Thailand Statistician
July 2017; 15(2): 111-127
http://statassoc.or.th
Contributed paper

Stratified Adaptive Cluster Sampling with Spatially Clustered Secondary Units

Mena Patummasut*[a] and John J. Borkowski [b]
[a] Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
[b] Department of Mathematical Sciences, Montana State University, Montana 59715, USA.
*Corresponding author; e-mail: mena_patummasut@yahoo.com

Received: 28 April 2016
Accepted: 12 December 2016

Abstract

This study introduces the use of adaptive cluster sampling with spatially clustered secondary units when the population study area can be stratified into smaller areas or strata based on available prior information, which can increase the precision of estimation. The method by which secondary units are adaptively added when the primary units are formed by a spatial cluster of secondary units is described, and has the advantage of saving cost and time of travelling and observing units in the sample compared to the standard approach of stratified adaptive cluster sampling with an initial sample random sample per stratum.

An unbiased estimator of the mean and its variance by applying the Horvitz-Thompson estimator is presented, and the advantages and disadvantages of stratified adaptive cluster sampling with spatially clustered secondary units in comparison to stratified adaptive cluster sampling based on a stratified random sample are described.

Keywords: Stratified adaptive cluster sampling, stratified sampling, Horvitz-Thompson estimator, spatial sampling.

1. Introduction

Unlike conventional sampling designs, an adaptive sampling design is a sampling design in which values of the variable of interest observed during the survey are considered before selecting the next units to include in the sample. That is, the procedure adaptively selects the sampling units based on values of the variable of interest (Thompson and Seber 1996). For spatially clustered populations such as rare species of plants and animals, it is often more appropriate and efficient to use adaptive cluster sampling (ACS) (Thompson 1990). In ACS, an initial sample of units is selected by using a conventional sampling design. Whenever the value of the variable of interest of a sampled unit satisfies a condition, its neighboring units are also added to the sample and sampling continues until no sampled unit satisfies the condition. ACS has been widely used in real-world situations. For example, it is used for surveys of forest inventories (Roesch 1993; Talvitie et al. 2006), deforestation (Magnussen et al. 2005), plants (Philippi 2005), macroalgae (Glodberg et al. 2006), larval sea lampreys (Sullivan et al. 2008), mussels (Smith et al. 2003), waterfowl (Smith et al. 1995), herpetofauna (Noon et al. 2006), sediment load (Arabkhedri et al. 2010), and hydroacoustic surveys (Conners and Schwager 2002).

Adaptive cluster sampling that is applied to a stratified population is known as stratified ACS, and was proposed by Thompson (Thompson 1991). In stratified ACS, an initial stratified sample is selected. Whenever the value of the variable of interest of a sampled unit satisfies a pre-specified condition, its neighborhood is also added to the sample until no more units satisfying such condition are found. For example, stratified ACS has been used to estimate the total carbon dioxide flux released by magmatic and hydrothermal sources at the Mud Volcano area within Yellowstone National Park (Boomer et al. 2000).

The disadvantage of ACS is that if a sampling unit is large, it will be time consuming and the cost will be high to observe all sampled units. To overcome this drawback, adaptive cluster sampling with spatially clustered secondary units (ACS-SCSU) was proposed by introducing the approach of adaptively adding units in the sample under ACS (Patummasut and Borkowski 2014). In ACS-SCSU, a primary unit is partitioned into smaller secondary units, and an initial sample of primary units is selected by simple random sampling without replacement. For each primary unit in the initial sample, neighboring secondary units are added to the sample and observed whenever the variable of interest of a secondary unit satisfies a specific condition. This procedure continues until no adaptively sampled secondary units satisfy a condition. ACS-SCSU is a special case of adaptive cluster sampling with primary and secondary units presented by Thompson (Thompson 1991a).

The set of all secondary units that satisfy the condition in the neighborhood of each other is called a network, while the secondary units that were adaptively sampled but did not satisfy the condition are called edge units. A network with its corresponding edge units is called a cluster. Units that do not satisfy the condition, including edge units, are considered networks of size one. Since secondary units are adaptively added to the sample instead of primary units, this yields a smaller final sample size, taking less time and cost to travel and observe all sampled units. Thus, ACS-SCSU is a time and cost-effective design, and it is appropriate when sampling a spatially aggregated population.

Two-stage ACS is one of the sampling designs allowing less cost and time of traveling observing sampled units. Under two-stage ACS, a sample of primary units is selected, then a subsample of secondary units within each sampled primary unit is taken. For this sampling design, there are two stages of sampling, while ACS-SCSU is one-stage sampling, leading less complicated and easier in estimation.

When prior information about a population is available, stratification can be done based on such prior information or based on the vicinity of the units when spatial correlation is assumed to be exist. The population can be stratified by grouping similar units into a stratum and dissimilar units into different strata. The advantage of stratification is that the variance of an estimator can be reduced (Lohr 1999). Since adaptive sampling can be used with conventional stratified sampling such as stratified ACS, and it is known that stratification can increase precision in estimation, this paper applies ASC-SCSU with stratified sampling, and it is called stratified ACS-SCSU.

