt{\widehat{\text{var}}(\hat{\tau}^{(2)})}$	$\sqrt{\widehat{\text{var}}(\hat{\tau}^{(2)})}$ srs	$\widehat{\text{var}}(\hat{\delta}^{(1,2)})$	$\widehat{\text{var}}(\hat{\delta}^{(1,2)})$ srs
0	2574 (362)	366 (96)	366 (96)	244 (30)	244 (30)
25	2578 (319)	334 (90)	366 (96)	301 (62)	242 (37)
50	2579 (309)	300 (82)	364 (96)	352 (74)	360 (80)
75	2565 (276)	260 (67)	363 (96)	396 (75)	446 (86)
100	2566 (221)	218 (34)	364 (100)	418 (75)	515 (94)

and so as the proportion retained from $s^{(1)}$ decreases the variance $\widehat{var}(\hat{\delta}^{(1,2)})$ increases. The increase in the variance as n_2 increases is less when the units in $s_2^{(2)}$ are selected with inclusion probability proportional to $\hat{\mu}_i^{(2)}$ than when $s_2^{(2)}$ is selected using *srswor*. This is because in the new design the variance of $\hat{\tau}_2^{(2)}$ reduces as n_2 increases, whereas it remains constant under *srswor*.

3.3 Efficiency measure

From table 1 it is clear that the efficiency of $\hat{\tau}^{(2)}$ increases as $\frac{n_2}{n}$ increases. For a fixed n_1 however the efficiency of the sampling strategy will depend on the relationship between $y_i^{(t)}$ and $\hat{\mu}_i^{(t)}$. If $\mu_i^{(t)}$ were known for all units in the population, then the efficiency of the sampling strategy would depend on the distribution of $\mu_i^{(t)}$ and the variability of $y_i^{(t)}$ about $\mu_i^{(t)}$. Simulations have indicated that there is an increase in the relative efficiency of $\hat{\tau}^{(2)}$ when sampling with inclusion probability proportional to $\hat{\mu}_i^{(t)}$ compared to under *srswor* as $var(y_i^{(t)})$ decreases and $\mu_i^{(t)}$ becomes more right-skewed.

In practice the $\mu_i^{(t)}$ must be estimated. A simple heuristic measure b where $\hat{\mu}_i^{(t)} = \mu_i^{(t)b}$ describes how well $\hat{\mu}_i^{(t)}$ estimates $\mu_i^{(t)}$. Ignoring the superscript for time, assume that the y_i are generated from a Poisson distribution with mean μ_i such that

$$\log(\mu_i) = \sum_{j=0}^Q \beta_j x_{ij}$$

An estimate of μ_i , using the same auxiliary variables, will be of the form

$$\log(\hat{\mu}_i) = \sum_{j=0}^Q \hat{\beta}_j x_{ij}$$

If for simplicity we assume that parameter estimates are related in a linear fashion to the parameters so that

$$\hat{\beta}_j = a_j^* + b_j^* \beta_j$$

we can write

$$\log(\hat{\mu}_i) = \sum_{j=0}^Q (a_j^* + b_j^* \beta_j) x_{ij}$$

For convenience, in particular to obtain a tractable expression, we assume that the relationship between parameter estimates and parameters is such that $b_j^* = b$ for $j = 0, \dots, Q$ so

$$\log(\hat{\mu}_i) = \sum_{j=0}^Q a_j^* x_{ij} + b \log(\mu_i) \Rightarrow \hat{\mu}_i = a \mu_i^b$$

The parameter $a = \sum_{j=0}^Q a_j^* x_{ij}$ is a scaling parameter and will have no effect on inclusion probabilities if sampling with probability proportional to $\hat{\mu}_i^{(t)}$. A value of b close to 1 therefore indicates that $\hat{\mu}_i^{(t)}$ is a good estimate of $\mu_i^{(t)}$ when it is used to generate inclusion probabilities. As $|b - 1|$ increases so $\hat{\mu}_i^{(t)}$ estimates $\mu_i^{(t)}$ less well, this could be classed as model misspecification. This is, of course, a very large oversimplification as in practice b_j^* will vary for each auxiliary variable; the effect of b_j^* on $\hat{\mu}_i$ will depend on the relative size of the auxiliary variable; and b is likely to deviate from one when the set of auxiliary variables used to estimate $\hat{\mu}_i$ are not the set of auxiliary variables that actually describe μ_i .

The impact of model misspecification on $\widehat{var}(\hat{\tau}^{(2)})$ can be investigated using the simulated population. The underlying abundance $\mu_i^{(t)}$ can be found using the $y_i^{(t)}$ to obtain a model of $\mu_i^{(t)}$. Using $\hat{\mu}_i^{(t)} = \mu_i^{(t)b}$ for different values of $b = -0.5, 0, 0.5, 1, 1.5, 2$ a sample $s^{(2)}$ is selected where $s_2^{(2)}$ of n_2 units are selected with probability proportional to $\mu_i^{(t)b}$. This is repeated for each of the $B=1000$ $s^{(1)}$ samples for $n_2 = 0, 25, 50, 75$ and 100. For each value of b the median variance and the 2.5 and 97.5 percentiles of the variance are calculated. These are shown in figure 2.

When the model is well specified, that is b is close to one, the median variance decreases as the proportion selected with inclusion probability proportional to $\hat{\mu}_i^{(t)}$ increases. Furthermore when b is close to one, the 95% interval, representing the variability in $\widehat{var}(\hat{\tau}^{(2)})$ is much greater under *srswor* than the other sampling strategies. As misspecification, $|b - 1|$, increases, the median variance and the variability in $\widehat{var}(\hat{\tau}^{(2)})$ increases. In particular the strategy in which no units are retained from one survey to another has the greatest increase in variability. The combined strategy in which part of the sample is retained from one survey to another and part is selected with probability proportional to $\hat{\mu}_i^{(t)}$ makes the methods more robust to model misspecification.

