

DESIGN-BASED AND MODEL-BASED ESTIMATION IN ADAPTIVE CLUSTER SAMPLING

A thesis submitted in partial fulfillment of the
requirements for the degree

Dr. rer. pol.

to the

Faculty IV, Economic and Social Statistics Department
at the University of Trier

submitted by

MSc Esha Mohamed

Supervisor: Prof. Dr. Ralf Münnich

Co-supervisor: PD. Dr. Siegfried Gabler

Trier, April 2017

Table of Contents

Table of Contents	iv
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
1. Introduction	2
2. Review of the literature	5
2.1 Current practices in animals' surveys	5
2.2 Adaptive Cluster Sampling (ACS) design and its variants	6
2.3 Practical issues in implementing ACS	19
2.4 Product method of estimation	21
3. Design-based estimation	23
3.1 Stratified strip ACS: A case study in estimating population total of African elephants	23
3.1.1 Introduction	23
3.1.2 The design	23
3.1.3 The modified Des Raj's estimator	25
3.1.4 Efficiency comparison	26
3.1.5 A small population example	27
3.1.6 Simulation study	31
3.1.7 Simulations using real data	35
3.1.8 Discussion	36
3.2 ACS for negatively correlated data	38
3.2.1 Introduction	38
3.2.2 Estimators	39
3.2.3 Efficiency comparisons	43
3.2.4 A small population example	47

3.2.5	Simulation study	50
3.2.6	Simulations using real data	55
3.2.7	Discussion	56
3.3	ACS with clusters selected without replacement and stopping rule	58
3.3.1	Introduction	58
3.3.2	The modified Raj's estimator	58
3.3.3	A small population example	60
3.3.4	Simulation study	62
3.3.5	Simulations using real data	65
3.3.6	Discussion	69
4.	Model-based estimation	71
4.1	Introduction	71
4.2	Notations	72
4.3	Sample selection probability	72
4.4	The model	74
4.5	Fitting of the model	75
4.6	Simulation set-up and results	82
4.7	Design-based simulation	85
4.8	Discussion	89
5.	Comparison of methods	92
6.	Summary and outlook	98
7.	Zusammenfassung und Ausblick	102
	Bibliography	113
	Appendix	114
A.	Product Appendix	115

A.1 Derivation of Bias and MSE of product Hansen-Hurwitz (HH) estimator	115
A.2 Derivation of $MSE(\hat{\mu}_{phh})$ in terms of network population values	116
A.3 Relative efficiency	117
B. MCMC Diagnostics	123
B.1 Gelman-Rubin convergence diagnostics	123
B.2 Autocorrelation of parameters	123
B.3 Trace Plots	123

List of Figures

Figure 2.1	Distribution of artificial population in a region of interest.	7
Figure 2.2	A figure to display how unit(s) can be observed more than once.	10
Figure 2.3	A figure to display the ACS without replacement of clusters (Dryver and Thompson, 2007) in which units are observed only once.	12
Figure 2.4	Plot of strip ACS with individual strips (Primary Sampling Unit (PSU)) along the x-axis and Secondary Sampling Unit (SSU) along the y-axis. The initial strips are colored light gray, adaptively added units are colored dark gray and edge units colored black.	15
Figure 2.5	Ordinary Stratified ACS	17
Figure 3.1	Plot of stratified strip ACS with initial sampled strip colored light gray, adaptively added units dark gray and edge units black.	24
Figure 3.2	Simulated populations under different parameter setting. .	32
Figure 3.3	Results of the relative efficiency for the different parameters and for the two initial sample sizes considered.	34
Figure 3.4	Results of the expected final sampling fraction for the different parameters and for the two initial sample sizes considered.	34
Figure 3.5	Distribution of the elephant population in which the region is divided into 4 strata. The strata are represented by thick boarder line. Map obtained from Kenana et al. (2013b). . .	35
Figure 3.6	Population distribution of the study (black *) and the auxiliary variable (grey +) when clusters are more (first row) and less (second row) compact and when the level of correlation is low (first column) and high (second column).	51

Figure 3.7	Relative efficiency when Horvitz-Thompson (HT) is used for an initial sample of size $n = 15$ for the different parameters when $\rho_{w_{xy}}$ is highly negative	53
Figure 3.8	Relative efficiency from the product HH for an initial sample of size $n = 15$ for the different parameters when $\rho_{w_{xy}}$ is highly negative	53
Figure 3.9	Relative Bias (RB) of the proposed product HT (left) and HH (right) for low (first row), intermediary (second row) and high (third row) correlation; highly rare (nop = 5) and less rare (nop = 10) and from more compact to less compact clusters (from first to third column).	54
Figure 3.10	Simulated population with first row $s = 0.00025$ and second row $s = 0.0005$. First, second and third columns are for $\lambda_2 = 50, 80$ and 150 respectively.	63
Figure 3.11	Results of the relative efficiency for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.	64
Figure 3.12	Results of the expected final sampling fraction for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.	65
Figure 3.13	Results of the relative bias for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.	66
Figure 3.14	<i>Top panel:</i> Distribution of hartebeest. <i>Bottom panel:</i> Distribution of oryx. Maps obtained from Kenana et al. (2013b).	67
Figure 3.15	Results of the relative efficiency	68

Figure 3.16 Results of the expected final sampling fraction.	69
Figure 4.1 Population distribution of study (gray *) and auxiliary variables (black .) with some values of α and θ that were used. α controls the number of non empty units and θ controls the number of groups of animals.	80
Figure 4.2 Posterior density plot for the predicted population total for one population in one of the strata. The solid line represents the true population value while the dotted lines represent the 95% HPD interval.	84
Figure 4.3 Bias (first row) and Mean Squared Error (MSE) (second row) for different values of α and θ . α controls the number of non-empty units and θ controls the number of groups of animals.	84
Figure 4.4 <i>Top panel</i> : Distribution of the study variable (elephants) obtained from an aerial survey that was conducted in October (dry season) 2013. <i>Bottom panel</i> : Distribution of the auxiliary variable (elephants) obtained from an aerial survey conducted in April (wet season) 2013. The first stratum is on the top left, second stratum is on the top right, third and fourth strata on bottom left and right respectively. Maps obtained from Kenana et al. (2013b) and Kenana et al. (2013a) respectively.	87
Figure 4.5 Box plots of the prediction bias (left) and the prediction MSE (right) for the proposed model and the modified model due to Rapley and Welsh (2008) (RaW)	89
Figure 4.6 Box plots of the bias (first column) and MSE (second column) under the RaW model (first row) and under the proposed model (second row) for each stratum.	90

-
- Figure 5.1 Box plots showing distribution of the expected final sample size under the design-based stratified strip ACS (StrtStrpD), the ACS for negatively correlated data (Product), the Adaptive Cluster Sampling with Clusters selected without replacement and Stopping Rule (ACS^{c-sr}) (StopRule) and the model-based stratified strip ACS (StrtStrpM). 94
- Figure 5.2 Scatter plots for the distribution of the estimated population total against the final sample size under the different proposed strategies. The labels 1 and 2 represent the ACS for negatively correlated data with HT (Product.ht) and HH (Product.hh) estimators respectively. The labels 3, 4 and 5 represent ACS^{c-sr} (StopRule), stratified strip ACS under design-based (StrtStrpD) and model-based (StrtStrpM) respectively. 96
- Figure 5.3 Scatter plots for the distribution of the estimated MSE of the estimated population total against the final sample size under the different proposed strategies. The labels 1 and 2 represent the ACS for negatively correlated data with HT (Product.ht) and HH (Product.hh) estimators respectively. The labels 3, 4 and 5 represent ACS^{c-sr} (StopRule), stratified strip ACS under design-based (StrtStrpD) and model-based (StrtStrpM) respectively. . . . 97
- Figure A.1 Relative efficiency when HT is used for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative 118
- Figure A.2 Relative efficiency when product HH is used for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative 119

Figure A.3 Results of the relative efficiency between the product HT, the product HH and the product Simple Random Sampling Without Replacement (SRSWOR) for an initial sample of size $n = 15$ and a high level of correlation coefficient.	120
Figure A.4 Relative bias for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative	121
Figure A.5 Relative bias when the initial sample of size $n = 15, 65$ and 100 (top, middle and bottom plot respectively) for the different parameters and different levels of correlation coefficient.	122
Figure B.1 Gelman-Rubin scale reduction factor plots.	124
Figure B.2 Trace plots for the artificial data.	125
Figure B.3 Trace plots for combined chains for the elephant data. . .	125
Figure B.4 Plots of the autocorrelation function.	126
Figure B.5 Autocorrelation function for the elephant data.	126

List of Tables

Table 3.1	An artificial population with the number of animals in each unit y_i represented on the left hand side while the transformed population w_i is presented on the right hand side. Second row with numbers 1, 2, 3, 4 are labels for each strip. Here $N_q^* = 4$, $M_q^* = 5$, $N_q = 20$, $n_q = 2$ and $C = \{y_i : y_i \geq 50\}$	28
Table 3.2	All possible ordered samples where samples are drawn without replacement of clusters. First column lists strips selected in the initial sample while strips whose units were adaptively added are included after semicolon. $\hat{v}(\hat{\tau})$ is the variance under non-adaptive.	30
Table 3.3	Allocations of samples to different strata according to optimal (Neyman, 1934) and optimal according to box-constraint (Gabler et al., 2012) allocations for some of the parameters considered and different initial sample sizes. q_1, \dots, q_4 represent the first, ..., fourth stratum respectively. . .	33
Table 3.4	Allocation of samples to different strata for elephant population under the proposed design	36
Table 3.5	Relative efficiency (RE) and expected final sampling fraction ($E(fn)$) for the elephant data. RE and $E(fn)$ are defined in equation (3.10).	36
Table 3.6	y and x populations and quantities to compute the ACS estimators	47
Table 3.7	All possible samples with their estimates. Units before the semicolon are the initial units while units after the semicolon are the adaptively added units.	49
Table 3.8	Values on the left (correlation coefficient) and the right (cv.r) hand side of equations (2.31) and (3.25)	55
Table 3.9	Results of the relative efficiency (RE) and relative bias (RB) for different initial sample sizes on the real data	56

Table 3.10	A small artificial population	60
Table 3.11	Results of the relative bias under ACS^{c-sr} design	69
Table 4.1	True versus estimated parameters for the artificial populations. α controls the number of non-empty units, θ controls the number of groups of animals while β_0 and β_1 is the intercept and the slope used to generate counts.	85
Table 4.2	Initial values set for parameters in each model. RaW represents the model due to Rapley and Welsh (2008) under stratified strip setting. α controls the number of non-empty units, θ controls the number of groups of animals while β_0 and β_1 is the intercept and the slope used to generate counts	86
Table 4.3	Correlation at network level for different strata. Difference is the total difference of the within network variance and the stratum variance of the study variable.	88
Table 4.4	Results of the comparison between model-based and design-based estimates using the bias and the MSE.	88
Table 5.1	Standard deviation and sample size allocated in each stratum	92
Table 5.2	ρ is the correlation coefficient at the unit (first row) and network (second row) level. cv.r is the right hand side of equation (3.25) (first row) and (2.31) (second row)	93
Table 5.3	Results of the coefficient of variation under stratified strip ACS (StrtStrpD), ACS for negatively correlated data with HT (Product.ht) and with HH (Product.hh) estimator, ACS^{c-sr} (StopRule) and under model-based estimation (StrtStrpM) . . .	93
Table 5.4	Results of the relative bias under stratified strip ACS (StrtStrpD), ACS for negatively correlated data with HT (Product.ht) and with HH (Product.hh) estimator, ACS^{c-sr} (StopRule) and under model-based estimation (StrtStrpM) . . .	94

List of Abbreviations

ACF	Autocorrelation Function.
ACS	Adaptive Cluster Sampling.
ACS ^c	Adaptive Cluster Sampling with Clusters selected without replacement.
ACS ⁿ	Adaptive Cluster Sampling with Networks selected without replacement.
ACS ^{c-sr}	Adaptive Cluster Sampling with Clusters selected without replacement and Stopping Rule.
CITES	Convention on International Trade in Endangered Species.
CV	Coefficient of Variation.
HH	Hansen-Hurwitz.
HPD	Highest Posterior Density.
HT	Horvitz-Thompson.
MCMC	Markov Chain Monte-Carlo.
MSE	Mean Squared Error.
PCP	Poisson Cluster Process.
PSU	Primary Sampling Unit.
RB	Relative Bias.
RRMSE	Relative Root Mean Squared Error.
SAE	small area estimation.
SRSWOR	Simple Random Sampling Without Replacement.
SSU	Secondary Sampling Unit.

Acknowledgement

I am incredibly grateful to my first supervisor Prof. Dr. Ralf Münnich for his support, guidance and encouragement throughout my PhD period. I am indeed grateful to my second supervisor PD. Dr. Siegfried Gabler for his support and the many discussions we had that led to the improvement of my work.

This research was funded by National Commission for Science, Technology and Innovation (NACOSTI, Kenya), Deutscher Akademischer Austauschdienst (DAAD) and the Rhineland-Palatinate research initiative via the Trier Centre of Sustainable Systems (TriCSS).

To Prof. Stefano Gattone, Prof. Arthur Dryver and Prof. Tonio Di Battista, I thank you very much for the collaborative work and the many discussions we had during the time. I thank Taita Taveta University (Kenya) for granting me study leave to pursue my PhD.

I am grateful to my colleagues in the department who supported me in settling down in Trier, going about the daily life and who ensured a nice working environment. I am especially indebted to Florian Ertz, Lisa Borsi and Dr. Thomas Zimmermann. My appreciation to the other international PhD students whom we spent most of our free time together. I am indeed thankful to my family back at home and the family that I came to have in Trier especially Maria and Willheim Franz.

1. Introduction

Adaptive cluster sampling (ACS) design was proposed by (Thompson, 2012, p. 319) as a suitable sampling design for populations that are rare, clustered, hidden or hard to reach. The design has the advantage of increasing the number of observations observed during a survey. It also results in more precise estimates of the population abundance or density for a given sample size or cost (Thompson, 2012, p. 319). The design is thus in a better place to inform management decisions on the conservation of rare, clustered, hidden or hard to reach populations.

A number of variations of the ACS design exists in literature. Some are briefly described in section (2.2), *see* Turk and Borkowski (2005) and Seber and Salehi (2013) for further review. Some of the recent work in the area include neighbourhood-free plots in the area of forestry (Yang et al., 2016), modeling data (collected by ACS) design at the unit level (Gonçalves and Moura, 2016), combining the ACS and the inverse sampling (Salehi et al., 2015), the use of two-stage design with complete allocation method (Moradi et al., 2014) and the comparison of the performance of different adaptive designs (Brown et al., 2013). Conditions under which the efficiency of the ACS design is increased relative to the non-adaptive design can be found in Gabler (2007), Turk and Borkowski (2005), Brown (2003), Thompson and Collins (2002) and Thompson and Seber (1996) among others.

The ACS design has broad applications in humans, animals and plants. In the context of animals Salehi et al. (2015) studies a variant of the ACS design on crabs, Smith et al. (1995) evaluated the performance of the ACS design on birds (Wintering waterfowl) using simulations based on real Waterfowl data, Green et al. (2010) performed a strip ACS to detect waterbirds colonies, Kissling et al. (2007) used systematic sampling design with ACS design to determine the distribution and the abundance of rare birds (Kittlitz's Murrelets) while Connors and Schwager (2002) used the ACS design to determine the stock size of fish (Rainbow smelt) using both simulations and field trial.

In human, Thompson and Collins (2002) introduced the ACS design targeting individuals with risk-related characters such as drug abusers and people at high risk of contracting HIV/AIDS. In plants Roesch (1993) used the ACS design to examine the pollution damage on trees. Yang et al. (2016) introduced a variant of ACS on forestry, Talvitie et al. (2006) applied the ACS design to assess damage on forest resulting from 2002 - 2003 drought in the city of Helsinki in Finland. Acharya et al. (2000) assessed rare tree species in Nepal using systematic ACS while Philippi (2005) used the ACS design to estimate abundance of some herbaceous species. Other areas of applications of the ACS design include soil pollution (Salehi et al., 2015)

This dissertation was motivated by a number of factors. These include the need to have more efficient estimates for populations that are rare and clustered. This entails the use of auxiliary information that is negatively correlated with the study variable and the use of the model-based approach to estimation. The other factor that motivated this dissertation is the cost of sampling which is measured by the number of units in the final sample. This consists of a combination of designs such as combining a stopping rule criterion and observe clusters once. *See* below and chapter 6 for details.

Section 2 provides a brief review on the practices that are currently used in surveying animals. The chapter also briefly reviews the ACS design and some of its extensions that have been used in this dissertation. A number of issues arise in the field implementation of the ACS design, this is also briefly covered in the section.

The aim was thus to examine the ACS design under design-based and model-based approaches. Within the design-based approach (chapter 3), three variations of the ACS design are developed. The first one which is given in section (3.1) is an extension of strip ACS proposed by (Thompson, 2012, p. 339) to a stratified setting in which networks are truncated at stratum boundaries. The proposed design is suitable for aerial or ship surveys. Improvement of the efficiency of the ACS when there exist an auxiliary

variable that is negatively correlated with the study variable is provided in section (3.2). Finally, cost reduction while improving on the efficiency of the ACS design is provided in section (3.3).

Model-based approaches provide more precise estimates than the design-based approaches, when the assumed model holds. Hence, chapter 4 examines model-based ACS under finite population setting. The chapter modeled data sampled through stratified strip ACS and predicted the population total. This chapter is an improvement of the work by Rapley and Welsh (2008) by incorporating an auxiliary information and by making use of the stratified strip ACS design. The stratified strip ACS design is described in section (3.1).

The above methods are compared using a simulation study in chapter 5. Finally summary and outlook are provided in chapter 6 with a German version in chapter 7.

2. Review of the literature

2.1 Current practices in animals' surveys

Strip Sampling is an aerial survey technique in which the region under investigation is divided into blocks of specified length and width. Then an aircraft flies at straight lines north-south or east-west direction (Norton-Griffiths, 1978, p. 22). The aircraft flies at constant height (approximately 150 m) and speed (157 km/h) (Green et al., 2010) with observers taking records. The method is common in surveying savanna wildlife such as African elephant. It has also been applied by Green et al. (2010) in birds that occupy open habitats.

Line transect is a ground technique in which the observer travels along lines or transects at a constant speed and count animals on either sides of the transect within a specified width. The method is common in African elephants and birds. In elephants, it could either involve counting elephants direct (Hedges, 2012, ch. 3) or their dung (Hedges, 2012, ch. 4). In birds, the transects are either randomly or systematically distributed (Bibby et al., 1992, p. 66).

Point transect on the other hand involves the observer walking along transects, making stops at specified points, and remain at the points for a specified amount of time (between 2 - 20 minutes) while counting species (Bibby et al., 1992, p. 95). The observation points are either randomly or systematically distributed in the region under investigation.

Sometimes the above designs are combined with stratification. In surveying rare birds, strategies that are used to improve the number of the targeted species to be observed include; using point transect technique and increase the waiting time at the observation points (Bibby et al., 1992, p. 102; Haselmayer and Quinn, 2000), repeat the counting process (Rosenstock et al., 2002) and undertaking an ACS design (Green et al., 2010; Kissling et al., 2007; Thompson, 2012, p. 339). This dissertation works at improving the number of species observed by using the ACS design, improve gains in efficiency and lowering sampling costs.

2.2 ACS design and its variants

The general idea behind the ACS design is first presented in this section while some of its variations are presented in the sub-section. Under the ACS design, the region of interest is partitioned into N non-overlapping units of equal sizes. Each unit in the population is associated with an observed variable of interest y_i , $i = 1, \dots, N$. An initial sample of size n units is selected by any of the probability sampling procedure. A condition or critical value is set before sampling begins. In this dissertation, the pre-specified condition is the presence of at least one animal in a selected unit *i.e.* $C = \{y_i : y_i > 0\}$. If any of the units selected in the initial sample meet the pre-specified condition, units in the neighbourhood of the selected unit are added to the sample and observed. If any of the added units meets the pre-specified condition, their neighboring units are added to the sample and observed. The procedure continues until there are no more units in the neighborhood that meet the pre-specified condition. Neighborhood is defined by spatial proximity.

An initial unit together with its adaptively added units that meet the pre-specified condition is called a network, units on the boundary of a network are known as edge units (Thompson, 2012, p. 319). A network together with its edge units forms a cluster. According to (Thompson, 2012, p. 319), an empty unit is a network of size one. Figure 2.1 shows the idea behind the ACS design on a region that is partitioned into $N = 400$ units. The units in the initial sample are colored light gray, adaptively added units are dark gray and edge units are colored black.

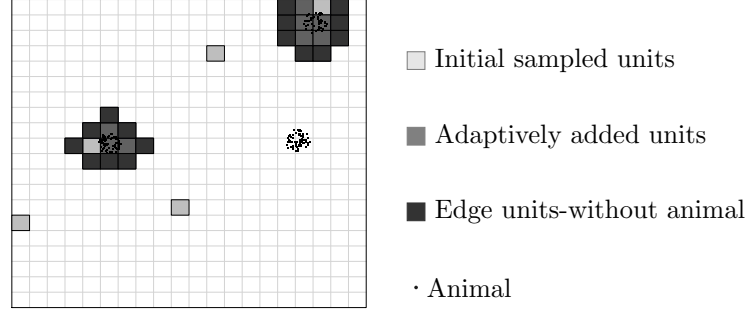


Fig. 2.1: Distribution of artificial population in a region of interest.

ACS without replacement of units

Let the population of units with observed variables be partitioned into K distinct networks and let the initial sample of size n be selected by SRSWOR. (Thompson, 2012, p. 319) modified the Horvitz-Thompson (HT) estimator to obtain an unbiased estimator of the population mean $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{k=1}^K y_k$ that he gave as

$$\hat{\mu}_{ht} = \frac{1}{N} \sum_{k=1}^K \frac{y_k J_k}{\alpha_k} \quad (2.1)$$

where y_k is the sum of y values in network k , J_k is an indicator variable that takes the value one if the selected network is in the sample and zero otherwise and α_k is the probability that the initial sample intersect network k defined by

$$\alpha_k = 1 - \frac{\binom{N - m_k}{n}}{\binom{N}{n}}. \quad (2.2)$$

The variance of $\hat{\mu}_{ht}$ is given by (Thompson, 2012, p. 319)

$$v(\hat{\mu}_{ht}) = \frac{1}{N^2} \sum_{j=1}^K \sum_{k=1}^K y_j y_k \frac{(\alpha_{jk} - \alpha_j \alpha_k)}{(\alpha_j \alpha_k)} \quad (2.3)$$

where α_{jk} is the probability that the initial sample intersects both networks j and k which (for $j \neq k$) is defined by

$$\alpha_{jk} = 1 - \left[\frac{\binom{N-m_j}{n}}{\binom{N}{n}} + \frac{\binom{N-m_k}{n}}{\binom{N}{n}} - \frac{\binom{N-m_j-m_k}{n}}{\binom{N}{n}} \right]. \quad (2.4)$$

in which m_k are the number of units in network k and $\alpha_{kk} = \alpha_k$. The unbiased estimator of $v(\hat{\mu}_{ht})$ is

$$\hat{v}(\hat{\mu}_{ht}) = \frac{1}{N^2} \sum_{j=1}^K \sum_{k=1}^K \frac{y_j y_k J_j J_k (\alpha_{jk} - \alpha_j \alpha_k)}{\alpha_{jk} \alpha_j \alpha_k}. \quad (2.5)$$

The population total is estimated by

$$\hat{\tau}_{ht} = N \times \hat{\mu}_{ht} \quad (2.6)$$

whose variance is

$$v(\hat{\tau}_{ht}) = N^2 \times v(\hat{\mu}_{ht}) \quad (2.7)$$

and can unbiasedly be estimated by

$$\hat{v}(\hat{\tau}_{ht}) = N^2 \times \hat{v}(\hat{\mu}_{ht}) \quad (2.8)$$

(Thompson, 2012, p. 319) also modified the Hansen-Hurwitz (HH) estimator to obtain an unbiased estimator of the population mean and total when samples are obtained without replacement which he gave as

$$\hat{\mu}_{hh} = \frac{1}{n} \sum_{k=1}^n w_{yk}, \quad \hat{\tau}_{hh} = N \times \hat{\mu}_{hh} \quad (2.9)$$

respectively where w_{yk} is the average of y values in the k -th network. The variance of $\hat{\mu}_{hh}$ is given by (Thompson, 2012, p. 319)

$$v(\hat{\mu}_{hh}) = \frac{N-n}{Nn} \sum_{k=1}^N \frac{(w_{yk} - \mu_y)^2}{N-1} \quad (2.10)$$

where μ_y is the population mean of the variable of interest. The variance of $\hat{\tau}_{hh}$ is given by

$$v(\hat{\tau}_{hh}) = N^2 \times v(\hat{\mu}_{hh}) \quad (2.11)$$

The unbiased estimator of the variance of the mean and total are (Thompson, 2012, p. 319)

$$\hat{v}(\hat{\mu}_{hh}) = \frac{N-n}{Nn} \sum_{k=1}^n \frac{(w_{yk} - \hat{\mu}_{hh})^2}{n-1}. \quad (2.12)$$

and

$$\hat{v}(\hat{\tau}_{hh}) = N^2 \times \hat{v}(\hat{\mu}_{hh}) \quad (2.13)$$

respectively.

ACS without replacement of cluster

Salehi and Seber (1997a) proposed an Adaptive Cluster Sampling with Networks selected without replacement (ACSⁿ) design in which networks are selected without replacement. In their proposed design, an initial unit i is selected at random. If the selected unit meets the pre-specified condition, neighbouring units are added to the sample and observed. If any of the added units meet the pre-specified condition, its neighbouring units are further added to the sample and observed. The process continues until a boundary of edge units (units that do not meet the pre-specified condition but were added in the sample by virtue of being in the neighbourhood of units that meet the pre-specified condition) is formed. The observed network is "removed" from the population. A second unit is selected from the remaining units and an adaptive search performed. The resulting network is "removed" from the population and a third unit selected from the remaining units. The procedure continues until n networks are observed.

This procedure ensures that networks are observed only once as opposed to the ACS without replacement of units (Thompson, 2012, p. 319) where network or networks can be observed more than once. The selection mechanism proposed by Salehi and Seber (1997a) has the advantage of savings on sampling costs as the observer does not have to travel and observe a network(s) that has already been observed. Even with this procedure, repeat observation of unit(s) can still occur if (Dryver and Thompson, 2007):

1. An edge unit of more than one network is observed more than once.
2. A unit selected in the initial sample is an edge unit of an adjacent selected network.
3. A unit that is already observed as an edge unit is selected in the subsequent steps as an initial unit.

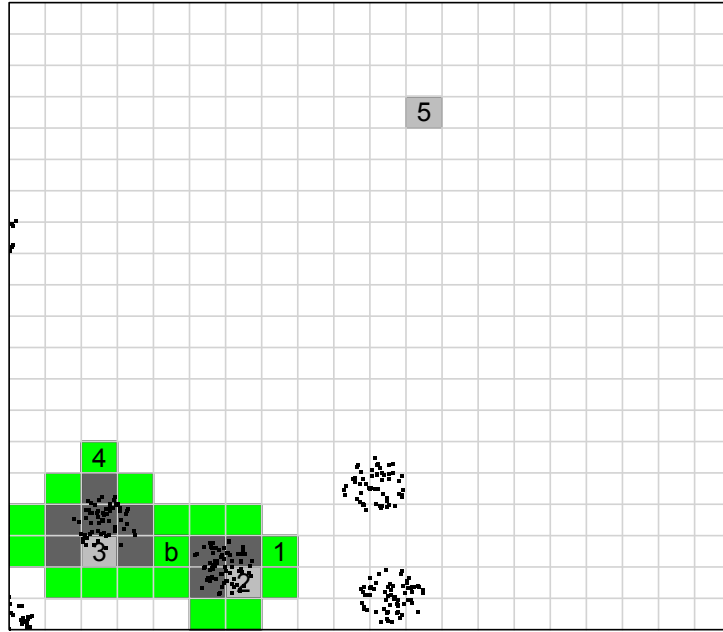


Fig. 2.2: A figure to display how unit(s) can be observed more than once.

Figure 2.2 shows when repeat observations can occur. To explain this,

suppose that an initial sample of size $n = 5$ was selected by SRSWOR. Let the selected units be the units with labels $1, \dots, 5$. The first unit selected does not meet the pre-specified condition and so sampling stops. The second and the third units selected meet the pre-specified condition, hence neighboring units (dark gray) are continuously added to the sample and observed until a boundary of units that do not meet the pre-specified condition (edge units colored green) are observed. Next, the fourth and fifth units are selected but like the first unit, they do not meet the pre-specified condition and so sampling stops.

The green colored unit labeled "b" is an edge unit for both the second and the third network hence it was observed twice (first point above). The green colored unit labeled "1" was observed twice; first as a unit in the initial sample and second as an edge unit of the second network (second point above). Finally, the unit labeled "4" which is colored green was also observed twice; as an edge unit to the third network and later selected as a unit in the initial sample (third point above). In this particular case, the cluster sizes in order of selection are $\{1, 12, 17, 1, 1\}$. Note that the three green colored units with labels "b", "1" and "4" that were observed twice did not add any new information during the second observation. They rather added cost as the observer will have to visit them again.

Due to the above reasons, Dryver and Thompson (2007) proposed an Adaptive Cluster Sampling with Clusters selected without replacement (ACS^c) design in which the entire cluster resulting from the selection of an initial unit is "removed" from the population before selecting the next unit. Figure 2.3 shows the same population as in figure 2.2 with the 1st, 2nd, 3rd and 5th initial units in the same positions as in figure 2.2. Under the ACS^c design, the first unit is selected and "removed" from the population as a network/cluster of size 1. The second unit is selected which leads to a cluster of size 11 is then also "removed" from the population. This cluster does not include the unit labeled "1" since it had already been "removed" from the population. The

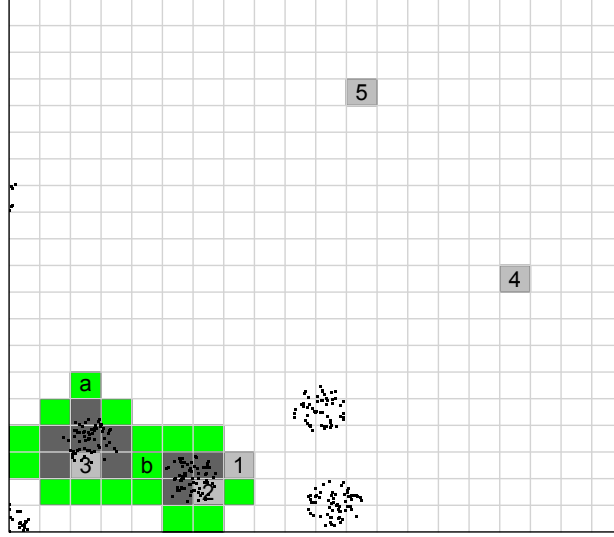


Fig. 2.3: A figure to display the ACS without replacement of clusters (Dryver and Thompson, 2007) in which units are observed only once.

third unit is selected and is removed from the population together with its resulting cluster of size 16. This cluster does not include the edge unit labeled "b" since it was removed following selection of the second unit. The previous position of the fourth units (now labeled "a") does not exist anymore and so the fourth unit will be selected from the remaining units as shown. Finally, the fifth unit maintains its position.

