Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Adaptive rectangular sampling: An easy, incomplete, neighbourhood-free adaptive cluster sampling design

by Bardia Panahbehagh

Release date: December 20, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2016
Vol. 42, No. 2, pp. 263-281
Statistics Canada, Catalogue No. 12-001-X

Adaptive rectangular sampling: An easy, incomplete, neighbourhood-free adaptive cluster sampling design

Bardia Panahbehagh[1]

Abstract

This paper introduces an incomplete adaptive cluster sampling design that is easy to implement, controls the sample size well, and does not need to follow the neighbourhood. In this design, an initial sample is first selected, using one of the conventional designs. If a cell satisfies a prespecified condition, a specified radius around the cell is sampled completely. The population mean is estimated using the π- estimator. If all the inclusion probabilities are known, then an unbiased π- estimator is available; if, depending on the situation, the inclusion probabilities are not known for some of the final sample units, then they are estimated. To estimate the inclusion probabilities, a biased estimator is constructed. However, the simulations show that if the sample size is large enough, the error of the inclusion probabilities is negligible, and the relative π- estimator is almost unbiased. This design rivals adaptive cluster sampling because it controls the final sample size and is easy to manage. It rivals adaptive two-stage sequential sampling because it considers the cluster form of the population and reduces the cost of moving across the area. Using real data on a bird population and simulations, the paper compares the design with adaptive two-stage sequential sampling. The simulations show that the design has significant efficiency in comparison with its rival.

Key Words: Adaptive cluster sampling; Adaptive two-stage sequential sampling; Primary and secondary sampling units; Inclusion probability.

1 Introduction

Adaptive cluster sampling (ACS) is an efficient design for rare and clustered populations (Thompson 1990; Thompson and Seber 1996). ACS was introduced for quadrat-based sampling, where the study area is usually partitioned into non-overlapping quadrats for sample selection. Depending on the situation, these are called “cells” or “secondary sampling units” (SSUs). In the first phase of the design, an initial sample is selected using one of the conventional designs, usually simple random sampling without replacement (SRSWOR). The term “conventional designs” (Thompson and Seber 1996) refers to designs in which the procedure for selecting the sample does not depend on any observation of the main variable, such as SRSWOR, stratified sampling and systematic sampling. If a rare event (a cell whose value is at least as large as the prespecified condition C) is found after the initial sample is obtained, then sampling continues in the neighbourhood of that location with the hope of observing more rare events. The process of searching the neighbourhood is continued until no more rare events are found. This design has been shown to be useful for estimating the parameters of highly clustered and rare populations (Smith, Brown and Lo 2004). However, ACS has some disadvantages, including the following two:

- The final sample size is random and uncontrollable.
- Neighbourhoods must be defined and followed. Following neighbourhoods in ACS means searching unit by unit and level by level around the initial rare events to find all the rare events

1. Bardia Panahbehagh, Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran. E-mail: panahbehagh@khu.ac.ir.

that are around them in different directions. The shape of the environment and the cluster may lead to confusion for the sampler.

To overcome the first problem, many designs, such as two-stage ACS and incomplete or restricted ACS (IACS), have been introduced.

Thompson (1991) introduced stratified ACS, and Salehi and Seber (1997) developed two-stage ACS. Two-stage ACS is designed to select a fixed number of primary sampling units (PSUs) by SRSWOR in the first stage, and then to select a fixed number of SSUs in each selected PSU, also by SRSWOR, in the second stage. The condition that is to be adapted and the neighbourhood are defined in terms of secondary units (rather than primary units). Salehi and Seber considered two schemes, depending on whether the clusters are allowed to overlap primary-unit boundaries or not, that would later be more desirable for controlling the final sample size. Some other related designs have also been introduced; as they are not related to the discussion in this paper, they are not mentioned.

Salehi and Smith (2005) made an essential change to two-stage ACS, known as two-stage sequential sampling, and Brown et al. (2008) introduced an adaptive version of two-stage sequential sampling (ATS). In ATS, the allocation of second-stage efforts among PSUs is based on preliminary information from the sampled PSUs. Additional survey efforts are directed to PSUs where the SSUs in the initial sample meet a prespecified criterion, or condition C (e.g., an individual from the rare population is present). More precisely, d times the number of units that satisfy condition C in the initial sample in a PSU is dedicated to the respective PSU as an additional sample using SRSWOR. Therefore, ATS could almost overcome the two problems, since, in this design, the final sample size is limited, and there is no need to define and follow the neighbourhood. But, ATS does not directly employ clustering of the population. This means that in ATS, the additional sample is a random sample of the whole respective PSU–ATS depends on the size, shape and location of the PSUs. Other developments in ATS are not essential (in other words, they have not changed the special aspects of ATS), so there is no need to mention them here.

IACS, Brown and Manly (1998) suggested a restricted version of ACS to control final sample size. They put a limit on the final sample size prior to sampling by selecting the initial sample sequentially. Chao and Thompson (1999) and Su and Quinn (2003) imposed a restriction on the number of neighbourhood levels beyond each unit that satisfies the condition in the initial fixed-size sample. All the neighbours of the units that satisfy the condition in the initial sample are in the first neighbourhood level. In turn, all the neighbours that are to be added based on the units in the first neighbourhood level are considered to be in the second neighbourhood level, and so on. In brief, a cluster, as defined in conventional ACS, is allowed in IACS to be truncated at a predetermined distance from the unit in the initial sample that intersects it. These authors introduced a biased estimator for the population mean.

Interestingly, Yang (2011) and Yang et al. (2011) introduced an adaptive plot design to overcome disadvantages of ACS in practice, and to use and define the neighbourhood, especially in tree-sampling surveys. They aimed to improve the practicability of ACS while maintaining some of its major characteristics. According to Yang et al. (2011), “the plot design is based on a conditional plot expansion: a larger plot (by a pre-defined plot size factor) is installed at a sample point instead of the smaller initial plot

if a pre-defined condition is fulfilled." Their target population was a tree population, and their aim was to estimate its density (the number of objects per hectare). Their design was not planned for quadrat-based sampling. In addition, they assumed that they could survey an additional area (after they finished sampling) to calculate the inclusion probability of the required tree, but this would be impossible or costly in other surveys.

Chao and Thompson (1999) introduced IACS for the first time. This design enables ACS to function without measuring all members of a cluster. They introduced the design in graph-theory form. Because it uses neighbourhoods like ACS, the design is complicated to manage, and the calculation of inclusion probabilities is also complicated.

In this paper, a manageable version of ACS, which has positive aspects of the designs discussed above, is proposed. Adaptive rectangular sampling (ARS) is a practical, efficient and easy-to-calculate adaptive design that is able to find rare events, does not need to follow the neighbourhood and controls the final sample size well.

ARS is introduced in Section 2 of the paper. Section 3 contains a real case study and some simulations, and Section 4 concludes the paper and provides a complete discussion of the advantages and disadvantages of the design.

2 Adaptive rectangular sampling

Suppose a total population of N units partitioned into M primary sampling units (PSUs), each containing N_h secondary sampling units (SSUs). Let $\{(h,j), h = 1,2,\ldots,M; j = 1,2,\ldots,N_h\}$ denote the j^{th} unit in the h^{th} primary unit, with the associated measurement or count y_{hj}. Then, $\tau_h = \sum_{j=1}^{N_h} y_{hj}$ is the total of the y values for the h^{th} PSU, and $\mu = 1/N \sum_{h=1}^{M} \tau_h$ is the population mean.

Adaptive rectangular sampling (ARS) can be performed in a two-stage procedure. The first stage of the ARS design consists of selecting a conventional random sample, s_0, of size m, with M PSUs.

The first phase of the second stage consists of selecting an initial conventional sample, s_{1h}, of size n_{1h} in the h^{th} PSU, where $h \in s_0$.