4 Use of strategy for a long-term monitoring strategy

The original motivation for these designs was to use them in a long-term monitoring strategy which aims to estimate both $\tau^{(t)}$ and $\delta^{(t',t)}$. A basic monitoring strategy would be of the form:

1. Survey 1:
 - (a) Take a sample $s^{(1)}$ using *srswor* of n units from U
 - (b) Develop a model $\zeta^{(1)}$ that uses the data in $s^{(1)}$
2. Survey 2:
 - (a) Predict $\hat{\mu}_i^{(2)} \forall i \in U$ using the model $\zeta^{(1)}$ and relevant covariates
 - (b) Take a sample $s^{(2)}$ of n units of which
 - i. $n_1^{(2)}$ units are selected using *srswor*, from U or $s^{(1)}$
 - ii. $n_2^{(2)}$ are selected using $\pi p \hat{\mu}^{(2)}$, from $s_1^{(2)}$ or $s_c^{(1)}$

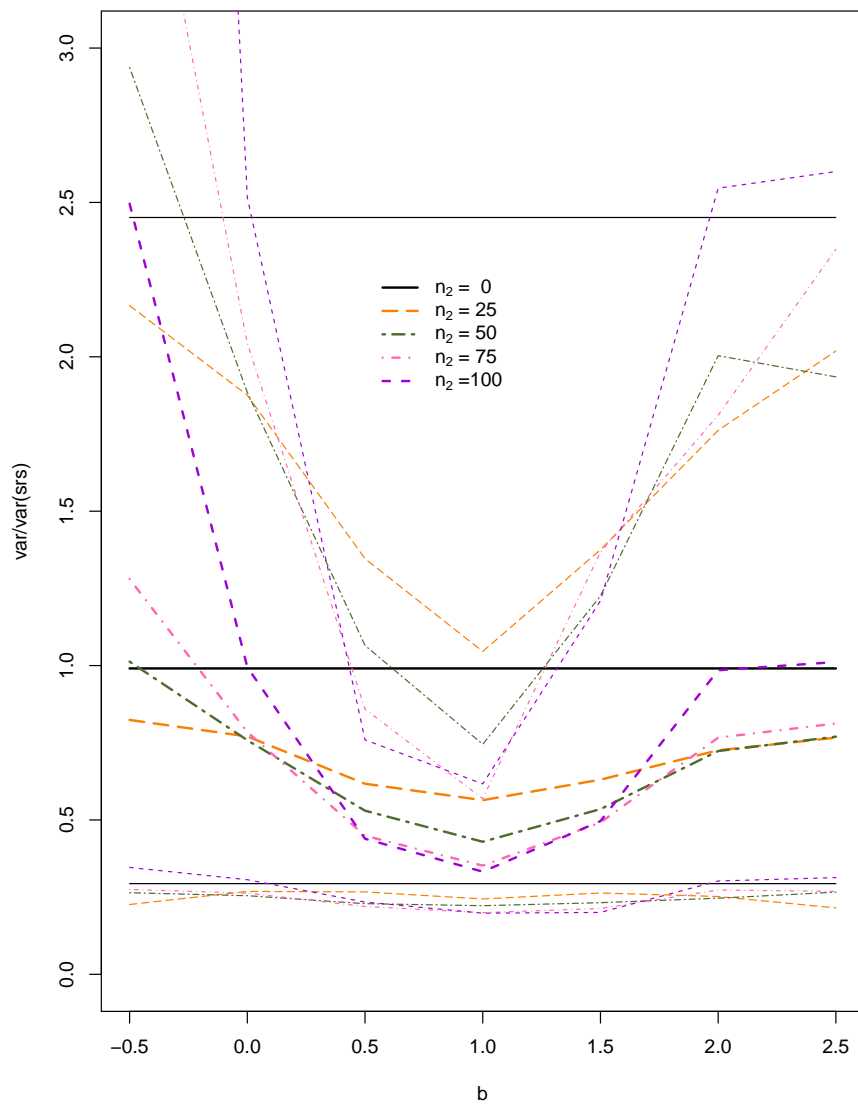


Fig. 2 Results from 1000 simulations of $\widehat{\text{var}}(\hat{\tau}^{(2)})$ for varying values of n_2 , where $n = 100$ and where $\mu_i^{(t)}$ is estimated to be $\mu_i^{(t)b}$ for different values of b compared to when $s^{(2)}$ is selected using simple random sampling. Thick lines are median results and thin lines the 2.5 and 97.5 percentiles.

- (c) Develop a model $\zeta^{(2)}$ using the data from $s^{(2)}$, and possibly the data from $s^{(1)}$.
3. Survey $t > 2$:
- (a) Predict $\hat{\mu}_i^{(t)} \forall i \in U$ using the model $\zeta^{(t-1)}$ and relevant covariates
- (b) Take a sample $s^{(t)}$ of n units of which
- i. $n_1^{(t)}$ units are selected, using *srswor*, from U or $s_1^{(t-1)}$
 - ii. $n_2^{(t)}$ units are selected using $\pi p \hat{\mu}^{(t)}$ from $s_{1c}^{(t)}$ or $s_{1c}^{(t-1)}$
- (c) Develop model $\zeta^{(t)}$ that uses the data in $s^{(t)}$, and possibly the data from $s^{(t-1)}, \dots, s^{(1)}$

Two key decisions in implementing the monitoring strategy are the proportion of the sampling effort that is allocated to retaining units from previous surveys and which data are used to construct the model $\zeta^{(t)}$.

4.1 Simulation

A monitoring strategy for 10 surveys was simulated for the same survey area as described in section 3. Populations were generated for the ten surveys using the inhomogeneous Poisson process with intensity $\lambda_{\mathcal{L}}$, as previously described in the on-line information. The intensity was therefore the same for all ten surveys. Hence the expected population total remained constant for all surveys, although the realised population total, $\tau^{(t)}$, did differ between the surveys.

Allocation of sampling effort In this simulation the overall sample size remains fixed each year at $n = 100$. The allocation of effort to retaining units would depend on the relative importance of $\tau^{(t)}$ and $\delta^{(t',t)}$ and the relative costs of sampling new units compared to retaining old units. Here it is assumed that the costs are the same for retaining units or sampling new ones. Furthermore the same proportion of units are retained in each survey so that the proportion of the sample retained in surveys $t = 3, \dots, 10$ is equal to the proportion retained in survey $t = 2$ so that $n_1^{(t)} = n_1^{(2)}$ for $t = 3, \dots, 10$.

Constructing $\zeta^{(t)}$. Two strategies were considered for constructing the model $\zeta^{(t)}$ from which $\hat{\mu}_i^{(t)}$ is predicted. First, the model can be constructed using the data from all previous surveys $s^{(1)}, \dots, s^{(t)}$. In this case it would be necessary to allow for the possibility of a temporal trend in the construction of $\zeta^{(1)}$. Second, only the data from the last survey $s^{(t)}$ could be used to construct the model. This strategy might be appropriate when the relationship between μ_i and x_i is changing rapidly, relative to the interval of the surveys and data from previous surveys are only marginally useful in estimating $\hat{\mu}_i^{(10)}$.