Comparing figure 2.2 and 2.3, units are clearly observed just once under the ACS^c design. Further, the ACS^c design results in lower variance compared to the ACS without replacement of units (Dryver and Thompson, 2007). The final sample size is however not always lower; it depends on the succeeding units selected.

Raj (1956) proposed an estimator that depends on the order of selection

which is given as

$$\begin{aligned}\hat{\tau}_{Raj} &= \frac{1}{n} \sum_{i=1}^n z_i^* \\ z_1^* &= \frac{y_1}{p_1}, \quad \text{for } i \geq 2 \\ z_i^* &= \sum_{j=1}^{i-1} y_j + \frac{y_i}{p_i} \times \left(1 - \sum_{j=1}^{i-1} p_j \right).\end{aligned}$$

The variance of $\hat{\tau}_{Raj}$ is given by (Raj, 1956)

$$v(\hat{\tau}_{Raj}) = \frac{1}{n^2} \sum_{i=1}^n v(z_i^*) \quad (2.14)$$

whose unbiased estimate is

$$\hat{v}(\hat{\tau}_{Raj}) = \frac{1}{n(n-1)} \sum_{i=1}^n (z_i^* - \hat{\tau}_{Raj})^2 \quad (2.15)$$

Dryver and Thompson (2007) modified the above Des Raj's estimator to obtain an unbiased estimator of the population total as

$$\hat{\tau}_{Raj.clust} = \sum_{i=1}^n c_i z_i \quad (2.16)$$

where $c_i = \frac{1}{n}$ are weights such that $\sum_{i=1}^n c_i = 1$ and $z_1 = \frac{y_{k1}}{p_1}$. For $i = 2, \dots, n$

$$z_i = \sum_{j \in s_{i-1}} y_{kj} + \frac{y_{ki}}{p_i}$$

with $p_i = \frac{m_{ki}}{N - \sum_{j \in s_{i-1}} m_{kj}}$ in which m_{ki} are the number of units in network k at step i and s_{i-1} is the set of units (including adaptively added and edge units) observed up to step $i - 1$. Its unbiased estimator of the variance is (Dryver and Thompson, 2007)

$$\hat{v}(\hat{\tau}_{Raj.clust}) = \frac{1}{n(n-1)} \sum_{i=1}^n (z_i - \hat{\tau}_{Raj.clust})^2 \quad (2.17)$$

ACS with a data-driven stopping rule

One of the challenges of the ACS design is the randomness of the final sample size. A number of variations have been proposed to overcome this challenge. These include stratification (Thompson, 2012, p. 353), order statistics (Thompson, 1996), restricted ACS (Brown and Manly, 1998; Salehi and Seber, 2002), inverse ACS (Christman and Lan, 2001) and stopping rule (Su and Quinn II, 2003; Gattone and Di Battista, 2011).

The ACS with a data-driven stopping rule, that was proposed by Gattone and Di Battista (2011) viewed ACS as a set of steps. The step $r = 0$ consists of units selected in the initial sample. If any of these units satisfies the pre-specified condition C , neighbouring units are added to the sample and observed. This forms the first step of sampling $r = 1$. If any of the added units meet the pre-specified condition C , their neighbouring units are added to the sample and observed. This is the second step of sampling $r = 2$ and so on.

From the second step of sampling, neighbouring units will be added to the sample if and only if (Gattone and Di Battista, 2011)

$$C_i^r = \left\{ \frac{s_i^{2(r)} \left(1 - \frac{1}{m_i^{r-1}}\right)}{s_i^{2(r-1)} \left(1 - \frac{1}{m_i^r}\right)} > 1 \right\} \quad (2.18)$$

is satisfied. $s_i^{2(r)}$ is an estimate of the variance of the within networks (within network variance) at the r -th step of the aggregative procedure and m_i^r is the number of units adaptively sampled after r steps from the i -th unit of the initial sample. If C_i^r is not verified then sampling is stopped, and the i -th network is truncated at the $(r-1)$ -th step. Units aggregated at the r -th step are units that activate the stopping rule and they are considered as stopping edge units (Gattone and Di Battista, 2011). The design however introduces bias to the modified HH and HT estimators.

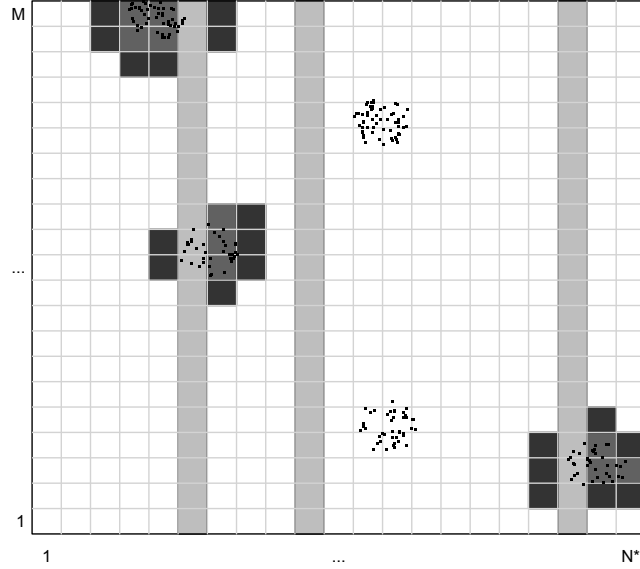


Fig. 2.4: Plot of strip ACS with individual strips (PSU) along the x-axis and SSU along the y-axis. The initial strips are colored light gray, adaptively added units are colored dark gray and edge units colored black.

Strip adaptive cluster sampling

Strip ACS was proposed by (Thompson, 2012, p. 339) when conducting aerial surveys of say walruses or polar bears or in ship surveys of whales. If an animal is observed on a selected strip, neighboring units are observed and added to the sample. The process continues until a boundary of edge units is reached.

Let N^* be the number of strips or PSU such that each strip has M SSU. Let n be the initial sample size (number of strips) selected and let $N = N^*M$ be the total number of SSU in the population. Figure 2.4 shows a study region with $N^* = 21$ strips each with $M = 21$ SSU giving a total of $N = 441$ SSU. An initial sample of size $n = 3$ is selected by SRSWOR. The selected units are colored light gray, adaptively added unit dark gray while edge units are black. The modified HT estimator is (Thompson, 2012, p. 339),

$$\hat{\tau}_{\text{strip}} = \sum_{k=1}^K \frac{y_k J_k}{\pi_k^*} \quad (2.19)$$

where J_k is an indicator variable that takes the value one if a selected network is in the sample and zero otherwise, K is the number of distinct networks in the population and π_k^* is the probability that the initial sample intersect network k defined by

$$\pi_k^* = 1 - \frac{\binom{N^* - b_k}{n}}{\binom{N^*}{n}} \quad (2.20)$$

in which b_k are the number of PSU that intersect network k . The unbiased estimator of the variance of $\hat{\tau}_{\text{strip}}$ is given by (Thompson, 2012, p. 339)

$$\hat{v}(\hat{\tau}_{\text{strip}}) = \sum_{j=1}^K \sum_{k=1}^K \frac{y_j y_k J_j J_k (\pi_{jk}^* - \pi_j^* \pi_k^*)}{\pi_{jk}^* \pi_j^* \pi_k^*} \quad (2.21)$$

where π_{jk}^* is the probability that the initial sample intersects both networks j and k defined by

$$\pi_{jk}^* = 1 - \left[\frac{\binom{N^* - b_j}{n}}{\binom{N^*}{n}} + \frac{\binom{N^* - b_k}{n}}{\binom{N^*}{n}} - \frac{\binom{N^* - b_j - b_k + b_{jk}}{n}}{\binom{N^*}{n}} \right] \quad (2.22)$$

in which b_{jk} are the number of PSU that intersect both networks j and k .

Stratified adaptive cluster sampling

(Thompson, 2012, p. 353) proposed an ACS design in which the initial sample is selected by stratified random sampling. Consider a region that is partitioned into H strata each with N_q units such that the entire population has a total of $N = \sum_{q=1}^H N_q$ units. For each unit i in stratum q , there is an associated observed value y_{qi} such that y_q is the sum of the y -values in stratum q and the

total number of observations in the population is $\tau = \sum_{q=1}^H y_q$. If n_q is the sample size in stratum q , then the total sample size is $n = \sum_{q=1}^H n_q$.

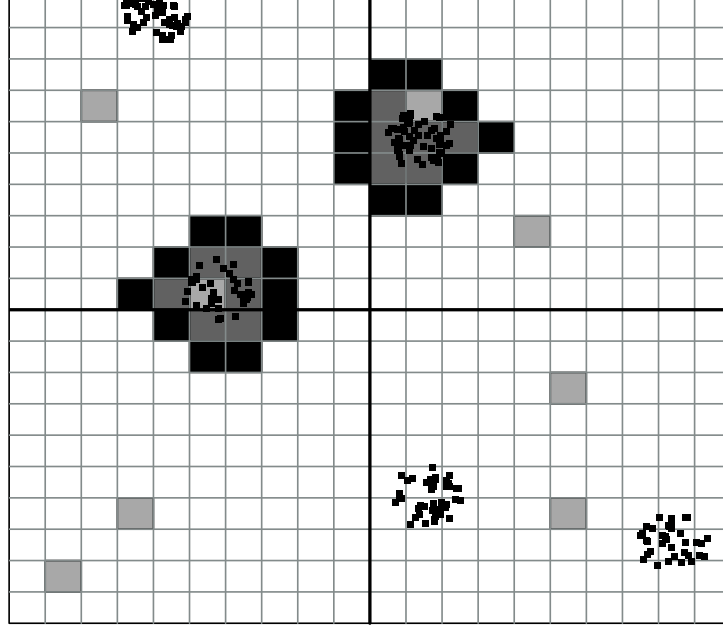


Fig. 2.5: Ordinary Stratified ACS

Figure 2.5 shows a region with $H = 4$ strata each with $N_q = 100$ units in which an initial sample of size $n_q = 2$ selected from each strata by SRSWOR. Units selected in the initial sample are filled with light gray color. Adaptively added units are filled in dark gray color while edge units are colored black.

To estimate the parameter of interest, (Thompson, 2012, p. 353) ignored stratum boundaries and partitioned the entire population into K distinct networks. Let y_k be the sum of the y -values in the k -th network and a_{qk} be the number of units in stratum q that intersect network k . (Thompson, 2012, p. 353) modified the HT estimator to obtain an unbiased estimator of the population total which is given by

$$\hat{\tau}_{\text{strat}} = \sum_{k=1}^K \frac{y_k J_k}{\pi_k} \quad (2.23)$$

whose unbiased estimator of the variance is

$$\hat{v}(\hat{\tau}_{\text{strat}}) = \sum_{j=1}^K \sum_{k=1}^K \frac{y_j y_k J_j J_k (\pi_{jk} - \pi_j \pi_k)}{\pi_{jk} \pi_j \pi_k} \quad (2.24)$$

where π_k is the probability that the initial sample intersect network k ,

$$\pi_k = 1 - \prod_{q=1}^H \frac{\binom{N_q - a_{kq}}{n_q}}{\binom{N_q}{n_q}} \quad (2.25)$$

and π_{jk} is the probability that the initial sample intersect both networks j and k ,

$$\pi_{jk} = 1 - \left[\prod_{q=1}^H \frac{\binom{N_q - a_{jq}}{n_q}}{\binom{N_q}{n_q}} + \prod_{q=1}^H \frac{\binom{N_q - a_{kq}}{n_q}}{\binom{N_q}{n_q}} - \prod_{q=1}^H \frac{\binom{N_q - a_{jq} - a_{kq}}{n_q}}{\binom{N_q}{n_q}} \right] \quad (2.26)$$

in which a_{kq} and a_{jq} are the number of units in strata q intersecting network j and k respectively. J_k is an indicator variable that takes the value one if the k -th network is in the sample and 0 otherwise.

Allocation of samples strata

Under the classical stratified sampling design, the strategies used to allocate sample sizes to different strata include equal allocation, optimal allocation (Neyman, 1934) and optimal allocation with box-constraint (Gabler et al., 2012; Münnich et al., 2012). If n_q is the sample size in stratum q , then

n_q with equal allocation is obtained as

$$n_q = \frac{n}{L}.$$

Under optimal allocation (Neyman, 1934), if the aim is to minimize the variance for a fixed total sample size then n_q is obtained as

$$n_q = \frac{nW_qS_q}{\sum_{q=1}^H W_qS_q} \quad (2.27)$$

where $W_q = \frac{N_q}{N}$ and S_q is the stratum standard deviation. In practice, S_q can be estimated from a pilot survey or previous studies. In this research however, S_q is calculated from the available population data.

Gabler et al. (2012) and Münnich et al. (2012) proposed optimal allocation with box-constraint method to optimize

$$\min_{n_q} \|\text{RRMSE}_{<.>}(\hat{\tau})\|_2 = \sqrt{\sum_{q=1}^H \text{RRMSE}(\hat{\tau}_{<.>})} \quad (2.28)$$

such that $L_q \leq n_q \leq U_q, q = 1, \dots, H$

where Relative Root Mean Squared Error (RRMSE) is defined as

$$\text{RRMSE}(\hat{\tau}) = \frac{\sqrt{\frac{1}{R} \sum_{i=1}^R (\hat{\tau}_i - \tau)^2}}{\tau}.$$

L_q and U_q are the lower and upper limits of the stratum sample size. Gabler et al. (2012) and Münnich et al. (2012) ensured that each strata had a sample size that falls between pre-defined lower (L_q) and upper (U_q) limits while minimizing the variance for a fixed total sample size.

2.3 Practical issues in implementing ACS

To the best of our knowledge, published work on real applications of the ACS design is still relatively low. This is partly because the design is still relatively

new and also because of some challenges faced when undertaking the design in the field. Smith et al. (2004, p. 78) discussed some of the challenges and ways to overcome them. Below is an updated brief discussion following their work.

Efficiency gain from using the ACS design requires some prior knowledge that the population under study is rare and clustered. In fact the HH estimator with the ACS design (equation (2.9)) is more efficient than the sample mean under SRSWOR when the within network variance is greater than the population variance (Thompson, 2012, p. 319). Such knowledge can, however, be obtained from previous studies on the targeted species or by conducting a pilot survey.

The randomness of the final sample size makes it difficult to determine the cost and the time required for a survey prior to the actual survey. There are strategies proposed in literature that can be used to reduce the final sample size. Some of these strategies include two-stage ACS (Salehi and Seber, 1997b), restricted ACS in which networks are sampled until pre-defined final sample size is reached (Brown and Manly, 1998), two-stage restricted ACS (Rocco, 2008), use of a stopping rule (Gattone and Di Battista, 2011) and making use of order statistics and a stopping rule (Su and Quinn II, 2003). The use of stratification, truncate networks at stratum boundaries and use different different critical values in each strata have also been proposed as ways of limiting the final sample size (Brown and Manly, 1998; Connors and Schwager, 2002). Restricted ACS is suitable when the cost of sampling is very high or if there are logistical challenges such as the distance to be covered by an aircraft before running out of fuel (Brown and Manly, 1998).

How to perform an adaptive search in the field may not be easy especially during aerial or ship surveys. The question is the definition of the neighborhood and how to fly the aircraft or ride the ship while observing neighboring units. Green et al. (2010) defined neighborhood as the four adjacent units. The aircraft is then flown in a squared pattern to survey neighboring units until no more units meet the condition of interest. Taking a squared route instead of a circle has the advantage of maintaining height of the aircraft and enable observers

make observations on both sides of the aircraft as it turns (Green et al., 2010).

Conners and Schwager (2002) defined neighborhood as segments of the transect that are parallel to the transect being surveyed. This makes field survey easier as it is easier to move along straight transects. It however has the disadvantage that more edge units are included in the sample hence increasing the cost of survey. The other important thing is to keep track of the units already surveyed while observing neighboring units (Smith et al., 2004, p. 91).

Consider sampling of birds using point transect. The norm is that the observer, on arrival at the specified point, starts counting directly or gives time for birds to settle before counting begins. Suppose that the selected point or unit meets the condition of interest and the observer is to move to the neighboring unit. This movement may make some birds fly out to maybe adjacent units. Defining neighborhood as the adjacent units is likely to lead to double counting. Smith et al. (2004, p. 93) suggested defining neighborhood as the units that are not directly adjacent to the surveyed unit. *i.e* neighboring units are units that are l meters away from the observed unit. This would work in surveying, say, rare breeding birds or birds that are not easily detectable whose colonies are relatively large as to stretch to the defined neighboring units.

The other challenges or factors that need to be put into consideration during survey of animals in general include general flight planning, piloting and observation. Norton-Griffiths (1978, p: 29-64) gives details on the practical issues to consider while designing and undertaking an aerial survey. For imperfect detectability, *see* Thompson and Seber (1994) and Naddeo and Pisani (2005).

2.4 Product method of estimation

Consider a population that is divided into N units and for each value y_i of the study variable, there is a value of the auxiliary variable x_i , $i = 1, 2, \dots, N$. The aim is to estimate the population total τ_y of the variable of interest y , given

that the population total τ_x of the auxiliary variable x is known. Further, the correlation between the study and the auxiliary variable is highly negative.

Consider drawing a sample of size n by SRSWOR. Let \bar{y} and \bar{x} be the estimator of the population mean of the study (μ_y) and the auxiliary (μ_x) variable respectively. The usual product estimator of the population total τ_y is given by (Murthy, 1964)

$$\hat{\tau}_p = N \times \hat{\mu}_p, \text{ where } \hat{\mu}_p = \frac{\bar{y}\bar{x}}{\mu_x}. \quad (2.29)$$

$\hat{\tau}_p$ is biased and its bias is given by

$$B(\hat{\tau}_p) = \frac{N(N-n)}{n} \frac{S_{xy}}{\mu_x}$$

while its MSE is given by

$$MSE(\hat{\tau}_p) = \frac{N(N-n)}{n} [S_y^2 + \bar{R}^2 S_x^2 + 2\bar{R} S_{xy}], \quad (2.30)$$

with $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$, $S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$ and $\bar{R} = \frac{\mu_y}{\mu_x}$.

Let the coefficient of variation of the study y and the auxiliary x variable be $CV_y = \frac{S_y}{\mu_y}$ and $CV_x = \frac{S_x}{\mu_x}$ respectively in which S_y is the standard deviation for y and S_x is the standard deviation for x . The MSE of the product estimator can also be written as

$$MSE(\hat{\tau}_p) = N^2 \mu_y^2 (CV_y^2 + CV_x^2 + 2\rho_{xy} CV_y CV_x)$$

with ρ_{xy} denoting the correlation coefficient between y and x . Murthy (1964) analyzed the relative precision of the product estimator and showed that it is more efficient than the sample mean if

$$\rho_{xy} < -\frac{1}{2} \frac{CV_x}{CV_y}. \quad (2.31)$$

3. Design-based estimation

3.1 Stratified strip ACS: A case study in estimating population total of African elephants

3.1.1 Introduction

African elephants are species of conservation concern. They are listed on Appendix I of the Convention on International Trade in Endangered Species (CITES) hence banning the international trade in ivory and other elephant products. Due to their concern, the Elephant Programme was established in Kenya in 1989 to protect the elephants in Kenya from becoming extinct caused by poachers. Some of its tasks is to monitor population totals of the elephants, their trend and carry out research. Elephants are both solitary and clustered animals (Khaemba and Stein, 2000) which provides a justification to explore the ACS design in the estimation of their totals.

In this chapter, strip ACS design proposed (Thompson, 2012, p. 339) is extended to a stratified setting. The sampling process considered in this chapter is sampling without replacement of clusters (Dryver and Thompson, 2007). The estimator due to Raj (1956) is modified to produce unbiased estimates under the proposed design. Simulations were carried out using artificial and real data.

The design and its proposed estimator is described in sub-sections (3.1.2) and (3.1.3) respectively. Its efficiency comparison and cost function are presented in sub-section (3.1.4). A small population example to demonstrate the calculation of the estimates is presented in sub-section (3.1.5). A Monte-Carlo simulation using artificial data is presented in sub-section (3.1.6) while simulations using real data are presented in sub-section (3.1.7). Finally, comments on the results are given in sub-section (3.1.8).

3.1.2 The design

Consider again figure 2.4 but now divided into $H = 3$ strata. Stratification can be based on habitat type or administrative convenience. In each stratum,

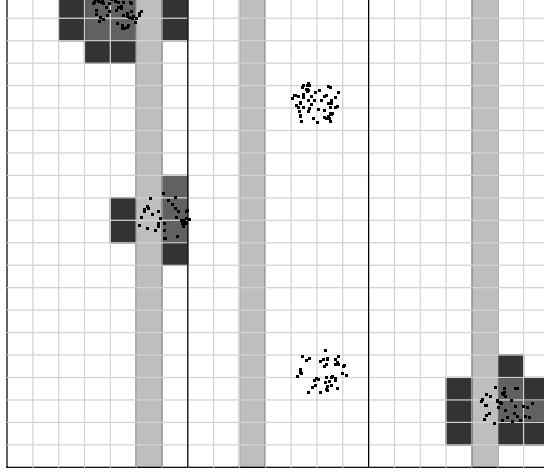


Fig. 3.1: Plot of stratified strip ACS with initial sampled strip colored light gray, adaptively added units dark gray and edge units black.

the first strip is selected at random and an adaptive search performed. The observed strip with its associated cluster(s) is "removed" from the population. A second strip is selected from the remaining strips and an adaptive search performed. Again, the observed strip with associated cluster(s) is "removed" from the population. The procedure continues until n_q strips are sampled in each stratum.

In stratified ACS, (Thompson, 2012, p. 353) ignored stratum boundaries during network formation. In this study, networks are truncated at stratum boundaries. Figure 3.1 shows that the proposed design contributes to lower sampling cost (measured in terms of final sample size) compared to the stratified design proposed by (Thompson, 2012, p. 353) in which networks are allowed to cross over stratum boundaries. The design further reduces sampling cost by avoiding visiting units that have previously been observed.

3.1.3 The modified Des Raj's estimator

Maintaining notations in the review section, let H be the number of strata. For each strata, let N_q^* be the number of strips or PSU, M_q^* the number of SSU in each strip, N_q the number of units and n_q the sample size (number of strips) such that $\sum_{q=1}^H n_q = n$ and $\sum_{q=1}^H N_q = N$. Let m_{kq} be the network size and w_{kq} the average of y -values in network k of stratum q .

When clusters are selected without replacement, following Raj (1956), the unbiased estimator of the population total is

$$\hat{\tau}_c = \sum_{q=1}^H \hat{\tau}_q, \quad \hat{\tau}_q = \sum_{i=1}^{n_q} c_{qi} z_{qi} \quad (3.1)$$

where

$$z_{q1} = \frac{\sum_{j \in s_{q1}} w_{qj}}{p_{q1}}, \quad p_{q1} = \frac{M_{q1}^*}{N_q} \quad (3.2)$$

in which s_{q1} are the SSU that are in the PSU that was selected at the first draw in stratum q , p_{q1} is the probability of selecting a strip of size M_{q1}^* at the first draw in stratum q in which M_{q1}^* is the number of SSU in the selected PSU and $c_{qi} = \frac{1}{n_q} \forall i$ are weights such that $\sum_{i=1}^{n_q} c_{qi} = 1$. For $i = 2, \dots, n$

$$z_{qi} = \sum_{j \in s_{i-1}} w_{qj} + \frac{\sum_{l=1}^{M_{qi}^* - M_{qi-1}^*} w_{ql}}{p_{qi}} \quad (3.3)$$

where $\sum_{l=1}^{M_{qi}^* - M_{qi-1}^*} w_{ql}$ represents the SSU values (in terms of network averages) in the N_q^* -th strip that have not been observed at draw i . The probability p_{qi} of selecting a strip of size M^* at draw i in stratum q is given by

$$p_{qi} = \frac{M_{qi}^*}{N_q - \sum_{j=1}^{i-1} t_{qj}} \quad (3.4)$$

with M_{qi}^* being the number of SSU, that have not been selected, in a strip that is selected at draw i while $\sum_{j=1}^{i-1} t_{qj}$ are the number of SSU (number of strips and their adaptively added units including their edge units) observed up to $i-1$ draw.

The variance of 3.1 is

$$v(\hat{\tau}_c) = \sum_{q=1}^H v(\hat{\tau}_q) = \sum_{q=1}^H \left(\frac{1}{n_q^2} \sum_{i=1}^{n_q} v(z_{qi}) \right)$$

whose unbiased estimator is

$$\hat{v}(\hat{\tau}_c) = \sum_{q=1}^H \hat{v}(\hat{\tau}_q) = \sum_{q=1}^H \left(\frac{1}{n_q(n_q - 1)} \sum_{i=1}^{n_q} (z_{qi} - \hat{\tau}_q)^2 \right) \quad (3.5)$$

3.1.4 Efficiency comparison

Let n^* be the sample size from a SRSWOR design. For an adaptive strategy to have a lower variance (when HH estimator is adopted) compared to the sample mean of a SRSWOR, the following condition must be satisfied (Thompson, 2012, p. 319):

$$\frac{n^* - n}{nn^*} \sigma_y^2 < \frac{N - n}{Nn(N - 1)} \sum_{k=1}^K \sum_{j=1}^{m_k} (y_j - w_k)^2 \quad (3.6)$$

where w_k is the average of observations in network k i.e. $w_k = \frac{1}{m_k} \sum_{j=1}^{m_k} y_j$, m_k is the network size, n is the initial sample size for the adaptive design and $\sigma_y^2 = \frac{\sum_{i=1}^N (y_i - \mu_y)^2}{N - 1}$ is the finite population variance.

Extending the above to the stratification, it implies that the stratified ACS will have lower variance than the classical stratified random sampling of fixed size n^* if

$$\frac{n_q^* - n_q}{n_q n_q^*} \sigma_{qy}^2 < \frac{N_q - n_q}{N_q n_q (N_q - 1)} \sum_{k=1}^{K_q} \sum_{j=1}^{m_{kqk}} (y_{jq} - w_{kq})^2, \quad q = 1, \dots, H \quad (3.7)$$

i.e. the stratified ACS design will result in lower variance compared to the classical stratified random sampling if, in each stratum, the within network variance is greater than the stratum population variance.

Consider a classical sampling design in which units are selected with probability proportional to size and without replacement. Assuming that the

sampling fraction is negligible, then (Salehi and Seber, 1997a; Chaudhuri and Stenger, 2005, p. 29).

$$v(\hat{\tau}_{Raj}) \leq v(\hat{\tau}_{hh}) \quad (3.8)$$

The above relation applies under ACSⁿ (Salehi and Seber, 1997a). Further note that the ACS^c has lower variance compared to the ACSⁿ (Dryver and Thompson, 2007). Hence the inequality given in (3.8) also holds under ACS^c design.

Cost function

During an aerial survey, temporary fuel stations can be distributed in the study region. This is necessary to save on time and cost as the aircraft will not have to travel to the base station to refuel (Norton-Griffiths, 1978, p. 48). With this in mind, the cost equation under the proposed design is given by

$$C = \sum_{q=1}^H C_q = \sum_{q=1}^H (c_0 + c_1 n_q + c_2 (v_q - M_q^*) + c_3 e_q) \quad (3.9)$$

where c_0 is a fixed cost, c_1 is the cost of surveying a strip selected in the initial sample, c_2 is the cost of surveying the adaptively added units (that meet the pre-specified condition) where v is the number of units in the initial sample and the adaptively added units (excluding the edge units). Finally c_3 is the cost of surveying edge units in which e are the number of edge units. Note that an empty unit that is selected in the initial sample is a network of size one (Thompson, 2012, p. 319).

3.1.5 A small population example

In this sub-section, the computation of the estimates are illustrated in one stratum using an artificial population that is presented in table 3.1. The same procedure applies to remaining strata. On the left hand side of the table are the counts in each unit while on the right hand side is the same data represented as network averages.

y_i				w_i			
1	2	3	4	1	2	3	4
80	3	8	0	75	3	8	0
70	4	2	1	75	4	2	1
5	5	60	2	5	5	80	2
6	2	80	100	6	2	80	80
1	0	1	4	1	0	1	4

Tab. 3.1: An artificial population with the number of animals in each unit y_i represented on the left hand side while the transformed population w_i is presented on the right hand side. Second row with numbers 1, 2, 3, 4 are labels for each strip. Here $N_q^* = 4$, $M_q^* = 5$, $N_q = 20$, $n_q = 2$ and $C = \{y_i : y_i \geq 50\}$.

In this particular stratum, there are $N_q^* = 4$ strips each with $M_q^* = 5$ SSU. The total number of SSU in this stratum is $N_q = 20$. Let the sample size n_q be 2 strips and the condition to adaptively add neighboring units be $C = \{y_i : y_i \geq 50\}$. The population total of y is 434. Table 3.2 gives results over all possible samples.

The table shows that z_1 is unbiased and this follows from selecting a strip of size one. To further explain the calculation, suppose that the first strip selected is that of column 3 and the second strip selected is that of column 2 whose results are given in row 8 of table 3.2. Then p_1 , p_2 and $p(s)$ are obtained as

$$\begin{aligned}
 p_1 &= \frac{M_{q1}^*}{N_q} \\
 &= \frac{5}{20} \\
 p_2 &= \frac{M_{qi}^*}{N_q - \sum_j^{i-1} t_{qj}} \\
 &= \frac{3}{20 - 10} = \frac{3}{10} \\
 p(s) &= p_1 \times p_2 \\
 &= \frac{5}{20} \times \frac{3}{10} = \frac{3}{40}
 \end{aligned}$$

in which p_1 represent the probability of selecting a strip of size 5 at the first draw, p_2 is the probability of selecting a strip of size 3 (remaining SSU in the selected strip) at the second draw given that the third strip was selected at the first draw and the probability of the specified sample is $p(s)$. The estimate of the population total is

$$\begin{aligned}
 z_{q1} &= \frac{\sum_{j \in s_{q1}} w_{qj}}{p_{q1}} \\
 &= (8 + 2 + 80 + 80 + 1) \times \frac{20}{5} \\
 &= 684 \\
 z_{q2} &= \sum_{j \in s_1} w_{qj} + \frac{\sum_{l=1}^{M_{qi}^* - M_{q1}^*} w_{ql}}{p_{qi}} \\
 &= (8 + 2 + 80 + 80 + 1 + 5 + 2 + 2 + 80 + 4) + (3 + 4 + 0) \times \frac{10}{3} \\
 &= 287.33 \\
 \hat{\tau}_q &= \frac{1}{n_q} \sum_{i=1}^{n_q} z_{qi} \times p(s) \\
 &= \frac{684 + 287.33}{2} \times \frac{3}{40} \\
 &= 485.67 \times \frac{3}{40} \\
 &= 36.43
 \end{aligned}$$

samples	v	p_1	p_2	$p(s)$	z_1	z_2	$\hat{\tau}_c$	$\hat{v}(\hat{\tau}_c)$	$\hat{v}(\hat{\tau})$
1,2;	10	1/4	3/13	3/52	648.00	199.33	24.44	6.16	6,724
2,1;	10	1/4	1/3	1/12	56.00	500.00	23.17	2,028.00	53,824
1,3;4,2	17	1/4	5/13	5/52	648.00	613.60	60.65	3,724.06	4,096
3,1;4,2	17	1/4	1/2	1/8	684.00	588.00	79.50	5,100.50	4,096
1,4;3,2	18	1/4	5/13	5/52	648.00	395.20	50.15	737.86	53,824
4,1;3,2	18	1/4	5/9	5/36	348.00	548.60	62.26	28.40	6,724
2,3;4	13	1/4	1/3	1/12	56.00	527.00	24.29	1,692.19	6,724
3,2;4	13	1/4	3/10	3/40	684.00	287.33	36.43	200.21	53,824
2,4;3	14	1/4	1/3	1/12	56.00	275.00	13.79	6,007.69	4,096
4,2;3	14	1/4	1/3	1/12	348.00	278.00	26.08	1,220.08	4,096
3,4;2	12	1/4	1/5	1/20	684.00	269.00	23.83	90.31	53,824
4,3;2	12	1/4	1/9	1/36	348.00	329.00	9.40	253.34	6,724
Expectation	14			1	434.00		434.00	21,088.80	21,548

Tab. 3.2: All possible ordered samples where samples are drawn without replacement of clusters. First column lists strips selected in the initial sample while strips whose units were adaptively added are included after semicolon. $\hat{v}(\hat{\tau})$ is the variance under non-adaptive.