2. Stratified Adaptive Cluster Sampling with Spatially Clustered Secondary Units and Technical Notation

Assume a population is partitioned into L strata, of which stratum h contains N_h equal-sized rectangular primary units, and let N be the total number of primary units in the population. Each primary unit u_{hi} ($h = 1, 2, 3, ..., L,\ i = 1, 2, 3, ..., N_h$), the i^{th} unit in stratum $h,$ is divided into 2 rows and 2 columns forming 4 equal-sized rectangular secondary units u_{hij} ($j = 1, 2, 3, 4$). Let m be the number of secondary units in each primary unit. In this research, $m = 4.$ Thus, there are $N = 4\sum_{h=1}^{L} N_h$ secondary units in the study region. Let y_{hij} be the value of the variable of interest corresponding to each secondary unit u_{hij}. The parameter of interest in this paper is the population mean $\mu = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j=1}^{m} y_{hij}$ or the population total $\tau = N\mu.$

2.1. Stratified adaptive cluster sampling with spatially clustered secondary units design

In stratified ACS-SCSU, an initial stratified sample of primary units of size n is obtained by taking an initial sample of primary units of size n_h from stratum h using simple random sampling without replacement. Thus, the total number of secondary units in the initial sample is $n = \sum_{h=1}^{L} n_h.$ Then for each sampled primary unit u_{hi} ($i = 1, 2, ..., n_h$), neighboring secondary units are added to the sample and observed whenever the variable of interest y_{hij} of a sampled secondary unit u_{hij} satisfies a specific condition $C,$ which is determined a priori by the researcher. It is typically an interval for the variable of interest of a secondary unit. That is, $y_{hij} > c$ where c is a constant defining the condition C. This procedure continues until no adaptively sampled secondary units satisfy C. The set of all secondary units satisfying C as a result of u_{hij} being in the initial sample form a network while the secondary units that were adaptively sampled but did not satisfy C are edge units. Together these units form a cluster. Note that the selection in one stratum may result in the addition of units from other strata to the sample if adaptive sampling is allowed to cross stratum boundaries. In this case, a cluster may intersect more than one stratum. Stratified ACS-SCSU is a special case combining of adaptive cluster sampling with primary and secondary units (Thompson 1991a) with stratified adaptive cluster sampling (Thompson 1991b).

For instance, suppose that the population is composed of $N = 100$ primary units, and each is divided into 2 rows and 2 columns forming 4 secondary units. Thus, there are 400 secondary units in the population. Suppose the population is stratified in to $L = 2$ equal-sized strata, each having $N_h = 50$ primary units or 200 secondary units as shown Figure 1. To illustrate a stratified ACS-SCSU scheme, suppose $n_h = 2$ (and, hence, $n = 4$). The stratified initial sample of 4 primary units is obtained by selecting an initial sample of 2 primary units from each stratum using simple random sampling without replacement. The obtained stratified initial sample is shown in Figure 2. Here, the condition C is defined to be $y_{hij} > 0.$ That is a neighboring secondary unit is adaptively added when a sampled secondary unit has value of the variable of interest $y_{hij} > 0.$ The neighborhood for a secondary unit u_{hij} in this example is defined to be the four adjacent units on the left, right, above and below unit u_{hij}.

0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	4	3	2	0	18	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	6	10	15	3	5	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	14	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	13	0	0	0	5	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0	0	9	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0

Figure 1 A population stratified into 2 strata formed by the vertical center line. The solid lines within each stratum form the 50 within-stratum primary units. The dotted lines form the 4 secondary units within each primary unit

For the initial sample in Figure 2, because the leftmost sampled primary unit of stratum 1 and the rightmost sampled primary unit of stratum 2 both contain four secondary units with $y_{hij} = 0$ (and which do not satisfy C), no additional unit will be adaptively added to the sample in the neighborhoods of these primary units. On the other hand, the rightmost sampled primary unit of stratum 1 contains 4 secondary units with y_{hij} values of 0, 2, 3 and 15. Because the three secondary units with values of 2, 3 and 15 satisfy condition C, their neighboring secondary units are adaptively sampled and observed. This procedure continues until no adaptively sampled secondary units having y_{hij} greater than zero are found. The resulting cluster whose network consists of 12 dark grey secondary units with 17 edge units in light grey is shown in Figure 3. Notice that this cluster intersects both stratum 1 and 2. Finally, for the rightmost sampled primary units of stratum 2, four secondary units with values y_{hij} of 2, 5, 6 and 14 satisfy condition C, so their neighboring secondary units are adaptively sampled. This procedure continues until no adaptively sampled secondary units have y_{hij} greater than zero. This results in the second cluster whose network consists of 8 dark grey secondary units with 10 edge units in light grey as shown in Figure 3.

0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	4	3	2	0	18	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	6	10	15	3	5	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	14	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	13	0	0	0	5	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0	0	9	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0

Figure 2 A stratified initial sample of 2 primary sampling units per stratum

0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	4	3	2	0	18	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	6	10	15	3	5	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	14	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	13	0	0	0	5	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0	0	9	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0

Figure 3 A final sample of adaptively sampled secondary units

2.2. Estimation

The Horvitz-Thompson estimator (Horvitz and Thompson 1952) is applied in the stratified ACS-SCSU design, but it is based on network inclusion probabilities. Let K be the number of distinct networks in the population without regard to stratum boundaries. Let y_k be the total of the *y*-values in the k^{th} network of the population. Let α_k be the probability that network k is included in the final sample. α_k is called the inclusion probability of network k.

For stratified ACS-SCSU, α_k is the probability at least one of the n primary units in the initial stratified sample intersects network k. Let x_{hk} be the number of primary units in stratum h that intersect network k. Using complementary probability, the probability that network k is included in the final sample is

$$\alpha_k = 1 - \prod_{h=1}^{L} \frac{\binom{N_h - x_{hk}}{n_h}}{\binom{N_h}{n_h}}. \tag{1}$$