Sampling strategies used in simulation In survey 1 a simple random sample $s^{(1)}$ of $n = 100$ units was selected. This was repeated $B = 1000$ times. The model $\zeta^{(1)}$ was obtained using the method described in section 3.1. In survey

$t = 2$ the subsample $s_1^{(2)}$ was selected using *srswor* from $s^{(1)}$ and in surveys $t = 3, \dots, 10$ the subsample $s_1^{(t)} = s_1^{(2)}$. Thus the subsamples $s_2^{(2)}$ was selected from $s_c^{(1)}$ and subsample $s_2^{(t)}$ for $t > 2$ from $s_{1c}^{(2)}$. For each of the initial 1000 samples three difference scenarios were considered for selecting samples in future surveys $s^{(t)}$ for $t = 2, \dots, 10$.

1. $s_2^{(t)}$ is selected with inclusion probability proportional to $\hat{\mu}_i^{(t)}$. The model $\zeta^{(t-1)}$ from which $\hat{\mu}_i^{(t)}$ is calculated is created using data from survey $t - 1$ only
2. $s_2^{(t)}$ is selected with inclusion probability proportional to $\hat{\mu}_i^{(t)}$. The model $\zeta^{(t-1)}$ from which $\hat{\mu}_i^{(t)}$ is calculated is created using data from all previous surveys $1, \dots, t - 1$.
3. $s_2^{(t)}$ is selected using *srswor* .

Because covariate values remained fixed throughout the survey and no temporal trend was assumed $\hat{\mu}_i^{(t)}$ was obtained from the model $\zeta^{(t-1)}$ by predicting $\hat{\mu}_i^{(t-1)} = \hat{\mu}_i^{(t)}$ for all units in the survey region. As previously Sunter's method (Sunter, 1977a,b) was used to select units with inclusion probability proportional to predicted abundance.

4.2 Results

The ratios $r_\tau^{(10)} = \frac{\widehat{var}(\hat{\tau}^{(10)})}{var(\hat{\tau}^{(10)})}$ and $r_\delta^{(1,10)} = \frac{\widehat{var}(\hat{\delta}^{(1,10)})}{var(\hat{\delta}^{(1,10)})}$ were calculated where $var(\hat{\tau}^{(10)})$ and $var(\hat{\delta}^{(1,10)})$ are the known variances obtained when new samples $s^{(1)}, \dots, s^{(10)}$ were selected using simple random sampling without replacement. The estimated variances were calculated for each of the three scenarios described above.

Figure 3 shows the distribution of these ratios for $n_2 = 0, 25, 50, 75$ and 100. When at least part of the sample is retained from $s^{(1)}$, ie $n_2 < 100$, both the mean and spread of the estimated variance of the total, $\widehat{var}(\hat{\tau}^{(10)})$, decreases as $n_2^{(10)}$ increases when $s_2^{(10)}$ is selected with inclusion probability proportional to $\hat{\mu}_i^{(10)}$. Furthermore the ratio is generally less than 1 suggesting that the method is more efficient than selecting $s^{(t)}$ using *srswor*. The decrease in variance as $n_2^{(10)}$ increases is not observed when $s_2^{(10)}$ is selected using *srswor*. In comparison the estimate of $\hat{\delta}^{(1,10)}$ is generally better when $s_2^{(2)}$ is selected using *srswor* than when it is selected using inclusion probability proportional to $\hat{\mu}_i^{(t)}$.

When $n_2 = 100$, and so all of the sample is selected with inclusion probability proportional to $\hat{\mu}_i^{(10)}$, the variability in $\widehat{var}(\hat{\tau}^{(10)})$ can be very large. In some simulations $\widehat{var}(\hat{\tau}^{(10)})$ was more than five times the known variance from selecting $s^{(10)}$ using *srswor*. This is because the sample $s^{(t)}$ has two purposes. First it must provide an estimate of $\hat{\tau}^{(t)}$ but second it must provide data to construct a model $\zeta^{(t)}$ to predict the $\hat{\mu}_i^{(t+1)}$ that are then used to determine

inclusion probabilities for $s^{(t+1)}$, the following survey. A sample selected with inclusion probability proportional to $\hat{\mu}_i^{(t)}$ does not tend to give good coverage of the sample space because it favours units with large values of $\hat{\mu}_i^{(t)}$. The model $\zeta^{(t+1)}$ would then be constructed using a restricted part of the sample space so that predictions $\mu_i^{(t+1)}$ may be extrapolations from the model for some parts of the sample space, those where $\hat{\mu}_i^{(t)}$ was low, and so may be poorly estimated. Thus the inclusion probabilities in survey $t + 1$ will be less highly correlated with $y_i^{(t+1)}$ leading to less efficient estimation of $\hat{\tau}^{(t+1)}$, because of poor estimation of $\hat{\tau}_2^{(t+1)}$. This effect is especially large when only data from the previous survey are used to estimate $\hat{\mu}_i^{(t)}$.

In a long-term monitoring study each survey needs a sample that can estimate $\tau^{(t)}$ and $\delta^{(t',t)}$ efficiently and contributes to estimating $\hat{\mu}_i^{(t+1)}$ effectively to ensure a precise estimate of $\tau^{(t+1)}$. The strategy proposed here gives a balance between increasing the precision of $\hat{\tau}^{(t)}$ and $\delta^{(t',t)}$ in survey t and providing coverage of the sample space to estimate $\mu_i^{(t+1)}$ sufficiently accurately that selecting part of the sample with inclusion probability proportional to $\hat{\mu}_i^{(t+1)}$ in survey $t + 1$ will lead to high precision in the estimate of $\tau^{(t+1)}$.

5 Discussion

The design-based sampling strategy developed in this paper uses information from previous surveys to change the design of the next survey. The mechanism by which the design adapts is more general than existing design-based adaptive sampling strategies where the sample $s^{(t)}$ is adapted to estimate $\hat{\tau}^{(t)}$ (Thompson and Seber, 1996; Jolly and Hampton, 1990; Brown et al, 2008), or $\hat{\tau}^{(t+1)}$ (Haines and Pollock, 1998). In these designs, rules for adding new sample units to the initial sample $s^{(t)}$ are based on the observed $y_i^{(t)}$. The new units are typically adjacent (defined by a neighbourhood for the specific situation) to the units in $s^{(t)}$. In the strategy proposed here adaptation is based on model-based predictions, $\hat{\mu}_i^{(t)} \forall i \in U$, so that new units can be selected from anywhere in the survey region. The model and therefore future survey design can continue to adapt through time as data from each survey become available although inference remains design-based.