3.1.6 Simulation study

A range of rare and clustered artificial populations were generated using Poisson Cluster Process (PCP) (Diggle, 2013, p. 101). In the PCP, first, the number of parents (nop) are generated as a realization of a Poisson process with mean λ_1 . Parents are then randomly distributed within the region of interest. Each parent has an average number of offsprings that are simulated according to a Poisson process with mean λ_2 . The variation of the locations of offsprings relative to their parents is controlled by a parameter s .

The artificial populations were generated using **R** package **Splancs** (Rowlingson and Diggle, 2015). The range of parameter values used were: $\text{nop} = \lambda_1 \in \{5, 10\}$, $\lambda_2 \in \{50, 60, 80, 100, 120, 150\}$ and $s \in \{0.00025, 0.0005, 0.001\}$. The above populations were generated in a region that was partitioned into $N = 30 \times 30 = 900$ units. Some of the simulated populations are given in figure (3.2).

For each population, $R = 20,000$ samples were drawn and estimates obtained following the formula shown above. An initial sample of size $n \in \{8, 12\}$ strips was used. Samples were allocated to different strata using optimal allocation (Neyman, 1934) and optimal allocation according to box-constraint (Gabler et al., 2012; Münnich et al., 2012). The standard deviation in equation 2.27 was evaluated at the network level using the available population total in each stratum. Table (3.3) presents the distribution of samples to different strata under the two optimal allocation strategies. For the optimal allocation with box-constraint, each stratum was set to have a minimum sample of size 2 and the maximum sample size was set to the total number of strips in the particular stratum.

From the table (table (3.3)), the method under Neyman allocation allocated some strata with one or non samples. In which case the particular strata may not be included in the variance estimation or its variance estimation may not be direct. For this reason, optimal allocation (Neyman, 1934) was removed

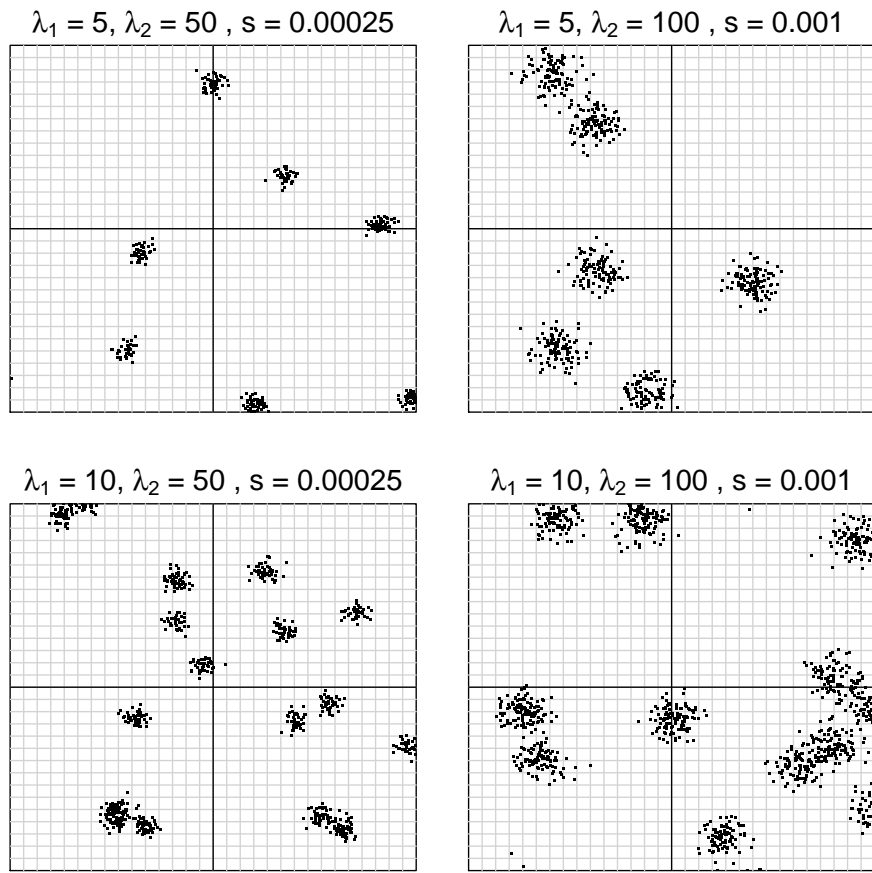


Fig. 3.2: Simulated populations under different parameter setting.

				Neyman optimal				Box-constraint			
n	λ_1	λ_2	s	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
8	5	50	0.00025	1	1	6	0	2	2	2	2
	10	80	0.00100	1	2	3	2	2	2	2	2
12	5	50	0.00025	1	1	10	0	2	2	6	2
	10	80	0.00100	1	3	5	3	2	3	4	3

Tab. 3.3: Allocations of samples to different strata according to optimal (Neyman, 1934) and optimal according to box-constraint (Gabler et al., 2012) allocations for some of the parameters considered and different initial sample sizes. q_1, \dots, q_4 represent the first, ..., fourth stratum respectively.

from the analysis and evaluation was performed under optimal allocation with box-constraint.

The comparison was made with its non-adaptive counterpart (stratified strip design). Under the non-adaptive design, the sample size was set by rounding off the expected final sample size v^* of the adaptive design. The evaluation was performed using relative efficiency (RE) and the expected final sampling fraction (fn). The RE and the fn were obtained as

$$RE = \frac{\hat{v}(\hat{\tau})}{\hat{v}(\hat{\tau}_c)}, \quad fn = \frac{v^*}{N} \quad (3.10)$$

If $RE > 1$ it implies that the proposed strategy is efficient relative to its non-adaptive counterpart.

Figure (3.3) shows results of the relative efficiency. From the figure, the adaptive strategy has helped in the improvement in efficiency relative to non-adaptive stratified strip design. Under the populations considered, the relative efficiency increases with increase in the number of units in a network as well as cluster size. This of course determines the within network variance. The results of the expected final sampling fraction (figure (3.4)) shows that increasing cluster size increases the final sampling effort.

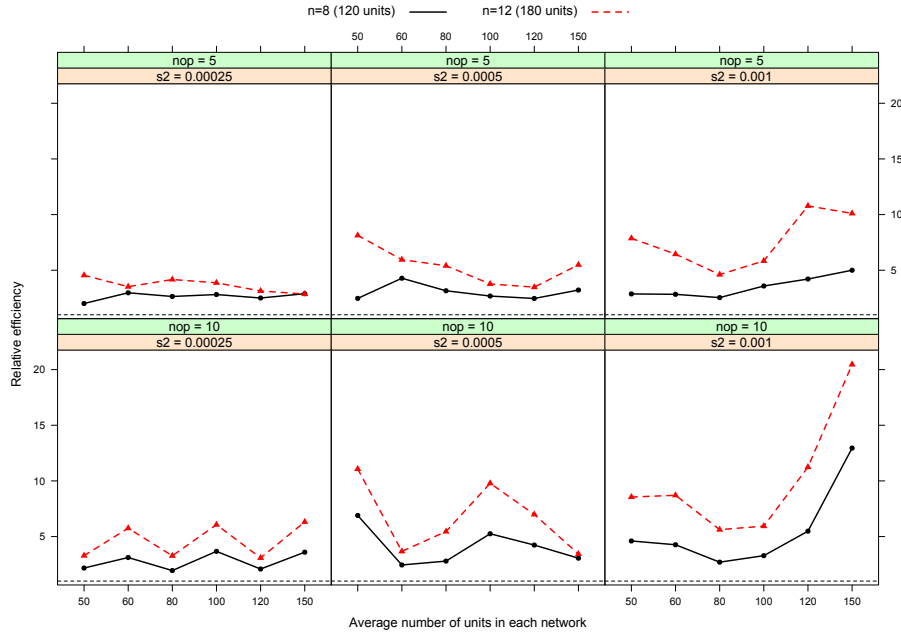


Fig. 3.3: Results of the relative efficiency for the different parameters and for the two initial sample sizes considered.

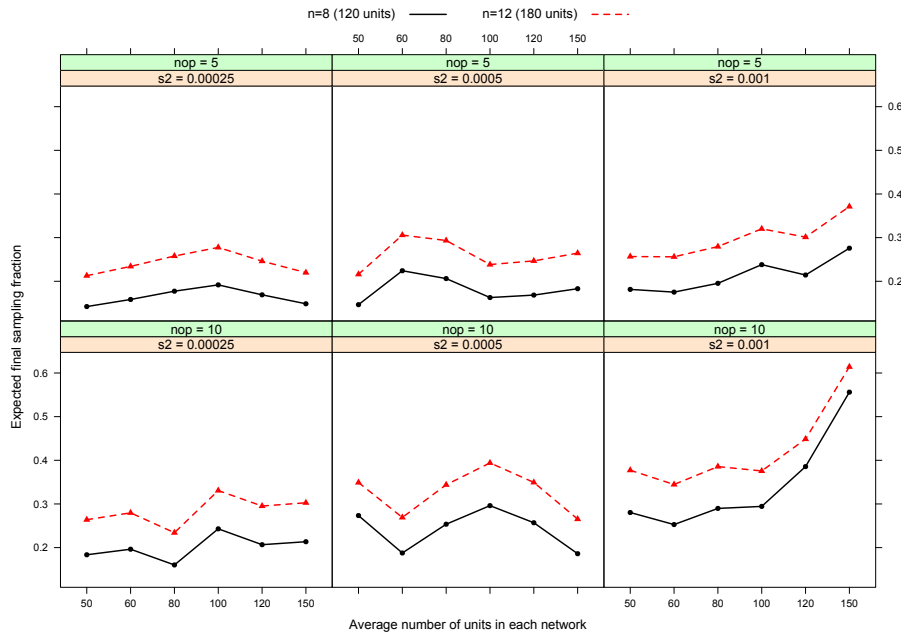


Fig. 3.4: Results of the expected final sampling fraction for the different parameters and for the two initial sample sizes considered.

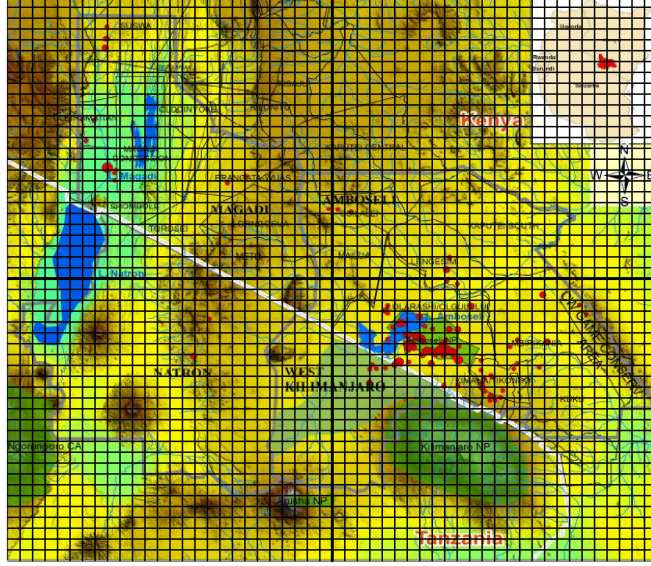


Fig. 3.5: Distribution of the elephant population in which the region is divided into 4 strata. The strata are represented by thick boarder line. Map obtained from Kenana et al. (2013b).

3.1.7 Simulations using real data

The proposed design was applied to a real data which was the distribution and counts of elephants across Amboseli-West Kilimanjaro/Magadi-Natron cross border landscape. The landscape covers parts of Kenya and Tanzania between $1^{\circ} 37' S$ and $3^{\circ} 13' S$ and between $35^{\circ} 49' E$ and $38^{\circ} 00' E$ covering an area of $25,623 km^2$. The data was obtained from an aerial survey that was conducted in October (dry season) 2013. Details on how the survey was conducted can be found in Kenana et al. (2013b).

A grid was laid on a map that was produced from the survey as shown in figure 3.5 and an effort was made to determine the counts in each unit. The region had $N = 2,304$ units such that there are 48 row and 48 columns. The number of strata was $H = 4$, each stratum had $N_q^* = 24$ strips (PSU) and each strip had $M_q^* = 24$ SSU. The number of SSU in each stratum is $N_q = 576$. The

initial sample size under ACS was $n = 8$ and 12 strips.

Strategies that were used to allocate samples sizes to different strata were optimal (Neyman, 1934) and optimal allocation according to box-constraint (Gabler et al., 2012; Münnich et al., 2012) in which the stratum standard deviation was evaluated at the network level. Table (3.4) shows distribution of sample sizes to different strata. Like in the simulated data, optimal allocation was omitted from the analysis. 10,000 sampling simulation were carried out. For the non-adaptive stratified strip design, the sample size used was derived by rounding off the expected final sample size under the ACS design.

	Optimal				Box-constraint			
n	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
8	2	0	0	6	2	2	2	2
12	3	0	1	8	2	2	2	6

Tab. 3.4: Allocation of samples to different strata for elephant population under the proposed design

Results of the relative efficiency and the expected final sampling fraction are given in table (3.5). From the table, the proposed design is more efficient compared to the non adaptive stratified strip sampling. As expected, efficiency increases with increase in the initial sample size.

n	RE	$E(fn)$
8 (192 units)	1.354	0.101
12 (288 units)	2.206	0.153

Tab. 3.5: Relative efficiency (RE) and expected final sampling fraction ($E(fn)$) for the elephant data. RE and $E(fn)$ are defined in equation (3.10).

3.1.8 Discussion

ACS is an efficient design when the population under study is rare and clustered. Greater efficiency is achieved when there are few but relatively large networks and the size of initial sample is close to the final sample (Brown, 2003). The relatively large networks would ensure that condition (3.7) is satisfied.

(Thompson, 2012, p. 353) ignored stratum boundaries when forming networks. In this study, the networks were truncated at the stratum boundaries. Truncating the networks at the stratum boundaries helps in maintaining the stratum boundaries from the design to the analysis stage. The other advantage of this approach is that, in practice, if there exist field stations, its easier and faster for each field station to conduct sampling and perform estimation within its jurisdiction and later combine the results.

Also if the networks are truncated at stratum boundaries as opposed to allowing networks to cross over boundaries *i.e* ignoring stratum boundaries as proposed by (Thompson, 2012, p. 353), helps in reducing the final sample size hence the sampling costs when there exists networks at boundaries.

When there are no networks that span over two or more strata, the proposed stratification and the stratification due to (Thompson, 2012, p. 353) results in equivalent results. The stratification due to (Thompson, 2012, p. 353) however results in lower variances compared to the proposed stratification when there exist networks that spread over at least two strata and with greater within network variance.

When it comes to stratification, the decision to either use the proposed approach and truncate the networks at the stratum boundaries or to follow (Thompson, 2012, p. 353) and allow the networks to cross over the boundaries, will depend on the level of precision desired and the amount of resources available. The resources could be in terms of equipments required to carry out the survey such as the number of light aircraft with their observation equipments, time available and the number of field experts available for the field survey.

The current study considers ordered samples that made use of the Des Raj estimator. However, re-ordering the samples and making use of the Murthy estimator (Murthy, 1957) results in more efficient estimates but computationally intensive. Further research involves exploration on the use of re-ordered samples in the analysis.

3.2 ACS for negatively correlated data

3.2.1 Introduction

It is common for researchers to use auxiliary information that is cheaper to obtain to improve on the efficiency of their estimators. The auxiliary information could be obtained from satellite images, previous census or registers. Classical estimators that make use of auxiliary information include ratio, regression and product estimators. From the three, which estimator to use to achieve more efficient estimates, depends on the correlation between the study variable y and the auxiliary variable x . Further, how far the regression line passes through the neighborhood of the origin must also be considered.

The ratio estimator is efficient when the correlation between the study and the auxiliary variable is highly positive and the regression line passes through the neighborhood of the origin. The conditions for the product estimator are similar to the ratio estimator but with the correlation being highly negative. The regression estimator on the other hand is more efficient than the ratio or the product estimator when the regression line passes away from the neighbourhood of the origin. Otherwise, it is similar to the ratio or to the product estimator.

In the context of rare and clustered populations that make use of the ACS design, Chutiman et al. (2013) studied the use of two auxiliary information under a stratified setting. Chao (2004a) studied the use of auxiliary information that is positively correlated with the variable of interest. He suggested a generalized ratio estimator based on the modified HT estimator. Dryver and Chao (2007) proposed a different type of ratio estimator that is based on the modified HH estimator. Chao et al. (2011) and Lin and Chao (2014) used Rao-Blackwell theorem to improve the efficiency of the ratio estimator under the ACS design. Chao (2004b) proposed the regression estimator under ACS design. Chutiman et al. (2013) gave a summary form of the ACS estimators in the presence of auxiliary variable(s).

In this sub-section, the situation when there is a negative correlation between

the study and the auxiliary variable is examined. Suppose that the abundance of wildlife species and the average amount of rainfall represents the study and the auxiliary variable respectively in an ecological or environmental study. Suppose further that an average amount of 500mm of rainfall is adequate to ensure good forage for the wildlife survival *i.e* the given park/area has approximately 1,500 animals of type *A* if the average amount of rainfall is 500mm. As the average amount of rainfall increases, the abundance of wildlife will decrease from death caused by lack of food (Prins, 1996, p. 12) or some get stuck in the mud.

A point to note in studies involving negative correlation between the study and the auxiliary variable is that the point of origin is not always the literal $\{0,0\}$. It is rather a known constant and the model is shifted accordingly. In the example above, the point of origin is $\{500,1500\}$. This sub-section proposes two product estimators and their associated variance estimators under the ACS design.

The proposed estimators are described in sub-section (3.2.2). Situations in which the product ACS is efficient relative to the ACS and to the product estimator under SRSWOR are presented in sub-section (3.2.3). An example using a small population to demonstrate the calculation of the estimates is presented in sub-section (3.2.4). A simulation study using artificial and real data are presented in sub-sections (3.2.5) and (3.2.6) respectively. Some comments on the results are finally given in sub-section (3.2.7).

3.2.2 Estimators

The product estimator due to Murthy (1964) is modified by incorporating the modified HT and HH estimators that were proposed by (Thompson, 2012, p. 319). The interest is to estimate the population total of the study variable given that the auxiliary variable for all units in the sample and the population total of auxiliary variable is available. Further there exists a negative correlation between the study and the auxiliary variable.

The product HH estimator

Let $\hat{\tau}_{hh_y}$ and $\hat{\tau}_{hh_x}$ be the modified HH estimators of y and x respectively as defined in equation (2.9). Further, let τ_x be the population total for x , then the product HH estimator is given by

$$\hat{\tau}_{phh} = \frac{\hat{\tau}_{hh_y} \hat{\tau}_{hh_x}}{\tau_x} \quad (3.11)$$

Lemma 1. *The bias of $\hat{\tau}_{phh}$ is*

$$B(\hat{\tau}_{phh}) = \frac{cov(\hat{\tau}_{hh_{xy}})}{\tau_x} \quad (3.12)$$

where $cov(\hat{\tau}_{hh_{xy}})$ is the covariance between x and y under the HH estimator given by

$$cov(\hat{\tau}_{hh_{xy}}) = \frac{N(N-n)}{n} \cdot \frac{1}{N-1} \sum_{k=1}^N (w_{yk} - \mu_y)(w_{xk} - \mu_x). \quad (3.13)$$

The MSE of $\hat{\tau}_{phh}$ to the first degree of approximation is

$$MSE(\hat{\tau}_{phh}) \approx \tau_y^2 \left(\frac{v(\hat{\tau}_{hh_y})}{\tau_y^2} + \frac{v(\hat{\tau}_{hh_x})}{\tau_x^2} + 2 \frac{cov(\hat{\tau}_{hh_{xy}})}{\tau_y \tau_x} \right). \quad (3.14)$$

where $v(\hat{\tau}_{hh_y})$ and $v(\hat{\tau}_{hh_x})$ are the variances of the HH estimator as defined in (2.11).

Proof. The proofs for the bias and the MSE are provided in appendix A.1. \square

The estimate of the bias and the MSE are, respectively

$$\hat{B}(\hat{\tau}_{phh}) = \frac{\widehat{cov}(\hat{\tau}_{hh_{xy}})}{\tau_x} \quad (3.15)$$

and

$$\widehat{MSE}(\hat{\tau}_{phh}) \approx \tau_y^2 \left(\frac{\hat{v}(\hat{\tau}_{hh_y})}{\hat{\tau}_{hh_y}^2} + \frac{\hat{v}(\hat{\tau}_{hh_x})}{\hat{\tau}_{hh_x}^2} + 2 \frac{\widehat{cov}(\hat{\tau}_{hh_{xy}})}{\hat{\tau}_{hh_y} \hat{\tau}_{hh_x}} \right) \quad (3.16)$$

in which

$$\widehat{cov}(\hat{\tau}_{hh_{xy}}) = \frac{N(N-n)}{n} \cdot \frac{1}{n-1} \sum_{k=1}^n (w_{yk} - \hat{\mu}_{hh_y})(w_{xk} - \hat{\mu}_{hh_x})$$

while $\hat{v}(\hat{\tau}_{hh_y})$ and $\hat{v}(\hat{\tau}_{hh_x})$ are the unbiased variance estimators of $v(\hat{\tau}_{hh_y})$ and $v(\hat{\tau}_{hh_x})$ respectively whose formulas are given in equation (2.13).

Thompson and Seber (1996) and Dryver and Chao (2007) viewed ACS as sampling without replacement of n networks with inclusion probability m_k/N . With this definition, lemma 2 provides an alternative formulation for the MSE.

Lemma 2. *Let $w_{y1}, w_{y2}, \dots, w_{yN}$ be the population network means of the y -values and let $w_{x1}, w_{x2}, \dots, w_{xN}$ be the population network means of the x -values. The MSE of $\hat{\tau}_{phh}$, to the first degree of approximation, is (Gattone et al., 2016b)*

$$\begin{aligned} MSE(\hat{\tau}_{phh}) &\approx \frac{N(N-n)}{n} \cdot \mu_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) \\ &\approx \left(\frac{N-n}{Nn} \right) \tau_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) \end{aligned} \quad (3.17)$$

where CV_{w_y} and CV_{w_x} are the coefficients of variations at the network level for the variables y and x respectively and $\rho_{w_{xy}}$ is the correlation coefficient at the network level

Proof. The proof for lemma 2 is provided in appendix A.2. \square

The variance of $\hat{\tau}_{phh}$ can be approximated by

$$v(\hat{\tau}_{phh}) \approx \frac{N(N-n)}{n} \cdot \frac{1}{N-1} \sum_{k=1}^N (w_{yk} + \bar{R}w_{xk})^2$$

where $\bar{R} = \frac{\mu_y}{\mu_x}$. This variance is estimated by

$$\hat{v}(\hat{\tau}_{phh}) \approx \frac{N(N-n)}{n} \cdot \frac{1}{n-1} \sum_{k=1}^n (w_{yk} + \hat{R}_{hh}w_{xk})^2.$$

in which $\hat{R}_{hh} = \frac{\hat{\mu}_{hh_y}}{\hat{\mu}_{hh_x}}$.

The product HT estimator

The product HT estimator is given by

$$\hat{\tau}_{pht} = \frac{\hat{\tau}_{ht_y} \hat{\tau}_{ht_x}}{\tau_x} \quad (3.18)$$

where $\hat{\tau}_{ht_y}$ and $\hat{\tau}_{ht_x}$ are the HT estimators of τ_y and τ_x , respectively as defined in equation (2.6).

Lemma 3. *The bias of $\hat{\tau}_{pht}$ is*

$$B(\hat{\tau}_{pht}) = \frac{cov(\hat{\tau}_{ht_{xy}})}{\tau_x} \quad (3.19)$$

where $cov(\hat{\tau}_{ht_{xy}})$ is the covariance between $\hat{\tau}_{ht_y}$ and $\hat{\tau}_{ht_x}$ given by

$$cov(\hat{\tau}_{ht_{xy}}) = \sum_{j=1}^K \sum_{k=1}^K \Delta_{jk} \frac{y_j^*}{\alpha_j} \frac{x_k^*}{\alpha_k}$$

with $\Delta_{jk} = \alpha_{jk} - \alpha_j \alpha_k$ while y_j^* is the sum of y -values in network j and x_k^* is the sum of x -values in network k . Its MSE to the first degree of approximation is

$$MSE(\hat{\tau}_{pht}) \approx v(\hat{\tau}_{ht_y}) + R^2 v(\hat{\tau}_{ht_x}) + 2R cov(\hat{\tau}_{ht_{xy}}) \quad (3.20)$$

where $R = \frac{\tau_y}{\tau_x}$ while $v(\hat{\tau}_{ht_y})$ and $v(\hat{\tau}_{ht_x})$ are the variances of the HT estimators defined in equation (2.7).

Proof. The proof is the same as that provided in appendix A.1 by replacing HH with HT in the notation. \square

Like in the product HH estimator above, the alternative formulation for the MSE with HT estimator is

$$MSE(\hat{\tau}_{pht}) \approx \tau_y^2 \left(CV_{ht_y}^2 + CV_{ht_x}^2 + 2\rho_{ht_{xy}} CV_{ht_y} CV_{ht_x} \right). \quad (3.21)$$

which can be estimated by

$$\widehat{MSE}(\hat{\tau}_{pht}) \approx \hat{v}(\hat{\tau}_{ht_y}) + \hat{R}_{ht}^2 \hat{v}(\hat{\tau}_{ht_x}) + 2 \hat{R}_{ht} \widehat{cov}(\hat{\tau}_{ht_{xy}})$$

where $\hat{R}_{ht} = \frac{\hat{\tau}_{ht_y}}{\hat{\tau}_{ht_x}}$ and $\widehat{cov}(\hat{\tau}_{ht_{xy}})$ is the unbiased estimator of $cov(\hat{\tau}_{ht_{xy}})$ given by

$$\widehat{cov}(\hat{\tau}_{ht_{xy}}) = \sum_{j=1}^K \sum_{k=1}^K \check{\Delta}_{jk} \frac{z_j y_j^*}{\alpha_j} \frac{z_k x_k^*}{\alpha_k}$$

with $\check{\Delta}_{jk} = \frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk}}$ in which z is an indicator variable that takes the value 1 if the selected unit is in the sample and 0 otherwise. The variance of $\hat{\tau}_{pht}$ can be approximated by

$$v(\hat{\tau}_{pht}) \approx \sum_{j=1}^K \sum_{k=1}^K \Delta_{jk} \frac{(y_j^* + R x_j^*)}{\alpha_j} \frac{(y_k^* + R x_k^*)}{\alpha_k}$$

whose estimate is

$$\hat{v}(\hat{\tau}_{pht}) \approx \sum_{j=1}^K \sum_{k=1}^K \check{\Delta}_{jk} \frac{z_j (y_j^* + \hat{R}_{ht} x_j^*)}{\alpha_j} \frac{z_k (y_k^* + \hat{R}_{ht} x_k^*)}{\alpha_k}. \quad (3.22)$$

3.2.3 Efficiency comparisons

The product ACS estimators are theoretically compared in terms of efficiency with respect to the standard ACS estimators (Thompson, 2012, p. 319) and with respect to the product estimator under SRSWOR (Murthy, 1964). Since the comparisons are based on the first order of approximation of the MSE of the suggested estimators, a simulation study will be provided to support the theoretical findings.

Comparisons with estimators under the standard ACS

The condition under which the proposed product estimator is more efficient than the standard ACS (Thompson, 2012, p. 319) is analyzed Using equations (3.17) and (2.11) , it follows that $MSE(\hat{\tau}_{phh}) < v(\hat{\tau}_{hh})$ (*i.e* the product

estimator $\hat{\tau}_{phh}$ is more efficient than the standard ACS estimator $\hat{\tau}_{hh}$) when

$$\frac{N(N-n)}{n} \mu_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) < \frac{N(N-n)}{n} S_{w_y}^2. \quad (3.23)$$

Note that

$$\begin{aligned} CV_{w_y}^2 &= \frac{S_{w_y}^2}{\mu_y} \\ \Rightarrow S_{w_y}^2 &= \mu_y^2 CV_{w_y}^2 \end{aligned} \quad (3.24)$$

Substituting equation (3.24) in equation (3.23), we have

$$\begin{aligned} \frac{N(N-n)}{n} \mu_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) &< \frac{N(N-n)}{n} \mu_y^2 CV_{w_y}^2 \\ 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} &< -CV_{w_x}^2 \\ \rho_{w_{xy}} &< -\frac{1}{2} \frac{CV_{w_x}^2}{CV_{w_y} CV_{w_x}} \\ \rho_{w_{xy}} &< -\frac{1}{2} \frac{CV_{w_x}}{CV_{w_y}} \end{aligned} \quad (3.25)$$

In a similar manner using equations (3.21) and (2.7), the product HT estimator $\hat{\tau}_{pht}$ is more efficient than the standard ACS estimator $\hat{\tau}_{ht}$, (*i.e.* $MSE(\hat{\tau}_{pht}) < v(\hat{\tau}_{ht})$) if

$$\rho_{ht_{xy}} < -\frac{1}{2} \frac{CV_{ht_x}}{CV_{ht_y}}. \quad (3.26)$$

Relations 3.25 and 3.26 are similar to condition (2.31) which compares the product estimator with the expansion estimator. However, conditions (3.25) and (2.31) are more easily interpretable since they are directly linked to the population values of y and x . This is true since $\rho_{w_{xy}}$ measures the correlation between y and x at the network level while ρ_{xy} measures the correlation at the unit level. Similarly, CV_{w_x} and CV_{w_y} are the coefficient of variations of the population network values w_{xk} and w_{yk} while CV_x and CV_y are the coefficient of variations of the population unit values x and y .

For the product HT estimator, the correlation coefficient $\rho_{ht_{xy}}$ and the

coefficient of variations CV_{ht_y} and CV_{ht_x} depend on the network totals y_j^* and the inclusion probabilities α_j . Hence, condition (3.26) also relates to the correlation and the variation of x and y at the network level. We will now focus on $\hat{\tau}_{hh}$ but the same considerations hold also for $\hat{\tau}_{ht}$.

In (3.25), if $CV_{w_x} = 2 CV_{w_y}$, the product HH estimator will always be less efficient than the standard ACS estimator regardless the value of the correlation $\rho_{w_{xy}}$. If at the network level, y and x have the same coefficient of variation then the product HH estimator is better than the standard ACS estimator if $\rho_{w_{xy}} \leq -0.5$. As the variability of the x variable increases, more negative correlation $\rho_{w_{xy}}$ is necessary to have the product HH estimator more efficient than the standard ACS estimator. As the variability of the x variable decreases, less negative correlation $\rho_{w_{xy}}$ is necessary to have the product HH estimator more efficient than the standard ACS estimator. When, for example, $CV_{w_x} = 2 CV_{w_y}$ then regardless of the correlation $\rho_{w_{xy}}$, the product HH estimator will always be less efficient than the standard ACS estimator. When $CV_{w_y} = 2 CV_{w_x}$, then $\rho_{w_{xy}}$ has to be less than -0.25 in order to have the HH product estimator more efficient than the standard ACS estimator.