In the second phase, all the SSUs around those in s_{1h} that satisfy condition C with the radius R are adaptively added, where $R \in \{1,2,\ldots,M_d\}$ and M_d is the maximum diameter of each PSU. Here, the definition of radius is different from the conventional definition. "Radius," for a cell, is defined based on all cells around it. For example, $R = 1$ refers to the first-level nearest neighbourhood, which consists of the eight SSUs around the cell, and $R = 2$ refers to the nearest and the second-nearest neighbourhoods, which consist of all 24 SSUs around the cell (8 SSUs for $R = 1$, plus 16 SSUs added for $R = 2$). If the cell is in a corner or close to a border of the PSU, these numbers are reduced (see Figure 2.1). Therefore, in the h^{th} PSU, there is an additional sample, s_{2h}, of random size n_{2h}. Now, the final sample is $s_h = s_{1h} \cup s_{2h}$, of random size $n_h = n_{1h} + n_{2h}$, inside the h^{th} PSU. This procedure can be performed under either an overlapping scheme and a non-overlapping scheme.

An estimator for the population mean

The π- estimator (the Horvitz–Thompson estimator) is an estimator for the population mean that requires inclusion probabilities for all sampled units. If all the inclusion probabilities are available, the following are used to estimate the population mean:

$$\hat{\mu} = \frac{1}{N}\sum_{h\in s_0}\frac{\hat{\tau}_h}{\pi_h},$$

where π_h is the inclusion probability of the h^{th} PSU, and

$$\hat{\tau}_h = \sum_{j\in s_h}\frac{y_{hj}}{\pi_{hj}},$$

where π_{hj} is the inclusion probability of the SSU (h, j). To use the π- estimator, π_{hj} must be calculated for all $\{(h, j); h \in s_0, j \in s_h\}$.

In addition, the variance of the estimator is

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{N^2}\left[\sum_{h=1}^{M}\sum_{h'=1}^{M}\left(\frac{\pi_{hh'} - \pi_h\pi_{h'}}{\pi_h\pi_{h'}}\right)\tau_h\tau_h' + \sum_{h=1}^{M}\frac{\mathrm{Var}(\hat{\tau}_h)}{\pi_h}\right],$$

where $\pi_{hh'}$ is the joint inclusion probability of the h^{th} and h'^{th} PSUs. An unbiased estimator for the above is

$$\widehat{\mathrm{Var}}(\hat{\mu}) = \frac{1}{N^2}\left[\sum_{h\in s_0}\sum_{h'\in s_0}\left(\frac{\pi_{hh'} - \pi_h\pi_h'}{\pi_h\pi_h'}\right)\frac{\hat{\tau}_h\hat{\tau}_h'}{\pi_{hh'}} + \sum_{h\in s_0}\frac{\widehat{\mathrm{Var}}(\hat{\tau}_h)}{\pi_h}\right],$$

where $\widehat{\mathrm{Var}}(\hat{\tau}_h)$ is

$$\widehat{\mathrm{Var}}(\hat{\tau}_h) = \sum_{j\in s_{h1}}\sum_{j'\in s_{h1}}\left(\frac{\pi_{hjj'} - \pi_{hj}\pi_{hj'}}{\pi_{hj}\pi_{hj'}}\right)\frac{y_{hj}y_{hj'}}{\pi_{hjj'}},$$

where $\pi_{hjj'}$ is the joint inclusion probability of (h, j) and (h, j') in the h^{th} PSU.

It is easy to calculate the inclusion probabilities for the first stage, especially when simple random sampling without replacement (SRSWOR) is used. In this situation, it is easy to see that

$$\pi_h = \frac{m}{M}, \quad \pi_{hh'} = \frac{m(m-1)}{M(M-1)};\ h \neq h', \quad \pi_{hh} = \pi_h.$$

To calculate π_{hj}, it is necessary to know how many of the cells around cell (h, j) within radius R satisfy condition C, because selecting them as the initial sample leads to selecting the cells around them as the final sample. It is necessary to introduce some new notations. In ARS, with the radius R, B_{hj} represents the event of unit (h, j) being selected as the final sample, and A_{hj} represents the event of unit (h, j) satisfying condition C and being selected as the initial sample. In addition, s_{hj} represents all the units that satisfy condition C and that would be adaptively added to the sample if A_{hj} occurs, including unit (h, j),

with the size f_{hj}. The size f_{hj} is partitioned as $f_{hj} = f_{hj1} + f_{hj2}$, where the former indicates the number of cells in s_{hj} that are available in the final sample (s_h) and the latter is defined as $f_{hj2} = f_{hj} - f_{hj1}$. However, no information about its units is available. $F_{\cdot}$ is defined like $f_{\cdot}$, but for all units (those satisfying condition C and those not satisfying condition C).

In addition, let $f_{hjj'}$ be the size of $s_{hjj'} = s_{hj} \cup s_{hj'}$, and $f_{hjj'1}$, $f_{hjj'2}$, and $F_{\cdot}$ be defined the same.

The sample s_{hj}, with the size f_{hj}, contains all the cells that lead to the selection of unit (h, j), and $s_{hjj'}$, with the size $f_{hjj'}$, contains all the cells that lead to the selection of at least one of (h, j) and (h, j') as the final sample.

Theorem 1: In ARS, with the radius R, for the h^{th} PSU and using SRSWOR to select the initial sample of size n_{1h},

$$\pi_{hj} = 1 - \frac{\binom{N_h - f_{hj}}{n_{1h}}}{\binom{N_h}{n_{1h}}} \tag{2.1}$$

$$\pi_{hjj'} = 1 - \frac{\binom{N_h - f_{hj}}{n_{1h}}}{\binom{N_h}{n_{1h}}} - \frac{\binom{N_h - f_{hj'}}{n_{1h}}}{\binom{N_h}{n_{1h}}} + \frac{\binom{N_h - f_{hjj'}}{n_{1h}}}{\binom{N_h}{n_{1h}}}. \tag{2.2}$$

For the proof of the theorem, see the Appendix.

Here, only one problem arises: f_{hj} is known only in the initial sample that satisfies the condition. However, other samples (those that are adaptively added) have partial information about f_{hj} (i.e., f_{hj1}). Let $\mathrm{Bin}(n, p)$ stand for a binomial distribution based on the independent Bernoulli variable n with parameter p. To estimate f_{hj}, f_{hj2} represents the number of successes (the number of units that satisfy condition C) in F_{hj2} trials (by searching F_{hj2} units); the independency and identicality (iid) of the trials are assumed. The latter assumption (iid) is for simplifying the calculations. This, of course, leads to bias in estimating some of the inclusion probabilities and, so, to bias for the respective π- estimator. With all the above assumptions, F_{hj2} would be considered a random variable with a binomial distribution, as follows:

$$f_{hj2} \sim \mathrm{Bin}\left(F_{hj2}, p_{hj}\right),$$

where p_{hj} is the probability of satisfying condition C for all cells in the h^{th} PSU around the j^{th} cell with the radius R. Then,

$$E\left(f_{hj}\right) = f_{hj1} + E\left(f_{hj2}\right) = f_{hj1} + p_{hj}F_{hj2}.$$

Calculating π_{hj} leads to two situations:

- If the cell satisfies condition C and belongs to the initial sample, or is adaptively added and is located in a place in the final sample that contains complete information about the area around it, then it is possible to calculate the inclusion probability precisely from the information in the final sample.
- If the final sample does not contain enough information to calculate the inclusion probability, two strategies are proposed:
    - There is partial information about f_{hj}, this means that everything is known about F_{hj1}, and only F_{hj2} needs to be investigated. For F_{hj2}, only knowledge about how many of the cells satisfy the condition is required; there is no need for exact knowledge about y. For example, if condition C is defined as $y > 0$, it is necessary to know only how many cells of F_{hj2} are nonempty. If this is easy to investigate, the exact inclusion probabilities for all of the units in the sample can be calculated.
    - It is not possible to calculate the inclusion probabilities, π_{hj} can be estimated as (see equation 2.1)

$$\hat{\pi}_{hj} = 1 - \frac{\binom{N_h - E\left(f_{hj}\right)}{n_{1h}}}{\binom{N_h}{n_{1h}}} = 1 - \frac{\binom{N_h - \left(f_{hj1} + p_{hj}F_{hj2}\right)}{n_{1h}}}{\binom{N_h}{n_{1h}}}. \qquad (2.3)$$

Using p_{hj}, or, in other words, assuming different probabilities for different cells, leads to tedious calculations. Estimating p_{hj} can be done based on the spatial pattern of the population. For example, in the case of clustered populations, it may be reasonable to assume two kinds of p_{hj}, one for units in the sample satisfying condition C and one for units not satisfying condition C, so that greater probability is provided for the units satisfying condition C. A wide class of spatial patterns can be assumed in estimating p_{hj}, but, here, to have a simple and understandable strategy, $p_{hj} = p_h$ is assumed for all units in the h^{th} PSU. Therefore, p_h is the probability of satisfying condition C for units in the h^{th} PSU. It may be possible to guess p_h from previous information or to estimate it without bias from the initial sample as the portion of the units in the initial sample in the h^{th} PSU that satisfy condition C.