Let E_i be the event that network i is included in the final sample. Thus, $P(E_j) = \alpha_j$ and $P(E_k) = \alpha_k$ for network j and k respectively. Let α_{jk} be the probability distinct networks j and k are both included in the final sample. α_{jk} is called the joint inclusion probability of networks j and k with

$$\begin{aligned}\alpha_{jk} = P(E_j \cap E_k) &= P(E_j) + P(E_k) - P(E_j \cup E_k) \\ &= \alpha_j + \alpha_k - P(E_j \cup E_k)\end{aligned} \tag{2}$$

where $P(E_j \cup E_k)$ is the probability that at least one of networks j and k is included in the final sample. Using complementary probability

$$P(E_j \cup E_k) = 1 - \prod_{h=1}^{L} \frac{\binom{N_h - x_{hi} - x_{hj} + x_{hjk}}{n_h}}{\binom{N_h}{n_h}}, \tag{3}$$

where x_{hjk} is the number of primary units in stratum h which intersect both networks j and k. Hence, substitution yields

$$\alpha_{jk} = 1 - \left(\prod_{h=1}^{L} \frac{\binom{N - x_{hj}}{n_h}}{\binom{N_h}{n_h}} + \prod_{h=1}^{L} \frac{\binom{N - x_{hk}}{n_h}}{\binom{N_h}{n_h}} - \prod_{h=1}^{L} \frac{\binom{N_h - x_{hi} - x_{hj} + x_{hjk}}{n_h}}{\binom{N_h}{n_h}} \right). \tag{4}$$

Define Z_k be the indicator variable having value 1 if network k is included in the final sample and zero otherwise. Using calculated inclusion probabilities, the Horvitz-Thompson estimator is implemented in the stratified ACS-SCSU design, and the unbiased estimator of the population mean μ is

$$\hat{\mu}_{SSC} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k Z_k}{\alpha_k}, \tag{5}$$

where K is the total number of networks in the population. First, note that for (5) it is not necessary to know K because if network k is not included in the final sample, then $Z_k = 0$, and it contributes zero to (5). Second, even though different primary units in the initial sample might intersect the same network, only the distinct networks observed in the final sample are utilized in the formula. Finally, one network can intersect more than one stratum, but this does not affect (5).

Applying the results of Horvitz and Thompson (1952), the variance of $\hat{\mu}_{SSC}$ is

$$v(\hat{\mu}_{SSC}) = \frac{1}{N^2}\sum_{j=1}^{K}\sum_{k=1}^{K} y_j y_k \left(\frac{\alpha_{jk}}{\alpha_j \alpha_k} - 1\right). \tag{6}$$

Note that $\alpha_{kk} = \alpha_k$. An estimator of this variance is

$$\hat{v}(\hat{\mu}_{SSC}) = \frac{1}{N^2}\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{y_j y_k Z_j Z_k}{\alpha_{jk}} \left(\frac{\alpha_{jk}}{\alpha_j \alpha_k} - 1\right). \tag{7}$$

$\hat{v}(\hat{\mu}_{SSC})$ is unbiased if all joint inclusion probabilities α_{jk} are greater than zero.

2.3. An Example

In Section 2.1, stratified ACS-SCSU scheme was demonstrated by applying it in the small population in Figure 1, and the obtained final sample is shown in Figure 3. Next, the calculations used for estimation will be illustrated. The sample consists of 2 networks with a total of y-values greater than zero. The left and the right networks have a total y-values of 74 and 40 respectively, that is, $y_1 = 74$ and $y_2 = 40.$ The number of strata in the population is $L = 2.$ Each stratum consists of 50 primary units, thus $N_h = 50.$ The number of primary units in the population is $N = 100,$ and the initial sample size in stratum h is $n_h = 2,$ so the initial sample size is $n = 4.$ The number of primary units in stratum 1 that intersect network 1 is 3, thus $x_{11} = 3.$ The number of primary units in stratum 2 that intersect network 1 is 1, thus $x_{21} = 1.$ From (1), the probability that network 1 is included in the sample is

$$\alpha_1 = 1 - \prod_{h=1}^{2} \frac{\binom{N_h - x_{h1}}{n_h}}{\binom{N_h}{n_h}} = 1 - \frac{\binom{N_1 - x_{11}}{n_1}}{\binom{N_1}{n_1}} \frac{\binom{N_2 - x_{21}}{n_2}}{\binom{N_2}{n_2}} = 1 - \frac{\binom{50-3}{2}}{\binom{50}{2}} \frac{\binom{50-1}{2}}{\binom{50}{2}} = 0.1529.$$

There are no primary units in stratum 1 that intersect network 2, thus $x_{12} = 0.$ The number of primary units in stratum 2 that intersect network 2 is 4, thus $x_{22} = 4.$ Hence, the probability that network 2 is included in the sample is

$$\alpha_2 = 1 - \prod_{h=1}^{2} \frac{\binom{N_h - x_{h1}}{n_h}}{\binom{N_h}{n_h}} = 1 - \frac{\binom{N_1 - x_{12}}{n_1}}{\binom{N_1}{n_1}} \frac{\binom{N_2 - x_{22}}{n_2}}{\binom{N_2}{n_2}} = 1 - \frac{\binom{50-0}{2}}{\binom{50}{2}} \frac{\binom{50-4}{2}}{\binom{50}{2}} = 0.1551.$$

By using (5), the estimate of the mean is

$$\hat{\mu}_{SSC} = \frac{1}{N}\sum_{k=1}^{K} \frac{y_k Z_k}{\alpha_k} = \frac{1}{100}\left(\frac{74}{0.1529} + \frac{40}{0.1551}\right) = 7.4187.$$