The use of the auxiliary variable would be recommended when the correlation is present at the network level because this will transfer to the value of $\rho_{w_{xy}}$. For rare and clustered populations, it would be difficult to find a situation in which $CV_{w_y} < CV_{w_x}$. The use of the ACS product estimator is recommended as long as the study variable y has a negative correlation with the auxiliary variable x at the network level and that the study variable y is rare and clustered.

Comparison with product estimator under SRSWOR

Another useful comparison is with the product estimator under SRSWOR (Murthy, 1964). Writing the MSE of $\hat{\tau}_p$ given in equation (2.30) in terms of the variances of y and x and the covariance between them, we have

$$MSE(\hat{\tau}_p) = v(\hat{\tau}_y) + R^2 v(\hat{\tau}_x) + 2 R cov(\hat{\tau}_{xy}).$$

From (3.20) we have $MSE(\hat{\tau}_{pht}) < MSE(\hat{\tau}_p)$, (*i.e.* $\hat{\tau}_{pht}$ is more efficient than $\hat{\tau}_p$) when

$$v(\hat{\tau}_{ht_y}) + R^2 v(\hat{\tau}_{ht_x}) + 2R \text{cov}(\hat{\tau}_{ht_{xy}}) < v(\hat{\tau}_y) + R^2 v(\hat{\tau}_x) + 2R \text{cov}(\hat{\tau}_{xy}). \quad (3.27)$$

or, equivalently when

$$CV_{ht_y}^2 + CV_{ht_x}^2 + 2\rho_{ht_{xy}} CV_{ht_y} CV_{ht_x} < CV^2(\hat{\tau}_y) + CV^2(\hat{\tau}_x) + 2\rho_{xy} CV(\hat{\tau}_y) CV(\hat{\tau}_x). \quad (3.28)$$

A similar condition can be obtained for the product HH estimator $\hat{\tau}_{phh}$. By recalling (3.14), we have that $MSE(\hat{\tau}_{phh}) < MSE(\hat{\tau}_p)$, (*i.e.* $\hat{\tau}_{phh}$ is more efficient than $\hat{\tau}_p$) when

$$v(\hat{\tau}_{hh_y}) + R^2 v(\hat{\tau}_{hh_x}) + 2R \text{cov}(\hat{\tau}_{hh_{xy}}) < v(\hat{\tau}_y) + R^2 v(\hat{\tau}_x) + 2R \text{cov}(\hat{\tau}_{xy}) \quad (3.29)$$

or, equivalently, by noting that $\rho_{w_{xy}} = \rho_{hh_{xy}}$ when

$$CV_{hh_y}^2 + CV_{hh_x}^2 + 2\rho_{w_{xy}} CV_{hh_y} CV_{hh_x} < CV^2(\hat{\tau}_y) + CV^2(\hat{\tau}_x) + 2\rho_{xy} CV(\hat{\tau}_x) CV(\hat{\tau}_y). \quad (3.30)$$

Thus the efficiency of the proposed product ACS estimators depends on the conjoint effects of three factors:

- The efficiency of ACS vs SRSWOR for the y variable *i.e.* $v(\hat{\tau}_{ht_y})$ vs $v(\hat{\tau}_y)$ or $v(\hat{\tau}_{hh_y})$ vs $v(\hat{\tau}_y)$
- The efficiency of ACS vs SRSWOR for the x variable *i.e.* $v(\hat{\tau}_{ht_x})$ vs $v(\hat{\tau}_x)$ or $v(\hat{\tau}_{hh_x})$ vs $v(\hat{\tau}_x)$
- The size of the correlation at the network level, $\rho_{ht_{xy}}$ or $\rho_{w_{xy}}$ and of the correlation at the unit level, ρ_{xy} .

These three factors interact with each other. Consider the situation in which the population y values are rare and clustered so to have ACS more efficient than SRSWOR. In such a case, we would have $v(\hat{\tau}_{ht_y}) < v(\hat{\tau}_y)$ or, equivalently

Unit i	Network i	α_k	y_i	w_{yk}	y_k^*	x_i	w_{xk}	x_k^*
1	1	0.4	1	1	1	7	7	7
2	2	0.4	0	0	0	8	8	8
3	3	0.4	2	2	2	6	6	6
4	4	0.7	10	505	1,010	1	3	6
4	4	0.7	1,000	505	1,010	5	3	6

Tab. 3.6: y and x populations and quantities to compute the ACS estimators

$CV_{ht_y}^2 < CV^2(\hat{\tau}_y)$. In the presence of negative correlation between y and x , the spread of the auxiliary variable may be such that the HT estimator $\hat{\tau}_{ht_x}$ is less efficient than $\hat{\tau}_x$, *i.e.* $v(\hat{\tau}_{ht_x}) > v(\hat{\tau}_x)$. The final effect on the efficiency of $\hat{\tau}_{pht}$ over $\hat{\tau}_p$ will be positive if the efficiency gain of ACS over SRSWOR for the y variable will overcome the efficiency loss of ACS over SRSWOR for the x variable.

Finally, one has to consider the values of $\rho_{ht_{xy}}$ and ρ_{xy} . Keeping the previous two factors fixed, a higher correlation at the network level than at the unit level will lead to a positive effect in terms of efficiency of $\hat{\tau}_{pht}$ over $\hat{\tau}_p$.

3.2.4 A small population example

To shed light on the computations, two small populations of size $N = 5$ were considered. The populations are shown in Table 3.6 where the condition to adaptively add adjacent units is $C = \{y : y \geq 5\}$. The example is adapted from Gattone et al. (2016b) in which the y values were obtained from (Thompson, 2012, p. 319) while the x values were generated such that the correlation at the unit level is equal to $\rho_{xy} = -0.09$ and the correlation at the network level is equal to $\rho_{w_{xy}} = -0.95$. The population totals for y and x values are 1,013 and 27 respectively.

The population values are such that condition (2.31) is not satisfied since $\rho_{xy} = -0.09$ is greater than $-\frac{1}{2} \frac{CV_x}{CV_y} = -0.11$. Hence the product estimator is not suitable for this population under SRSWOR. On the other hand, both conditions (3.25) and (3.26) are satisfied (due to the high value of correlation

at the network level) since $\rho_{w_{xy}} = -0.95$ is less than $-\frac{1}{2} \frac{CV_{wx}}{CV_{wy}} = -0.16$ and $\rho_{ht_{xy}} = -0.69$ is less than $-\frac{1}{2} \frac{CV_{ht_x}}{CV_{ht_y}} = -0.22$. Hence the product ACS estimators are useful for this population.

Table 3.7 lists the $\binom{5}{2} = 10$ possible samples of the ACS design when an initial sample of size $n = 2$ is selected by SRSWOR. The estimators $\hat{\tau}_y, \hat{\tau}_x, \hat{\tau}_{hh_y}, \hat{\tau}_{hh_x}, \hat{\tau}_{ht_y}, \hat{\tau}_{ht_x}, \hat{\tau}_{phh}$ and $\hat{\tau}_{pht}$ were computed for each possible sample. The table confirms the above analysis that the τ_p is inefficient relative to SRSWOR for this population while both τ_{pht} and τ_{phh} are efficient relative to SRSWOR. From the table, the most efficient estimator of τ_y is the HH product estimator $\hat{\tau}_{phh}$.

Samples	$\hat{\tau}_y$	$\hat{\tau}_{h_y}$	$\hat{\tau}_{ht_y}$	$\hat{\tau}_x$	$\hat{\tau}_{h_x}$	$\hat{\tau}_{ht_x}$	$\hat{\tau}_p$	$\hat{\tau}_{phh}$	$\hat{\tau}_{pht}$
1,2	2.50	2.50	2.50	37.50	37.50	37.50	3.47	3.47	3.47
1,3	7.50	7.50	7.50	32.50	32.50	32.50	9.03	9.03	9.03
1,4;3,5	27.50	1265.00	1445.36	20.00	25.00	26.07	20.37	1171.30	1395.65
1,5;4,2	2502.50	1265.00	1445.36	30.00	25.00	26.07	2780.56	1171.30	1395.65
2,3	5.00	5.00	5.00	35.00	35.00	35.00	6.48	6.48	6.48
2,4;3,5	25.00	1262.50	1442.86	22.50	27.50	28.57	20.83	1285.88	1526.83
2,5;4,3	2500.00	1262.50	1442.86	32.50	27.50	28.57	3009.26	1285.88	1526.83
3,4;5	30.00	1267.50	1447.86	17.50	22.50	23.57	19.44	1056.25	1264.00
3,5;4	2505.00	1267.50	1447.86	27.50	22.50	23.57	2551.39	1056.25	1264.00
4,5;3	2525.00	2525.00	1442.86	15.00	15.00	8.57	1402.78	1402.78	458.05
$\hat{\tau}$	1,013.00	1,013.00	1,013.00	27.00	27.00	27.00	982.36	844.86	885.00
Bias	0.00	0.00	0.00	0.00	0.00	0.00	-30.64	-168.14	-128.00
MSE	1,490,383.50	571,539.75	435,460.29	54.75	39.75	57.43	1,562,491.78	339,489.59	429,504.93

Tab. 3.7: All possible samples with their estimates. Units before the semicolon are the initial units while units after the semicolon are the adaptively added units.

3.2.5 Simulation study

Artificial populations were generated on a square region that was partitioned into $30 \times 30 = N = 900$ units. The study variables y were generated according to a PCP (Diggle, 2013, p. 101). In the PCP, first, the number of parents (nop) are generated from a Poisson process with mean λ_1 and parents are randomly distributed within the study region. For each parent, the number of offsprings are generated according to a Poisson process with mean λ_2 . The variation of the location of the offsprings relative to their parents is controlled by the parameter $s2$. The parameters setting were $\lambda_1 \in \{5, 10\}$, $\lambda_2 \in \{10, 20, 30, 40, 50, 60, 80, 100, 120, 150\}$ and $s2 \in \{0.00025, 0.0005\}$. The variables were simulated using **R** package **Splancs** (Rowlingson and Diggle, 2015).

The auxiliary variables x were simulated such that three levels of correlations existed with the study variable at the network level; low ($-0.3 < \rho_{w_{xy}} \leq 0$), intermediate ($-0.6 < \rho_{w_{xy}} \leq -0.3$), and high ($-0.9 < \rho_{w_{xy}} \leq -0.6$). The auxiliary variables were simulated using **R** package **Ecodist** (Goslee and Urban, 2007) with some modifications.

The combination of the above parameters enabled the examination of the proposed estimators on populations with different degrees of rarity and clusteredness; from rare and clustered populations to less rare and less clustered populations and different levels of network correlation between x and y . Some of the simulated population are shown on figure 3.6.

The initial sample size under the ACS design were $n \in \{15, 65, 100\}$. Under SRSWOR, the size of the sample was set to the expected final sample size of the adaptive design. For each population, $R = 50,000$ samples were drawn and the values of the estimators $\hat{\tau}_{pht}$, $\hat{\tau}_{phh}$, $\hat{\tau}_{ht}$, $\hat{\tau}_{hh}$ and $\hat{\tau}_p$ computed.

Evaluation of the estimators was done using the relative efficiency and the relative bias which were computed as follows:

- Relative efficiency of the product ACS estimators

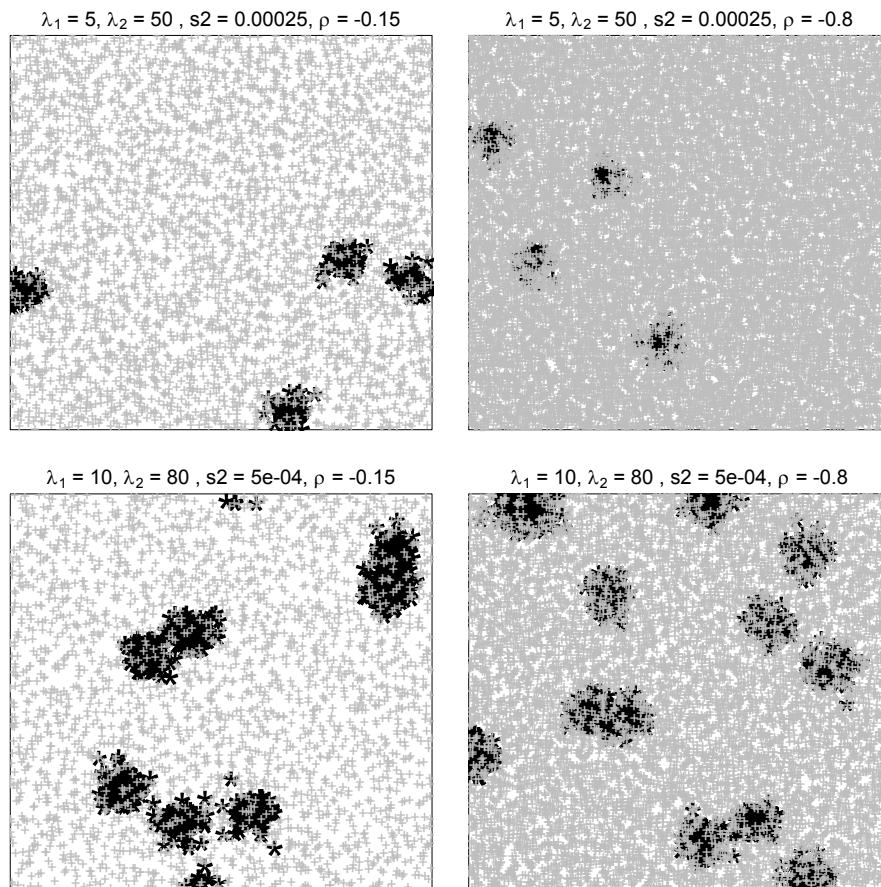


Fig. 3.6: Population distribution of the study (black *) and the auxiliary variable (grey +) when clusters are more (first row) and less (second row) compact and when the level of correlation is low (first column) and high (second column).

$$RE_{pht} = \frac{MSE(\hat{\tau})}{MSE(\hat{\tau}_{pht})} \text{ and } RE_{phh} = \frac{MSE(\hat{\tau})}{MSE(\hat{\tau}_{phh})}$$

- Relative efficiency of standard ACS estimators

$$RE_{ht} = \frac{MSE(\hat{\tau})}{MSE(\hat{\tau}_{ht})} \text{ and } RE_{hh} = \frac{MSE(\hat{\tau})}{MSE(\hat{\tau}_{hh})}$$

- Relative efficiency of the product estimator under SRSWOR

$$RE_p = \frac{MSE(\hat{\tau})}{MSE(\hat{\tau}_p)}.$$

- Relative bias of the product ACS estimators

$$RB_{pht} = 100 \times \frac{1}{R} \sum_{i=1}^R \frac{\hat{\tau}_{pht_i} - \tau_y}{\tau_y},$$

$$RB_{phh} = 100 \times \frac{1}{R} \sum_{i=1}^R \frac{\hat{\tau}_{phh_i} - \tau_y}{\tau_y}$$

- Relative bias of the product estimator under SRSWOR

$$RB_p = 100 \times \frac{1}{R} \sum_{i=1}^R \frac{\hat{\tau}_{p_i} - \tau_y}{\tau_y}.$$

Results of the relative efficiency under the product HT estimator when the initial sample size is $n = 15$ and when the correlation between the study and auxiliary variable is high are presented in figure 3.7. From the figure, the product HT has higher efficiency gain compared to the standard HT estimator. The gain in efficiency is higher for populations in which the number of clusters are few ($nop = 5$) and the clusters are more compact ($s2 = 0.00025$). When the number of clusters increases but clusters remain more compact, higher gain in efficiency of the product HT estimator is achieved in populations in which the average number of units in a network is less than 80. Populations with less compact clusters ($s2 = 0.0005$) have less gain in efficiency of the product ACS relative to the standard ACS design compared to populations with more compact clusters.

Figure (3.8) shows results of the product HH estimator. The figure shows that the results are similar to the results under the HT estimator. However the product HT estimator has higher efficiency gain compared to the product HH estimator. This is shown in appendix (A.3).

Figure (A.1) in the appendix shows the relative efficiency of the HT estimator for the different levels of correlation coefficient. The figures shows that the gain in efficiency of the product HT over the standard HT increases with increase

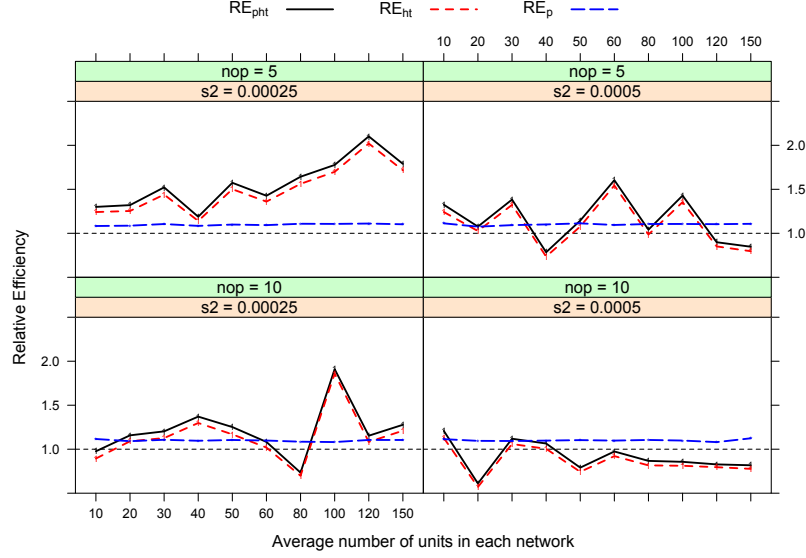


Fig. 3.7: Relative efficiency when HT is used for an initial sample of size $n = 15$ for the different parameters when $\rho_{w_{xy}}$ is highly negative

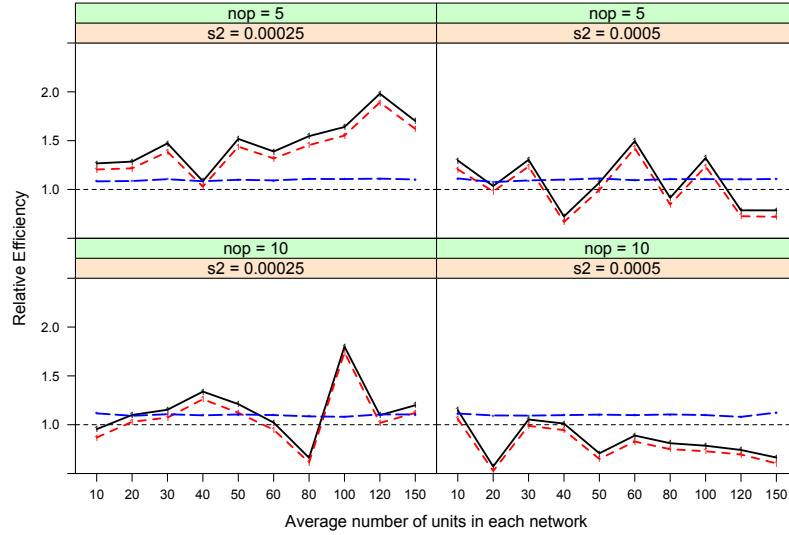


Fig. 3.8: Relative efficiency from the product HH for an initial sample of size $n = 15$ for the different parameters when $\rho_{w_{xy}}$ is highly negative

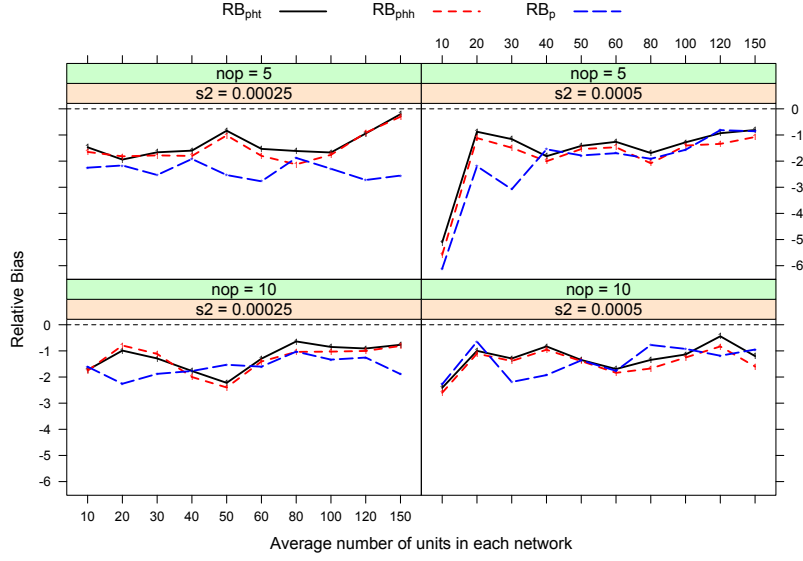


Fig. 3.9: RB of the proposed product HT (left) and HH (right) for low (first row), intermediary (second row) and high (third row) correlation; highly rare ($nop = 5$) and less rare ($nop = 10$) and from more compact to less compact clusters (from first to third column).

in the level of correlation coefficient. The same behavior can be seen with the product HH estimator as shown in appendix (A.2).

Results of the relative bias of the product under ACS (HT and HH) and the product under SRSWOR when the level of correlation between the study and the auxiliary variable is high for an initial sample of size $n = 15$ are presented in figure (3.9). The figure shows that, in general, the product ACS has lower bias compared to the product SRSWOR when the populations under study have few clusters ($nop = 5$). Figure (A.4) in the appendix shows the relative bias for the different levels of correlation coefficient when the initial sample is of size $n = 15$. The figure shows that the relative bias of the product ACS increases with the increase in the level of negative correlation. On the other hand, the relative bias decreases with increase in initial sample size (figure A.5).

Variable	Unit level		Network level	
	cv.r	ρ_{xy}	cv.r	$\rho_{w_{xy}}$
Oryx and temp	-0.0014	-0.0432	-0.0021	-0.0582
Wildbeest and temp	-0.0014	-0.0110	-0.0031	-0.0236

Tab. 3.8: Values on the left (correlation coefficient) and the right (cv.r) hand side of equations (2.31) and (3.25)

3.2.6 Simulations using real data

The performance of the proposed estimators was evaluated by performing simulations using two real data sets. The data sets for the study variable were the distribution and counts of oryx and wildbeest across Amboseli-West Kilimanjaro/Magadi-Natron cross border landscape. The landscape covers parts of Kenya and Tanzania between $1^{\circ} 37' S$ and $3^{\circ} 13' S$ and between $35^{\circ} 49' E$ and $38^{\circ} 00' E$ covering an area of $25,623 km^2$. The data were obtained from an aerial survey that was conducted in October (dry season) 2013. Details on how the survey was conducted can be found in Kenana et al. (2013b).

A grid was laid on maps that were produced from the survey and effort was made to determine the counts in each unit. The region was partitioned into $N = 2,304$ units such that there are 48 row and 48 columns. Temperature was the auxiliary data that was considered for each quadrant over the study area. This data was obtained from **R** package **raster** (Hijmans, 2015).

The data sets were verified to see if they meet conditions (2.31), (3.25) and (3.26) that are necessary for the efficiency of the product estimator under both adaptive and non-adaptive design. Table 3.8 presents values of the left and right hand side of the equations. From the table, both conditions are satisfied for both data sets although the correlation is very weak. Gattone et al. (2016b) presented data from a similar survey that was conducted in the year 2010 which had relatively higher negative correlation.

Table 3.9 presents results of the relative efficiency and the relative bias.

	n	RE_{pht}	RE_{ht}	RE_{phh}	RE_{hh}	RE_p	RB_{pht}	RB_{phh}	RB_p
Oryx	30	2.09	2.08	1.93	1.92	1.01	-1.16	-0.61	0.26
	50	2.03	2.02	1.87	1.86	1.01	-0.43	-0.50	1.80
	100	2.18	2.17	1.82	1.81	1.00	-1.10	-1.19	-0.07
	150	2.39	2.38	1.78	1.77	1.00	-0.06	-0.03	-1.04
	200	2.61	2.61	1.79	1.78	1.00	-0.01	-0.18	-0.60
Wildbeest	30	3.06	3.05	2.55	2.53	1.00	-0.01	0.02	0.39
	50	2.99	2.99	2.20	2.19	1.00	-0.20	-0.23	-0.96
	100	4.49	4.50	2.34	2.33	1.00	0.13	0.35	0.61
	150	6.55	6.59	2.38	2.37	1.00	-0.06	0.06	-0.14
	200	9.83	9.90	2.49	2.48	1.00	0.07	0.46	-0.81

Tab. 3.9: Results of the relative efficiency (RE) and relative bias (RB) for different initial sample sizes on the real data

The subscript p , ht and hh represent the product, HT and HH estimators respectively. From the table, the proposed method has contributed to a slight gain in efficiency compared to the standard ACS design.

3.2.7 Discussion

The simulations results are based on simulated data assuming varying network level correlations. Network level correlation does not guarantee unit level correlation, nor is unit level negative correlation required for the product estimator under ACS. Thus if unit level correlation is very close to zero and this is not the case on a network level then the ACS product estimators may be appropriate whereas the SRSWOR product estimators are, in this case, not appropriate. On the other hand if there exists a negative unit level correlation and a positive network correlation, then the ACS product estimator is not appropriate and the researcher should not use it. This can happen if the auxiliary variable is rare and clustered and often found together with the variable of interest. In this case it could be possible for the two variables to be negatively correlated on a unit level but be positively correlated on a network level. The researcher should be cautious of this scenario even if it is unlikely to

occur. Ideally, the researcher would have an estimate or at least a concept for the network level correlation from a prior study on the same attributes of interest to determine if the product ACS estimator is the best estimator to use. We note the product estimator with HH estimator under the ACS design that was done by Chaudhury and Hanif (2015) and Chutiman et al. (2013).

3.3 ACS with clusters selected without replacement and stopping rule

3.3.1 Introduction

The current chapter extends the theoretical understanding of ACS by considering the conjoint use of the without replacement selection of clusters and a stopping rule in which the adaptive sampling phase is stopped whenever the stopping rule is verified. We refer to the new design as ACS^{c-sr}. The design limits further the sampling effort in form of traveling expenses by avoiding repeat observations (Dryver and Thompson, 2007) and by reducing the final sample size (Gattone and Di Battista, 2011).

The focus is on the data-driven stopping rule that proposed by Gattone and Di Battista (2011) to control the final sample size when prior knowledge of the population structure is not available. The stopping rule alleviates the lack of prior information and enlarges situations in which ACS could be more efficient than conventional sampling designs (Gattone and Di Battista, 2011). However, it introduces bias into the estimators.

Since samples are drawn one at a time, the Des Raj estimator is modified to suit to the proposed design as shown in sub-section (3.3.2). A Monte-Carlo simulation study using artificial data is given in sub-section (3.3.4). Simulations to evaluate the performance of the proposed estimators using real wildlife data is given in sub-section (3.3.5). Sub-section (3.3.6) provides some comments on the results and suggests openings for future research.

3.3.2 The modified Raj's estimator

Following the notations of Gattone et al. (2016a), let Ψ_i^w be the i -th truncated network whose size is m_i^w and let Ψ_i^c be the i -th truncated cluster. An i -th truncated cluster contains an i -th truncated network and the stopping edge units. The stopping edge units are units that activate the stopping rule (Gattone and Di Battista, 2011).

The modified Raj estimator for the population total is given by

$$\hat{\tau}^{c-sr} = \frac{1}{n} \sum_{i=1}^n t_i^{c-sr} \quad (3.31)$$

where

$$t_1^{c-sr} = \frac{\hat{y}_{1.}^w}{p_1^*} \quad (3.32)$$

and for $i = 2, \dots, n$

$$t_i^{c-sr} = \sum_{j=1}^{i-1} \hat{y}_{j.} + \frac{\hat{y}_{i.}^w}{p_i^*} \left(1 - \sum_{j=1}^{i-1} p_j^* \right) \quad (3.33)$$

with

$$\begin{aligned} p_i^* &= \frac{m_i^w}{N} \\ \hat{y}_{i.}^w &= \sum_{j \in \Psi_i^w} y_j \\ \hat{y}_{j.} &= \frac{\sum_{j \in \Psi_i^c} y_j}{f_j} \end{aligned}$$

where $\hat{y}_{i.}^w$ is the sum of the units in the i -th truncated network, $\hat{y}_{j.}$ is an estimate of the total in network j and $f_j = \frac{m_j^w}{m_j^c}$.

The variance can be approximated by

$$\text{var}(\hat{\tau}^{c-sr}) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(t_i^{c-sr}) + V \quad (3.34)$$

that can be estimated by

$$\widehat{\text{var}}(\hat{\tau}^{c-sr}) = \frac{1}{n(n-1)} \sum_{i=1}^n (t_i^{c-sr} - \hat{\tau}^{c-sr})^2 + \hat{V} \quad (3.35)$$

with

$$\hat{V} = \sum_{i=1}^n c_i \quad (3.36)$$

in which

$$c_1 = z_1 \quad (3.37)$$

and for $i = 1, \dots, n$

$$c_i = \sum_{j=1}^{i-1} z_j + \frac{z_i}{p_i^*} \left(1 - \sum_{j=1}^{i-1} p_j^* \right) \quad (3.38)$$

where

$$z_i = \frac{1}{m_i^c - 1} \sum_{j \in \Psi_i^c} (y_j - \bar{y}_i^c)^2 \quad (3.39)$$

with

$$\bar{y}_i^c = \frac{1}{m_i^c} \sum_{j \in \Psi_i^c} y_j. \quad (3.40)$$

\hat{V} is the Des Raj's estimator of the quantity $\sum_{i=1}^N \sigma_i^2$ where σ_i^2 is the within-network variance of the i -th network. The estimate of the variance is the sum of two parts: the first part is equal to the estimate of the variance calculated on the assumption that the network totals y_j have been measured without error while the second part is obtained from the networks total estimate itself by substituting the estimated variances for the estimates of the totals of the networks (Gattone et al., 2016a).

3.3.3 A small population example

In this sub-section, the design and estimation are demonstrated with a small artificial population that is given in table (3.10). Let the initial sample be of size $n = 2$ and the condition to adaptively add neighboring units be $C = \{y : y > 0\}$.

0	0	0	0	0
2	0	0	0	0
10	0	0	③	1
22	1	0	13	10
2	0	0	3	②

Tab. 3.10: A small artificial population

Let the first unit to be selected be the unit whose value is 3 (the circled

unit). At the first step of sampling, units with values 13 and 1 are included in the sample. These units together with the initial unit are colored light gray. At the second step of sampling, neighboring units with values $\{0, 0, 0, 0, 3, 10\}$ are added to the sample. These units are colored dark gray. The second step units whose values are zero are the edge units.

Applying condition (2.18), we note that

$$\begin{aligned} s_1^{2(1)} &= \frac{\sum_{j \in \Psi_i^w} (y_j - w_1^{(1)})^2}{m_1^{(1)}} \\ &= \frac{1}{3}((3 - 5.67)^2 + (13 - 5.67)^2 + (1 - 5.67)^2) \\ &= 27.56 \end{aligned}$$

where $w_1^{(1)}$ is the mean of values in the network associated with the first unit selected after the first step of sampling. $s_1^{2(2)}$ is calculated in a similar manner. Hence, we have

$$\frac{s_1^{2(2)}}{s_1^{2(1)}} \times \frac{1 - \frac{1}{m_1^{(1)}}}{1 - \frac{1}{m_1^{(2)}}} = \frac{20.89}{27.56} \times \frac{1 - \frac{1}{3}}{1 - \frac{1}{5}} = 0.5685 < 1 \quad (3.41)$$

hence the sampling stops at this step and units added at the second step whose values meet the condition C form the stopping edge units. In table (3.10), the stopping rule edge units are the units whose values are $\{3, 10\}$. Units selected up to the first step of sampling form a truncated network (*i.e* units whose cells are colored light gray).