Estimating p_h based on the initial sample is a common procedure in adaptive designs (for example, see Brown et al. [2008]). For rare populations, such estimations might be imprecise. Practically, however, this is not a serious problem in ARS, because, for initial-sample units that satisfy the condition, it is possible to calculate inclusion probabilities without error (p_h is not required). Furthermore, p_h is not required in adaptively added units with $y = 0$. For some adaptively added units with $y > 0$, p_h has an insignificant role in calculating inclusion probabilities. The example in the next subsection and the simulation results in Section 3 confirm such assertions.

For $\pi_{hjj'}$ (as for π_{hj}), if, depending on the final sample, there is enough information to calculate f_{hj}, $f_{hj'}$ and $f_{hjj'}$, then it is enough to use equation 2.2. If there is partial information about $f_{hjj'}$, then

$$f_{hjj'2} \sim \text{Bin}\left(F_{hjj'2}, p_h\right),$$

and then

$$E\left(f_{hjj'}\right) = f_{hjj'1} + E\left(f_{hjj'2}\right) = f_{hjj'1} + p_h F_{hjj'2}.$$

And, it is enough to replace the respective "$f_{\cdot}$"s with "$E\left(f_{\cdot}\right)$"s in equation 2.2. Without the assumption of $p_{hj} = p_h$, $p_{hjj'}$ should perhaps be used instead of p_h for estimating $f_{hjj'}$. This would make the calculations more difficult. A reduction in precision (assuming constant p_h for all the units in the h^{th} PSU) allows such a simple sampling strategy to be presented.

Discussing an example can help clarify all of the above formulas and calculations.

Discussion of an example

For this example, see the top-left PSU in the right plot in Figure 2.1, where $N = 112$, $N_h = 28$, $h = 1$ and $n_{11} = 2$. Assume that it is necessary to calculate π_{hj} for two units in Figure 2.1, with $R = 1$. First, for the initial sample with $y = 6$, it is easy to see that $F_{hj} = 6$, that it has five cells around it plus itself, and that five of them satisfy the condition $\left(f_{hj} = 5\right)$. This information is available at the end of the sampling in the final sample. Therefore,

$$\pi_{y=6} = 1 - \frac{\binom{28-5}{2}}{\binom{28}{2}} \simeq 0.33.$$

For an adaptively added sample, like $y = 248$, as discussed earlier, there is partial information (see Figure 2.2). Here, $F_{hj} = 9$ (the blue cells in part B) and $F_{hj1} = 6$ (the orange cells in part C). From the final sample, $f_{hj1} = 5$ is also known (the positive response in the orange cells in part C). In addition, $F_{hj2} = 3$ (the blue cells in part C), but f_{hj2} is not known. To estimate it, p_1 would be estimated from the initial sample as $p_1 = 1/2 = 0.5$ (see the green cells in the first PSU in Figure 2.1), then

$$\hat{\pi}_{y=248} = 1 - \frac{\binom{28-(5+0.5\times 3)}{2}}{\binom{28}{2}} \simeq 0.42.$$

But, according to the population, the following can be calculated:

$$\pi_{y=248} = 1 - \frac{\binom{28-7}{2}}{\binom{28}{2}} \simeq 0.44,$$

which is very close to the estimation.

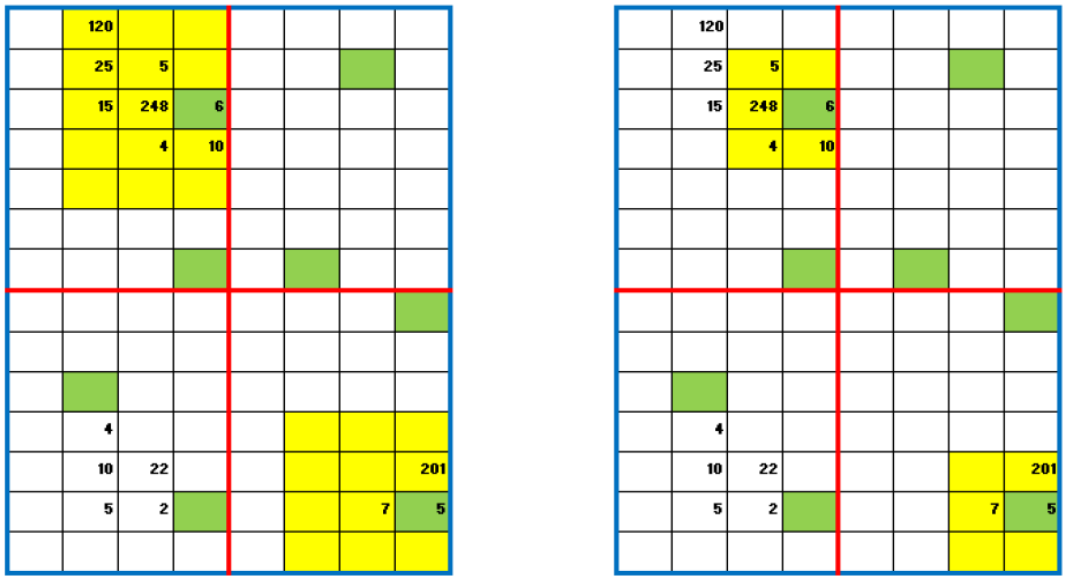

Figure 2.1 $N = 112$, $N_h = 28$, $M = 4$, $m = 4$ and $n_{1h} = 2$. **The green SSUs are the initial sample, and the yellow cells are the adaptively added sample. The right plot indicates non-overlapping ARS with** $R = 1$, **and the left plot indicates** $R = 2$, **where condition** C **is defined as** $y > 0$. **Numbers show respective** y **values for the cells.**

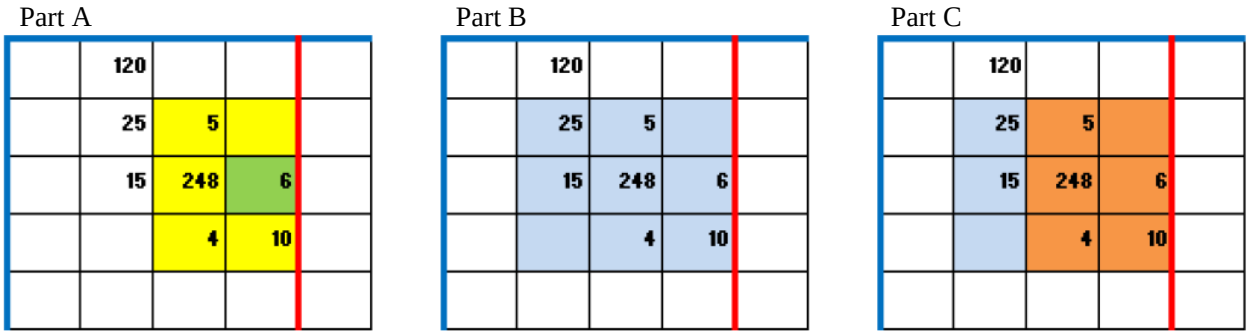

Figure 2.2 Inclusion probability for $y = 248$. **Part A indicates a part of the final sample. Part B (the blue cells) indicates the cells that must be known for the inclusion probability of** $y = 248$. **In part C, the orange cells are those that must be known and for which information is available, and the blue cells are those that must be known but for which no information is available.**