The number of primary units in stratum 1 which intersect both networks1 and 2 is 0, thus $x_{112} = 0.$ The number of primary units in stratum 2 which intersect both networks1 and 2 is 0, thus $x_{212} = 0.$

$$\alpha_{jk} = 1 - \left(\prod_{h=1}^{L} \frac{\binom{N - x_{hj}}{n_h}}{\binom{N_h}{n_h}} + \prod_{h=1}^{L} \frac{\binom{N - x_{hk}}{n_h}}{\binom{N_h}{n_h}} - \prod_{h=1}^{L} \frac{\binom{N_h - x_{hj} - x_{hk} + x_{hjk}}{n_h}}{\binom{N_h}{n_h}} \right)$$

$$\alpha_{12} = 1 - \left(\frac{\binom{N_1 - x_{11}}{n_1}}{\binom{N_1}{n_1}} \frac{\binom{N_2 - x_{21}}{n_2}}{\binom{N_2}{n_2}} + \frac{\binom{N_1 - x_{12}}{n_1}}{\binom{N_1}{n_1}} \frac{\binom{N_2 - x_{22}}{n_2}}{\binom{N_2}{n_2}} \right.$$

$$\left. - \frac{\binom{N_1 - x_{11} - x_{12} + x_{112}}{n_1}}{\binom{N_1}{n_1}} \frac{\binom{N_2 - x_{21} - x_{22} + x_{212}}{n_2}}{\binom{N_2}{n_2}} \right)$$

$$= 1 - \left(\frac{\binom{50-3}{2}}{\binom{50}{2}} \frac{\binom{50-1}{2}}{\binom{50}{2}} + \frac{\binom{50-0}{2}}{\binom{50}{2}} \frac{\binom{50-4}{2}}{\binom{50}{2}} - \frac{\binom{50-3-1+0}{2}}{\binom{50}{2}} \frac{\binom{50-0-4+0}{2}}{\binom{50}{2}} \right)$$

$$= 0.0218.$$

From (7), the estimate of the variance estimator is

$$\hat{v}(\hat{\mu}_{SSC}) = \frac{1}{N^2} \sum_{j=1}^{K} \sum_{k=1}^{K} \frac{y_j y_k Z_j Z_k}{\alpha_{jk}} \left(\frac{\alpha_{jk}}{\alpha_j \alpha_k} - 1 \right)$$

$$= \frac{1}{N^2} \left[\frac{y_1^2}{\alpha_1} \left(\frac{1}{\alpha_1} - 1 \right) + \frac{y_2^2}{\alpha_2} \left(\frac{1}{\alpha_2} - 1 \right) + \frac{y_1 y_2}{\alpha_{12}} \left(\frac{\alpha_{12}}{\alpha_1 \alpha_2} - 1 \right) \right]$$

$$= \frac{1}{100^2} \left[\frac{74^2}{0.1529} \left(\frac{1}{0.1529} - 1 \right) + \frac{40^2}{0.1551} \left(\frac{1}{0.1551} - 1 \right) \right.$$

$$\left. + \frac{(74^2)(40^2)}{0.0218} \left(\frac{0.0218}{(0.1529)(0.1551)} - 1 \right) \right]$$

$$= 21.7788.$$

The number of secondary units in the sample, which is found by counting the number of dark-shaded and light-shaded secondary units in Figure 3, is 54. This is equivalent to 54/4 = 13.5 primary units in terms of sampling effort. On the other hand, under stratified ACS with the same initial sample, the number of primary and secondary units in the final sample are 26 and 104, respectively, as shown in Figure 4. For stratified ACS, neighboring primary units are adaptively added to the sample instead of secondary units, thus stratified ACS requires a much larger sample size than stratified ACS-SCSU. Thus, ACS-SCSU is much more cost-effective. It can be concluded that stratified ACS-SCSU gives less cost and time spent travelling and observing units in the final sample than stratified ACS for this example. In other words, the cost and time of travelling and observing units in the stratified ACS sample is twice more expensive than that of stratified ACS-SCSU sample.

0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	4	3	2	0	18	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	6	10	15	3	5	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	14	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	13	0	0	0	5	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0	0	9	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0

Figure 4 A final sample of stratified ACS whose initial sample is shown in Figure 2

3. Simulation Study and Result

To investigate the performance of stratified ACS-SCSU compared to stratified ACS of primary units, California redwood trees data (Strauss 1975) was used in a simulation. The simulation is composed of 1000 iterations, and for the i^{th} iteration, the value for the corresponding estimator $\hat{\mu}_i$ under both sampling designs is calculated. The formula used to estimate an estimator variance is the sample variance of the estimates:

$$\hat{v}(\hat{\mu}) = \frac{1}{999}\sum_{i=1}^{1000}(\hat{\mu}_i - \bar{\mu})^2, \tag{8}$$

where $\bar{\mu}$ is the sample mean of the 1000 $\hat{\mu}_i$ values (Dryver and Thompson 2005). Note that the modified Horvitz-Thompson estimator based on initial intersection probabilities is used to estimate the population mean for stratified ACS.

The effective sample size υ_i, which is the number of primary units in the final sample, of stratified ACS-SCSU were calculated for i^{th} iteration, and the estimated expected effective sample size of primary units under stratified ACS-SCSU is calculated by the formula

$$\upsilon = \frac{1}{1000}\sum_{i=1}^{1000}\upsilon_i, \tag{9}$$

while the expected sample size of stratified ACS for each initial sample size is calculated from

$$\upsilon = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\pi_{hi}, \tag{10}$$

which is the sum of inclusion probabilities in the population where

$$\pi_{hi} = 1 - \prod_{g=1}^{L} \frac{\binom{N_g - m_{ghi} + a_{ghi}}{n_g}}{\binom{N_g}{n_g}} \tag{11}$$

and m_{ghi} is the number of (primary) units in the intersection of stratum g with the network that contains unit u_{hi}, and a_{ghi} is the total number of units in the intersection of stratum g with distinct networks having unit u_{hi} as an edge unit.

The California redwood trees study area consists of 400 primary units, and each contains 4 secondary units. In the simulation, the population was stratified into 2 equal-sized strata (each stratum consists of 200 primary units) as shown in Figure 5, and into 4 equal-sized strata (each stratum consists of 100 primary units) as shown in Figure 6. To compare stratified ACS to ACS-SCSU, three conditions C_0, C_1 and C_2 were used. That is, adaptively sample if $y > 0$ for C_0, adaptively sample if $y > 1$ for C_1, and adaptively sample if $y > 2$ for C_2. For the 1000 simulated samples, $\hat{v}(\hat{\mu})$ and υ for stratified ACS and stratified ACS-SCSU were compared.