The strategy provides a simple evidence-based mechanism for determining inclusion probabilities when there are multiple potential auxiliary variables that could affect $y_i^{(t+1)}$. Observations from previous surveys are used to construct a model $\zeta^{(t)}$ that describes the relationship between observed $y_i^{(t)}$ and auxiliary variables for $i \in s^{(t)}$, either for the most recent survey or all previous surveys $t = 1, \dots, t$. From this model $\zeta^{(t)}$, predicted expected abundance for the next survey time $\hat{\mu}_i^{(t+1)} \forall i \in U$ can be obtained and this determines the inclusion probabilities for part of the sample $s^{(t+1)}$. Inclusion probabilities therefore have the potential to vary smoothly over the survey region -

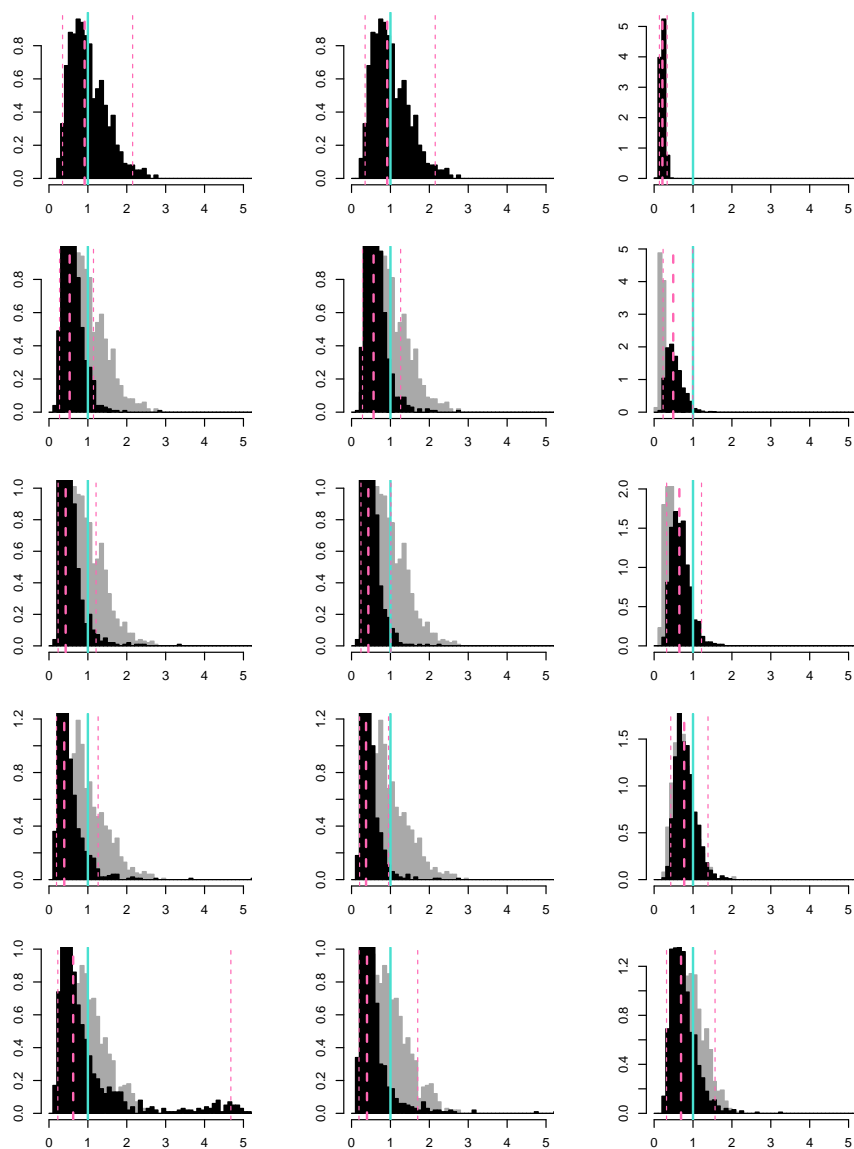


Fig. 3 Distributions from 1000 simulations of ratio $r_\tau^{(10)}$ (first two columns) and $r_\delta^{(1,10)}$ (last column), compared to results taking a new simple random sample. Top to bottom: $n_2 = 0, 25, 50, 75, 100$. First column shows results of $r_\tau^{(10)}$ when only units from the last survey are used to estimate $\hat{\mu}_i^{(10)}$. Second and third columns show results of $r_\tau^{(10)}$ and $r_\delta^{(1,10)}$ when data from all surveys are used to estimate $\hat{\mu}_i^{(10)}$. Black are results when $s_2^{(2)}$ is selected with inclusion probability proportional to $\hat{\mu}_i^{(10)}$. Grey are results when $s_2^{(2)}$ is selected using simple random sampling. Grey solid line indicates one and dashed lines are the 0.025, 0.50 and 0.975 quantiles of the black distribution. Results are truncated at 5 times the variance from simple random sampling. This occurs in the bottom row ($n_2 = 100$) only.

unlike stratification - and the effects of many auxiliary variables are easily incorporated in the design.

Estimates of $\tau^{(t)}$ using the strategy developed in this paper were more precise than those obtained from standard rotating panel designs; where both the retained and the new part of the sample are selected using *srswor*. However for rotating panel designs, model-assisted estimators are often used to improve the efficiency of $\hat{\tau}^{(t)}$ and $\delta^{(t',t)}$. These estimators include observed $y_i^{(t)}$ from previous surveys. An advantage of these estimators is that the efficiency of other parameters, including those using data other than just $y_i^{(t)}$ can be improved whilst the design, determined at the start of the monitoring program, remains inefficient. A disadvantage of the proposed strategy is that parameters that are functions of variables other than $y_i^{(t)}$ may be poorly estimated unless $y_i^{(t)}$ is correlated with them.

In the proposed sampling strategy part of the sample is retained from one survey to another and the rest is selected with inclusion probability proportional to predicted abundance, $\hat{\mu}_i^{(t)}$. As the proportion of the sample that is not retained increases, the precision of $\hat{\tau}^{(t)}$ increases. The strategy is robust if at least some units are retained from one survey to another. A strategy in which everything is selected with probability proportional to predicted abundance is not recommended - even for one survey. This is because the variability in $\widehat{var}(\hat{\tau}^{(t)})$ is large if $\hat{\mu}_i^{(t)}$ is poorly estimated. The importance of using a robust strategy is particularly important in a long-term monitoring project where the implications for estimating $\hat{\mu}_i^{(t+1)}$ in the future must also be considered.

The strategy is flexible and so the proportion retained from one survey to another does not need to remain fixed throughout a long-term monitoring programme. Each survey in a monitoring programme could be tailored to meet both the short-term objectives of that survey and the long-term objectives of the monitoring programme. For example some surveys might be aimed at providing good estimates of $\mu_i^{(t)}$ and others at precise estimates of $\tau^{(t)}$ and others of $\delta^{(1,t)}$. Further exploration of this and development of alternative estimators needs to be done.

Acknowledgements This work was partly funded by an EPSRC studentship. Professor Steve Buckland and Professor David Elston were very generous with their time both as stimulating PhD supervisors and with reading and commenting on a previous version of this paper. Some of the R functions were based on functions written by Dr David Borchers at St Andrews University. Comments from an anonymous referee significantly improved the readability of the text.