Now, assume that the goal is to calculate a joint probability, $\pi_{y=6,y'=5}$ (see Figure 2.3). Here, according to equation 2.2, $f_{hj} = 5$ (part B), $f_{hj'} = 3 + f_{hj'2}$, $F_{hj'2} = 5$ (the blue cells in part C), $f_{hjj'} = 5 + f_{hjj'2}$ and $F_{hjj'2} = 5$ (part D), and $E\left(f_{hjj'2}\right) = 5 \times 0.5 = 2.5$. From the information in the sample,

$$\hat{\pi}_{y=6,y'=5} = 1 - \frac{\binom{28-5}{2}}{\binom{28}{2}} - \frac{\binom{28-(3+0.5\times 5)}{2}}{\binom{28}{2}} + \frac{\binom{28-(5+0.5\times 5)}{2}}{\binom{28}{2}} \simeq 0.22.$$

With complete information about the population,

$$\pi_{y=6,y'=5} = 1 - \frac{\binom{28-5}{2}}{\binom{28}{2}} - \frac{\binom{28-6}{2}}{\binom{28}{2}} + \frac{\binom{28-8}{2}}{\binom{28}{2}} \simeq 0.22,$$

which shows no error to two decimal places. By writing a code for "$f_{\cdot}$"s and "$F_{\cdot}$"s, the π- estimator can be calculated easily.

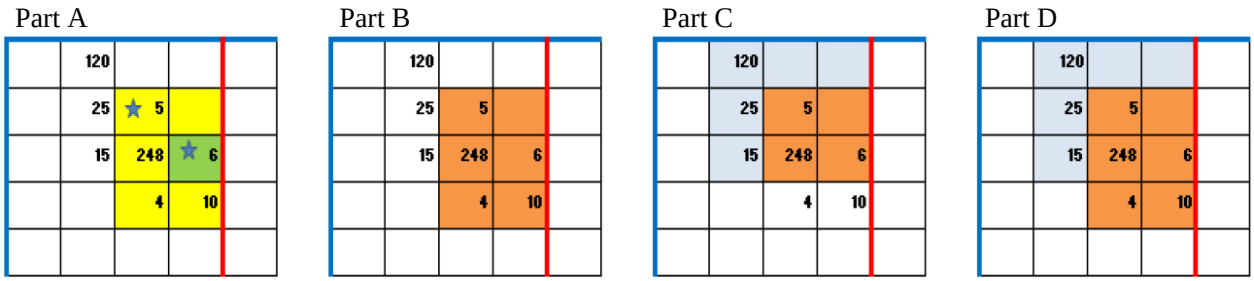

Figure 2.3 Joint inclusion probability for $y = 6$ and $y' = 5$. Coloured cells indicate the information that is required; orange cells indicate available information and blue cells indicate the information that needs to be estimated from the final sample.

3 A real case study and simulations

In this section, adaptive rectangular sampling (ARS) is evaluated and compared with adaptive two-stage sequential sampling (ATS) and two-stage simple random sampling without replacement (TSS). Here, ARS is not compared with adaptive cluster sampling (ACS) for two reasons: first, Salehi and Smith (2005) compared ATS with two-stage ACS, and, second, it is not fair to compare ARS with ACS or even incomplete ACS because ACS needs to define, use and follow the neighbourhood, while ARS does not.

If ATS is a design free of neighbourhood, then ARS satisfies this condition too, because, if a sampler can recognize the border of the cells, or, in other words, can distinguish secondary sampling units (SSUs), the sampler can also recognize an SSU with its radius area. In addition, the area to be surveyed may be specified before samples are taken. Based on a map of the area, it is possible to use SRSWOR for the SSUs

and the area around them if the SSUs satisfy the condition. Because the sampler need not return to the area to take the second phase of the sample (according to the ATS process), ARS seems to be easier and less costly than ATS. For a better comparison, the cost factor should be taken into consideration.

The comparison is done using two kinds of data: a real case study and simulation cases.

Here, efficiency is defined as

$$\text{eff}(.) = \frac{\text{MSE}(\bar{y}_{\text{tss}})}{\text{MSE}(.)},$$

where $\bar{y}_{\text{tss}}$ is the conventional mean estimator in TSS, MSE stands for mean square error and "." stands for one of the following:

- $\hat{\mu}_{\text{ATS}}$: Murthy's estimator in ATS, which is unbiased, and which will be referred to as "ATS". This estimator, for the mean of the h^{th} primary sampling unit (PSU), would be presented as

$$\hat{\mu}_{\text{ATS}} = \hat{q}_1 \bar{y}_{hc} + (1 - \hat{q}_1)\bar{y}_{hc'}, \tag{3.1}$$

  where $\hat{q}_1$ is the proportion of the units that satisfy condition C in the initial sample, and $\bar{y}_{hc}$ and $\bar{y}_{hc'}$ are the means of the final-sample units satisfying condition C and not satisfying condition C, respectively, in the h^{th} PSU. In this estimator, the first portion of the unit satisfying condition C is estimated from the initial sample, and the respective value is adapted to the mean of the sample satisfying condition C to construct the estimator.
- $\hat{\mu}_{\text{ARS}}$: the π- estimator in ARS, which is unbiased.
- $\hat{\mu}_{\text{ARS}.\hat{\pi}}$: the π- estimator in ARS, with π estimated from the final sample using equation 2.3. It is not unbiased, and its relative bias is defined as $\text{Rbias}.\hat{\mu}_{\text{ARS}.\hat{\pi}} = \left|\left(\mu_y - \hat{\mu}_{\text{ARS}.\hat{\pi}}\right)/\mu_y\right|$.

From now on, the acronym "ARS" refers to both $\hat{\mu}_{\text{ARS}}$ and $\hat{\mu}_{\text{ARS}.\hat{\pi}}$.

Furthermore, two formulas are used for the error in estimating inclusion probabilities in $\hat{\mu}_{\text{ARS}.\hat{\pi}}$:

- $\hat{\mu}_{D.Inclu}$: this shows the mean of the difference between real inclusion probabilities and the respective estimation (i.e., the mean of $e_{hj} = \pi_{hj} - \hat{\pi}_{hj}$ for all the sample units)
- $\hat{\mu}_{AD.Inclu}$: this shows the mean of the absolute difference between real inclusion probabilities and the respective estimation (i.e., the mean of $\left|e_{hj}\right| = \left|\pi_{hj} - \hat{\pi}_{hj}\right|$ for all the sample units).

A non-overlapping scheme is used in this section.

A real case study on a blue-winged teal population

Smith, Conroy and Brakhage (1995) used a population of blue-winged teal to evaluate ACS. The population comes from comprehensive counts, which were made from helicopters from December 13 to 15, 1992, in central Florida. The blue-winged teal population is extremely clustered, with a total of $N = 200$

units (Figure 3.1). A simulation study found ACS to be efficient for this population, in the sense that the variance of the estimator is smaller than in simple random sampling (Smith, Conroy and Brakhage 1995).

The population was partitioned into $M = 8, 4, 2$ PSUs. ARS, ATS and TSS were performed in the population with different values for m, n_{1h}, R and d (a multiple in ATS that indicates the size of the additional sample in the second phase for units satisfying condition C), with 25,000 simulations for each combination of values. For a fair comparison, d was chosen in such a way that the expected final sample sizes for ATS and ARS were almost the same. For TSS, the sample size in each simulation was the same as that for ARS. The expected sample sizes were calculated using Monte Carlo simulations. It is notable that in ARS, if two or more adaptively added samples overlapped, the overlap was measured once. Practically, if there is overlap in the sample, the relevant cells must be sampled and measured only once.