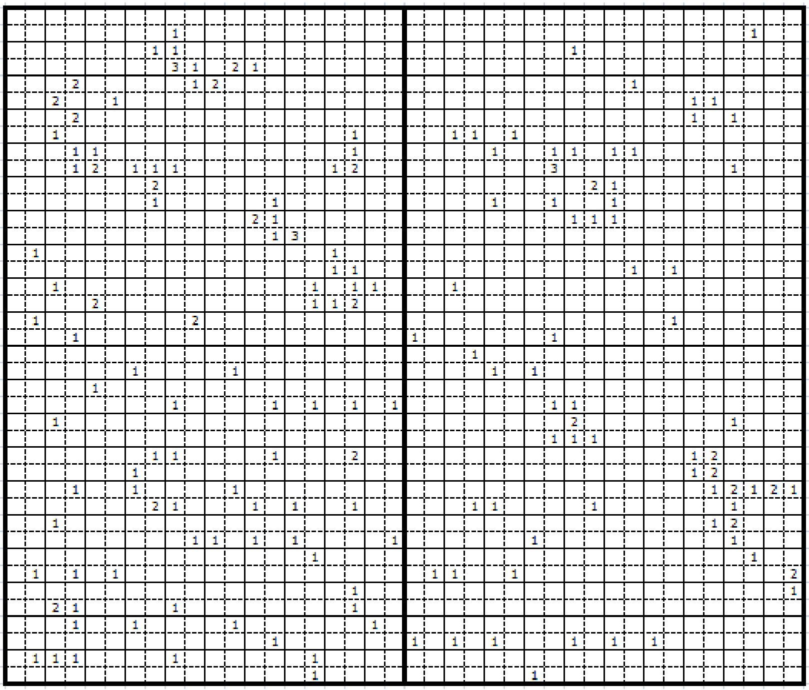

Figure 5 California redwood trees data partitioned into 2 strata of 200 primary units and 800 secondary units

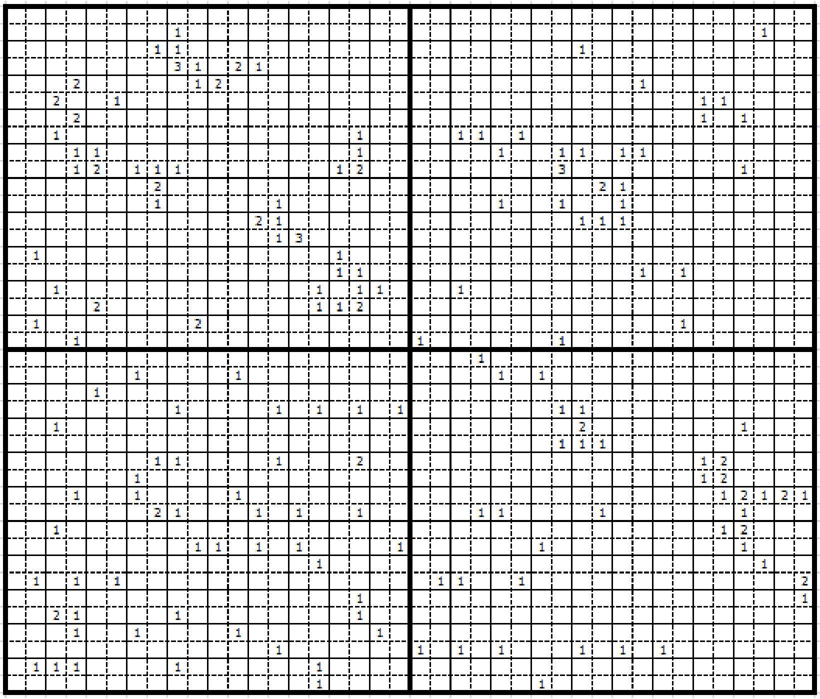

Figure 6 California redwood trees data partitioned into 4 strata of 100 primary units and 400 secondary units

The simulation results for the California redwood trees population with 2 and 4 strata were presented in Table 1 and 2, respectively. These results indicated that, for conditions C_0 and C_1, stratified ACS had a smaller average estimated variance than stratified ACS-SCSU for each initial sample size n, and the estimated relative efficiencies (ratio of the estimated variances: R. E.) were less than one. Hence, stratified ACS is more efficient than stratified ACS-SCSU when considering the same value of the initial sample size. However, for condition C_2, the estimated relative efficiencies for some initial sample sizes are greater than one. For these cases, stratified ACS-SCSU is not efficient than stratified ACS. These comparisons, however, are misleading because the effective sample size υ under SACS is always larger (often much larger) than υ under SACS-SCSU. The effective sample size υ indicates the true sampling effort generated by the researcher instead of the initial sample size. Therefore, it is appropriate to compare variances between two ACS designs when the effective sample sizes are similar (Turk and Borkowski 2005).

Note that for the same initial sample size n in the population with 2 and 4 strata, υ is smaller for stratified ACS-SCSU than stratified ACS for both conditions C_0 with $y > 0$, C_1 with $y > 1$ and C_2 with $y > 2$. Thus, stratified ACS-SCSU will, on average, be more efficient in terms of cost and time spent to observe the units in the final sample.

The average estimated variances in Table 1 and 2 were plotted against the estimated expected effective sample size in the graphs in Figure 7. For the population with 2 and 4 strata and for both conditions, the graph of the estimated variances for stratified ACS-SCSU was below the graph for stratified ACS. This implies that the estimated variances of the estimator of the mean for stratified ACS-SCSU are smaller, on average, than the estimated variances under stratified ACS for the same average effective sample size. Thus, when considering the same final sample size, stratified ACS-SCSU is more efficient than stratified ACS.