Then the network together with its edge units and stopping rule edge units are "removed" from the population and a second unit is selected at random from the remaining units. Suppose the second unit selected is the second circled unit with value 2. Though the unit meets the condition of interest, note that the neighboring units had already been observed and "removed" from the population. Thus the second unit sampled is a network of size one with no edge units.

To obtain the estimate of the population total, we proceed as follows: the

number of units in the truncated networks and clusters are

$$m_1^w = 3, m_2^w = 1, m_1^c = 9, m_2^c = 1$$

and f_j is

$$f_1 = \frac{3}{9} = 0.3333, f_2 = \frac{1}{1} = 1.$$

The probability of drawing the first and the second networks are

$$p_1^* = \frac{3}{25} = 0.12, p_2^* = \frac{1}{25} = 0.04$$

while the y -values in the truncated networks and clusters are

$$y_1^w = \{3, 1, 13\}, y_2^w = \{2\}, y_1^c = \{3, 1, 13, 10, 3, 0, 0, 0, 0\}, y_2^c = \{2\}.$$

Hence the estimate of y -values in the truncated network \hat{y}_j is

$$\hat{y}_{1.} = \frac{30}{0.3333} = 90.009, \hat{y}_{2.} = \frac{2}{1} = 2.$$

Finally t_i^{c-sr} is

$$t_1^{c-sr} = \frac{17}{0.12} = 141.6667, t_2^{c-sr} = 90.009 + \frac{2}{0.04}(1 - 0.12) = 134.009.$$

Thus the modified Des Raj estimate of the population total based on the selected sample is given by

$$\hat{\tau}^{c-sr} = \frac{141.6667 + 134.009}{2} = 137.8379$$

3.3.4 Simulation study

Monte-Carlo simulation of the proposed design was performed using a range of rare and clustered artificial populations that were generated using PCP (Diggle, 2013, p. 101). In the PCP, first, the number of parents are

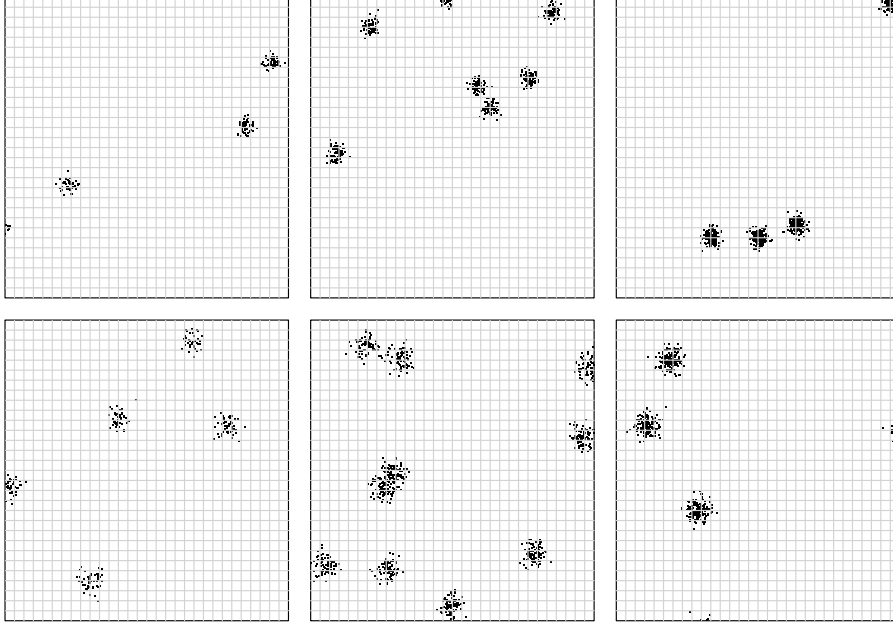


Fig. 3.10: Simulated population with first row $s = 0.00025$ and second row $s = 0.0005$. First, second and third columns are for $\lambda_2 = 50, 80$ and 150 respectively.

generated as a realization of a Poisson process with mean λ_1 . Parents are then randomly distributed within the region of interest. Each parent has an average number of offsprings that are simulated according to a Poisson process with mean λ_2 . The variation of the locations of offsprings relative to their parents is controlled by a parameter s .

The populations were generated in **R** package **splancs** (Rowlingson and Diggle, 2015) whose parameters setting were $\lambda_1 = 5$, $\lambda_2 \in \{50, 60, 80, 100, 120, 150\}$ and $s \in \{0.00025, 0.0005\}$. The above populations were generated in a region that was partitioned into $N = 30 \times 30 = 900$ units. Some of the simulated populations are shown on figure 3.10.

A varied initial samples of sizes $n \in \{20, 30, 50, 100, 120\}$ were considered and 10,000 sampling simulations were performed on each population. For the adaptive designs, the critical value was $C = \{y : y > 0\}$. The designs considered for comparison were $\text{ACS}^{\text{c-sr}}$, ACS^{c} and stratified ACS. The stratified ACS had

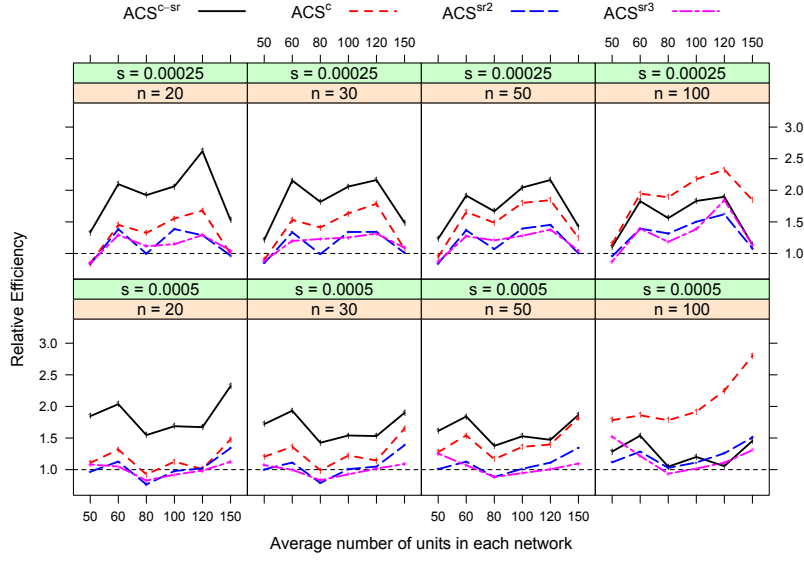


Fig. 3.11: Results of the relative efficiency for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.

the region partitioned into two and three strata with networks truncated at stratum boundaries. This is to achieve the limitation of the final sample size.

Evaluation was based on the relative efficiency of ACS^{c-sr} and ACS^c versus SRSWOR and on stratified ACS versus the classical stratified sampling. The expected final sample size v was used to evaluate estimates under the non-adaptive designs. The final sampling fraction $fn = \frac{v}{N}$ of the ACS designs was also evaluated. Finally, since the use of the stopping rule results in biased estimates (Gattone and Di Battista, 2011; Su and Quinn II, 2003), the relative bias under ACS^{c-sr} design was evaluated.

Figure (3.11) presents results of the relative efficiency. From the figure, the proposed design helped in achieving gains in efficiency compared to the other designs considered in this study. This gain in efficiency was observed across all the populations considered with different cluster sizes and different cluster compactness and for smaller initial sample sizes ($n < 100$).

Results of the expected final sampling fraction are shown in figure (3.12).

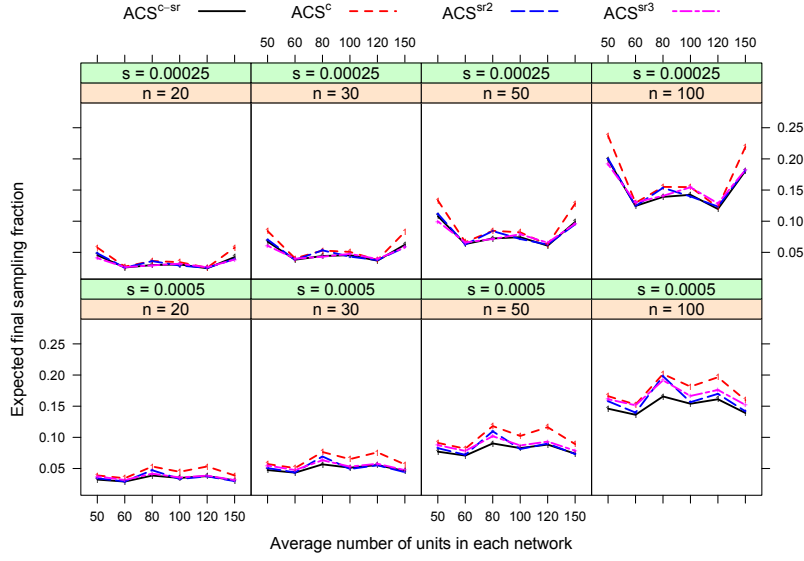


Fig. 3.12: Results of the expected final sampling fraction for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.

From the figure, the proposed design achieved the reduction in the final sampling fraction, for all populations considered, compared to the ACS^c design. The comparison in the reduction of the final sampling fraction between the proposed design ACS^{c-sr} and the stratified ACS are similar especially for smaller initial sample sizes $n < 50$. However, the proposed design reduced the final sample size more when clusters are less compact and higher initial sample size used.

All populations considered had values of the absolute bias that was less than 25% (figure 3.13). The relative bias however increases with increase in the initial sample size.

3.3.5 Simulations using real data

The proposed design was applied to three real data sets which were the distribution and counts of hartebeest, elephants and oryx across Amboseli-West Kilimanjaro/Magadi-Natron cross border landscape. The landscape covers parts of Kenya and Tanzania between $1^\circ 37' S$ and $3^\circ 13' S$ and between $35^\circ 49' E$

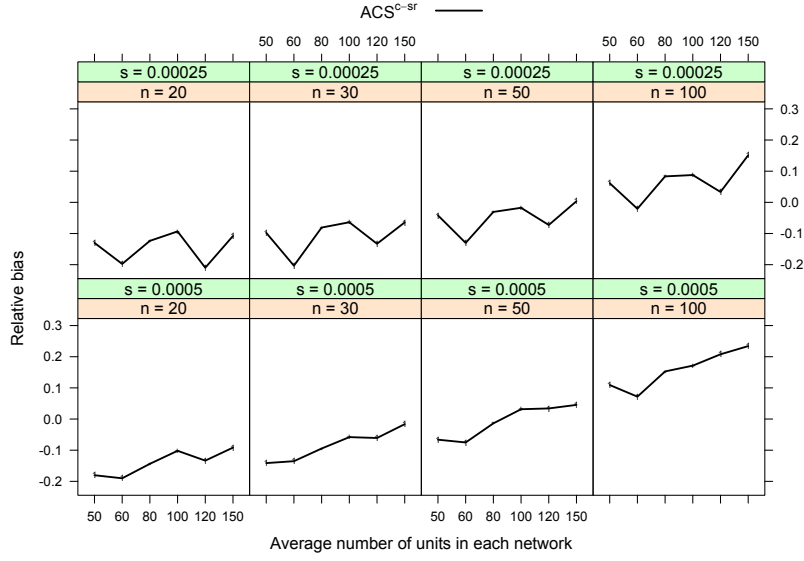
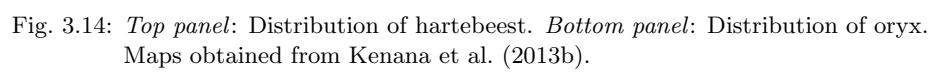


Fig. 3.13: Results of the relative bias for different initial sample sizes (across columns) for more and less compact clusters (first and second rows respectively). The different number of units in each network are represented on the x-axis.

and $38^{\circ} 00' E$ covering an area of $25,623 \text{ km}^2$. The data was obtained from an aerial survey that was conducted in October (dry season) 2013. Details on how the survey was conducted can be found in Kenana et al. (2013b). Figure 3.14 presents the distribution of two of the populations studied.

A grid was laid on maps that were produced from the survey as shown in the figure and an effort was made to determine the counts in each unit. The region had $N = 2,304$ units such that there were 48 rows and 48 columns. A range of initial sample sizes ($n \in \{30, 50, 100, 120, 150\}$) were used and for each population 10,000 sampling simulations were done.

Figure 3.15 presents results of the relative efficiency for the three data sets with different initial sample sizes. From the figure, the proposed design results in gain in efficiency across the populations when the initial sample size is small. Increasing the initial sample size decreases the gain in efficiency under the proposed design. From the populations and the designs considered, the ACS^c results in higher gain in efficiency if the initial sample size is increased.



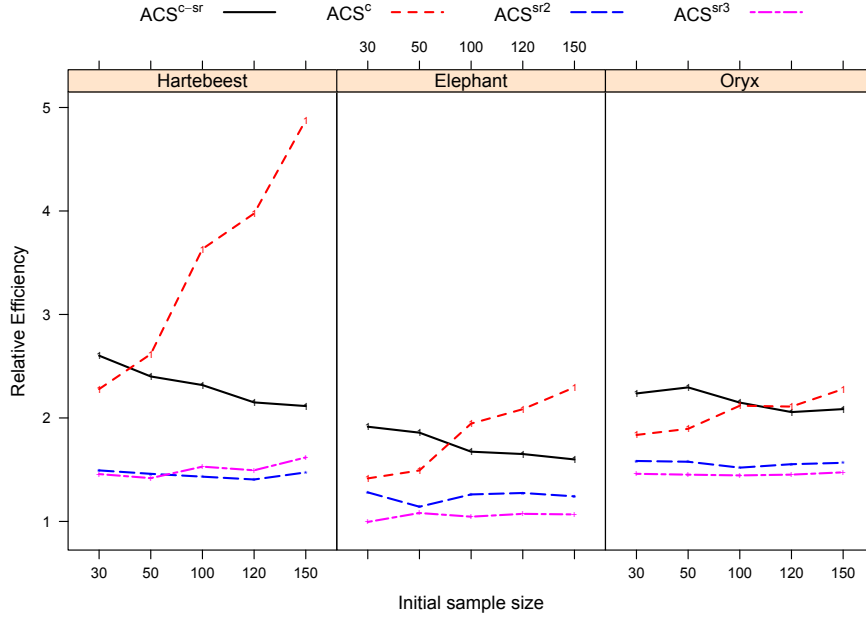


Fig. 3.15: Results of the relative efficiency

These results are in agreement with results of the artificial populations that were presented in sub-section (3.3.4).

Results on the expected final sampling fraction are shown on figure 3.16. The figure shows that the proposed design has helped in limiting the expected final sampling fraction relative to the ACS^C design across the three real populations considered. In comparison to the stratified ACS in which networks are truncated at stratum boundaries, the reduction in the expected final sampling fraction depends on the presence of networks that cross over stratum boundaries. When the hartbeest data is partitioned into three strata, the largest network crosses over two strata hence the stratification with three strata resulted in the lowest expected final sampling fraction. This was not observed in the elephant and oryx data.

Since ACS^{C-sr} is biased (Gattone et al., 2016a), table 3.11 presents results of its relative bias for the three real data sets. The table shows that the absolute bias decreases as the sample size increases for elephant and oryx data. The opposite was observed with hartbeest data.

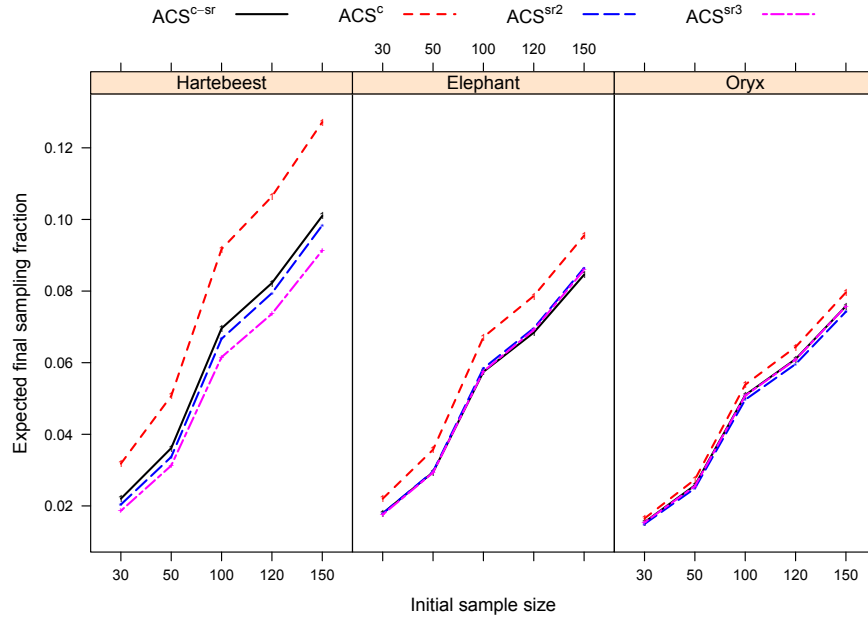


Fig. 3.16: Results of the expected final sampling fraction.

n	Hartebeest	Elephant	Oryx
30	-0.039	-0.142	-0.113
50	0.009	-0.110	-0.073
100	0.092	-0.022	-0.031
120	0.116	0.015	-0.020
150	0.151	0.051	-0.015

Tab. 3.11: Results of the relative bias under ACS^{c-sr} design

3.3.6 Discussion

A good sampling design is one that has lower mean squared error and lower sampling cost among other factors. The current chapter aimed at achieving these two factors. The proposed design with its proposed estimators helped in improving the gains in efficiency as well as reducing the final sampling fraction. The rate of improvement depends on the combination of the degree of rarity and the cluster compactness which are general factors that determines efficiency under the ACS design.

Any variation of ACS to be considered to limit the sampling effort requires the specification of some particular value. This specification requires prior knowledge of the population characteristics (Gattone and Di Battista, 2011; Gattone et al., 2016a). The proposed design achieved this limitation of the sampling effort while increasing gains in efficiency without the need for prior knowledge of the population characteristics.

As opposed to design-based sampling in general, the size of the initial sample had a negative effect on the gain in efficiency under the proposed design. In that, the efficiency decreases as the initial sample size increases; this could partly be contributed by the relative bias. Although the absolute relative bias under both simulated and real data was below 25%, this bias increases with increase in the initial sample size under the artificial populations considered. With the real data, this same scenario was observed with hartbeest data. The elephant and oryx data had their biases decrease as the initial sample size increases.

Obtaining the bias analytically may not be direct since the truncated network formed depends on the initial unit sampled (Gattone et al., 2016a). This and the order of the bias however, deserves further investigation as it may help in understanding the effect of the increase in the initial sample size together with the population characteristics on the efficiency of the proposed estimator.

4. Model-based estimation for rare and clustered population with application to prediction of population total of African elephants

4.1 Introduction

Rapley and Welsh (2008) modeled data collected through ACS, with an initial sample selected by SRSWOR, in which networks are sampled without replacement. Following their work, the current chapter models data collected through stratified strip ACS design (that was described in section 3.1). The sample selection mechanism is that of clusters selected without replacement since it results in much lower sampling costs compared to sampling without replacement of networks (Dryver and Thompson, 2007). The model is modified by including auxiliary variable with an aim of improving the bias and the MSE.

In ACS studies that make use of auxiliary variable, Chao (2004b) set the condition to adaptively add neighboring units to depend on the auxiliary variable. In this study, one critical assumption is that the auxiliary variable (that is known for all units in the population) has a high correlation with the study variable at the network level. That is where the auxiliary variable is more concentrated, the study variable is also more concentrated such that $C = \{y_i : y_i > c\} \approx C^* = \{x_i : x_i > d\}$. Such an approximation assumption has also been used by Salehi et al. (2015).

In this chapter, the interest is to predict the population total of African elephants. Notations are described in section (4.2). Definition of the informative and the non-informative sampling and its importance to the present chapter is given in section (4.3). The proposed modified model is given in section (4.4). Fitting of the model using the Bayesian approach is given in section (4.5). Model-based and design-based simulations are given in sections (4.6) and (4.7) respectively. Further comments are finally given in section (4.8).

4.2 Notations

Let there be $q = 1, \dots, H$ strata. The data sampled in each stratum is modeled separately and inference performed in each stratum. The predicted population total is obtained by summing the predicted totals in each stratum. For simplification of notations, the model presented below is for one stratum and the same applies to the remaining strata.

Let the number of PSU be L such that each PSU has M SSU. The total number of SSU in the study region is $N = LM$. It is assumed that all PSU have the same number of SSU *i.e* all strips are of equal length. Denote the number of units that meet the condition of interest (the number of non-empty units in this case) by N_y , the number of non-empty networks by K and the number of units in each non-empty network by M_k such that $N_y = \sum_{k=1}^K M_k$. Since an empty unit is a network of size one (Thompson, 2012, p. 319), there are $N - N_y + K = A$ networks in the study region. Let the sum of the values of the study and the auxiliary variables in each network be denoted by Y_k and X_k respectively. Then $Y = \sum_{k=1}^K Y_k$ and $X = \sum_{k=1}^K X_k$ are the total values for the study and the auxiliary variables in the study region. The objective is to predict the population total of the variable of interest in the study region. This objective can be represented as

$$\hat{Y} = \sum_{k \in s} y_k + \sum_{k \notin s} \hat{y}_k$$

4.3 Sample selection probability

A sampling mechanism is said to be non-informative about the parameter Θ if (Valliant et al., 2000, p. 39; Chambers and Clark, 2012, p. 10)

$$p(s|Y, X; \Theta) = p(s|X; \Theta)$$

otherwise, it is informative *i.e* a sampling mechanism is non-informative about the finite population parameter Θ if the rule for selecting units in the sample depends on the auxiliary variable. When the rule for selecting units in the sample depends on the variable of interest, then the sampling mechanism is said to be informative.

The ACS design is informative since units to be include in the sample depend on the observed variable of interest (Gabler, 2007) *i.e* (Thompson and Seber, 1996, p. 52)

$$p(s|Y; \Theta) \neq p(s)$$

Under informative sampling, to make inference on the population parameters using the sampled data, the joint distribution of the model used to generate the population data and the sample selection process is used. This is followed by integrating (in the case of continuous outcomes) or summing (for discrete outcomes) over units that are not in the sample (Chambers et al., 2012, p. 27; Valliant et al., 2000, p. 39; Pfeffermann et al. (1998)).

With the above in mind and suppose that $s = s_1, \dots, s_n$ is the ordered sample, then the conditional probability of selecting the ordered sample is

$$p(s|N_y, K, M_k) = \prod_{i=1}^n \frac{M_{s_i}}{\sum_{j=1}^{N-n_y+\kappa} z_j - \sum_{k=0}^{i-1} (z_{s_k} + e_{s_k})} \quad (4.1)$$

where M are the number of SSU in a selected strip, z is the size of the j -th network and e represents the number of edge units. The denominator represents the sum of the sizes of all networks in the population minus the sum of the sizes of the networks and the edge units (or the sizes of clusters) that are sampled up to the previous step of sampling. Note that a unit that does not meet the condition of interest is a network of size one (Thompson, 2012, p. 319).

4.4 The model

The population model under the ACS design that makes use of the auxiliary information is presented below. The model takes the form

$$N_y | \alpha \sim \text{Truncated binomial}(N, \alpha), N_y = 1, \dots, N \quad (4.2a)$$

$$K | N_y, \theta \sim \text{Truncated binomial}(n_y, \theta), K = 1, \dots, n_y \quad (4.2b)$$

$$M_k | N_y, K \sim \mathbf{1}_\kappa + \text{Multinomial}(n_y - \kappa, \frac{1}{\kappa} \mathbf{1}_\kappa),$$

$$M_k = 1, \dots, n_y - \kappa, \sum_{k=1}^{\kappa} M_k = n_y \quad (4.2c)$$

$$Y_k | X_k, K, M_k, \lambda \sim \text{independent truncated Poisson}(\lambda) \quad (4.2d)$$

where $\lambda = m_k \exp(\beta_0 + \beta_1 X_k)$, with $m_k, k = 1, \dots, \kappa$ an offset. The distributions given in equations (4.2a) - (4.2c) are the same as those from the model provided by Rapley and Welsh (2008). The distribution in equation (4.2d) is the modified version of the model of Rapley and Welsh (2008).

Let n be the size of the sample and s the ordered sample s_1, \dots, s_n . Writing the variables in form of the sampled and the unsampled components, we have $N_y = (N_{y0}, N_{y1})$, $K = (K_0, K_1)$, $M_k = (M_{k0}, M_{k1})$, $Y_k = (Y_{k0}, Y_{k1})$ and $X_k = (X_{k0}, X_{k1})$ in which the sampled components are represented by subscript 0 and the unsampled components represented by subscript 1. Further denote the observed data variables as $D_0 = \{N_{y0}, K_0, M_{k0}, Y_{k0}, X_{k0}\}$ while the unobserved as $D_1 = \{N_{y1}, K_1, M_{k1}, Y_{k1}, X_{k1}\}$ so that $D = \{D_0, D_1\}$. The vector of parameters is denoted by $\Theta = \{\alpha, \theta, \lambda\}$. The likelihood for the data which is the joint density of s and D conditional Θ is

$$f(s, D | \Theta) = p(s | N_y, K, M_k) f(D | \Theta) \quad (4.3)$$

which is given by

$$\begin{aligned}
P(s, N_y, K, M_k, X_k, Y_k | \alpha, \theta, \lambda) = \\
p(s | N_y, K, M_k) \times \binom{N}{n_y} \left(\frac{\alpha^{n_y} (1 - \alpha)^{N - n_y}}{1 - (1 - \alpha)^N} \right) \times \\
\binom{n_y}{\kappa} \left(\frac{\theta^\kappa (1 - \theta)^{n_y - \kappa}}{1 - (1 - \theta)^{n_y}} \right) \times \left((n_y - \kappa)! \prod_{k=1}^{\kappa} \frac{1}{(m_k - 1)!} \left(\frac{1}{\kappa} \right)^{m_k - 1} \right) \times \\
\prod_{k=1}^{\kappa} \left(\frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \right) \quad (4.4)
\end{aligned}$$

The joint distribution of the observed variables D_0 is obtained by summing equation (4.3) or equation (4.4) over the non-sampled variables D_1 .

The above model is solved using the Bayesian approach in a manner that is almost similar to (Rapley and Welsh, 2008). Prior information is incorporated in the model and samples are drawn from the posterior distribution to obtain population estimates. The joint posterior distribution is

$$\begin{aligned}
P(\alpha, \theta, \lambda | s, N_y, K, M_k, X_k, Y_k) = p(s | N_y, K, M_k) \times \\
\left[\binom{N}{n_y} \left(\frac{\alpha^{n_y} (1 - \alpha)^{N - n_y}}{1 - (1 - \alpha)^N} \right) \times \left(\frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \alpha^{a_1 - 1} (1 - \alpha)^{b_1 - 1} \right) \right] \times \\
\left[\binom{n_y}{\kappa} \left(\frac{\theta^\kappa (1 - \theta)^{n_y - \kappa}}{1 - (1 - \theta)^{n_y}} \right) \times \left(\frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)\Gamma(b_2)} \theta^{a_2 - 1} (1 - \theta)^{b_2 - 1} \right) \right] \times \\
\left[\left((n_y - \kappa)! \prod_{k=1}^{\kappa} \frac{1}{(m_k - 1)!} \left(\frac{1}{\kappa} \right)^{m_k - 1} \right) \right] \times \\
\left[\prod_{k=1}^{\kappa} \left(\frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \right) \times p(\lambda) \right] \quad (4.5)
\end{aligned}$$

where $\lambda = m_k \exp(\beta_0 + \beta_1 x_k)$, $p(\lambda)$ is a prior distribution on the parameter $\beta = \{\beta_0, \beta_1\}$ and a_1, a_2, b_1, b_2 are shape parameters of their respective distribution.

4.5 Fitting of the model

Gibbs sampler was used to obtain samples from the posterior distribution given in equation (4.5) with the following steps.

1. Use information from the auxiliary variable to set the initial values of N_{y1}, K_1, M_{k1} and Y_{k1}
2. Draw α from the marginal distribution $[\alpha|N_y]$
3. Draw θ from the marginal distribution $[\theta|N_y, K]$
4. Draw β from the marginal distribution $[\beta|K, M_k, X_k, Y_k]$
5. Update λ using $\lambda = m_k \exp(\beta_0 + \beta_1 X_k)$
6. Draw $(N_{y1}, K_1, M_{k1}, Y_{k1})$ from the marginal distribution $[N_{y1}, K_1, M_{k1}, Y_{k1}|N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda]$
7. Repeat steps 2 to 6 R number of iterations.

Drawing α

The marginal distribution of $[\alpha|N_y]$ is given by the second term on the right hand side of equation (4.5). We follow Rapley and Welsh (2008) by sampling using Metropolis-Hastings in which the acceptance ratio is

$$\text{ratio}_\alpha = \frac{f(\alpha^*|N_{y0}, N_{y1})q(\alpha|\alpha^*)}{f(\alpha|N_{y0}, N_{y1})q(\alpha^*|\alpha)} \quad (4.6)$$

where

$$f(\alpha^*|N_{y0}, N_{y1}) = \frac{\alpha^{*n_{y0}+n_{y1}+a_1-1}(1-\alpha^*)^{N-n_{y0}-n_{y1}+b_1-1}}{1-(1-\alpha^*)^N}$$

is the distribution to be sampled from and

$$q(\alpha^*|\alpha) \sim \text{beta}(n_{y0} + n_{y1} + a_1 - 1, N - n_{y0} - n_{y1} + b_1 - 1)$$

is the proposed distribution in which α^* and α represent the proposed and the previous value of α respectively.

Drawing θ

The marginal distribution of $[\theta|N_y, K]$ is given by the third term on the right hand side of equation (4.5). We also follow Rapley and Welsh (2008) by drawing samples using Metropolis-Hastings in which the acceptance ratio is

$$\text{ratio}_\theta = \frac{f(\theta^*|N_{y0}, N_{y1}, K_0, K_1)q(\theta|\theta^*)}{f(\theta|N_{y0}, N_{y1}, K_0, K_1)q(\theta^*|\theta)} \quad (4.7)$$

where

$$f(\theta^*|N_{y0}, N_{y1}, K_0, K_1) = \frac{\theta^{*\kappa_0 + \kappa_1 + a_2 - 1}(1 - \theta^*)^{n_{y0} + n_{y1} - \kappa_0 - \kappa_1 + b_2 - 1}}{1 - (1 - \theta^*)^{n_{y0} + n_{y1}}}$$

and

$$q(\theta^*|\theta) \sim \text{beta}(\kappa_0 + \kappa_1 + a_2 - 1, n_{y0} + n_{y1} - \kappa_0 - \kappa_1 + b_2 - 1)$$

with θ^* and θ being the proposed and the previous value of θ respectively.