A Estimating $\tau^{(t)}$ and $\delta^{(t',t)}$ and their variances

Details of estimators for $\tau^{(t)}$ and $\delta^{(t',t)}$ and their variances are described below for the general sampling strategy in which in survey t the sample $s_1^{(t)}$ is selected using simple random sampling without replacement from $s_1^{(t')}$ and $s_2^{(t)}$ is selected with inclusion probability

proportional to $\hat{\mu}_i^{(t)}$ from $s_{1c}^{(t')} = U - s_1^{(t')}$. When $t' = 1$ and $t = 2$ then, $s_{1c}^{(t')} = s^{(1)}$ and $s_{1c}^{(t')} = s_c^{(1)}$. More generally $s_{1c}^{(t')} \subseteq s^{(1)}$.

A.1 Estimating $\tau^{(t)}$ and its variance

For ease of notation the superscript (t) will be omitted in this section so only superscripts relating to the previous survey t' are included.

Let $\tau = \omega\tau_1 + (1-\omega)\tau_2$ where $\omega = \frac{n_1}{n}$ and τ_k is estimated using the data from s_k only, then an unbiased estimator of τ_1 can be obtained by using the standard Horvitz-Thompson estimator and its variance is estimated using the Sen-Yates Grundy variance estimator so that:

$$\begin{aligned}\hat{\tau}_1 &= \sum_{s_1} \frac{y_i}{\pi_i} = \frac{N}{n_1} \sum_{s_1} y_i \\ \text{var}(\hat{\tau}_1) &= \frac{1}{2} \sum_U \sum_U (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 = \frac{N(N-n_1)}{n_1} \frac{\sum_U (y_i - \bar{y}_U)^2}{N-1} \\ \widehat{\text{var}}(\hat{\tau}_1) &= \frac{1}{2} \sum_{s_1} \sum_{s_1} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 = \frac{N(N-n_1)}{n_1} \frac{\sum_{s_1} (y_i - \bar{y}_{s_1})^2}{n_1-1}\end{aligned}$$

where $\bar{y}_U = \sum_U \frac{y_i}{N}$ and $\bar{y}_{s_1} = \sum_{s_1} \frac{y_i}{n_1}$.

An unbiased estimator of τ_2 is

$$\hat{\tau}_2 = \sum_{s_2} \frac{y_i}{\pi_i^\dagger} \text{ where } \pi_i^\dagger = Pr(i \in s_2 | s_1^{(t')}) Pr(i \in s_{1c}^{(t')}) = \pi_{i_2 | s_1^{(t')}} \pi_{i_{1c}}^{(t')}$$

If $I_{i_k} = \begin{cases} 1 & \text{if } i \in s_k \\ 0 & \text{otherwise} \end{cases}$ and $I_{i_k}^{(t')} = \begin{cases} 1 & \text{if } i \in s_k^{(t')} \\ 0 & \text{otherwise} \end{cases}$ then by taking expectations over $s_{1c}^{(t')}$, and over s_2 given that $s_{1c}^{(t')}$ has been selected it can be shown that $\hat{\tau}_2$ is an unbiased estimator of τ_2

$$\begin{aligned}E[\hat{\tau}_2] &= E_{s_{1c}^{(t')}} \left[E_{2|s_{1c}^{(t')}} \left[\sum_{i \in s_2} \frac{y_i}{\pi_{i_2 | s_{1c}^{(t')}} \pi_{i_{1c}}^{(t')}} \right] \right] \\ &= E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \frac{y_i}{\pi_{i_2 | s_{1c}^{(t')}} \pi_{i_{1c}}^{(t')}} E_{2|s_{1c}^{(t')}} [I_{i_2 | s_{1c}^{(t')}}^{(t')}] \right] \\ &= E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \frac{y_i}{\pi_{i_{1c}}^{(t')}} \right] = \sum_U \frac{y_i}{\pi_{i_{1c}}^{(t')}} E_{s_{1c}^{(t')}} [I_{i_{1c}}^{(t')}] = \tau\end{aligned}$$

Joint inclusion probabilities are required to calculate the variance. For simplicity the notation for these are denoted

$$\pi_{ij}^\dagger = Pr(i \& j \in s_2 | s_1^{(t')}) Pr(i \& j \in s_{1c}^{(t')}) = \pi_{(ij)_2 | s_1^{(t')}} \pi_{(ij)_{1c}}^{(t')}$$

and specific differences are

$$\Delta_{ij}^{(t)\dagger} = \pi_{ij}^{(t)\dagger} - \pi_i^{(t)\dagger} \pi_j^{(t)\dagger}$$

and

$$\Delta_{(ij)_k}^{(t')} = \pi_{(ij)_k}^{(t')} - \pi_{i_k}^{(t')} \pi_{j_k}^{(t')}$$

The variance of $\hat{\tau}_2$ is

$$\begin{aligned} \text{var}(\hat{\tau}_2) &= \text{var}_{s_{1c}^{(t')}} \left(E_{2|s_{1c}^{(t')}}[\hat{\tau}_2] \right) + E_{s_{1c}^{(t')}} \left[\text{var}_{2|s_{1c}^{(t')}}(\hat{\tau}_2) \right] \\ &= \text{var}_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \frac{y_i}{\pi_{i_{1c}}^{(t')}} \right] + E_{s_{1c}^{(t')}} \left[E_{2|s_{1c}^{(t')}} \left[\sum_{s_2} \sum_{s_2} \frac{y_i y_j}{\pi_i^\dagger \pi_j^\dagger} \right] - E_{2|s_{1c}^{(t')}} \left[\sum_{s_2} \frac{y_i}{\pi_i^\dagger} \right] E_{2|s_{1c}^{(t')}} \left[\sum_{s_2} \frac{y_j}{\pi_j^\dagger} \right] \right] \\ &= \sum_U \sum_U \left(\pi_{(ij)_{1c}}^{(t')} - \pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')} \right) \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} \\ &\quad + E_{s_{1c}^{(t')}} \left[E_{2|s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \sum_{s_{1c}^{(t')}} \frac{y_i y_j}{\pi_i^\dagger \pi_j^\dagger} \left(I_{(ij)_{2|s_{1c}^{(t')}}} - I_{i_{2|s_{1c}^{(t')}}} I_{j_{2|s_{1c}^{(t')}}} \right) \right] \right] \\ &= \sum_U \sum_U \Delta_{(ij)_{1c}}^{(t')} \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} + E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \sum_{s_{1c}^{(t')}} \Delta_{(ij)_{2|s_{1c}^{(t')}}} \frac{y_i}{\pi_i^\dagger} \frac{y_j}{\pi_j^\dagger} \right] \end{aligned}$$