Results are presented in Tables 3.1 and 3.2. For information about the MSEs of the estimators, $\mathrm{MSE}(\bar{y}_{\mathrm{tss}})$ is presented in the results. With this MSE and the efficiency of the estimators, the MSEs of the other estimators are easy to calculate. The results are noteworthy: ARS was better than ATS in all situations. ARS, unlike ATS, was also always more efficient than TSS. The efficiency of ARS was sometimes seven or eight times that of TSS, whereas this number was at most around two and a half for ATS. For more than 55% of the cases, the efficiency of ARS was greater than 2, whereas this was true of less than 5% of the cases for ATS.

The relative bias of $\hat{\mu}_{\mathrm{ARS}.\hat{\pi}}$ is acceptable for most of the cases; it may be unacceptable for a few cases with a little sample size. For around 61% of the cases, the relative bias was less than 0.03, and, for around 92% of them, the relative bias was less than 0.07.

Efficiency improved by increasing the radius R, and a larger radius R was proper for larger PSUs. In this population, there are two important clusters at the top of the population plot. With $R = 2$, selecting one of the cells in a large PSU as the initial sample led to the selection of all of them. That is why $R = 2$, with a large enough initial sample size, showed such significant efficiency.

In addition, the number of PSUs in the first stage was important, and the results indicate that more PSUs lead to efficiency improvements. As discussed before, the efficiency of ATS depends on the size, shape and location of the PSUs. When the population could not be partitioned into some empty and full PSUs, ATS was not as efficient (see populations 2 and 4). But as ARS uses the cluster form of the population, it is not as dependent on PSUs and could even perform in a population with one PSU, which would be meaningless for ATS.

In addition, for Population 1, $\hat{\mu}_{AD.Inclu} = 0.025$ and $\hat{\mu}_{D.Inclu} = 0.002$; for Population 2, $\hat{\mu}_{AD.Inclu} = 0.023$ and $\hat{\mu}_{D.Inclu} = 0.005$; for Population 3, $\hat{\mu}_{AD.Inclu} = 0.024$ and $\hat{\mu}_{D.Inclu} = -0.004$; and, for Population 4, $\hat{\mu}_{AD.Inclu} = 0.025$ and $\hat{\mu}_{D.Inclu} = -0.008$. The mean of the inclusion probabilities for these simulations was around 0.22. According to $\hat{\mu}_{D.Inclu}$ and $\hat{\mu}_{AD.Inclu}$, the errors in estimating the inclusion probabilities seem to be almost negligible. This is why $\hat{\mu}_{\mathrm{ARS}.\hat{\pi}}$ was almost unbiased. The relative bias of $\hat{\mu}_{\mathrm{ARS}.\hat{\pi}}$ showed acceptable precision for "$\hat{\pi}_j$"s.

Table 3.1
Efficiency of the estimators, with $N = 200$, $C = 0$, $M = 8,4$

Population 1	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	Rbias.$\hat{\mu}_{ARS.\hat{\pi}}$	MSE$(\bar{y}_{tss})$
	1	4	1	6	7	1.32	2.05	1.62	0.13	81,076
			3	6	19	1.32	1.84	1.64	0.06	22,657
			5	5	30	1.25	1.79	1.69	0.03	12,865
			7	4	40	1.18	1.76	1.71	0.03	8,729
		8	1	6	13	1.22	1.92	1.51	0.14	35,171
			3	6	38	1.33	1.99	1.80	0.06	10,324
			5	5	59	1.35	2.13	2.04	0.04	5,502
			7	4	79	1.34	2.46	2.43	0.03	3,502
	2	4	1	15	10	2.43	3.02	3.04	0.02	71,670
			3	12	27	1.61	2.17	2.24	0.02	16,835
			5	9	40	1.33	1.99	2.05	0.00	9,219
			7	7	51	1.23	1.94	1.97	0.00	6,463
		8	1	15	20	2.37	3.00	3.01	0.02	32,290
			3	12	54	1.69	2.53	2.61	0.02	7,030
			5	9	80	1.52	2.98	3.05	0.02	3,570
			7	7	101	1.45	3.81	3.85	0.01	2,270
Population 2	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	Rbias.$\hat{\mu}_{ARS.\hat{\pi}}$	MSE$(\bar{y}_{tss})$
	1	2	4	6	13	1.06	1.90	1.58	0.11	36,208
			8	5	25	0.94	1.81	1.61	0.07	15,775
			10	5	30	1.03	1.82	1.66	0.05	12,434
			15	5	42	1.10	1.81	1.72	0.03	7,852
		4	4	6	26	1.00	1.92	1.60	0.11	16,911
			8	5	49	0.98	2.00	1.79	0.07	7,339
			10	5	60	1.00	2.05	1.89	0.06	5,475
			15	4	84	1.03	2.39	2.31	0.03	3,180
	2	2	4	12	20	1.29	2.51	2.52	0.00	28,341
			8	11	36	1.04	2.07	2.08	0.00	11,024
			10	10	42	1.03	2.02	2.02	0.01	8,521
			15	8	55	1.03	1.96	1.96	0.01	5,352
		4	4	12	40	1.04	2.32	2.32	0.02	11,725
			8	11	71	0.99	2.34	2.35	0.00	4,507
			10	10	84	1.01	2.59	2.60	0.00	3,338
			15	8	111	1.04	3.41	3.42	0.00	1,891

Table 3.2
Efficiency of the estimators, with $N = 200$, $C = 0$, $M = 4,2$

Population 3	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	Rbias.$\hat{\mu}_{ARS.\hat{\pi}}$	MSE$(\bar{y}_{tss})$
	1	2	4	6	14	1.11	1.73	1.40	0.18	34,451
			8	5	25	1.05	1.71	1.51	0.10	16,795
			10	5	31	1.09	1.69	1.56	0.08	13,671
			15	5	43	1.16	1.61	1.56	0.03	9,819
		4	4	6	27	1.26	2.18	1.73	0.18	15,357
			8	5	51	1.26	2.78	2.48	0.10	6,759
			10	5	62	1.34	3.18	2.94	0.08	5,078
			15	4	86	1.39	4.80	4.72	0.03	2,964
	2	2	4	12	21	1.30	1.90	1.95	0.05	26,117
			8	11	36	1.13	1.64	1.69	0.03	12,047
			10	10	43	1.11	1.56	1.60	0.02	9,872
			15	8	55	1.12	1.48	1.49	0.01	7,686
		4	4	12	42	1.43	2.55	2.64	0.05	10,818
			8	11	72	1.51	3.43	3.66	0.02	4,275
			10	10	85	1.55	4.29	4.57	0.01	3,183
			15	8	110	1.83	9.44	9.87	0.00	1,856
Population 4	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	Rbias.$\hat{\mu}_{ARS.\hat{\pi}}$	MSE$(\bar{y}_{tss})$
	1	1	10	6	17	0.93	1.68	1.32	0.13	27,053
			15	5	24	0.90	1.69	1.43	0.09	17,860
			20	5	31	0.91	1.63	1.45	0.08	13,802
			30	4	44	0.92	1.58	1.50	0.02	9,952
		2	10	6	34	0.98	2.16	1.69	0.12	11,851
			15	5	49	0.95	2.52	2.11	0.10	7,409
			20	5	62	0.97	2.91	2.58	0.06	5,153
			30	4	87	1.01	4.56	4.41	0.03	2,940
	2	1	10	14	27	1.01	1.70	1.69	0.02	18,908
			15	13	37	0.93	1.50	1.51	0.02	12,092
			20	10	45	0.84	1.42	1.43	0.01	9,338
			30	8	58	0.88	1.35	1.35	0.00	7,329
		2	10	14	53	1.08	2.38	2.41	0.02	7,512
			15	13	73	1.06	2.80	2.92	0.01	4,317
			20	11	90	1.03	3.59	3.79	0.00	2,933
			30	8	116	1.03	7.60	8.00	0.00	1,672

Artificial populations

The spatial pattern was generated with an R code following the Poisson cluster process (Brown 2003). The number of clusters was selected from a Poisson distribution, and cluster centres were randomly located throughout the site. Individuals within the cluster were located around the cluster centre at a random distance, following an exponential distribution, and in a random direction, following a uniform distribution. The parameters of the code were changed to generate three different populations. With the addition of the population in the example subsection of the paper, this subsection uses four artificial populations to evaluate ARS (see Figure 3.2):

- Population 5: rare and not clustered,
- Population 6: not rare, but clustered,
- Population 7: not rare and not clustered,
- Population 8: rare and clustered.