Drawing β

The marginal distribution of $[\beta|K, M_k, X_k, y_k]$ can be obtained from the last term of equation (4.5). Since log-linear model has no conjugate prior (Hoff, 2009, p. 173), a bivariate normal prior is assigned to the parameter β . Metropolis is used to draw samples with a bivariate normal as the proposed distribution. The acceptance ratio is

$$\text{ratio}_\beta = \frac{f(\beta^*|\kappa_0, \kappa_1^*, m_{k0}, m_{k1}^*, X_k, y_{k0}, y_{k1}^*)}{f(\beta|\kappa_0, \kappa_1, m_{k0}, m_{k1}, X_k, y_{k0}, y_{k1})} \quad (4.8)$$

where

$$f(\beta^*|\kappa_0, \kappa_1^*, m_{k0}, m_{k1}^*, X_k, y_{k0}, y_{k1}^*) = \prod_{k=1}^{\kappa_0 + \kappa_1^*} \left(\frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \right) \times \prod_{j=1}^2 \text{dnorm}(\beta_j^*, 0, \sigma^2).$$

Drawing $N_{y1}, K_1, M_{k1}, Y_{k1}$

The probability for the unsampled units at each step of sampling is

$$p(s_1|N_y, K, M_k) = \prod_{i=1}^n \frac{1}{\sum_{j=1}^{N-n_{y0}-n_{y1}+\kappa_0+\kappa_1} z_j - \sum_{k=0}^{i-1} (z_{s_k} + e_{s_k})} \quad (4.9)$$

The joint marginal distribution of the unsampled variables N_{y1} , K_1 , M_{k1} and Y_{k1} is

$$\begin{aligned} & [N_{y1}, K_1, M_{k1}, Y_{k1} | N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda] = \\ & p(s_1|N_y, K, M_k) \times \binom{N}{n_y} \left(\frac{\alpha^{n_y} (1-\alpha)^{N-n_y}}{1 - (1-\alpha)^N} \right) \times p(\alpha) \times \\ & \binom{n_y}{\kappa} \left(\frac{\theta^\kappa (1-\theta)^{n_y-\kappa}}{1 - (1-\theta)^{n_y}} \right) \times p(\theta) \times \left((n_y - \kappa)! \prod_{k=1}^{\kappa} \frac{1}{m_k - 1} \left(\frac{1}{\kappa} \right)^{m_k - 1} \right) \times \\ & \prod_{k=1}^{\kappa} \left(\frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \right) \times p(\lambda) \end{aligned} \quad (4.10)$$

where $p(\alpha)$, $p(\theta)$, and $p(\lambda)$ are prior distributions as shown in equation (4.5). An attempt was done to draw samples of N_{y1} , K_1 , M_{k1} and Y_{k1} separately. This could not however guarantee the relation $\sum_{k=1}^{k_1} m_{k1} = n_{y1}$ in all samples drawn. Hence, N_{y1} , K_1 , M_{k1} and Y_{k1} were drawn jointly as in Rapley and Welsh (2008).

Splitting equation (4.10) in terms of the observed and the unobserved variables, we have

$$\begin{aligned}
& [N_{y1}, K_1, M_{k1}, Y_{k1} | N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda] \\
& = p(s_1 | N_y, K, M_k) \times \\
& \quad \frac{N!}{(n_{y0} + n_{y1})! (N - n_{y0} - n_{y1})!} \cdot \frac{\alpha^{n_{y0}} \alpha^{n_{y1}} (1 - \alpha)^{N - n_{y0}} (1 - \alpha)^{-n_{y1}}}{1 - (1 - \alpha)^N} \\
& \quad \times \frac{(n_{y0} + n_{y1})!}{(\kappa_0 + \kappa_1)! (n_{y0} + n_{y1} - \kappa_0 - \kappa_1)!} \cdot \frac{\theta^{\kappa_0} \theta^{\kappa_1} (1 - \theta)^{n_{y0} - \kappa_0} (1 - \theta)^{n_{y1} - \kappa_1}}{1 - (1 - \theta)^{n_{y0} + n_{y1}}} \\
& \quad \times ((n_{y0} + n_{y1} - \kappa_0 - \kappa_1)! \prod_{k=1}^{\kappa_0 + \kappa_1} \frac{1}{m_k - 1} \left(\frac{1}{\kappa_0 + \kappa_1} \right)^{\sum_{k=1}^{\kappa_0 + \kappa_1} m_k - 1}) \\
& \quad \times \prod_{k=1}^{\kappa_0 + \kappa_1} \left(\frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \right) \times p(\lambda) \times p(\alpha) \times p(\theta)
\end{aligned} \tag{4.11}$$

On further simplification, equation (4.11) becomes

$$\begin{aligned}
& [N_{y1}, K_1, M_{k1}, Y_{k1} | N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda] \\
& = p(s_1 | N_y, K, M_k) \times \frac{\alpha^{n_{y1}} (1 - \alpha)^{-n_{y1}}}{(N - n_{y0} - n_{y1})!} \times \frac{\theta^{\kappa_1} (1 - \theta)^{n_{y1} - \kappa_1}}{(\kappa_0 + \kappa_1)! 1 - (1 - \theta)^{n_{y0} + n_{y1}}} \\
& \quad \times \prod_{k \notin s}^{\kappa_1} \frac{1}{m_k - 1} \left(\frac{1}{\kappa_0 + \kappa_1} \right)^{\sum_{k=1}^{\kappa_0 + \kappa_1} m_k - 1} \times \prod_{k \notin s}^{\kappa_1} \frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))}.
\end{aligned} \tag{4.12}$$

Metropolis-Hastings is used to draw sample from equation (4.12). The high correlation at the network level between the study and the auxiliary variable comes into play when constructing the proposed distribution. Figure 4.1 shows distribution of artificial data of the study and the auxiliary variable under different parameter settings. The correlation at the network level between the study and the auxiliary variable was set to $0.6 \leq \rho < 0.9$. The condition to add neighboring units on the auxiliary variable was $C^* = \{x_i : x_i > 2\}$. This ensured that $C = \{y_i : y_i > 0\} \approx C^* = \{x_i : x_i > 2\}$ is satisfied. As the figure portrays, with such high correlation, it becomes easier to use the auxiliary information with the given critical value to determine the number of non-empty units N_y , the number of non-empty networks K and the number of units within each non-empty network M_k .

With this information, N_{y1} is generated by sampling from a discrete uniform distribution centered on the previous value of N_{y1} with support $N_{y1} \pm 2$ and

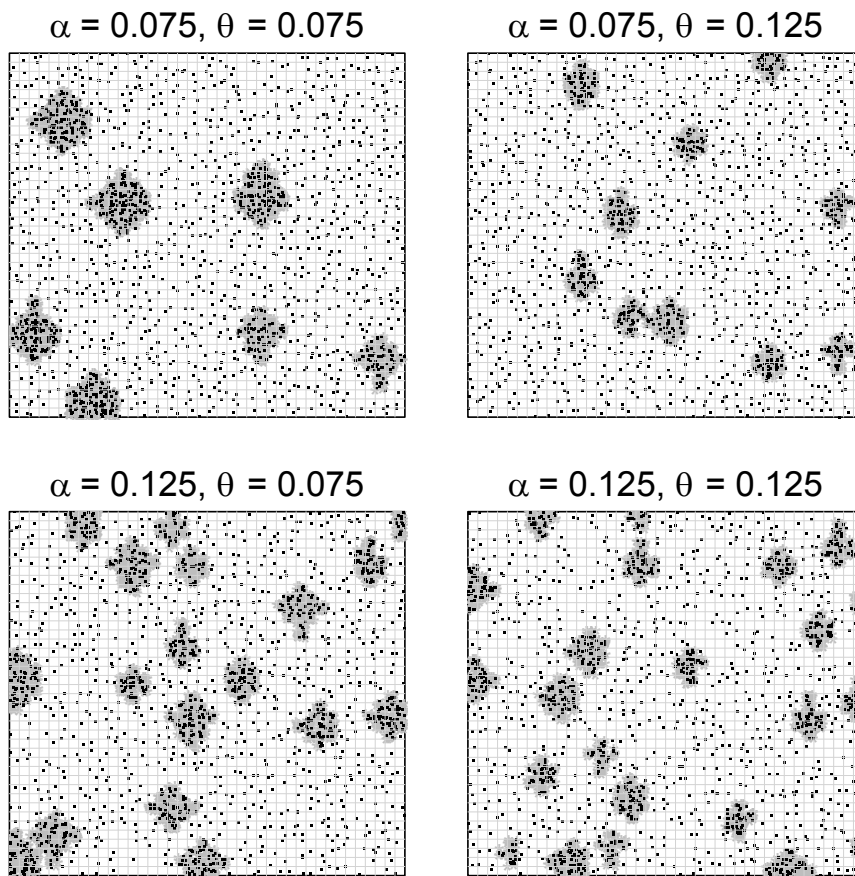


Fig. 4.1: Population distribution of study (gray *) and auxiliary variables (black .) with some values of α and θ that were used. α controls the number of non empty units and θ controls the number of groups of animals.

N_{y1} lies between the initial value of N_{y1} and the initial value of N_{y1} plus some allowance, in this study, 5. i.e $N_{y1}^1 \leq N_{y1} \leq N_{y1}^1 + 5$.

Given N_{y1} , K_1 is also generated according to a discrete uniform distribution centered on the previous value of K_1 with support $K_1 \pm 2$ and $1 \leq K_1 \leq K_1^1$ where K_1^1 is the initial value of K_1 . With N_{y1} and K_1 values obtained, M_{k1} and Y_{k1} were generated as follows

$$\begin{aligned} M_{k1}|N_{y1}, K_1 &\sim \mathbf{1}_{\kappa_1} + \text{Multinomial}(n_{y1} - \kappa_1, \frac{1}{\kappa_1} \mathbf{1}_{k1}) \\ Y_{k1}|X_{k1}, K_1, M_{k1}, \lambda &\sim \text{independent truncated Poisson}(\lambda) \end{aligned} \quad (4.13)$$

where $\lambda = m_{k1} \exp(\beta_0 + \beta_1 X_{k1})$. The proposed distribution is then

$$\begin{aligned} [N_{y1}, K_1, M_{k1}, Y_{k1}]_{prop} = \\ C \times (n_{y1} - \kappa_1)! \prod_{k \notin s}^{\kappa_1} \frac{1}{(m_{k1} - 1)!} \left(\frac{1}{\kappa_1} \right)^{\sum_{k \notin s}^{\kappa_1} m_{k1} - 1} \times \prod_{k \notin s}^{\kappa_1} \frac{\lambda^{y_k} \exp(-\lambda)}{y_k! (1 - \exp(-\lambda))} \end{aligned} \quad (4.14)$$

where C is a constant. The acceptance ratio becomes

$$\text{ratio} = \frac{f(D_1^*|D) q(D_1)}{f(D_1|D) q(D_1^*)} \quad (4.15)$$

where

$$\frac{f(D_1^*|D)}{q(D_1^*)} = \frac{[N_{y1}^*, K_1^*, M_{k1}^*, Y_{k1}^* | N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda]}{[N_{y1}^*, K_1^*, M_{k1}^*, Y_{k1}^*]_{prop}}$$

defines the distribution to be sampled from divide by the proposed distribution based on the current values of the random variables N_{y1} , K_1 , M_{k1} and Y_{k1} . Upon simplification, this gives

$$\begin{aligned}
\frac{f(D_1^*|D)}{q(D_1^*)} &= \log(C) + \log(p(s_1|N_y, K, M_k)) \\
&+ \log\left(\frac{1}{(N - n_{y0} - n_{y1}^*)!(\kappa_0 + \kappa_1^*)!(n_{y1}^* - \kappa_1)!}\right) \\
&+ n_{y1}^* \log\left(\frac{\alpha}{1 - \alpha}\right) + \log\left(\frac{\theta^{\kappa_1}(1 - \theta)^{(n_{y1}^* - \kappa_1)}}{1 - (1 - \theta)^{n_{y0} + n_{y1}^*}}\right) \\
&- \sum_{k \in s}^{\kappa_0} (m_k - 1) \log(\kappa_0 + \kappa_1^*) + \sum_{k \notin s}^{\kappa_1^*} (m_k - 1) \log \frac{\kappa_1^*}{\kappa_0 + \kappa_1^*}
\end{aligned} \tag{4.16}$$

where $(N_{y1}^*, K_1^*, M_{k1}^*, Y_{k1}^*)$ are the proposed values of the random variables.

On the other hand, the proposed distribution divide by the distribution to be sampled from based on the previous values of the random variables N_{y1}, K_1, M_{k1} and Y_{k1} becomes

$$\frac{q(D_1)}{f(D_1|D)} = \frac{[N_{y1}, K_1, M_{k1}, Y_{k1}]_{prop}}{[N_{y1}, K_1, M_{k1}, Y_{k1}|N_{y0}, K_0, M_{k0}, X_{k0}, Y_{k0}, X_{k1}, \alpha, \theta, \lambda_k]}$$

which is

$$\begin{aligned}
\frac{q(D_1)}{f(D_1|D)} &= \log(C) + \log \frac{1}{p(s_1|N_y, K, M_k)} \\
&+ \log((N - n_{y0} - n_{y1}^*)!(\kappa_0 + \kappa_1^*)!(n_{y1}^* - \kappa_1)!) \\
&+ n_{y1} \log\left(\frac{1 - \alpha}{\alpha}\right) + \log\left(\frac{1 - (1 - \theta)^{n_{y0} + n_{y1}}}{\theta^{\kappa_1}(1 - \theta)^{(n_{y1} - \kappa_1)}}\right) \\
&+ \sum_{k \in s}^{\kappa_0} (m_k - 1) \log(\kappa_0 + \kappa_1) + \sum_{k \notin s}^{\kappa_1} (m_k - 1) \log\left(\frac{\kappa_0 + \kappa_1}{\kappa_1}\right)
\end{aligned} \tag{4.17}$$

4.6 Simulation set-up and results

A range of artificial populations were simulated according to model (4.2) with parameters $\alpha, \theta \in \{0.075, 0.100, 0.125\}$ and $\beta = (1, 0.03)$. This allowed for the evaluation of the estimates under different population structure, from few

to numerous clusters. Recall that α controls the number of non-empty units in the study region and θ controls the number of groups of animals. The auxiliary variable was generated such that the correlation at the network level between the study and the auxiliary variable was $6 \leq \rho < 9$.

The study region was partitioned into $80 \times 80 = N = 6400$ units. The number of strata considered was $H = 4$ such that each strata had $L = 40$ PSU each with $M = 40$ SSU. The number of SSU in each strata was $N_q = 1600$ units. An initial sample consisted of size $n = 12$ PSU (strips). Samples were allocated to different strata by optimal allocation under box-constraint (Gabler et al., 2012; Münnich et al., 2012).

500 populations were simulated in each parameter setting and units drawn according stratified strip ACS. $R = 100,000$ samples were drawn from the posterior distribution, discarded the first 1,001 samples as "burn in" and thinned the chain by picking every 90-th sample. Convergence of the Markov Chain Monte-Carlo (MCMC) was monitored using **R** package **CODA**. Autocorrelation and trace plots for one of the samples in one of the strata are presented in the appendix in figures B.4 and B.2 respectively. The model was evaluated using the bias and the MSE obtained as

$$B(\hat{Y}) = \sum_{j=1}^{500} \frac{\hat{Y}_j - Y_j}{500}$$

$$\widehat{MSE}(\hat{Y}) = \sum_{j=1}^{500} \frac{(\hat{Y}_j - Y_j)^2}{500}$$

where \hat{Y}_j is the predicted population total while Y_j is the true population total.

Figure 4.2 shows a density plot of the predicted population total from the posterior samples of one chain. Also shown in the figure is the true population total and the 95% Highest Posterior Density (HPD) interval.

Box plots for the bias and the MSE are given in figure 4.3. From the figure, the predicted bias increases with increase in the number of cluster (which is controlled by the parameter θ). The figure further shows that the bias decreases

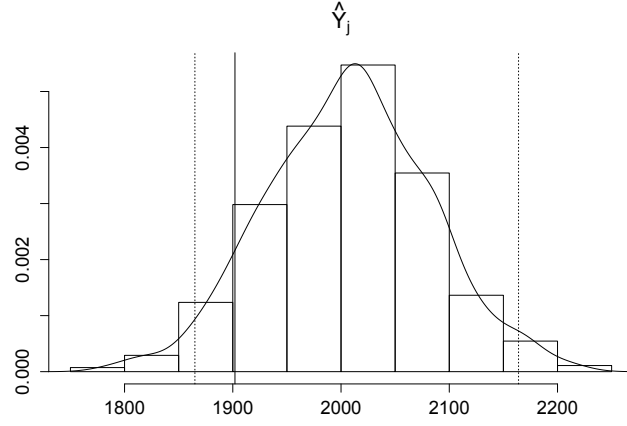


Fig. 4.2: Posterior density plot for the predicted population total for one population in one of the strata. The solid line represents the true population value while the dotted lines represent the 95% HPD interval.

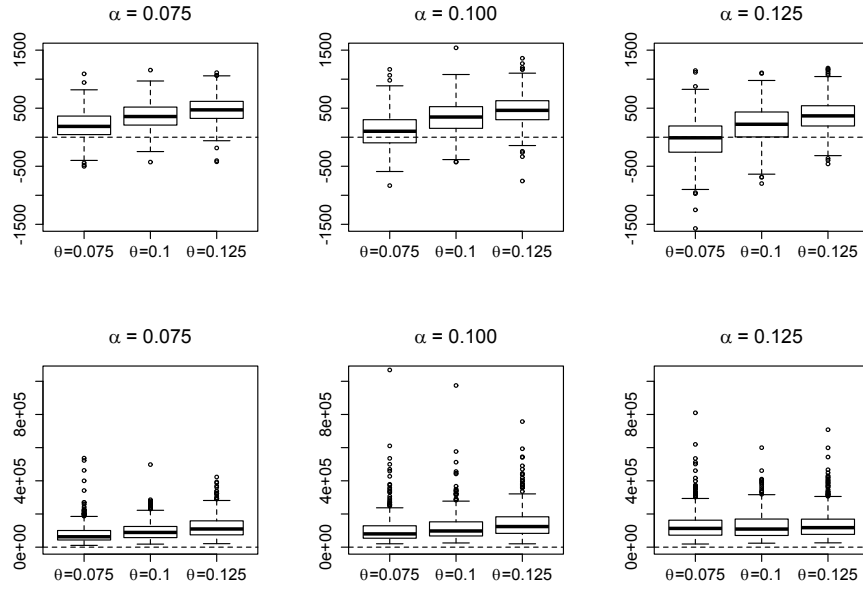


Fig. 4.3: Bias (first row) and MSE (second row) for different values of α and θ . α controls the number of non-empty units and θ controls the number of groups of animals.

as the number of non-empty units in the region increases (controlled by the parameter α). In other words, the bias is low if the population has few clusters ($\theta = 0.075$) and each cluster has relatively large network size ($\alpha = 0.0125$) and vice-versa. Finally, the MSE increases with increase in both α and θ *i.e* the MSE increases as the number of clusters as well as cluster sizes increase.

α	$\hat{\alpha}$	θ	$\hat{\theta}$	β_0	$\hat{\beta}_0$	β_1	$\hat{\beta}_1$
0.075	0.079	0.075	0.080	1.000	0.999	0.300	0.303
0.075	0.079	0.100	0.099	1.000	1.010	0.300	0.288
0.075	0.078	0.125	0.120	1.000	1.019	0.300	0.276
0.100	0.103	0.075	0.075	1.000	0.985	0.300	0.323
0.100	0.102	0.100	0.096	1.000	1.000	0.300	0.302
0.100	0.102	0.125	0.118	1.000	1.011	0.300	0.286
0.125	0.128	0.075	0.072	1.000	0.978	0.300	0.338
0.125	0.127	0.100	0.094	1.000	0.994	0.300	0.311
0.125	0.126	0.125	0.116	1.000	1.005	0.300	0.295

Tab. 4.1: True versus estimated parameters for the artificial populations. α controls the number of non-empty units, θ controls the number of groups of animals while β_0 and β_1 is the intercept and the slope used to generate counts.

Table 4.1 presents results of the true and estimated parameters for the simulated data. As the table shows, all parameters were estimated with values close to their true values with an absolute bias not more than 4%.

4.7 Design-based simulation

The proposed model was evaluated by performing a design-based simulation using elephant data. The study variable was the distribution and counts of elephants across Amboseli-West Kilimanjaro/Magadi-Natron cross border landscape. The landscape covers parts of Kenya and Tanzania between $1^\circ 37' \text{ S}$ and $3^\circ 13' \text{ S}$ and between $35^\circ 49' \text{ E}$ and $38^\circ 00' \text{ E}$ covering an area of $25,623 \text{ km}^2$. The data was obtained from an aerial survey that was conducted in October (dry season) 2013. Data of the auxiliary variable was the distribution and counts of elephants obtained from a similar survey covering

the same area conducted in April (wet season) 2013. Details on how the two surveys were conducted are found in Kenana et al. (2013b) and Kenana et al. (2013a).

Maps produced from the two surveys on the distribution of the study and the auxiliary variable are given in figure 4.4. A grid was laid on the maps as shown in the figure and effort was made to determine the counts in each quadrant.

The region was divided into four strata as shown in the figure. 500 samples each of size $n = 12$ were drawn from the data. Like in the model-based simulation, samples were allocated to strata by optimal allocation according to box-constraint. In each strata, the minimum and maximum sample size was set to 2 and 24 (24 is the total number of strips in each strata) respectively. Two parallel chains were ran with initial values given in table 4.2. For each chain $R = 100,000$ draws were made from the posterior distributions, discarded the first 1,001 draws and picked every 90-th draw. Diagnostics to determine the convergence of the MCMC under the proposed model is provided in appendix B.

	Proposed Model		RaW Model	
Parameters	chain 1	chain 2	chain 1	chain 2
α	0.05	0.08	0.08	0.05
θ	0.04	0.02	0.15	0.10
β_0	0.05	0.50	0.05	0.08
β_1	0.02	0.30	0.04	0.03

Tab. 4.2: Initial values set for parameters in each model. RaW represents the model due to Rapley and Welsh (2008) under stratified strip setting. α controls the number of non-empty units, θ controls the number of groups of animals while β_0 and β_1 is the intercept and the slope used to generate counts

Table 4.3 presents the correlation between the study variable (Elephant distribution and counts obtained from October 2013 survey) and the auxiliary variable (Elephant distribution and counts obtained from April 2013 survey) for each strata. Also presented in the table is the total difference of the within network variance relative to the stratum variance. Since all the differences are

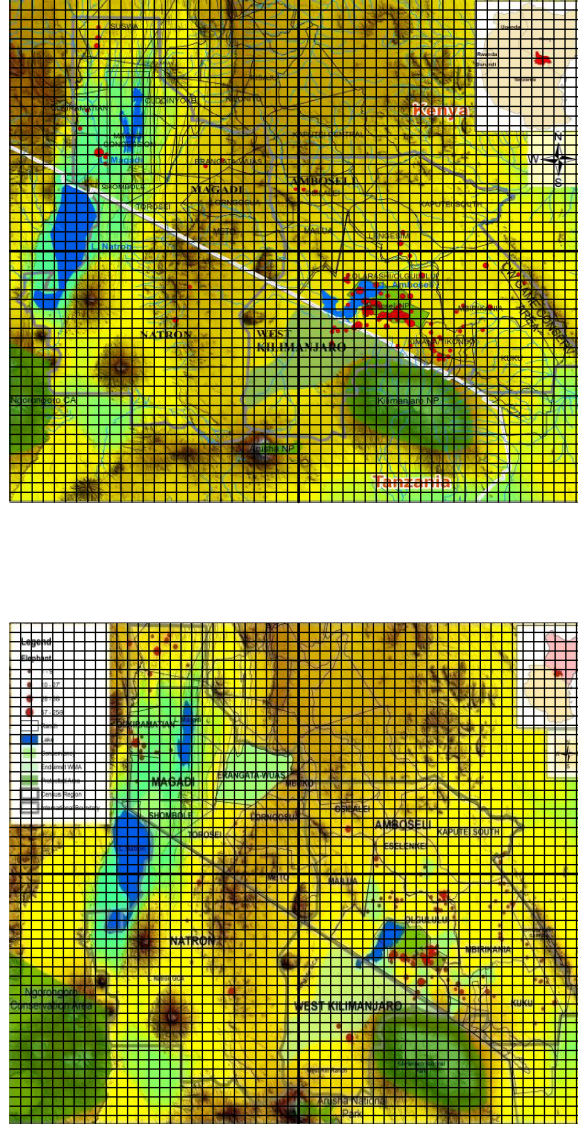


Fig. 4.4: *Top panel*: Distribution of the study variable (elephants) obtained from an aerial survey that was conducted in October (dry season) 2013. *Bottom panel*: Distribution of the auxiliary variable (elephants) obtained from an aerial survey conducted in April (wet season) 2013. The first stratum is on the top left, second stratum is on the top right, third and fourth strata on bottom left and right respectively. Maps obtained from Kenana et al. (2013b) and Kenana et al. (2013a) respectively.

positive, the data is suitable for the ACS design.

Strata position	Correlation	Difference
Strata 1	0.282	0.127
Strata 2	0.656	0.001
Strata 3	-0.006	0.002
Strata 4	-0.037	0.217

Tab. 4.3: Correlation at network level for different strata. Difference is the total difference of the within network variance and the stratum variance of the study variable.

A comparison was made between the proposed model and the model due to Rapley and Welsh (2008) modified to the stratified strip design and maintaining the order of the sampled units; herein denoted as RaW model. The comparison was based on the bias and the MSE. Figure 4.5 presents the resulting box plots. The high negative bias and the high MSE from the proposed model were greatly attributed by the fourth strata whose correlation at the network level was -0.037 as seen in table 4.3. To clearly see this, box plots on the contribution of the total bias and the total MSE by each stratum were constructed; these are shown in figure 4.6. *see* more comments in the discussion.

A comparison was also made between the model-based and the design-based estimates using the bias and the MSE. Estimates under the design-based estimation were obtained following estimation procedure presented in sub-section 3.1. These results are given in table 4.4

Type	Bias	MSE
Proposed Model	-91.39	69,195.25
RaW Model	492.56	102,786.05
Design-based estimates	0.00	255,302.80

Tab. 4.4: Results of the comparison between model-based and design-based estimates using the bias and the MSE.

From the table, the design-based estimator though unbiased, had the highest MSE compared to the two model-based estimators. As already seen,

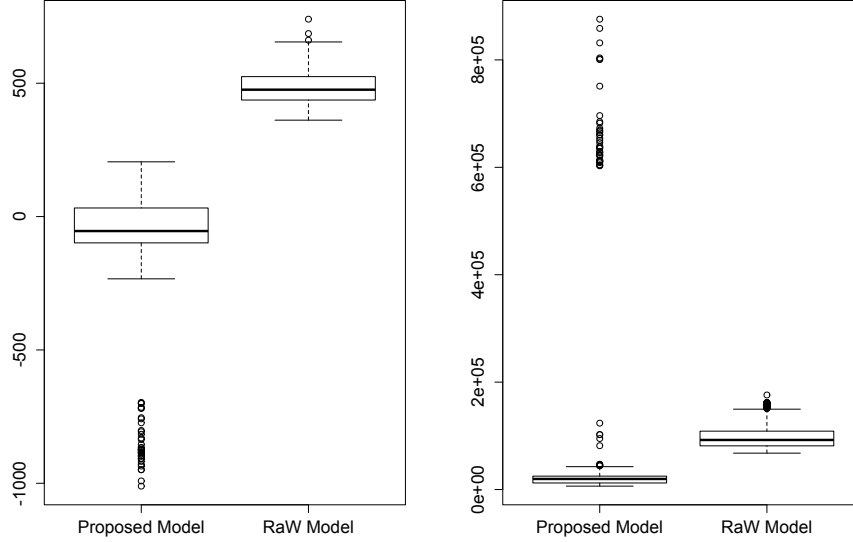


Fig. 4.5: Box plots of the prediction bias (left) and the prediction MSE (right) for the proposed model and the modified model due to Rapley and Welsh (2008) (RaW)

the proposed model performed better in terms of lower MSE compared to the RaW model.

4.8 Discussion

The current chapter extended the model due to Rapley and Welsh (2008) by incorporating auxiliary information, making use of stratified strip ACS (described in sub-section 3.1) and by sampling with clusters selected without replacement. Results basing on the Bayesian approach using the artificial and the real data showed that the proposed model has helped in efficiency gain of the estimates. This however depends on a combination of the following factors:

1. The relationship between the within network variance and the population variance in each stratum.
2. The correlation between the study and the auxiliary variable at the network level

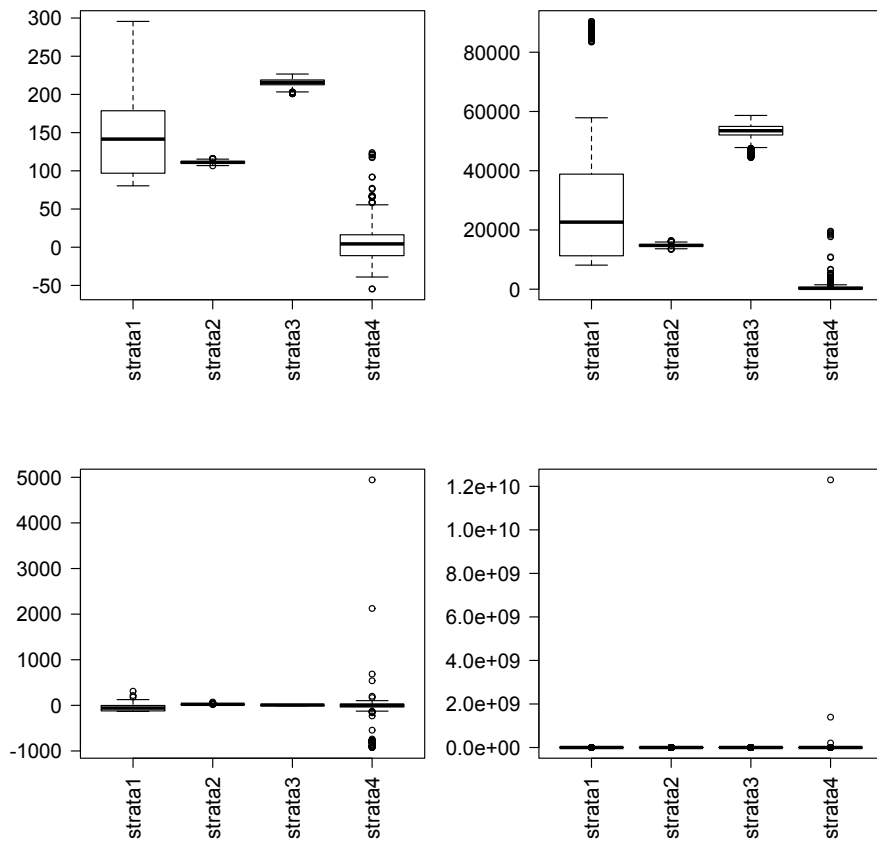


Fig. 4.6: Box plots of the bias (first column) and MSE (second column) under the RaW model (first row) and under the proposed model (second row) for each stratum.

3. The size of the initial sample relative to the final sample

Combining figure 4.6 and table 4.3, the following is noticed:

1. Under the ACS that does not make use of auxiliary information (RaW model), efficiency gain (in each stratum) increases with increase in the difference between the within network variance and the stratum variance. This can be seen in the first row of figure 4.6 and third column of table 4.3.
2. When the auxiliary information is introduced to the model as in the proposed model, efficiency gain depends further to the increase in correlation level between the study and the auxiliary variable (second row of figure 4.6 and second column of table 4.3).

In the simulations, the condition to adaptively add neighboring units was set at $C = \{y_i : y_i > 0\} \approx C^* = \{x_i : x_i > 2\}$. This means that the edge units have observed y -values equals to 0 hence it does not make a difference by modeling Y_k in the networks only rather than in the cluster. If however, the condition was set to, say, $C = \{y_i : y_i > 1\}$, it is then necessary to model the observed values in a cluster rather than in a network.