4. Discussion

The stratified ACS-SCSU is a time and cost-effective sampling design to apply in rare and clustered population partitioned into strata. In practice, sampling cost and time are important factors for the sampler to consider when choosing an appropriate sampling design to use. Thus, this motivates the researcher to develop time and cost-effective sampling designs such as adaptive web sampling (Thompson 2006), path sampling (Patummasut and Dryver 2012), stratified ACS (Thompson 1991b), and ASC-SCSU (Patummasut and Borkowski 2014). Additionally, it is known that stratification can increase the precision of the estimator (Lohr 1999), thus stratified ACS-SCSU would be a sampling design that can save time, money and effort of the researchers with a higher statistical precision for a spatially aggregated population.

To apply stratified ASC-SCSU in a geographical study area, the study area may be partitioned into smaller but similar areas according to a known variable such as soil or habitat type or plant species. Although a large area appears to be homogeneous, stratification based on geographical locations can be used to produce a sample aggregated over the entire area. Note that the overall variability in the estimation can be reduced when stratification leads to reduced variability within a stratum with larger variation across strata. Moreover, the most precise estimator will be obtained if the units in each stratum are as similar as possible (Thompson 2002).

For stratified ACS-SCSU, ACS-SCSU is applied in every stratum in the population providing good spatial coverage and, hence a representative sample of the population. On the other hand, a sample from two-stage ACS will consist of sampling additional secondary units from a certain area because all secondary units in neighboring primary units must be sampled, which results in increased sampling effort and cost, and often less efficient estimation.

In stratified ACS-SCSU, the unbiased estimator of the population mean and its variance were developed by applying the Horvitz-Thompson estimator based on network inclusion probabilities. Horvitz-Thompson estimation has been frequently used in estimation for a variety of sampling designs (Lucas and Seber 1977; Thompson 1990; Thompson 1991; Borkowski 2003; Akanisthanon et al. 2010; Mohammadi and Salehi 2011) because it is a general estimator of the population total or mean for any probability sampling design, and can be used as long as the inclusion probabilities can be provided (Horvitz and Thompson 1952).

It is known that the estimator can be improved by using the Rao-Blackwell method (Rao 1945; Blackwell 1947) and, when possible, using ratio estimation. For example, Salehi (1999), Dryver and Thompson (2005), and Chao et al. (2011) used the Rao-Blackwell method to improve estimators in ACS, while Chao (2004) and Dryver and Chao (2007) offered ratio estimators in ACS. For ratio estimation, auxiliary information is utilized to get the better estimation. In addition, Chutiman (2010) proposed the new ratio estimator in stratified ACS. Possibly, a ratio estimator for stratified ACS-SCSU could be developed based on the auxiliary information obtained from sampling survey, so this could be an interesting future research topic.

Table 1 Results from the simulation study on the California redwood trees data population with 2 strata under the condition C_0: $y > 0$, C_1: $y > 1$and C_2: $y > 2$ and the estimated relative efficiencies R. E.= $\hat{v}(\hat{\mu}_{SACS})$ / $\hat{v}(\hat{\mu}_{SSC})$

		$C_0: y > 0$					$C_1: y > 1$					$C_2: y > 2$				
		SACS		SACS-SCSU			SACS		SACS-SCSU			SACS		SACS-SCSU		
n	$n_{□}$	υ	$\hat{v}(\hat{\mu}_{SACS})$	υ	$\hat{v}(\hat{\mu}_{SSC})$	R.E.	υ	$\hat{v}(\hat{\mu}_{SACS})$	υ	$\hat{v}(\hat{\mu}_{SSC})$	R.E.	υ	$\hat{v}(\hat{\mu}_{SACS})$	υ	$\hat{v}(\hat{\mu}_{SSC})$	R.E.
2	1	12.49	0.3178	3.02	0.4075	0.78	3.77	0.3816	2.09	0.4635	0.82	2.66	0.4366	2.01	0.4318	0.99
4	2	24.39	0.1475	6.06	0.1709	0.86	7.49	0.2055	4.15	0.2133	0.96	5.30	0.1769	4.00	0.2293	1.30
6	3	35.74	0.0877	8.95	0.1078	.081	11.17	0.1274	6.21	0.1517	0.84	7.93	0.1321	5.99	0.1464	1.11
8	4	46.56	0.0675	11.79	0.0844	0.80	14.80	0.1054	8.25	0.1123	0.94	10.55	0.0916	7.95	0.1157	1.26
10	5	56.88	0.0555	14.65	0.0687	0.81	18.40	0.0742	10.31	0.0873	0.85	13.16	0.0773	9.93	0.1794	1.03
20	10	101.94	0.0240	28.18	0.0316	0.76	35.79	0.0374	20.25	0.0420	0.89	26.02	0.0395	19.58	0.0413	1.05
30	15	138.08	0.0163	40.38	0.0178	0.92	52.25	0.0241	29.93	0.0249	0.97	38.62	0.0252	29.00	0.0258	1.02
40	20	167.46	0.0104	51.97	0.0122	0.85	67.87	0.0158	39.28	0.0181	0.87	50.95	0.0174	38.17	0.0188	1.08
50	25	191.65	0.0073	62.89	0.0090	0.81	82.72	0.0125	48.50	0.0135	0.93	63.04	0.0134	47.05	0.0127	0.95
100	50	267.38	0.0023	106.87	0.0024	0.96	147.57	0.0052	89.65	0.0047	1.11	120.22	0.0064	87.65	0.0049	0.77

Table 2 Results from the simulation study on the California redwood trees data population with 4 strata under the condition C_0: $y > 0$, C_1: $y > 1$and C_2: $y > 2$ and the estimated relative efficiencies R. E.= $\hat{v}(\hat{\mu}_{SACS})$ / $\hat{v}(\hat{\mu}_{SSC})$