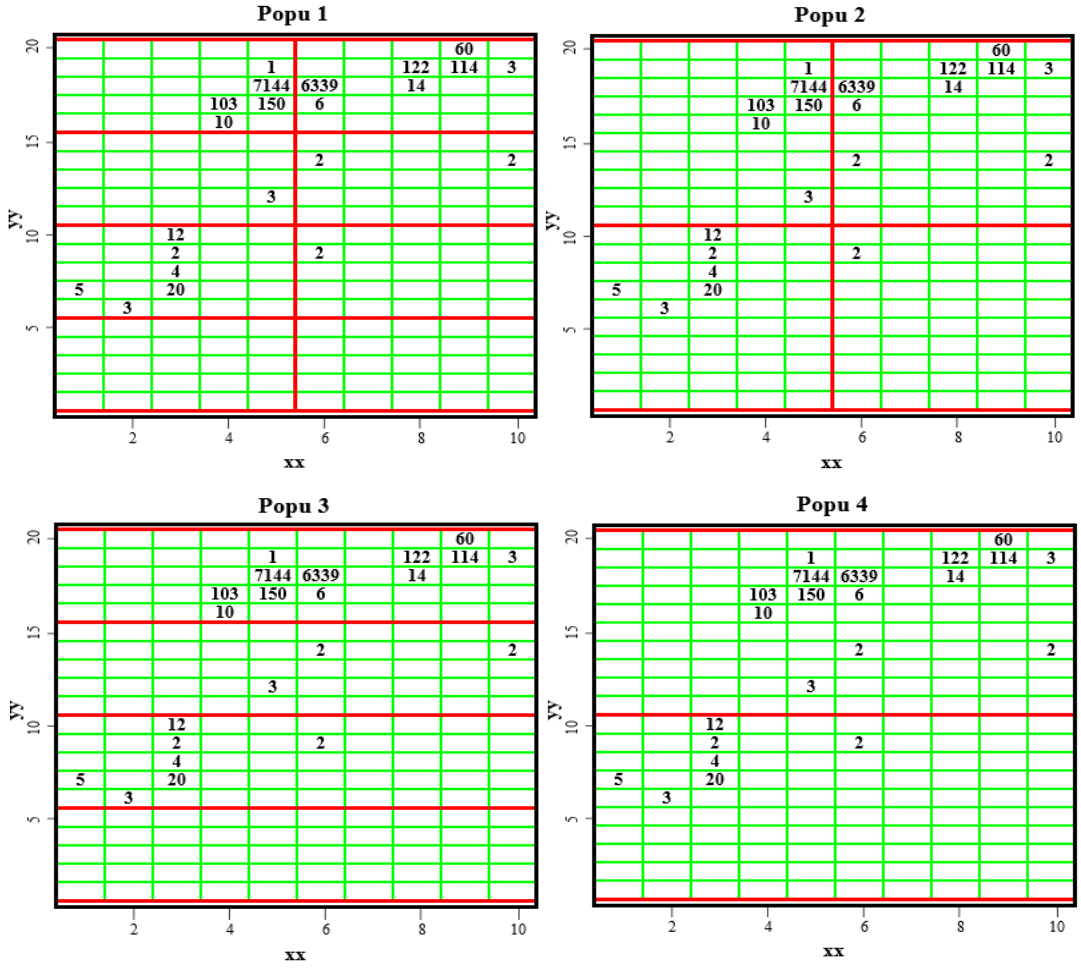

Figure 3.1 Bird populations. Numbers show the respective y values for the cells, with a mean of 70.61 and a variance of 453,709.52. Red lines indicate the borders of the PSUs.

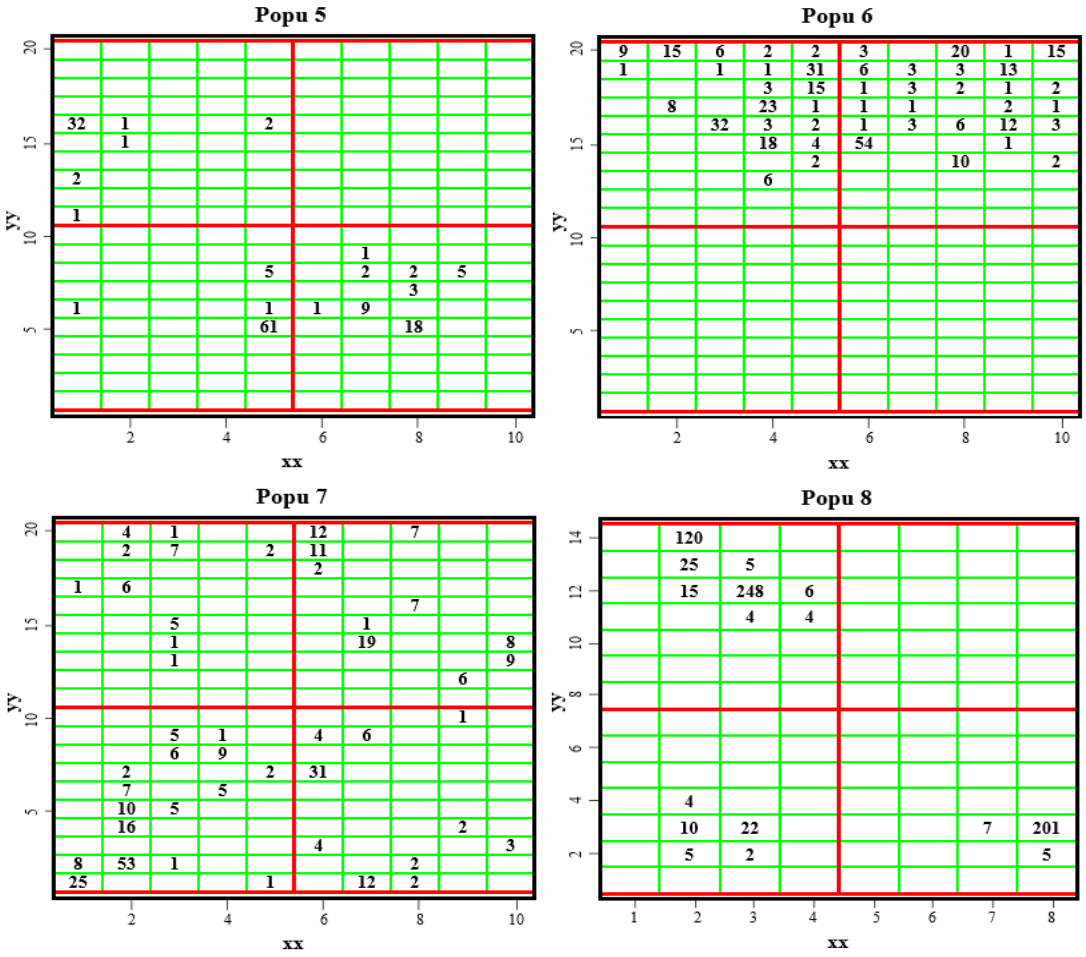

Figure 3.2 Artificial populations. Numbers show the respective y values for the cells, with 0.74, 1.76, 1.68 and 6.10 as the means, and 25.74, 35.34, 29.08 and 1,025.10 as the variances of the four populations, respectively. Red lines indicate the borders of the PSUs.