Although this cannot be calculated explicitly, its unbiased estimator

$$\widehat{\text{var}}(\hat{\tau}_2) = \sum_{s_2} \sum_{s_2} \frac{\Delta_{(ij)_{1c}}^{(t')}}{\pi_{ij}^\dagger} \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} + \sum_{s_2} \sum_{s_2} \frac{\Delta_{(ij)_{2|s_{1c}^{(t')}}}}{\pi_{(ij)_{2|s_{1c}^{(t')}}}} \frac{y_i}{\pi_i^\dagger} \frac{y_j}{\pi_j^\dagger}$$

can. By taking expectations of the estimated variance it can be shown that it is an unbiased estimator of $\text{var}(\hat{\tau}_2)$

$$\begin{aligned} E[\widehat{\text{var}}(\hat{\tau}_2)] &= E_{s_{1c}^{(t')}} \left[E_{2|s_{1c}^{(t')}} \left[\sum_{s_2} \sum_{s_2} \frac{\Delta_{(ij)_{1c}}^{(t')}}{\pi_{(ij)_{2|s_{1c}^{(t')}}} \pi_{(ij)_{1c}}^{(t')}} \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} \right] \right] \\ &\quad + E_{s_{1c}^{(t')}} \left[E_{2|s_{1c}^{(t')}} \left[\sum_{s_2} \sum_{s_2} \frac{\Delta_{(ij)_{2|s_{1c}^{(t')}}}}{\pi_{(ij)_{2|s_{1c}^{(t')}}}} \frac{y_i}{\pi_i^\dagger} \frac{y_j}{\pi_j^\dagger} \right] \right] \\ &= E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \sum_{s_{1c}^{(t')}} \frac{\Delta_{(ij)_{1c}}^{(t')}}{\pi_{(ij)_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} \right] + E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \sum_{s_{1c}^{(t')}} \Delta_{(ij)_{2|s_{1c}^{(t')}}} \frac{y_i}{\pi_i^\dagger} \frac{y_j}{\pi_j^\dagger} \right] \\ &= \sum_U \sum_U \Delta_{(ij)_{1c}}^{(t')} \frac{y_i y_j}{\pi_{i_{1c}}^{(t')} \pi_{j_{1c}}^{(t')}} + E_{s_{1c}^{(t')}} \left[\sum_{s_{1c}^{(t')}} \sum_{s_{1c}^{(t')}} \Delta_{(ij)_{2|s_{1c}^{(t')}}} \frac{y_i}{\pi_i^\dagger} \frac{y_j}{\pi_j^\dagger} \right] \end{aligned}$$

The covariance is defined as

$$\begin{aligned}
cov(\hat{\tau}_1, \hat{\tau}_2) &= E_{s_1^{(t')}} \left[E_{s|s_1^{(t')}} [\hat{\tau}_1 \hat{\tau}_2] \right] - E_{s_1^{(t')}} \left[E_{s|s_1^{(t')}} [\hat{\tau}_1] \right] E_{s_1^{(t')}} \left[E_{s|s_1^{(t')}} [\hat{\tau}_2] \right] \\
&= E_{s_1^{(t')}} \left[E_{s|s_1^{(t')}} \left[\sum_{s_1} \sum_{s_2} \frac{y_i}{\pi_{i_1|s_1^{(t')}}} \frac{y_j}{\pi_{j_1}^{(t')}}} \right] \right] \\
&\quad - E_{s_1^{(t')}} \left[\sum_{s_1} \frac{y_i}{\pi_{i_1|s_1^{(t')}}} \right] E_{s_1^{(t')}} \left[E_{s|s_1^{(t')}} \left[\sum_{s_2} \frac{y_j}{\pi_{j_1}^{(t')}}} \right] \right] \\
&= E_{s_1^{(t')}} \left[\sum_{s_1^{(t')}} \sum_{s_1^{(t')}} \frac{y_i}{\pi_{i_1}^{(t')}} \frac{y_j}{\pi_{j_{1c}}^{(t')}}} \right] - E_{s_1^{(t')}} \left[\sum_{s_1^{(t')}} \frac{y_i}{\pi_{i_1}^{(t')}}} \right] E_{s_{1c}^{(t')}} \left[\frac{y_j}{\pi_{j_{1c}}^{(t')}}} \right] \\
&= \sum_U \sum_U \pi_{i_1 j_{1c}}^{(t')} \frac{y_i}{\pi_{i_1}^{(t')}} \frac{y_j}{\pi_{j_{1c}}^{(t')}} - \sum_U \sum_U y_i y_j
\end{aligned}$$

Because $\pi_{i_1 j_{1c}}^{(t')} = \pi_{i_1}^{(t')} - \pi_{(ij)_1}^{(t')}$ is the probability that $i \in s^{(t')}$ and $j \in s_{1c}^{(t')}$ this can be rewritten as

$$cov(\hat{\tau}_1, \hat{\tau}_2) = - \sum_U \sum_U \left(\pi_{(ij)_1}^{(t')} - \pi_{i_1}^{(t')} \pi_{j_1}^{(t')} \right) \frac{y_i}{\pi_{i_1}^{(t')}} \frac{y_j}{\pi_{j_{1c}}^{(t')}}$$

so that an unbiased estimator of this covariance is:

$$\widehat{cov}(\hat{\tau}_1, \hat{\tau}_2) = \sum_{s_1} \sum_{s_1} \frac{\Delta_{(ij)_1}^{(t')}}{\pi_{(ij)_1}^{(t')}} \frac{y_i}{\pi_{i_1}^{(t')}} \frac{y_j}{\pi_{j_{1c}}^{(t')}}$$

Taking expectations of this estimator over $s_1|s_1^{(t')}$ and $s_1^{(t')}$ gives

$$E[\widehat{cov}(\hat{\tau}_1, \hat{\tau}_2)] = E_{s_1^{(t')}} \left[\sum_{s_1^{(t')}} \sum_{s_1^{(t')}} \frac{\Delta_{(ij)_1}^{(t')}}{\pi_{(ij)_1}^{(t')}} \frac{y_i}{\pi_{i_1}^{(t')}} \frac{y_j}{\pi_{j_{1c}}^{(t')}}} \right] = cov(\hat{\tau}_1, \hat{\tau}_2)$$

Because this estimator only uses data in s_1 then when $t' = 1$ and $t = 2$ this can be written as

$$\widehat{cov}(\hat{\tau}_1^{(2)}, \hat{\tau}_2^{(2)}) = - \frac{N}{n_1} \left(\sum_{s_1^{(2)}} y_i^{(2)2} - \frac{1}{n_1 - 1} \sum_{s_1^{(2)}} \sum_{\substack{s_1^{(2)} \\ i \neq j}} y_i^{(2)} y_j^{(2)} \right)$$