		C_0: $y > 0$					C_1: $y > 1$					C_2: $y > 2$				
		SACS		SACS-SCSU			SACS		SACS-SCSU			SACS		SACS-SCSU		
n	$n_{\square}$	v	$\hat{v}(\hat{\mu}_{SACS})$	v	$\hat{v}(\hat{\mu}_{SSC})$	R.E.	v	$\hat{v}(\hat{\mu}_{SACS})$	v	$\hat{v}(\hat{\mu}_{SSC})$	R.E.	v	$\hat{v}(\hat{\mu}_{SACS})$	v	$\hat{v}(\hat{\mu}_{SSC})$	R.E.
4	1	24.96	0.1455	5.98	0.1576	0.92	7.53	0.1803	4.13	0.1983	0.91	5.31	0.1797	4.02	0.1886	1.05
8	2	47.54	0.0681	11.87	0.0754	0.90	14.88	0.0863	8.25	0.1009	0.86	10.57	0.0964	8.03	0.1027	1.07
12	3	68.03	0.0455	17.56	0.0499	0.91	22.07	0.0617	12.36	0.0675	0.91	15.79	0.0625	12.05	0.0643	1.03
16	4	86.65	0.0318	23.70	0.0370	0.86	29.10	0.0421	16.46	0.0458	0.92	20.95	0.0495	16.06	0.0487	0.98
20	5	103.61	0.0253	29.42	0.0292	0.87	35.96	0.0345	20.58	0.0363	0.95	26.08	0.0357	20.08	0.0349	0.98
40	10	169.25	0.0104	56.72	0.0144	0.72	68.15	0.0158	41.14	0.0175	0.90	51.04	0.0176	40.15	0.0172	0.98
60	15	213.33	0.0025	82.36	0.0083	0.69	97.20	0.0098	61.53	0.0111	0.88	75.01	0.0110	60.20	0.0119	1.08
80	20	244.70	0.0035	106.81	0.0051	0.69	123.63	0.0067	81.99	0.0081	0.83	98.09	0.0075	80.23	0.0078	1.04
100	25	268.23	0.0023	130.41	0.0039	0.59	147.90	0.0049	102.31	0.0060	0.82	120.36	0.0059	100.3	0.0060	1.02

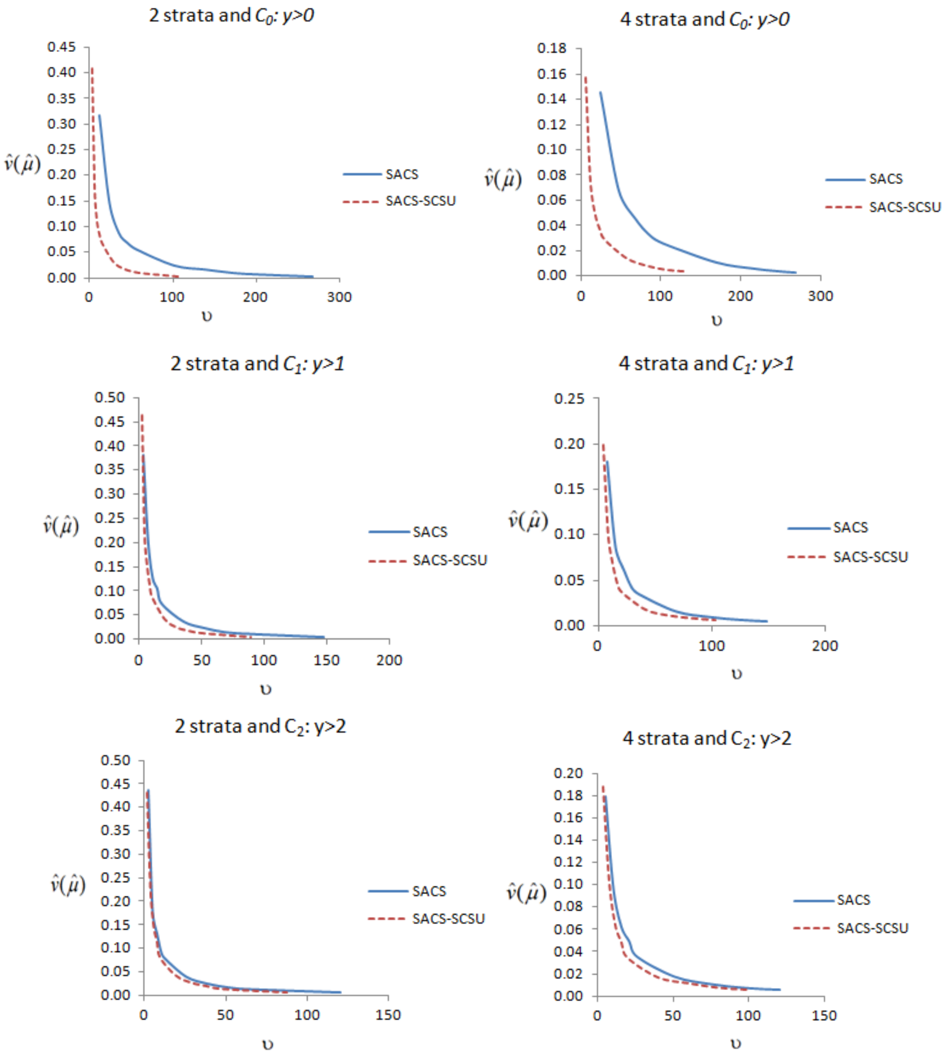

Figure 7 Graphs of average estimated variances from the simulation study on the California redwood trees data. SACS represents stratified ACS and SACS-SCSU represents stratified ACS-SCSU

5. Conclusions

Stratified ACS-SCSU, or ACS-SCSU with stratification, was proposed in this paper when prior information allows stratification which can increase precision in estimation. In stratified ACS-SCSU, an initial stratified sample of primary units is selected by taking the initial sample of primary units from each stratum using simple random sampling without replacement. Then for each sampled primary unit, neighboring secondary units are added to the sample and observed whenever the variable of interest of a secondary unit satisfies a specific condition. This procedure continues until no adaptively sampled secondary units satisfy the condition. The estimator of the mean for this design was obtained by applying the Horvitz-Thompson estimator based on network inclusion probabilities.