While Rapley and Welsh (2008) modeled the data collected by ACS design at the network level, Gonçalves and Moura (2016) instead modeled the data at the unit level. They showed that modeling the data at the unit level provides gain in efficiency as opposed to modeling at the network level. Extending the idea of Gonçalves and Moura (2016) by incorporating auxiliary information deserves further exploration.

Another possibility involves the extension of the proposed model to the problem of predicting population total of African elephants in all parks in Kenya by achieving further gain in efficiency and ensuring lower sampling costs. This can be solved by employing small area estimation (SAE) technique. The model, though, was applied to African elephants, can be applied to other populations provided they meet the above requirements.

5. Comparison of methods

It has been shown in the previous chapters that the proposed strategies outperform their non-adaptive counterparts. In this chapter, performance of the above proposed strategies are compared through simulations using real elephant data. The study and the auxiliary variable is the distribution and counts of elephants surveyed in October (dry season) and April (wet season) 2013 respectively. The initial sample size is 192 units across the designs. The methods are compared using Coefficient of Variation (CV), expected final sample size and RB. We start by exploring the data taking note of conditions that improve efficiency of the specified methods.

Stratified Strip ACS

The study region is partitioned into four strata with the standard deviation for each stratum shown in table 5.1. The initial sample size translates to eight strips. As the minimum sample size for each stratum is two, then each stratum receives the same sample size as shown in the table.

strata	sd_q	n_q
stratum 1	2.180	2.000
stratum 2	0.353	2.000
stratum 3	0.475	2.000
stratum 4	6.388	2.000

Tab. 5.1: Standard deviation and sample size allocated in each stratum

The application of optimal allocation according to box-constraint method was similar to equal allocation since the overall sample size was small. Noting that stratum standard deviations vary and with given sample sizes allocated to each stratum, gains in efficiency are not expected to be high. In each stratum, the variability within each network is higher than the stratum variability. Hence the population is suitable for adaptive sampling.

	ρ	cv.r
Unit level	0.559	-0.508
Network level	0.504	-0.471

Tab. 5.2: ρ is the correlation coefficient at the unit (first row) and network (second row) level. cv.r is the right hand side of equation (3.25) (first row) and (2.31) (second row)

ACS for negatively correlated data

Table 5.2 shows that the correlation coefficient at both the unit and the network level is positive with moderate values. Further, the conditions on equations (2.31) and (3.25) are both not fulfilled. For this population set, the auxiliary variable will not improve gains in efficiency both under adaptive and non-adaptive design i.e the product estimator under both adaptive and non-adaptive is not suitable for this population set.

Results

Variability of the methods

Results of the CV are given in table 5.3. The table confirms the preliminary findings above; both stratified strip ACS and ACS for negatively correlated data are highly variable. For this particular population, the model-based estimation was the least variable design. This is followed by the design with stopping rule.

StrtStrpD	Product.ht	Product.hh	StopRule	StrtStrpM
0.792	0.515	0.778	0.428	0.251

Tab. 5.3: Results of the coefficient of variation under stratified strip ACS (StrtStrpD), ACS for negatively correlated data with HT (Product.ht) and with HH (Product.hh) estimator, ACS^{c-sr} (StopRule) and under model-based estimation (StrtStrpM)

Bias of the methods

Results of the bias are given in table 5.4. From the table, the product HH estimator and the estimator due to the stopping rule had much higher bias. As

expected, the stratified strip ACS is unbiased.

StrtStrpD	Product.ht	Product.hh	StopRule	StrtStrpM
0.000	0.038	0.098	0.090	-0.075

Tab. 5.4: Results of the relative bias under stratified strip ACS (StrtStrpD), ACS for negatively correlated data with HT (Product.ht) and with HH (Product.hh) estimator, ACS^{c-sr} (StopRule) and under model-based estimation (StrtStrpM)

Expected final sample size

The cost of sampling under the studied methods is evaluated using the expected final sample size. Figure 5.1 shows the distribution of the expected final sample size across the proposed strategies.

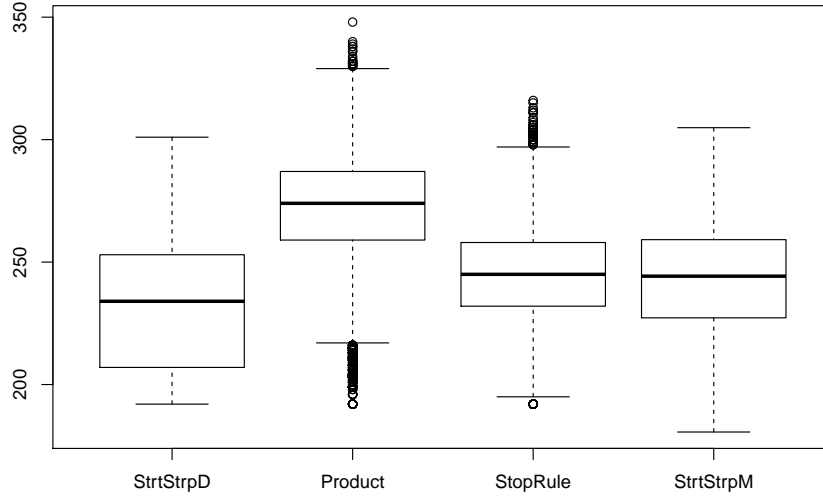


Fig. 5.1: Box plots showing distribution of the expected final sample size under the design-based stratified strip ACS (StrtStrpD), the ACS for negatively correlated data (Product), the ACS^{c-sr} (StopRule) and the model-based stratified strip ACS (StrtStrpM).

The design that made use of the negative correlation did not employ any techniques to reduce the final sample size. As explained under the specific subsections, the use of stratification and truncating networks at stratum boundaries helps in limiting the final sample size when there exists networks that cross over

stratum boundaries. The use of the stopping rule is also a strategy to reduce the final sample size. On the other hand, the use of without replacement of clusters results in higher final sample size (but this depends on the population structure). Hence the combination of the stopping rule and the without replacement of cluster is a trade-off between lower MSE and higher final sample size compared to the stopping rule of Gattone and Di Battista (2011).

Summary and conclusion

To summarize the chapter, a scatter plot of the estimated population total against the final sample size for the proposed methods is given in figure 5.2. The horizontal line is the true population total. The figure shows that the StrtStrpM (model-based stratified strip ACS labeled "5") that made use of auxiliary information that is positively correlated with the variable of interest, is the most preferable approach since its estimates of the population total are closest to the true population value. It achieved this with a relatively lower sampling cost as can be seen from its final sample size.

The figure further shows that the Product.ht and Product.hh (ACS for negatively correlated data with both HT and HH estimators labeled "1" and "2" respectively) estimators is not desirable. Recall that this design achieves gains in efficiency if the correlation between the study and the auxiliary variable is highly negative at the network level. This was however, not achieved from the preliminary investigation above. Note the way the distribution of their population estimates spread further away from the true population value. Further their sampling cost is high due to the high values of the final sample size. The product HH estimator is the least desirable in this case.

Comparing the StrtStrpD (stratified strip ACS labeled "4") and the StopRule (ACS^{c-sr} that employed a stopping rule and with clusters selected without replacement labeled "3"), first remember that these two designs do not make use of auxiliary information but they use different mechanism to

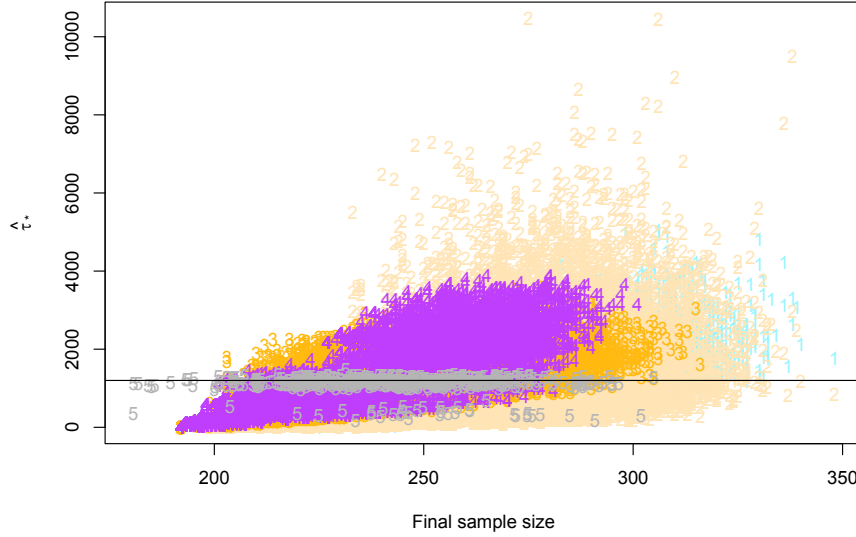


Fig. 5.2: Scatter plots for the distribution of the estimated population total against the final sample size under the different proposed strategies. The labels 1 and 2 represent the ACS for negatively correlated data with HT (Product.ht) and HH (Product.hh) estimators respectively. The labels 3, 4 and 5 represent ACS^{C-SR} (StopRule), stratified strip ACS under design-based (StrtStrpD) and model-based (StrtStrpM) respectively.

increase efficiency as well as reduce the sampling cost. The figure shows that the former design, even though has lower cost in terms of the final sample size, the spread of its estimated population total is higher compared to the latter. The above analysis is similar for the estimates of the MSE of the estimated population totals that is shown in figure 5.3.

Thus the choice of a specified design first depends on the characteristics of the population under study in terms of the degree of rarity and clusteredness. Further, if auxiliary variable is available, then the level of the correlation at the network level and the direction too matters. Other factors that affect the choice of a particular design include: if the aim is for the use of an unbiased estimator(s), the level of variability of the estimator to be accommodated and the amount of resources (such as budget, technical personnel, time and equipments) available.

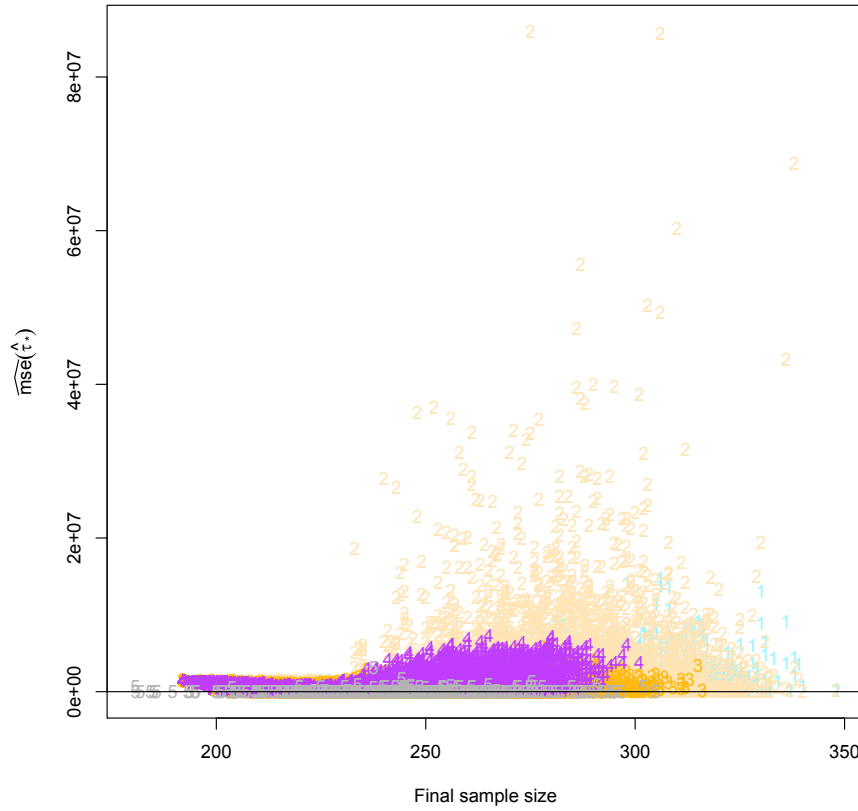


Fig. 5.3: Scatter plots for the distribution of the estimated MSE of the estimated population total against the final sample size under the different proposed strategies. The labels 1 and 2 represent the ACS for negatively correlated data with HT (Product.ht) and HH (Product.hh) estimators respectively. The labels 3, 4 and 5 represent ACS^{c-sr} (StopRule), stratified strip ACS under design-based (StrtStrpD) and model-based (StrtStrpM) respectively.

6. Summary and outlook

This dissertation looked at both design-based and model-based estimation for rare and clustered populations using the idea of the ACS design. The ACS design (Thompson, 2012, p. 319) starts with an initial sample that is selected by a probability sampling method. If any of the selected units meets a pre-specified condition, its neighboring units are added to the sample and observed. If any of the added units meets the pre-specified condition, its neighboring units are further added to the sample and observed. The procedure continues until there are no more units that meet the pre-specified condition. In this dissertation, the pre-specified condition is the detection of at least one animal in a selected unit.

In the design-based estimation, three estimators were proposed under three specific design setting. The first design was stratified strip ACS design that is suitable for aerial or ship surveys. This was a case study in estimating population totals of African elephants. In this case, units/quadrant were observed only once during an aerial survey. The Des Raj estimator (Raj, 1956) was modified to obtain an unbiased estimate of the population total. The design was evaluated using simulated data with different levels of rarity and clusteredness. The design was also evaluated on real data of African elephants that was obtained from an aerial census conducted in parts of Kenya and Tanzania in October (dry season) 2013. In this study, the order in which the samples were observed was maintained. Re-ordering the samples by making use of the Murthy's estimator (Murthy, 1957) can produce more efficient estimates. Hence a possible extension of this study. The computation cost resulting from the $n!$ permutations in the Murthy's estimator however, needs to be put into consideration.

The second setting was when there exists an auxiliary variable that is negatively correlated with the study variable. The Murthy's estimator (Murthy, 1964) was modified. Situations when the modified estimator is preferable was given both in theory and simulations using simulated and two

real data sets. The study variable for the real data sets was the distribution and counts of oryx and wildbeest. This was obtained from an aerial census that was conducted in parts of Kenya and Tanzania in October (dry season) 2013. Temperature was the auxiliary variable for two study variables. Temperature data was obtained from **R** package **raster**. The modified estimator provided more efficient estimates with lower bias compared to the original Murthy's estimator (Murthy, 1964). The modified estimator was also more efficient compared to the modified HH and the modified HT estimators of (Thompson, 2012, p. 319). In this study, one auxiliary variable is considered. A fruitful area for future research would be to incorporate multi-auxiliary information at the estimation phase of an ACS design. This could, in principle, be done by using for instance a multivariate extension of the product estimator (Singh, 1967) or by using the generalized regression estimator (Särndal et al., 1992).

The third case under design-based estimation, studied the conjoint use of the stopping rule (Gattone and Di Battista, 2011) and the use of the without replacement of clusters (Dryver and Thompson, 2007). Each of these two methods was proposed to reduce the sampling cost though the use of the stopping rule results in biased estimates. Despite this bias, the new estimator resulted in higher efficiency gain in comparison to the without replacement of cluster design. It was also more efficient compared to the stratified design which is known to reduce final sample size when networks are truncated at stratum boundaries. The above evaluation was based on simulated and real data. The real data was the distribution and counts of hartebeest, elephants and oryx obtained in the same census as above. The bias attributed by the stopping rule has not been evaluated analytically. This may not be direct since the truncated network formed depends on the initial unit sampled (Gattone et al., 2016a). This and the order of the bias however, deserves further investigation as it may help in understanding the effect of the increase in the initial sample size together with the population characteristics on the

efficiency of the proposed estimator.

Chapter four modeled data that was obtained using the stratified strip ACS (as described in sub-section (3.1)). This was an extension of the model of Rapley and Welsh (2008) by modeling data that was obtained from a different design, the introduction of an auxiliary variable and the use of the without replacement of clusters mechanism. Ideally, model-based estimation does not depend on the design or rather how the sample was obtained. This is however, not the case if the design is informative; such as the ACS design. In this case, the procedure that was used to obtain the sample was incorporated in the model. Both model-based and design-based simulations were conducted using artificial and real data. The study and the auxiliary variable for the real data was the distribution and counts of elephants collected during an aerial census in parts of Kenya and Tanzania in October (dry season) and April (wet season) 2013 respectively. Areas of possible future research include predicting the population total of African elephants in all parks in Kenya. This can be achieved in an economical and reliable way by using the theory of SAE.

Chapter five compared the different proposed strategies using the elephant data. Again the study variable was the elephant data from October (dry season) 2013 and the auxiliary variable was the elephant data from April (wet season) 2013. The results show that the choice of particular strategy to use depends on the characteristic of the population under study and the level and the direction of the correlation between the study and the auxiliary variable (if present).

One general area of the ACS design that is still behind, is the implementation of the design in the field especially on animal populations. This is partly attributed by the challenges associated with the field implementation, some of which were discussed in section 2.3. Green et al. (2010) however, provides new insights in undertaking the ACS design during an aerial survey such as how the aircraft should turn while surveying neighboring units.

A key point throughout the dissertation is the reduction of cost during a

survey which can be seen by the reduction in the number of units in the final sample (through the use of stopping rule, use of stratification and truncating networks at stratum boundaries) and ensuring that units are observed only once (by using the without replacement of cluster sampling technique).

The cost of surveying an edge unit(s) is assumed to be low in which case the efficiency of the ACS design relative to the non-adaptive design is achieved (Thompson and Collins, 2002). This is however not the case in aerial surveys as the aircraft flies at constant speed and height (Norton-Griffiths, 1978). Hence the cost of surveying an edge unit is the same as the cost of surveying a unit that meets the condition of interest. The without replacement of cluster technique plays a greater role of reducing the cost of sampling in such surveys. Other key points that motivated the sections in the dissertation include gains in efficiency (in all sections) and practicability of the designs in the specific setting.

Even though the dissertation focused on animal populations, the methods can as well be implemented in any population that is rare and clustered such as in the study of forestry, plants, pollution, minerals and so on.

7. Zusammenfassung und Ausblick

In der vorliegenden Dissertation wurden sowohl design-, als auch modellbasierte Schätzverfahren für seltene und gleichzeitig geclusterte Populationen, unter Verwendung des ACS Designs, betrachtet. Das ACS Design (Thompson, 2012, p. 319) beginnt mit einer, mittels Probability Sampling, gezogenen Stichprobe. Falls eines der gezogenen Einheiten ein zuvor spezifiziertes Kriterium erfüllt, werden dessen benachbarten Einheiten der Stichprobe hinzugefügt und beobachtet. Falls die neu hinzugefügten Elemente ebenfalls die zuvor festgelegte Bedingung erfüllen, werden deren angrenzenden Einheiten ebenfalls zur Stichprobe hinzugefügt. Dieses Vorgehen wird so lange wiederholt, bis keines der ausgewählten Elemente mehr dem Auswahlkriterium entspricht. Im Rahmen der vorliegenden Dissertation ist das Auswahlkriterium als das Erfassen von mindestens einem Tier, in der betrachteten Einheit, definiert.

Im Kontext der designbasierten Schätzung wurden jeweils drei Schätzer unter drei verschiedenen Designs untersucht. Das erste betrachtete Design ist stratified strip ACS, welches für Schiffs- oder auch Lufterhebungen geeignet ist. Anwendung fand dieses Vorgehen beispielsweise im Rahmen einer Fallstudie, welche den Totalwert afrikanischer Elefanten schätzte. In dieser Studie wurde jede Unit bzw. jeder Quadrant nur ein einziges mal während der Lufterhebung beobachtet. Der Des Raj Schätzer Raj (1956) wurde modifiziert, um unverzernte Schätzer für den wahren Populationstotalwert zu erhalten. Die Evaluation des Designs wurde anhand eines simulierten Datensatzes mit unterschiedlichen Seltenheitsgraden und Clusterungsintensitäten durchgeführt. Ferner wurde ein Datensatz mit realen Beobachtungen afrikanischer Elefanten aus einem Luftzensus, welcher im Oktober (Trockenzeit) 2013 in Teilen Kenias und Tansanias durchgeführt wurde, zur Bewertung herangezogen. In der vorliegenden Studie wurde die Reihenfolge, in welcher die Stichproben beobachtet wurden, nicht verändert. Allerdings können durch Neuordnung und unter Verwendung des Murthy Schätzers (Murthy, 1957), effizientere

Schätzwerte gewonnen werden - was eine denkbare Erweiterung dieser Arbeit darstellt. Die zusätzliche Berechnungszeit, die durch die $n!$ Permutation bei der Verwendung des Murthy's Schätzer entsteht, muss bei einer solchen Überlegung allerdings berücksichtigt werden.

Im zweiten betrachteten Design wurde eine Hilfsvariable verwendet, welche negativ mit der interessierenden Variable korreliert ist. Der Murthy Schätzer (Murthy, 1964) wurde hierfür modifiziert. Situationen, in denen der modifizierte Schätzer zu bevorzugen ist wurden sowohl theoretisch, als auch simulativ anhand synthetischer Daten und zwei realen Datensätzen dargestellt. Die Untersuchungsvariablen in den beiden realen Datensätzen waren jeweils das örtliche Vorkommen und die Anzahl von Gnus und Oryx Antilopen. Auch hier diente der Zensus aus Kenia und Tansania von 2013 als Datengrundlage. In beiden Studien wurde die Temperatur als Hilfsvariable herangezogen. Die Temperaturdaten wurden dem **R** Paket **raster** entnommen. Verglichen mit dem Originalschätzer nach Murthy (Murthy, 1964) lieferte der modifizierte Schätzer nicht nur effizientere Schätzwerte, sondern auch eine geringere Verzerrung. Darüber hinaus war der modifizierte Schätzer auch effizienter als die modifizierten HH und HT Schätzer nach (Thompson, 2012, p. 319). Im Rahmen dieser Arbeit wird lediglich eine Hilfsvariable betrachtet. Ein aussichtsreicher Bereich zukünftiger Forschung wäre es demnach, mehrere Hilfsvariablen in der Schätzphase eines ACS Designs zu berücksichtigen. Dies könnte grundsätzlich schon durch das Verwenden einer multivariaten Erweiterung des Produkt Schätzers (Singh, 1967) oder unter Zuhilfenahme des verallgemeinerten Regressionsschätzers (Särndal et al., 1992) erreicht werden.

Der dritte, im designbasierten Kontext, betrachtete Fall untersuchte die Verwendung eines Abbruchkriteriums (Gattone and Di Battista, 2011) in Verbindung mit dem Nichtzurücklegen gezogener Cluster (Dryver and Thompson, 2007). Beide Methoden wurden jeweils entwickelt, um die Erhebungskosten zu reduzieren - wenngleich die Verwendung des Abbruchkriteriums zu verzerrten Ergebnissen führt. Trotz dieser Verzerrung

resultierte der neue Schätzer in einem größeren Effizienzzuwachs als das alternative Design ohne Zurücklegen. Gleichzeitig war er ebenfalls effizienter als das stratifizierte Design, welches dafür bekannt ist, die endgültige Stichprobengröße zu verringern, wenn Netzwerke an Stratumgrenzen abgeschnitten werden. Die obige Evaluation basiert sowohl auf realen, wie auch synthetischen Daten. Die realen Daten, sowie die Anzahl und das örtliche Vorkommen von Elefanten und Oryx- und Kuhantilopen stammen aus dem bereits erwähnten Zensus zur Bestimmung des Wildbestands in Kenia und Tansania. Die Verzerrung, welche der Verwendung des Abbruchkriteriums zuzuschreiben ist, wurde nicht analytisch quantifiziert. Dieses Vorgehen mag nicht ganz korrekt sein, da das letztlich geformte, abgeschnittene Netzwerk von der ersten gezogenen Einheit abhängt (Gattone et al., 2016a). Weitere Nachforschungen bezüglich der Größenordnung der Verzerrungen sind daher erforderlich. Ziel ist es den Effekt, welcher die Erhöhung der ursprünglichen Stichprobengröße in Verbindung mit den Populationscharakteristika auf die Effizienz des vorgestellten Schätzers hat, besser nachvollziehen zu können.

In Kapitel vier wurden Daten modelliert, welche anhand des stratified strip ACS (Unterkapitel (3.1)) gewonnen wurden. Dies stellt insofern eine Erweiterung des Modells von Rapley and Welsh (2008) dar, als dass die Daten anhand eines anderen als dem von den Autoren verwendeten Design modelliert werden und zusätzlich eine Hilfsvariable eingeführt wird. Im Idealfall hängt die modellbasierte Schätzung nicht vom gewählten Design beziehungsweise der Art wie die Stichprobe generiert wurde, ab. Dies trifft jedoch nicht zu, falls das Design informativ ist - wie es beim ACS Design der Fall ist. Also wurde in dieser Erweiterung die Vorgehensweise zur Stichprobengenerierung in das Modell integriert. Sowohl modell- als auch designbasierte Simulationen wurden mit realen und künstlichen Daten durchgeführt. Im Falle des Wildbestandszensus, wäre ein potentiell denkbares Forschungsgebiet beispielsweise die Schätzung eines Populationstotalwertes aller in kenianischen Parks lebenden afrikanischen Elefanten. Dies kann über die Verwendung von

SAE Methoden erreicht werden.

In Kapitel 5 wurden die verschiedenen vorgeschlagenen Strategien anhand der Elefantendaten miteinander verglichen. Die Elefantendaten von Oktober (Trockenzeit) 2013 waren erneut die Studienvariable, während die Elefantendaten von April (Regenzeit) 2013 als Hilfsvariable fungierten. Die Resultate zeigen, dass die Wahl einer spezifischen Strategie sowohl von den Eigenschaften der Studienpopulation, als auch von der Stärke und der Richtung der Korrelation zwischen der Studienvariable und der Hilfsvariable abhängt (falls vorhanden).

Ein Bereich des ACS Designs, in welchem immer noch Forschungsbedarf besteht, ist die Anwendung im Feld, besonders in Hinblick auf Tierpopulationen. Dieser Umstand kann nur zum Teil den Herausforderungen im Bereich der Feldimplementierung zugeschrieben werden, wie sie zum Teil in Kapitel 2.3 beschrieben wurden. Green et al. (2010) hingegen liefert neue Erkenntnisse wie das ACS Design während einer Lufterhebung angewendet werden kann - beispielsweise wie das Flugzeug sich drehen soll während benachbarte Einheiten erhoben werden.

Die Kostenreduktion im Rahmen einer Erhebung stellt einen Kernpunkt der vorliegenden Dissertation dar. Eine solche ist an der verringerten Anzahl an Elementen in der finalen Stichprobe zu erkennen. Dies wurde wiederum durch die Verwendung eines Abbruchkriteriums, aber auch durch Stratifizierung und der Begrenzung von Netzwerken an Stratumgrenzen erreicht. Ebenso lassen sich die Kosten eines Surveys dadurch verringern, dass alle Einheiten nur ein einziges Mal observiert werden (durch das Ziehen von Clustern ohne Zurücklegen).

Die Kosten, eine Randeinheit zu erheben, werden als gering angenommen. In diesem Fall ist das ACS Design effizienter als das nicht adaptive Design (Thompson and Collins, 2002). Dieses Ergebnis trifft allerdings nicht für Lufterhebungen zu. Da das Flugzeug mit einer konstanten Geschwindigkeit, sowie Höhe fliegt, sind die Kosten eine Randeinheit zu erheben identisch mit denen einer interessierenden Einheit (Norton-Griffiths, 1978). In solchen

Erhebungen spielt das Ziehen von Clustern ohne Zurückzulegen eine wesentlich größere Rolle im Bereich der Kostenreduktion. Andere zentrale Punkte, welche die Kapitel dieser Dissertation motivierten, sind Effizienzzugewinne (in allen Kapiteln) und die Praktikabilität der Designs unter den spezifischen Rahmenbedingungen.

Obwohl der Fokus der vorliegenden Arbeit auf dem Schätzen von Tierpopulationen liegt, können die vorgestellten Methoden auch in allen anderen Populationen umgesetzt werden, welche in kleiner Zahl und geclustert vorkommen. Als alternative Anwendungsfelder bieten sich demnach die Forstwirtschaft, Pflanzen, Umweltverschmutzung, Mineralien und Ähnliches an.

Bibliography

- Acharya, B., Bhattarai, G., De Gier, A., and Stein, A. (2000). Systematic adaptive cluster sampling for the assessment of rare tree species in nepal. *Forest Ecology and Management*, 137(1):65–73.
- Bibby, C. J., Burgess, N. D., and Hill, D. A. (1992). *Bird census techniques*. Academic Press, London.
- Brown, J. A. (2003). Designing an efficient adaptive cluster sample. *Environmental and Ecological Statistics*, 10(1):95–105.
- Brown, J. A. and Manly, B. J. F. (1998). Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, 5(1):49–63.
- Brown, J. A., Moradi, M. S., Panahbehagh, B., and Smith, D. R. (2013). Adaptive survey design for sampling rare and clustered populations. *Mathematics and Computers in Simulations*, 93:108–116.
- Chambers, R. L. and Clark, R. G. (2012). *An introduction to model-based survey sampling with applications*. Oxford University Press.
- Chambers, R. L., Steel, D. G., Wang, S., and Welsh, A. H. (2012). *Maximum likelihood estimation for sample surveys*. CRC Press.
- Chao, C.-T. (2004a). Ratio estimation on adaptive cluster sampling. *Journal of Chinese Statistical Association*, 42(3):307–327.
- Chao, C.-T. (2004b). Regression estimation in adaptive cluster sampling; estimator, confidence interval and sample selection: Report.
- Chao, C.-T., Dryver, A. L., and Chiang, T.-C. (2011). Leveraging the rao–blackwell theorem to improve ratio estimators in adaptive cluster sampling. *Environmental and Ecological Statistics*, 18(3):543–568.
- Chaudhuri, A. and Stenger, H. (2005). *Survey sampling: theory and methods*. CRC Press.

-
- Chaudhury, M. S. and Hanif, M. (2015). Generalized exponential-cum-exponential estimator in adaptive cluster sampling. *Pakistan Journal of Statistics and Operation Research*, 11(4):553–574.
- Christman, M. C. and Lan, F. (2001). Inverse adaptive cluster sampling. *Biometrics*, 57:1096–1105.
- Chutiman, N., Chiangpradit, M., and Suraphee, S. (2013). A new estimator using auxiliary information in stratified adaptive cluster sampling. *Open Journal of Statistics*, 3:278–282.
- Connors, M. E. and Schwager, S. J. (2002). The use of adaptive cluster sampling for hydroacoustic surveys. *ICES Journal of Marine Science*, 59(6):1314–1325.
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press.
- Dryver, A. L. and Chao, C.-T. (2007). Ratio estimators in adaptive cluster sampling. *Environmetrics*, 18(6):607–620.
- Dryver, A. L. and Thompson, S. K. (2007). Adaptive sampling without replacement of clusters. *Statistical Methodology*, 4(1):35–43.
- Gabler, S. (2007). Adaptive and spatial sampling designs. *Encyclopedia of Statistics in Quality and Reliability*, pages 850–853.
- Gabler, S., Ganninger, M., and Münnich, R. T. (2012). Optimal allocation of the sample size to strata under box constraints. *Metrika*, 75(2):151–161.
- Gattone, S. A. and Di Battista, T. (2011). Adaptive cluster sampling with a data driven stopping rule. *Statistical methods & applications*, 20(1):1–21.
- Gattone, S. A., Mohamed, E., and Di Battista, T. (2016a). Adaptive cluster sampling with clusters selected without replacement and stopping rule. *Environmental and Ecological Statistics*, 23:453–468.