A.2 Estimating the variance of $\hat{\delta}^{(t',t)}$

The difference $\hat{\delta}^{(t',t)}$ is calculated as $\hat{\delta}^{(t',t)} = \hat{\tau}^{(t)} - \hat{\tau}^{(t')}$ and its variance is $var(\hat{\delta}^{(t',t)}) = var(\hat{\tau}^{(t)}) + var(\hat{\tau}^{(t')}) - 2cov(\hat{\tau}^{(t)}, \hat{\tau}^{(t')})$. The general form of the covariance term is

$$\begin{aligned}
cov(\hat{\tau}^{(t')}, \hat{\tau}^{(t)}) &= \omega^{(t')} \omega^{(t)} cov(\hat{\tau}_1^{(t')}, \hat{\tau}_1^{(t)}) + (1 - \omega^{(t')}) \omega^{(t)} cov(\hat{\tau}_2^{(t')}, \hat{\tau}_1^{(t)}) \\
&\quad + \omega^{(t')} (1 - \omega^{(t)}) cov(\hat{\tau}_1^{(t')}, \hat{\tau}_2^{(t)}) + (1 - \omega^{(t')}) (1 - \omega^{(t)}) cov(\hat{\tau}_2^{(t')}, \hat{\tau}_2^{(t)})
\end{aligned}$$

although this reduces to two terms, described in section A.2.1, when $t' = 1$ and $t = 2$. When it is not the case that $t' = 1, t = 2$ then the covariance is estimated using the strategy of Holmes and Skinner (2000) in which only the first term

$$\text{cov}(\hat{\tau}^{(t')}, \hat{\tau}^{(t)}) \approx \omega^{(t')} \omega^{(t)} \text{cov}(\hat{\tau}_1^{(t')}, \hat{\tau}_1^{(t)})$$

is estimated. This is a reasonable strategy because most of the covariance is expected to be due to a positive correlation between observations from the matched units between the two time periods; indeed the reason for retaining units in the sample from one survey to another is because the $y_i^{(t)}$ and $y_i^{(t')}$ are expected to be correlated. Other terms would be expected to contribute little to the covariance because they relate to the covariance between different units surveyed in the two surveys. At worst, the estimated covariance may be an underestimate of the true covariance so that the estimated variance overestimates the true variance, this is preferable to underestimating the true variance. Further work to obtain the full set of covariance terms is required; the covariance terms are complex and it is difficult to see how to derive their estimators.

Both $\hat{\tau}_1^{(t')}$ and $\hat{\tau}_1^{(t)}$ are estimated using the Horvitz-Thompson estimator and so the covariance and its estimator are of the form:

$$\text{cov}(\hat{\tau}_1^{(t')}, \hat{\tau}_1^{(t)}) = \sum_U \sum_U \frac{y_i^{(t')}}{\pi_{i_1}^{(t')}} \frac{y_j^{(t)}}{\pi_{j_1}^{(t)}} \left(\pi_{(ij)_1}^{(t',t)} - \pi_{i_1}^{(t')} \pi_{j_1}^{(t)} \right) \quad (1)$$

$$\widehat{\text{cov}}(\hat{\tau}_1^{(t')}, \hat{\tau}_1^{(t)}) = \sum_{s_1^{(t')}} \sum_{s_1^{(t)}} \frac{y_i^{(t')}}{\pi_{i_1}^{(t')}} \frac{y_j^{(t)}}{\pi_{j_1}^{(t)}} \left(\frac{\pi_{(ij)_1}^{(t',t)} - \pi_{i_1}^{(t')} \pi_{j_1}^{(t)}}{\pi_{(ij)_1}^{(t',t)}} \right) \quad (2)$$

The exact form of the inclusion probabilities will depend on the strategy used to select $s_1^{(t')}$ and $s_1^{(t)}$. The joint inclusion probability will be of the form $\pi_{(ij)_1}^{((t',t))} = Pr(j \in s_1^{(t)} | i \& j \in s_1^{(t')}) Pr(i \& j \in s_1^{(t')})$.

A.2.1 Estimating $\text{cov}(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)})$

When $t' = 1$ and $t = 2$ the covariance reduces to two terms so that $\text{cov}(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)}) = \text{cov}(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)})$ and the inclusion probabilities are:

$$\begin{aligned} \pi_i^{(1)} &= \frac{n}{N} & \pi_{j_1}^{(2)} &= \frac{n_1}{n} \\ \pi_{ij_1}^{(1,2)} &= Pr(j \in s_1^{(2)} | i \& j \in s_1^{(1)}) Pr(i \& j \in s_1^{(1)}) = \begin{cases} \frac{n_1}{n} \frac{n(n-1)}{N(N-1)} = \frac{n_1(n-1)}{N(N-1)} & i \neq j \\ \frac{n_1}{n} \frac{n}{N} = \frac{n_1}{N} & i = j \end{cases} \end{aligned}$$

and so the covariance and its estimator are

$$\begin{aligned} \text{cov}(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)}) &= \frac{N-n}{n} \left[\sum_U y_i^{(1)} y_i^{(2)} - \frac{1}{N-1} \sum_U \sum_{\substack{U \\ i \neq j}} y_i^{(1)} y_j^{(2)} \right] \\ \widehat{\text{cov}}(\hat{\tau}_1^{(1)}, \hat{\tau}_1^{(2)}) &= \frac{(N-n)N}{nn_1} \left[\sum_{s_1^{(2)}} y_i^{(1)} y_i^{(2)} - \frac{1}{n-1} \sum_{\substack{s_1^{(1)} \\ s_1^{(2)}}} \sum_{\substack{U \\ i \neq j}} y_i^{(1)} y_j^{(2)} \right] \end{aligned}$$