The simulation results support that stratified ACS-SCSU can be a time and cost-effective sampling design, and it is appropriate for sampling in a spatially aggregated population.

References

Akanisthanon J, Budsaba K, Borkowski JJ. k-tuple simple latin square sampling designs. Thail. Stat. 2010; 8: 93-107.

Arabkhedri M, Lai FS, Noor-Akma I, Mohamad-Roslan MK. An application of adaptive cluster sampling for estimating total suspended sediment load. Hydrol. Res. 2010; 41: 63-73.

Blackwell D. Conditional expectation and unbiased sequential estimation. Ann. Math Stat. 1947; 18: 105-110.

Boomer K, Werner C, Brantley S. CO_2 emissions related to the Yellowstone volcanic System. J. Geophys. Res. 2000; 105(B5): 10817-10830.

Borkowski JJ. Simple latin square sampling $\pm$k designs. Commun. Stat. Theory. 2003; 32: 215-237.

Chao CT. Ratio estimation on adaptive cluster sampling. J. Chinese Stat. Assoc. 2004; 42: 307-327.

Chao CT, Dryver AL, Chiang T. Leveraging the Rao-Blackwell theorem to improve ratio estimators in adaptive cluster sampling. Environ. Ecol. Stat. 2011; 18: 543-568.

Chutiman N. A new ratio estimator in stratified adaptive cluster sampling. Thail. Stat; 2010; 8: 223-233.

Conners ME, Schwager SJ. The use of adaptive cluster sampling for hydroacoustic surveys. ICES J. Mar. Sci. 2002; 59: 1314-1325.

Dryver AL, Thompson SK. Improving unbiased estimators in adaptive cluster sampling. J. R. Stat. Soc. Ser. B Stat. Methodol. 2005; 67: 157-166.

Dryver AL, Chao CT. Ratio estimators in adaptive cluster sampling. Environmetrics. 2007; 18: 607-620.

Goldberg NA, Heine JN, Brown JA. The application of adaptive cluster sampling for rare subtidal macroalgae. Mar. Biol. 2006; 151:1343-1348.

Horvitz DG, Thompson ME. A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc. 1952; 47: 663-685.

Lohr SL. Sampling: Design and analysis. California: Duxbury Press Publishing Company; 1999.

Lucas HA, Seber GAF. Estimating coverage and particle density using the line intercept method. Biometrika. 1977; 64: 618-622.

Magnussen S, Kurz W, Leckie DG, Paradine D. Adaptive cluster sampling for estimation of deforestation rates. Eur. J. For. Res. 2005; 124: 207-220.

Mohammadi M, Salehi MM. Horvitz-Thompson estimator of population mean under inverse sampling designs. Iranian Math. Soc. 2011; 1-14.

Noon BR, Ishwar NM, Vasudevan K. Efficiency of adaptive cluster and random sampling in detecting terrestrial herpetofauna in a tropical rainforest. Wildlife Soc. B. 2006; 34: 59-68.

Patummasut M, Borkowski JJ. Adaptive cluster sampling with spatially clustered secondary units. J. Appl. Sci. 2014; 14: 2516-2522.

Patummasut M, Dryver AL. A new sampling design for a spatial population: Path sampling. J. Appl. Sci. 2012; 12: 1355-1363.

Philippi T. Adaptive cluster sampling for estimation of abundances within local populations of low abundance plants. Ecology. 2005; 86: 1091-1100.

Rao CR. Information and accuracy attainable in estimation of statistical parameters. Bull. Calcutta Math. Soc.1945; 37: 81-91.

Roesch, F. Adaptive cluster sampling for forest inventories. Forest Sci. 1993; 39: 65-69.

Salehi M. Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling. Environ. Ecol. Stat. 1999; 6: 183-195.

Smith DR, Conroy MJ, Brakhage DH. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. Biometrics. 1995; 51: 777-788.

Smith DR, Villella RF, Lemarié DP. Application of adaptive cluster sampling to low-density populations of freshwater mussels. Environ. Ecol. Stat. 2003; 10: 7-15.

Strauss DJ. A model for clustering. Biometrika. 1975; 63: 467-475.

Sullivan WP, Morrison BJ, Beamish FWH. Adaptive cluster sampling: estimating density of spatially autocorrelated larvae of the sea lamprey with improved precision. J. Great Lakes Res. 2008; 34: 86-97.

Talvitie M, Leino O, Holopainen M. Inventory of sparse forest populations using adaptive cluster sampling. Silva Fenn. 2006; 40: 101-108.

Thompson SK. Adaptive cluster sampling. J. Amer. Statist. Assoc. 1990; 85: 1050-1059.

Thompson SK. Adaptive cluster sampling: Designs with primary and secondary units. Biometrics. 1991a; 47: 1103-1115.

Thompson SK. Stratified adaptive cluster sampling. Biometrika. 1991b; 78: 389-397.

Thompson SK. Sampling, New York: Wiely; 2002.

Thompson SK. Adaptive web sampling. Biometrics. 2006; 62: 1124-1234.

Thompson SK, Seber GAF. Adaptive sampling. New York: Wiley; 1996.

Turk P, Borkowski JJ. A review of adaptive cluster sampling: 1990-2003. Environ. Ecol. Stat. 2005; 12: 55-94.