Results are presented in Tables 3.3 and 3.4. The relative bias of $\hat{\mu}_{\text{ARS}.\hat{\pi}}$ was acceptable for all cases.

Table 3.3
Efficiency of the estimators, with $N = 200$, $C = 0$, $M = 4$

Population 5	R	m	n_1	d	$E(n)$	$\hat{\mu}_{\text{ATS}}$	$\hat{\mu}_{\text{ARS}}$	$\hat{\mu}_{\text{ARS}.\hat{\pi}}$	Rbias.$\hat{\mu}_{\text{ARS}.\hat{\pi}}$	MSE$(\bar{y}_{\text{tss}})$
	1	2	4	6	12	0.94	1.53	1.59	0.01	2.29
			8	5	23	0.88	1.48	1.55	0.01	1.01
			10	5	28	0.88	1.45	1.52	0.00	0.77
			15	4	40	0.86	1.50	1.55	0.01	0.48
		4	4	6	24	0.88	1.44	1.50	0.01	1.01
			8	5	46	0.82	1.45	1.52	0.00	0.44
			10	5	57	0.80	1.48	1.54	0.01	0.34
			15	4	81	0.83	1.66	1.72	0.01	0.19
	2	2	4	14	18	1.01	1.30	1.38	0.04	1.93
			8	13	33	0.90	1.11	1.18	0.04	0.76
			10	11	39	0.82	1.07	1.14	0.02	0.56
			15	8	52	0.77	1.12	1.17	0.02	0.34
		4	4	14	37	0.80	1.01	1.08	0.03	0.74
			8	13	66	0.78	1.04	1.06	0.04	0.29
			10	10	78	0.67	1.02	1.03	0.03	0.22
			15	8	104	0.66	1.07	1.12	0.02	0.12

Table 3.3 (continued)
Efficiency of the estimators, with $N = 200$, $C = 0$, $M = 4$

Population 6	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	$\text{Rbias.}\hat{\mu}_{ARS.\hat{\pi}}$	$\text{MSE}(\bar{y}_{tss})$
	1	2	10	5	18	0.98	1.09	1.09	0.02	1.53
			15	4	31	0.96	1.08	1.12	0.01	1.35
			20	4	37	1.00	1.12	1.16	0.01	1.25
			30	3	49	1.02	1.13	1.16	0.01	1.14
		4	10	5	35	1.05	1.28	1.34	0.02	0.28
			15	4	62	1.09	1.56	1.73	0.00	0.17
			20	4	74	1.19	1.76	1.95	0.01	0.12
			30	2	97	1.25	2.48	2.68	0.01	0.06
	2	2	10	10	28	0.87	1.07	1.11	0.01	1.34
			15	7	42	0.90	1.15	1.18	0.01	1.24
			20	6	48	0.93	1.16	1.19	0.01	1.18
			30	5	57	0.99	1.15	1.16	0.00	1.12
		4	10	10	56	0.80	1.30	1.40	0.01	0.46
			15	7	85	0.91	2.76	2.97	0.01	0.23
			20	6	95	1.01	4.19	4.43	0.01	0.18
			30	4	114	1.18	12.19	12.43	0.00	0.12

Table 3.4
Efficiency of the estimators, with $N = 200,128$, $C = 0$, $M = 4$

Population 7	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	$\text{Rbias.}\hat{\mu}_{ARS.\hat{\pi}}$	$\text{MSE}(\bar{y}_{tss})$
	1	2	4	6	19	0.74	0.69	0.68	0.00	1.72
			8	6	34	0.84	0.76	0.75	0.02	0.84
			10	5	41	0.79	0.77	0.76	0.01	0.69
			15	3	56	0.73	0.85	0.84	0.01	0.50
		4	4	6	38	0.72	0.70	0.70	0.02	0.68
			8	5	69	0.66	0.57	0.57	0.01	0.28
			10	5	83	0.72	0.68	0.69	0.00	0.21
			15	4	112	0.69	0.76	0.75	0.02	0.11
	2	2	4	15	33	0.76	0.68	0.70	0.02	1.13
			8	10	55	0.57	0.62	0.64	0.00	0.54
			10	9	63	0.61	0.70	0.72	0.01	0.45
			15	8	79	0.63	0.81	0.82	0.02	0.36
		4	4	13	66	0.54	0.51	0.51	0.01	0.36
			8	10	110	0.40	0.45	0.48	0.02	0.13
			10	9	126	0.38	0.47	0.49	0.01	0.09
			15	8	156	0.31	0.51	0.53	0.01	0.04
Population 8	R	m	n_1	d	$E(n)$	$\hat{\mu}_{ATS}$	$\hat{\mu}_{ARS}$	$\hat{\mu}_{ARS.\hat{\pi}}$	$\text{Rbias.}\hat{\mu}_{ARS.\hat{\pi}}$	$\text{MSE}(\bar{y}_{tss})$
	1	2	3	6	11	0.99	1.97	1.85	0.05	95.39
			5	5	17	0.93	1.88	1.84	0.01	55.29
			10	4	30	0.96	1.72	1.73	0.00	27.73
			13	3	36	0.96	1.60	1.61	0.00	22.02
		4	3	6	22	0.92	2.21	2.10	0.04	39.74
			5	5	35	0.86	2.49	2.49	0.02	21.13
			10	3	59	0.85	4.28	4.36	0.00	8.12
			13	3	71	1.00	6.22	6.27	0.00	5.24
	2	2	3	13	17	1.11	1.68	1.74	0.02	78.55
			5	10	25	0.88	1.50	1.53	0.01	40.42
			10	7	37	0.87	1.41	1.41	0.01	21.05
			13	5	42	0.90	1.34	1.34	0.00	17.92
		4	3	13	34	0.97	1.70	1.75	0.03	27.92
			5	10	49	0.80	1.82	1.87	0.01	13.18
			10	7	74	0.80	3.35	3.39	0.00	4.86
			13	5	83	0.80	5.03	5.06	0.00	3.24

In Population 5, ATS was not efficient at all (except in one situation), but ARS performed well and was always more efficient than both ATS and TSS. ARS was better with $R = 1$, relative to $R = 2$, because the population was not clustered and a large radius could have wasted the sample. However, because of a large cluster in the lower-right PSU, ARS was efficient for $R = 2$.

In Population 6, which was highly clustered but not rare, ATS was not as efficient as TSS in almost half of the cases, especially when the sample size was not very large. The performance of ARS was very good. Because of the large size of the clusters, ARS was better with $R = 2$ than $R = 1$, and, in one case, it was

12 times as efficient as TSS, while this number was 1.18 for ATS. With $R = 2$, if a nonempty cell was selected as the initial sample, many of the other nonempty cells would also be selected. With a large enough initial sample size, almost all of them would be selected, providing almost complete information on the population and higher efficiency in comparison with other designs.

In Population 7, which was an almost-ordinary population, ARS and ATS were not efficient. In this case, the population was not clustered, and ARS wasted the sample searching around cells with a response satisfying condition C that were almost all empty. In such situations, ATS is more efficient than ARS (this happened in some cases), since ATS spreads the additional sample size equally across all the PSUs.

Finally, in Population 8, which was rare and completely clustered, ATS was almost not efficient at all. Again, ARS performed very well; it was sometimes six times as efficient as TSS and ATS.

In addition, for Population 5, $\hat{\mu}_{AD.Inclu} = 0.025$ and $\hat{\mu}_{D.Inclu} = -0.004$; for Population 6, $\hat{\mu}_{AD.Inclu} = 0.020$ and $\hat{\mu}_{D.Inclu} = -0.010$; for Population 7, $\hat{\mu}_{AD.Inclu} = 0.027$ and $\hat{\mu}_{D.Inclu} = 0.006$; and, for Population 8, $\hat{\mu}_{AD.Inclu} = 0.016$ and $\hat{\mu}_{D.Inclu} = 0.004$. The means of the inclusion probabilities for the four populations were, respectively, 0.21, 0.31, 0.24 and 0.34. Again, the errors in estimating the inclusion probabilities were almost negligible.