It is also possible to estimate the covariance

$$\begin{aligned}
cov(\hat{\tau}^{(1)}, \hat{\tau}_2^{(2)}) &= E_{s^{(1)}} \left[E_{s_2^{(2)} | s_1^{(t')}} \left[\sum_{s^{(1)}} \sum_{s_2^{(2)}} \frac{y_i^{(1)}}{\pi_i^{(1)}} \frac{y_j^{(2)}}{\pi_{j|s^{(1)}}^{(2)} \pi_{j_c}^{(1)}} \right] \right] \\
&\quad - E_{s^{(1)}} \left[E_{s_2^{(2)} | s_1^{(t')}} \left[\sum_{s_1^{(t')}} \frac{y_i^{(1)}}{\pi_i^{(1)}} \right] \right] E_{s^{(1)}} \left[E_{s_2^{(2)} | s_1^{(t')}} \left[\sum_{s_2^{(2)}} \frac{y_j^{(2)}}{\pi_{j|s^{(1)}}^{(2)} \pi_{j_c}^{(1)}} \right] \right] \\
&= E_{s^{(1)}} \left[\sum_{s^{(1)}} \sum_{s_c^{(1)}} \frac{y_i^{(1)}}{\pi_i^{(1)}} \frac{y_j^{(2)}}{\pi_{j_c}^{(1)}} \right] - E_{s^{(1)}} \left[\sum_{s_1^{(t')}} \frac{y_i^{(1)}}{\pi_i^{(1)}} \right] E_{s^{(1)}} \left[\sum_{s_c^{(1)}} \frac{y_j^{(2)}}{\pi_{j_c}^{(1)}} \right] \\
&= \sum_U \sum_U \pi_{ij_c}^{(1)} \frac{y_i^{(1)}}{\pi_i^{(1)}} \frac{y_j^{(2)}}{\pi_{j_c}^{(1)}} - \sum_U y_i^{(1)} \sum_U y_j^{(2)} \\
&= \sum_U \sum_U \left(\pi_{ij}^{(1)} - \pi_i^{(1)} \pi_i^{(1)} \right) \frac{y_i^{(1)}}{\pi_i^{(1)}} \frac{y_j^{(2)}}{\pi_{j_c}^{(1)}}
\end{aligned}$$

This is the same form as $cov(\hat{\tau}_1, \hat{\tau}_2)$ so that

$$\begin{aligned}
cov(\hat{\tau}^{(1)}, \hat{\tau}_2^{(2)}) &= - \left(\sum_U y_i^{(1)} y_i^{(2)} - \frac{1}{N-1} \sum_U \sum_{i \neq j} y_i^{(1)} y_j^{(2)} \right) \\
\widehat{cov}(\hat{\tau}^{(1)}, \hat{\tau}_2^{(2)}) &= - \frac{N}{n_1} \left(\sum_{s_1^{(2)}} y_i^{(1)} y_i^{(2)} - \frac{1}{n_1-1} \sum_{s_1^{(2)}} \sum_{s_1^{(2)}}' y_i^{(1)} y_j^{(2)} \right)
\end{aligned}$$

B Estimators when $s_2^{(t)}$ is selected using simple random sampling

When $s_2^{(2)}$ is selected using simple random sampling without replacement then the inclusion probability $\pi_2^{(t)}$ can be calculated unconditionally because:

$$\begin{aligned}
\pi_2^{(t)} &= Pr(i \in s_2^{(2)} | i \in s_1^{(1)}) Pr(i \in s_1^{(1)}) + Pr(i \in s_2^{(2)} | i \in s_c^{(1)}) Pr(i \in s_c^{(1)}) \\
&= \frac{n_1}{n} \frac{n}{N} + \frac{n_2}{N-n} \frac{N-n}{N} = \frac{n}{N}
\end{aligned}$$

The same holds for the joint inclusion probabilities and so $\hat{\tau}^{(2)}$ and $\hat{\tau}^{(t)}$ when $t > 2$ can be estimated using the Horvitz-Thompson estimator with its corresponding variance and variance estimator. The covariance $cov(\hat{\tau}^{(t')}, \hat{\tau}^{(t)})$ required for calculating the variance of $\hat{\delta}^{(t',t)}$ is of the form of equation (1) and its corresponding estimator equation (2).

References

Binder DA, Hidiroglou MA (1988) Sampling in time. In: Handbook of Statistics, vol 6, Elsevier Science Publishers, North Holland, pp 187-211

- Brewer KRW, Hanif M (1982) Sampling with Unequal Probabilities, Lecture notes in Statistics, vol 15. Springer-Verlag, New York
- Brown JA, Salehi M M, Moradi M, Bell G, Smith DR (2008) An adaptive two-stage sequential design for sampling rare and clustered populations. *Population Ecology* 50:239–245
- Buckland ST, Elston DA (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30:478–495
- Chao MT (1982) A general purpose unequal probability sampling plan. *Biometrika* 69:653–656
- Dufour J, Gambino J, Kennedy B, Lindeyer J, Singh MP (1998) Methodology of the Canadian Labour Force Survey. Tech. Rep. 71-526-XPB, Statistics Canada
- Duncan GJ, Kalton G (1987) Issues of design and analysis of surveys. *International Statistical Review* 55:97–117
- Haines DE, Pollock KH (1998) Estimating the number of active and successful bald eagle nests: an application of the dual frame method. *Environmental and Ecological Statistics* 5:245–256
- Hansen MH, Madow WG, Tepping BJ (1983) An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78:776–793
- Holmes DJ, Skinner CJ (2000) Variance estimation for labour force survey estimates of level and change. UK Government Statistical Service Methodology Series No. 21
- Horvitz D, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47:663–685
- Jessen RJ (1942) Statistical investigation of a farm survey for obtaining farm facts. *Iowa Agricultural Station Research Bulletin* 304:54–59
- Jolly GM, Hampton I (1990) A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Science* 47:1282–1291
- Khaemba WM, Stein A (2000) Use of GIS for a spatial and temporal analysis of Kenyan wildlife with generalised linear modelling. *International Journal of Geographical Information Science* 14:833–853
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*, 2nd edn. Chapman and Hall, London
- Opsomer JD, Botts C, Kim JY (2003) Small area estimation in a watershed erosion assessment survey. *Journal of Agricultural, Biological, and Environmental Statistics* 8(2):139–152
- Overton WS, Stehman SV (1996) Desirable design characteristics for long-term monitoring of ecological variables. *Environmental and Ecological Statistics* 3:349–361
- Overton WS, White D, Stevens Jr DL (1990) Design report for EMAP. Environmental Monitoring and Assessment Program. EPA/600/3-91/053, U.S. Environmental Protection Agency
- Särndal CE, Swennsson B, Wretman J (1992) *Model Assisted Survey Sampling*. Springer-Verlag, New York
- Sen AR (1973) Theory and applications of sampling on repeated occasions with several auxiliary variables. *Biometrics*
- Skalski JR (1990) A design for long-term status and trends monitoring. *Journal of Environmental Management* 30:139–144
- Sunter A (1977a) List sequential sampling with equal or unequal probabilities without replacement. *Applied Statistics* 26:261–268
- Sunter A (1977b) Response burden, sample rotation, and classification renewal in economic surveys. *International Statistical Review* 45:209–222
- Thompson SK, Seber GAF (1996) *Adaptive Sampling*. Wiley, New York
- Tillé Y (2006) *Sampling Algorithms*. Springer series in Statistics, Springer
- Venables W, Ripley B (2002) *Modern Applied Statistics with S*, 4th edn. Springer, London
- Wood SN (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC
- Yates F, Grundy PM (1953) Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society, B* 15:235–261