Lastly, $\hat{\mu}_{\mathrm{ARS}.\hat{\pi}}$ showed significant efficiency, even higher than $\hat{\mu}_{\mathrm{ARS}}$ (sometimes for a larger sample size). Since $\hat{\mu}_{\mathrm{ARS}}$ is unbiased, $\hat{\mu}_{\mathrm{ARS}}$ is preferred when there is enough information to calculate it. When information for calculating "π_j"s is lacking, $\hat{\mu}_{\mathrm{ARS}.\hat{\pi}}$ is a very good alternative for estimating the population mean with almost no bias.

Costs are not discussed, because this factor favours ARS, which is a cheaper design in comparison with ATS and TSS. Since ARS is more efficient than the other designs without considering costs, it is obvious that with costs factored in, the efficiency of ARS would be higher again. On the other hand, if the costs of much travelling under ATS and TSS are almost the same as the cost of searching more cells to find whether they satisfy condition C (not measuring them exactly) to calculate the unbiased π-estimator, then the comparison is fair.

4 Discussion

The adaptive rectangular sampling (ARS) design is an adaptive design that is easy to manage, saves on travel, is easy to calculate, is neighbourhood-free, and controls the final sample size.

The design is adaptive because the final sample size depends on observed values, and the design is able to find rare clustered events.

It is easy to manage because it is straightforward in determining the places that should be investigated for the additional sample. The design uses the intuitive behaviour of field biologists – once they find a rare event, they want to search in the immediate neighbourhood. It is even easier than adaptive cluster sampling (ACS). In addition, ARS can perform in both a two-stage form and a one-stage form. With this design, unlike with adaptive two-stage sequential sampling (ATS), there is no need to worry about the size, location and shape of primary sampling units. Unlike ACS, incomplete ACS (IACS) and ATS, it is possible for ARS to indicate the entire potential sample that the samplers need to select before they start sampling.

As for the travel-saving feature of ARS, there is no difference between adaptive designs such as ACS, ARS and ATS in the first phase of the second stage, for selecting the initial sample. But, in the second phase, ARS travels between cells generally less than ACS and IACS (with its edge units) and, especially, much less than ATS and two-stage sampling, with equal sample sizes. Because of this feature, ARS would be appropriate for costly travelling surveys of clustered populations, regardless of its efficiency.

ARS is easy to calculate, because the inclusion probabilities for the final sample size are easy to calculate, and this means that the π- estimator can be used instead of Murthy's estimator. Murthy's estimator, equation 3.1, is strongly dependent on the size of the initial sample (and on estimating q); as initial samples could be small in some situations, this is a weakness of Murthy's estimator. Therefore, one of the advantages of ARS as a sequential design is its avoidance of Murthy's estimator and its use of the π- estimator instead. In addition, calculating the π- estimator in IACS is not very easy, because it is a little complicated to estimate "π_j"s (see Chao and Thompson 1999). As discussed in Section 2.1, it is easy to calculate or estimate "π_j"s in ARS, compared with the method used by Chao and Thompson (1999).

The design is neighbourhood-free, in the sense that it does not follow the neighbourhood as in ACS and IACS; this would be complicated for the sampler after certain steps. ARS is an easy design for samplers, especially for difficult environments. ACS has not yet been used on a routine basis in field surveys for forest inventory and biodiversity monitoring, as there are also practical difficulties in field implementation (Yang et al. 2011). A design like ARS may be more appropriate in such environments.

The design controls the final sample size well with the choice of radius R. This paper presents an easy version of ARS. ARS can be performed in different ways (e.g., someone could plan a design to sample around a cell instead of investigating all the cells around it). This is a suggestion for future work on ARS.

To use ARS, it is important to know that the population units are separated in a cluster form; otherwise, the design would waste the sample units. This is one of the disadvantages of ARS. An advantage of this design is its expansion of the definition of clusters. Because the designer can change the radius R, a cluster in ARS consists of units that are around each other even at a distance, and there is no need for them to be adjacent.

Compared with other designs, ARS has some of the same advantages. Like ACS, it takes advantage of clustering to find rare events; like ATS, it does not need to follow the neighbourhood. And, like IACS, it controls the final sample size well.

ARS is a new design, and it should be evaluated on real populations to enable researchers to find out its abilities, advantages and disadvantages.

Appendix

For π_{hj}, it is easy to see that

$$\pi_{hj} = P\left(B_{hj}\right) = P\left(\cup_{k\in s_{hj}} A_{hk}\right) = 1 - P\left(\cap_{k\in s_{hj}} A'_{hk}\right) = 1 - \frac{\binom{N_h - f_{hj}}{n_{1h}}}{\binom{N_h}{n_{1h}}},$$

and, for $\pi_{hjj'}$, using the fundamental probability principle,

$$\begin{aligned}
\pi_{hjj'} &= P\left(B_{hj} \cap B_{hj'}\right) = 1 - P\left(B_{hj} \cap B_{hj'}\right)' = 1 - P\left(B'_{hj} \cup B'_{hj'}\right) \\
&= 1 - \left[P\left(B'_{hj}\right) + P\left(B'_{hj'}\right) - P\left(B'_{hj} \cap B'_{hj'}\right)\right] \\
&= 1 - \left[1 - P\left(\cup_{k \in s_{hj}} A_{hk}\right) + 1 - P\left(\cup_{k \in s_{hj'}} A_{hk}\right) - \left(1 - P\left(\cup_{k \in s_{hjj'}} A_{hk}\right)\right)\right] \\
&= 1 - \frac{\binom{N_h - f_{hj}}{n_{1h}}}{\binom{N_h}{n_{1h}}} - \frac{\binom{N_h - f_{hj'}}{n_{1h}}}{\binom{N_h}{n_{1h}}} + \frac{\binom{N_h - f_{hjj'}}{n_{1h}}}{\binom{N_h}{n_{1h}}}.
\end{aligned}$$

References

Brown, J.A. (2003). Designing an efficient adaptive cluster sample. *Environmental and Ecological Statistics*, 10, 95-105.

Brown, J.A., and Manly, B.J.F. (1998). Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, 5, 49-63.

Brown, J.A., Salehi, M.M., Moradi, M., Bell, G. and Smith, D.R. (2008). An adaptive two-stage sequential design for sampling rare and clustered populations. *Population Ecology*, 50, 239-245.

Chao, C.T., and Thompson, S.K. (1999). Incomplete adaptive cluster sampling designs. In *Proceedings of the Survey Research Methods Section*, American Statistical Association, 345-350.

Salehi, M.M., and Seber, G.A.F. (1997). Two-stage adaptive cluster sampling. *Biometrics*, 53, 959-970.

Salehi, M.M., and Smith, D.R. (2005). Two-stage sequential sampling: A neighborhood-free adaptive sampling procedure. *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 84-103.

Smith, D.R., Brown, J.A. and Lo, N.C.H. (2004). Application of adaptive sampling to biological populations. In *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters*, (Ed., W.L. Thompson), Covelo, California: Island Press. 75-122.

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

Su, Z., and Quinn, T.J., II (2003). Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. *Environmental and Ecological Statistics*, 10, 17-41.

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

Thompson, S.K. (1991). Stratified adaptive cluster sampling. *Biometrika*, 78, 389-397.

Thompson, S.K., and Seber, G.A.F. (1996). *Adaptive Sampling*, New York: John Wiley & Sons, Inc.

Yang, H. (2011). *New Adaptive Plot Designs for Sampling Rare and Clustered Populations*. Ph.D. thesis at the Faculty of Forest Sciences and Forest Ecology of Georg-August-Universität Göttingen.

Yang, H., Kleinn, C., Fehrmann, L., Tang, S. and Magnussen, S. (2011). A new design for sampling with adaptive sample plots. *Environmental and Ecological Statistics*, 8, 223-237.