-
- Gattone, S. A., Mohamed, E., Dryver, A. L., and Münnich, R. T. (2016b). Adaptive cluster sampling for negatively correlated data. *Environmetrics*, 27:E103–E113.
- Gonçalves, K. C. M. and Moura, F. A. S. (2016). A mixture model for rare and clustered populations under adaptive cluster sampling. *Bayesian Analysis*, 11(2):519–544.
- Goslee, S. C. and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22:1–19.
- Green, M. C., Leberg, P., and Luent, M. (2010). Evaluation of aerial sampling methods for detecting waterbird colonies. *Journal of Field Ornithology*, 81(4):411–419.
- Haselmayer, J. and Quinn, J. S. (2000). A comparison of point counts and sound recording as bird survey methods in amazonian southeast peru. *The Condor*, 102(4):887–893.
- Hedges, S. (2012). *Monitoring elephant populations and assessing threats*. University Press: India.
- Hijmans, R. J. (2015). raster: Geographic data analysis and modeling. R package version 2.5-2.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer Texts in Statistics.
- Kenana, L. M., Bakari, S. K., Bitok, E., Machoke, N. M., Hamzah, H. K., Mukenka, J., Chepkwony, R. K., and Mwiu, S. N. (2013a). Total aerial count for Amboseli-West Kilimanjaro and Magadi-Natron cross border landscape, April 2013: A technical report for KWS, Kenya and TAWIRI, Tanzani.
- Kenana, L. M., Bakari, S. K., Bitok, E., Machoke, N. M., Hamzah, H. K., Mukenka, J., Mwiu, S. N., Kyale, D. M., and Cheptei, J. (2013b). Large

- mammal aerial total count for Amboseli-West Kilimanjaro and Magadi-Natron cross border landscape, October 2013: A technical report for Kenya Wildlife Service (KWS), Kenya and Tanzania Wildlife Research Institute (TAWIRI), Tanzani.
- Khaemba, W. M. and Stein, A. (2000). Use of gis for a spatial and temporal analysis of kenyan wildlife with generalised linear modelling. *International Journal of Geographical Information Science*, 14(8):833–853.
- Kissling, M., Kuletz, K., and Brockmann, S. (2007). Distribution and abundance of kittlitz’s murrelets along the outer coast of glacier bay national park and preserve. In *Proceedings of the Fourth Glacier Bay Science Symposium (JF Piatt and SM Gende, eds.)*. US Geological Survey, Reston, VA. *Scientific Investigations Report*, volume 5047, pages 124–128.
- Lin, F.-M. and Chao, C.-T. (2014). Variances and variance estimators of the improved ratio estimators under adaptive cluster sampling. *Environmental and Ecological Statistics*, 21(2):285–311.
- Moradi, M. M., Brown, J. A., and Guo, W. (2014). Adaptive cluster sampling in two-stage sampling. *Australian and Newzealand Journal of Statistics*, 56(4):347–357.
- Münnich, R. T., Sachs, E. W., and Wagner, M. (2012). Numerical solution of optimal allocation problems in stratified sampling under box constraints. *AStA Advances in Statistical Analysis*, 96(3):435–450.
- Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhyā: The Indian Journal of Statistics*, 18(3/4):379–390.
- Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(1):69–74.
- Naddeo, S. and Pisani, C. (2005). Two-stage adaptive cluster sampling. *Statistical Methods and Applications*, 14(1):3–10.

-
- Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625.
- Norton-Griffiths, M. (1978). Counting animals. handbooks on techniques currently used in african wildlife ecology. *African Wildlife Leadership Foundation, Nairobi*.
- Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8(4):1087–1114.
- Philippi, T. (2005). Adaptive cluster sampling for estimation of abundances within local populations of low-abundance plants. *Ecology*, 86(5):1091–1100.
- Prins, H. (1996). *Ecology and behaviour of the African buffalo: social inequality and decision making*. Chapman & Hall Wildlife Ecology and Behaviour Series.
- Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association*, 51(274):269–284.
- Rapley, V. E. and Welsh, A. H. (2008). Model-based inferences from adaptive cluster sampling. *Bayesian Analysis*, 3(4):717–736.
- Rocco, E. (2008). Two-stage restricted adaptive cluster sampling. *Metron-International Journal of Statistics*, 66(3):313–327.
- Roesch, F. A. (1993). Adaptive cluster sampling for forest inventories. *Forest Science*, 39(4):655–669.
- Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., and Carter, M. F. (2002). Landbird counting techniques: current practices and an alternative. *The Auk*, 119(1):46–53.
- Rowlingson, B. and Diggle, P. (2015). *splanes: Spatial and Space-Time Point Pattern Analysis*. R package version 2.01-37.

-
- Salehi, M. M., Moradi, M., Al Khayat, J. A., Brown, J., and Yousif, A. E. M. (2015). Inverse adaptive cluster sampling with unequal selection probabilities: Case studies on crab holes and arsenic pollution. *Australian & New Zealand Journal of Statistics*, 57(2):189–201.
- Salehi, M. M. and Seber, G. A. (2002). Theory & methods: Unbiased estimators for restricted adaptive cluster sampling. *Australian & New Zealand Journal of Statistics*, 44(1):63–74.
- Salehi, M. M. and Seber, G. A. F. (1997a). Adaptive cluster sampling with networks selected without replacement. *Biometrika*, 84(1):209–219.
- Salehi, M. M. and Seber, G. A. F. (1997b). Two-stage adaptive cluster sampling. *Biometrics*, 53(3):959–970.
- Seber, G. A. F. and Salehi, M. M. (2013). *Adaptive Sampling Designs*. Springer.
- Singh, M. P. (1967). Multivariate product method of estimation for finite populations. *Journal of the Indian Society of Agricultural Statistics*, 31:375–378.
- 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*. Island Press.
- Smith, D. R., Conroy, M. J., and Brakhage, D. H. (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, 51(2):777–788.
- Särndal, C. E., Swesson, B., and Wretman, J. H. (1992). *Model assisted survey sampling*. Springer: New York.
- Su, Z. and Quinn II, T. J. (2003). Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. *Environmental and Ecological Statistics*, 10(1):17–41.

-
- Talvitie, M., Leino, O., and Holopainen, M. (2006). Inventory of sparse forest populations using adaptive cluster sampling. *Silva Fennica*, 40(1):101–108.
- Thompson, S. and Seber, G. (1996). *Adaptive Sampling*. Wiley, New York.
- Thompson, S. K. (1996). Adaptive cluster sampling based on order statistics. *Environmetrics*, 7(2):123–133.
- Thompson, S. K. (2012). *Sampling*. John Wiley & Sons.
- Thompson, S. K. and Collins, L. M. (2002). Adaptive sampling in research on risk-related behaviors. *Drug and Alcohol Dependence*, 68:57–67.
- Thompson, S. K. and Seber, G. A. F. (1994). Detectability in conventional and adaptive sampling. *Biometrics*, 50(3):712–724.
- Turk, P. and Borkowski, J. J. (2005). A review of adaptive cluster sampling: 1990–2003. *Environmental and Ecological Statistics*, 12(1):55–94.
- Valliant, R., Dorfman, A. H., and Royall, R. M. (2000). *Finite population sampling and inference: A prediction approach*. John Wiley.
- Yang, H., Magnussen, S., Fehrmann, L., Mundhenk, P., and Kleinn, C. (2016). Two neighborhood-free plot designs for adaptive sampling of forests. *Environmental and Ecological Statistics*, 23(2):279–299.

APPENDIX

A. Product Appendix

The proofs below are according Gattone et al. (2016b).

A.1 Derivation of Bias and MSE of product HH estimator

To obtain the bias, the variance and the MSE we write

$$e_0 = \frac{\hat{\mu}_{hh_y} - \mu_y}{\mu_y} \Rightarrow \hat{\mu}_{hh_y} = \mu_y(e_0 + 1)$$

$$e_1 = \frac{\hat{\mu}_{hh_x} - \mu_x}{\mu_x} \Rightarrow \hat{\mu}_{hh_x} = \mu_x(e_1 + 1)$$

then

$$\begin{aligned} E(e_0) &= E(e_1) = 0, & E(e_0 e_1) &= \rho_{hh_{xy}} \frac{\sqrt{v(\hat{\mu}_{hh_y})v(\hat{\mu}_{hh_x})}}{\mu_y \mu_x} \\ E(e_0^2) &= \frac{v(\hat{\mu}_{hh_y})}{\mu_y^2}, & E(e_1^2) &= \frac{v(\hat{\mu}_{hh_x})}{\mu_x^2} \end{aligned}$$

where $v(\hat{\mu}_{hh_y})$ and $v(\hat{\mu}_{hh_x})$ are the variances of the HH estimators given in (2.10),

$$\rho_{hh_{xy}} = \frac{cov(\hat{\mu}_{hh_{xy}})}{\sqrt{v(\hat{\mu}_{hh_y})v(\hat{\mu}_{hh_x})}}$$

and

$$cov(\hat{\mu}_{hh_{xy}}) = \frac{N-n}{Nn} \frac{1}{N-1} \sum_{k=1}^N (w_{yk} - \mu_y)(w_{xk} - \mu_x).$$

The bias of $\hat{\mu}_{phh}$ is given by

$$\begin{aligned} B(\hat{\mu}_{phh}) &= E(\hat{\mu}_{phh} - \mu_y) \\ &= \mu_y E[(1 + e_0)(1 + e_1) - 1] \\ &= \mu_y E(e_0 + e_1 + e_0 e_1) \\ &= \mu_y \rho_{hh_{xy}} \frac{\sqrt{v(\hat{\mu}_{hh_y})v(\hat{\mu}_{hh_x})}}{\mu_y \mu_x} \\ &= \frac{cov(\hat{\mu}_{hh_{xy}})}{\mu_x}. \end{aligned} \tag{A.1}$$

while that of $\hat{\tau}_{phh}$ becomes

$$B(\hat{\tau}_{phh}) = \frac{cov(\hat{\tau}_{hh_{xy}})}{\tau_y} \quad (\text{A.2})$$

where

$$cov(\hat{\tau}_{hh_{xy}}) = \frac{N(N-n)}{n} \frac{1}{N-1} \sum_{k=1}^N (w_{yk} - \mu_y)(w_{xk} - \mu_x).$$

The MSE can be obtained by writing

$$\begin{aligned} MSE(\hat{\mu}_{phh}) &= E(\hat{\mu}_{phh} - \mu_y)^2 \\ &= \mu_y^2 E(e_0 + e_1 + e_0 e_1)^2 \\ &= \mu_y^2 E(e_0^2 + e_1^2 + 2e_0 e_1 + 2e_0^2 e_1 + 2e_0 e_1^2 + e_0^2 e_1^2). \end{aligned}$$

To the first order approximation, the MSE of $\hat{\mu}_{phh}$ is

$$\begin{aligned} MSE(\hat{\mu}_{phh}) &\approx \mu_y^2 \left(\frac{v(\hat{\mu}_{hh_y})}{\mu_y^2} + \frac{v(\hat{\mu}_{hh_x})}{\mu_x^2} + 2\rho_{hh_{xy}} \frac{\sqrt{v(\hat{\mu}_{hh_y})v(\hat{\mu}_{hh_x})}}{\mu_y \mu_x} \right) \\ &\approx \mu_y^2 \left(\frac{v(\hat{\mu}_{hh_y})}{\mu_y^2} + \frac{v(\hat{\mu}_{hh_x})}{\mu_x^2} + 2 \frac{cov(\hat{\mu}_{hh_{xy}})}{\mu_y \mu_x} \right) \quad (\text{A.3}) \end{aligned}$$

while the MSE of $\hat{\tau}_{phh}$ is

$$MSE(\hat{\tau}_{phh}) \approx \tau_y^2 \left(\frac{v(\hat{\tau}_{hh_y})}{\tau_y^2} + \frac{v(\hat{\tau}_{hh_x})}{\tau_x^2} + 2 \frac{cov(\hat{\tau}_{hh_{xy}})}{\tau_y \tau_x} \right) \quad (\text{A.4})$$

A.2 Derivation of $MSE(\hat{\mu}_{phh})$ in terms of network population values

Let $w_{y1}, w_{y2}, \dots, w_{yN}$ be the population network means of the y -values with variance $S_{w_y}^2 = \sum_{i=1}^N \frac{(w_{yi} - \mu_y)^2}{N-1}$ and let $w_{x1}, w_{x2}, \dots, w_{xN}$ be the population network means of the x -values with variance $S_{w_x}^2$. Furthermore, define the covariance between these two network populations as $S_{w_{xy}} = \frac{1}{N-1} \sum_{i=1}^N (w_{yi} - \mu_y)(w_{xi} - \mu_x)$. The HH estimators $\hat{\mu}_{hh_y}$ and $\hat{\mu}_{hh_x}$ simply correspond to the sample mean estimator applied to the transformed populations of y and x , respectively (Thompson and Seber, 1996; Dryver and Chao, 2007). By noting

that $v(\hat{\mu}_{hh_y}) = \frac{N-n}{Nn} S_{w_y}^2$, $v(\hat{\mu}_{hh_x}) = \frac{N-n}{Nn} S_{w_x}^2$ and $v(\hat{\mu}_{hh_{xy}}) = \frac{N-n}{Nn} S_{w_{xy}}$, we can rewrite A.3 as follows

$$\begin{aligned} MSE(\hat{\mu}_{phh}) &\approx \frac{N-n}{Nn} \mu_y^2 \left(\frac{S_{w_y}^2}{\mu_y^2} + \frac{S_{w_x}^2}{\mu_x^2} + 2 \frac{S_{w_{xy}}}{\mu_y \mu_x} \right) \\ &\approx \frac{N-n}{Nn} \mu_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) \quad (A.5) \end{aligned}$$

where $CV_{w_y} = \frac{S_{w_y}}{\mu_y}$, $CV_{w_x} = \frac{S_{w_x}}{\mu_x}$ and $\rho_{w_{xy}} = \frac{S_{w_{xy}}}{S_{w_y} S_{w_x}}$ is the correlation at network level between y and x . It follows that

$$MSE(\hat{\tau}_{phh}) \approx \frac{N-n}{Nn} \tau_y^2 \left(CV_{w_y}^2 + CV_{w_x}^2 + 2\rho_{w_{xy}} CV_{w_y} CV_{w_x} \right) \quad (A.6)$$

A.3 Relative efficiency

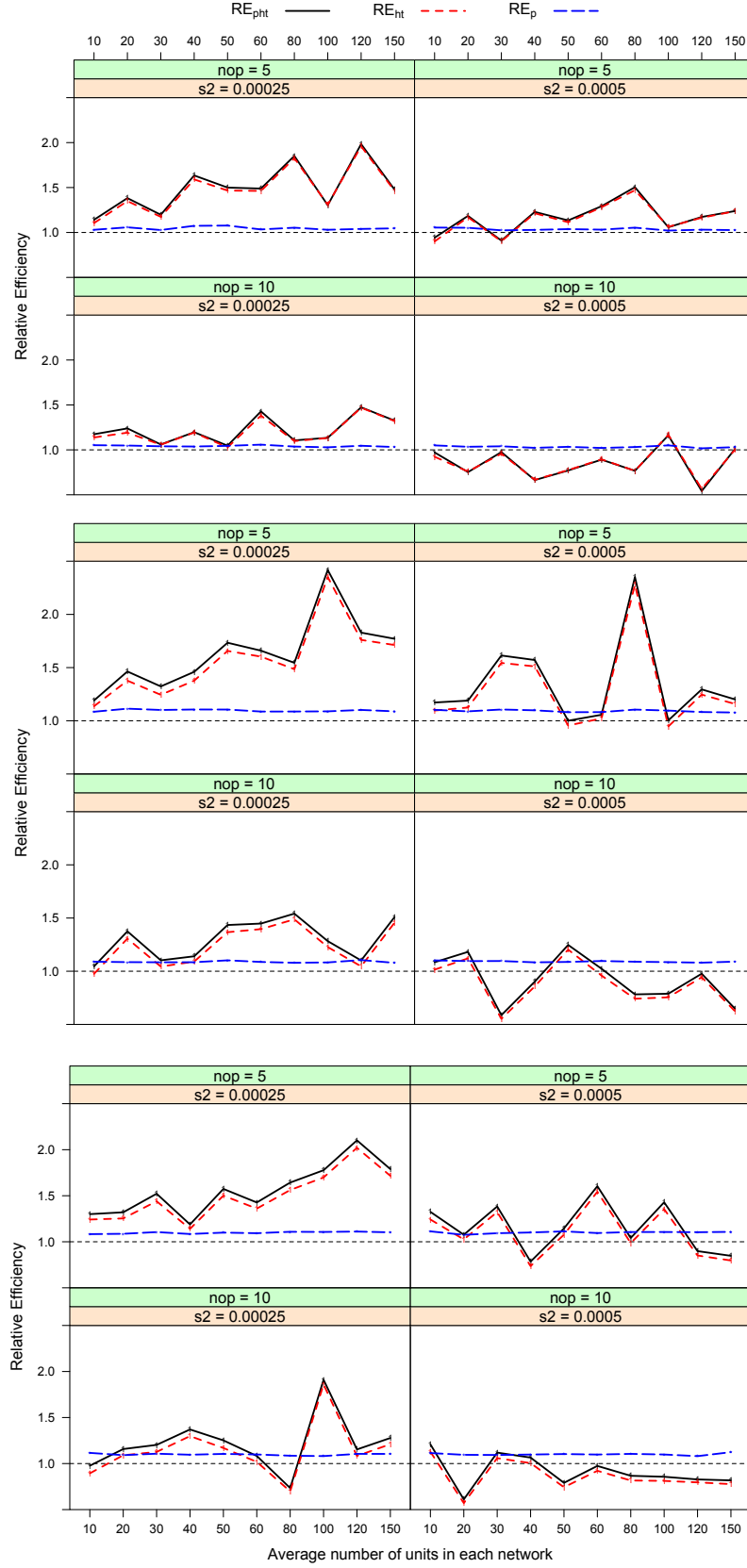


Fig. A.1: Relative efficiency when HT is used for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative

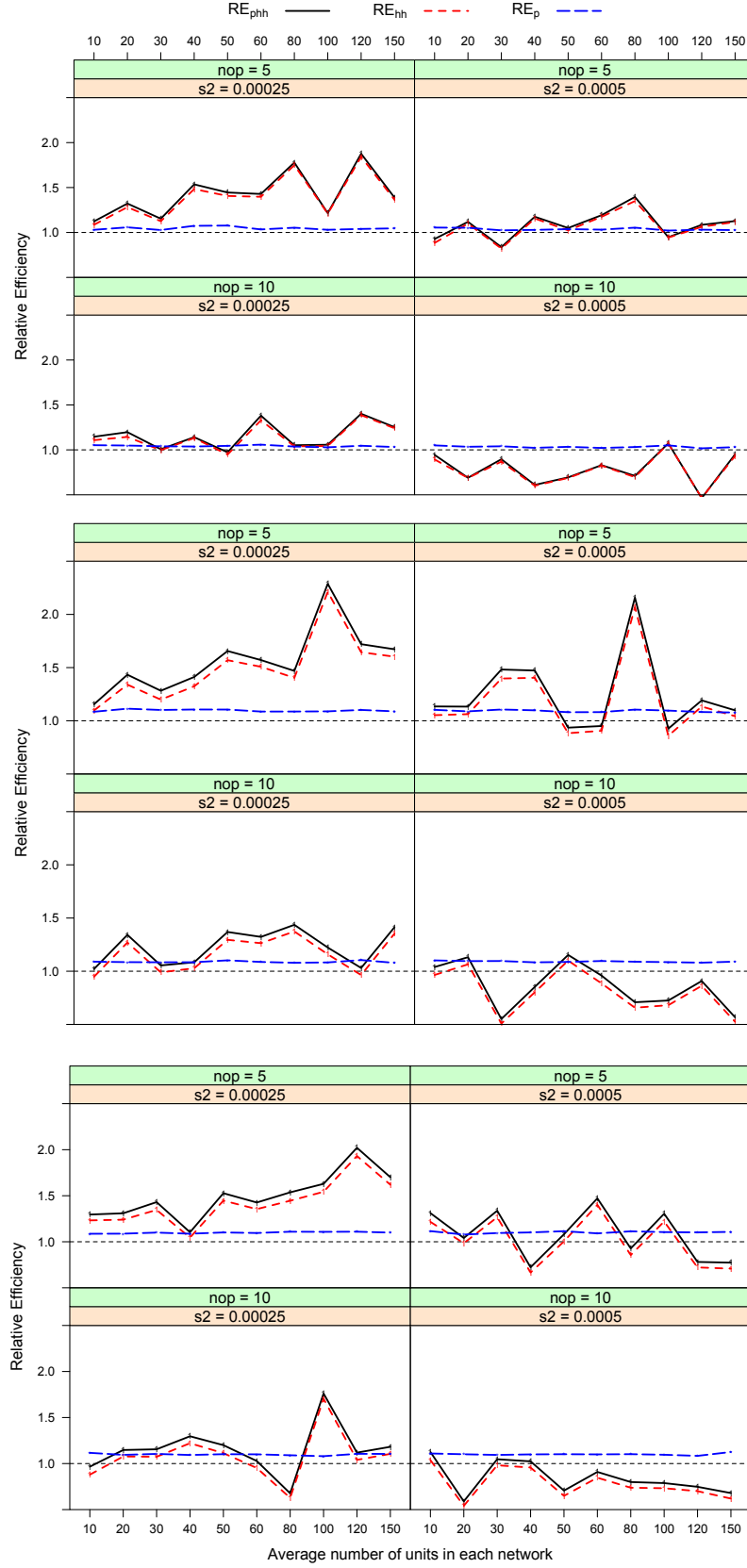


Fig. A.2: Relative efficiency when product HH is used for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative

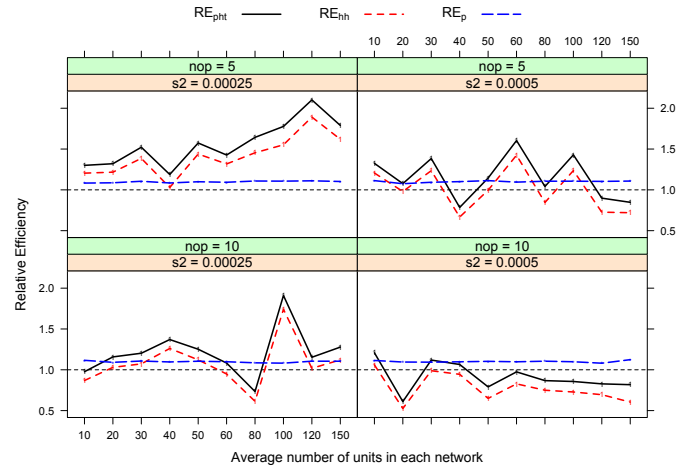


Fig. A.3: Results of the relative efficiency between the product HT, the product HH and the product SRSWOR for an initial sample of size $n = 15$ and a high level of correlation coefficient.

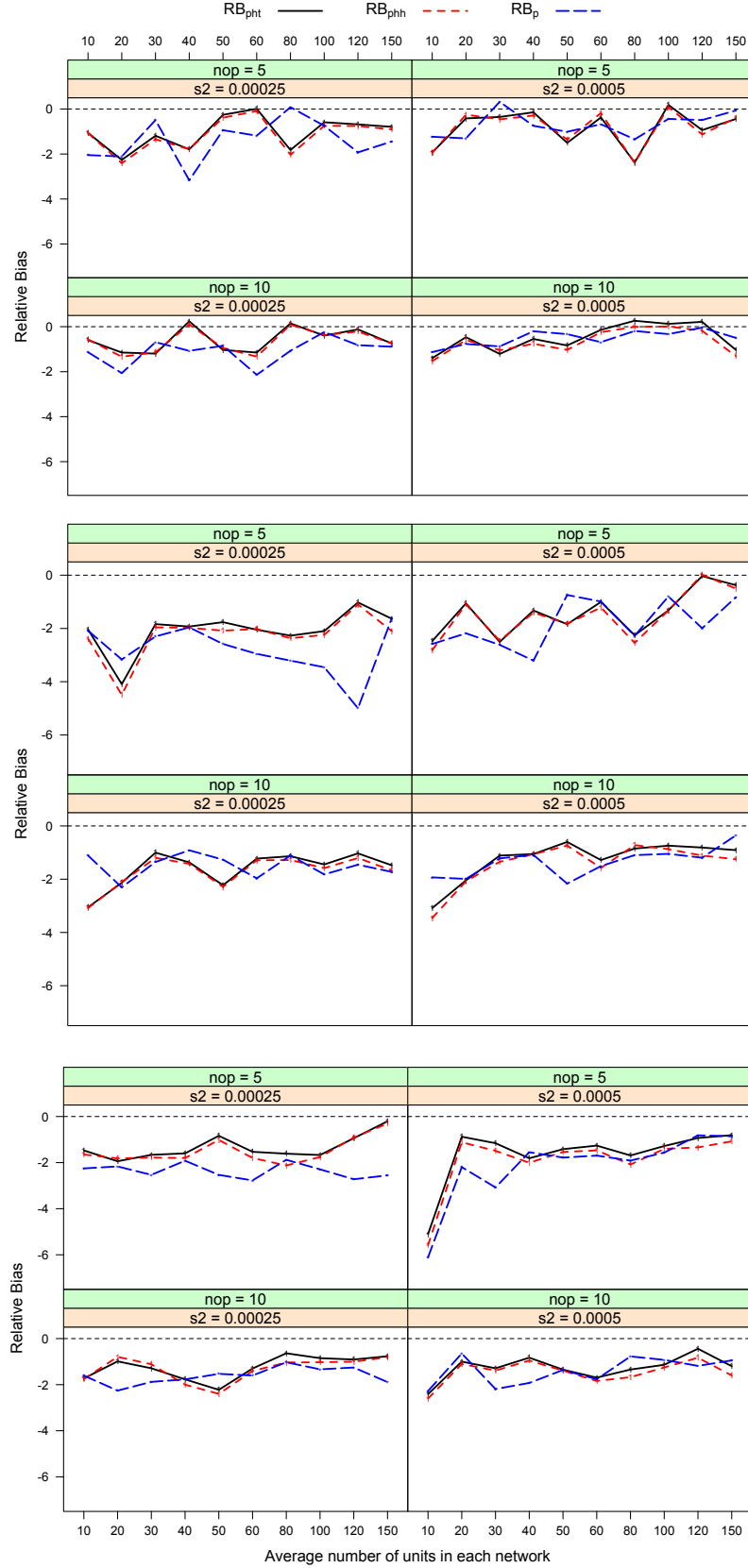


Fig. A.4: Relative bias for an initial sample of size $n = 15$ for the different parameters. Top plot is when $\rho_{w_{xy}}$ is low but negative, middle and bottom plots are for when $\rho_{w_{xy}}$ is intermediate and highly negative

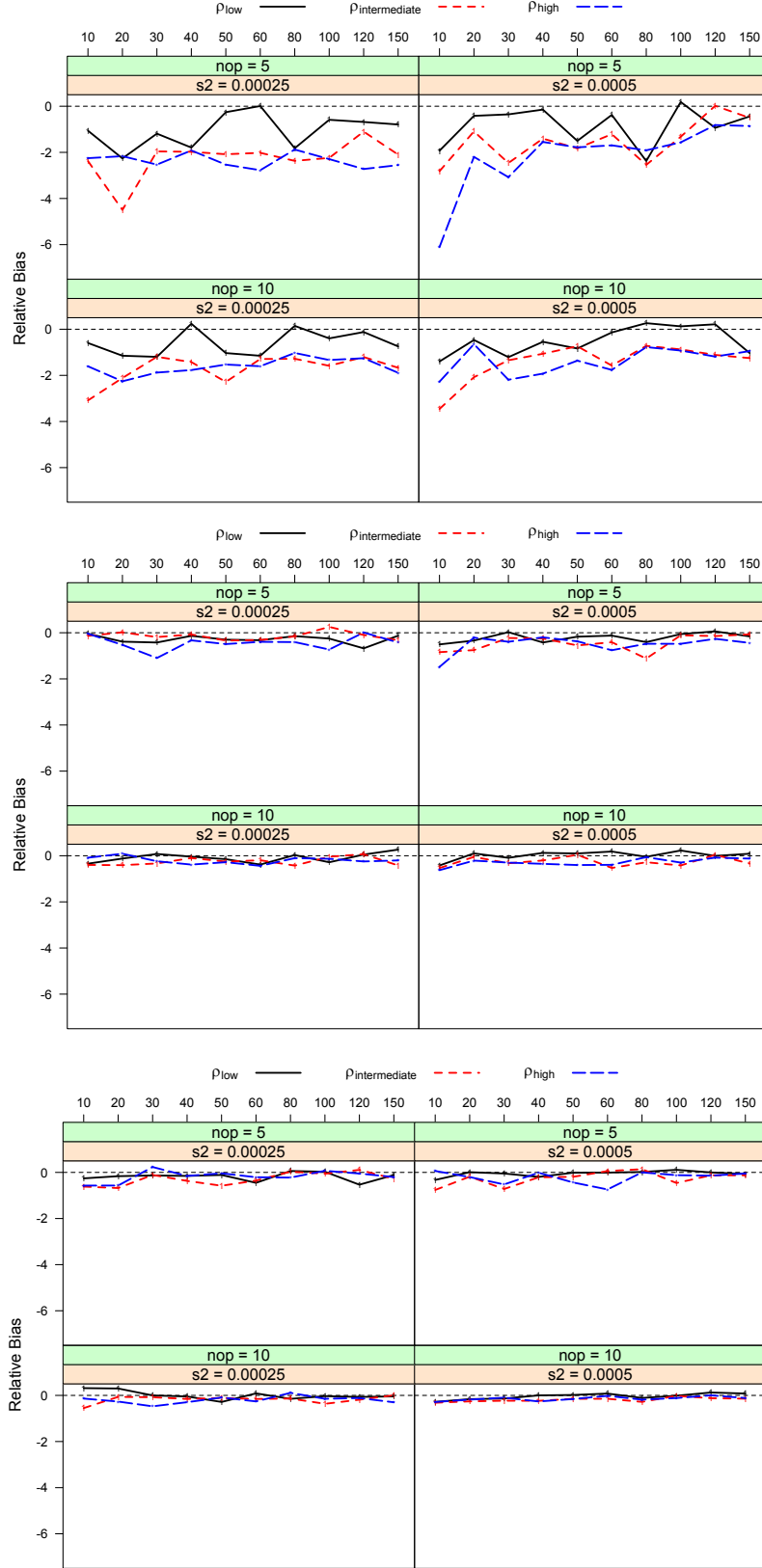


Fig. A.5: Relative bias when the initial sample of size $n = 15, 65$ and 100 (top, middle and bottom plot respectively) for the different parameters and different levels of correlation coefficient.

B. MCMC Diagnostics

The current appendix presents MCMC diagnostics for the artificial and the real data. This is done by employing a combination of methods to determine if the MCMC algorithms have converged and mixed well.

B.1 Gelman-Rubin convergence diagnostics

As already described in the chapter, two parallel chain were ran for the real data. The Gelman-Rubin convergence diagnostics is such that, when j chains are run with different starting values, the scale reduction factor is

$$\sqrt{\hat{R}} = \sqrt{\frac{\text{Total variance}}{\text{Within variance}}}$$

where the numerator is the total variance for the combined chain while the denominator is the variance within each chain. If the MCMC chain has converged, then $\sqrt{\hat{R}}$ is approximately one. Figure B.1 shows the resulting plots.

B.2 Autocorrelation of parameters

A plot of autocorrelation function is used to determine if the samples drawn from the posterior distribution are independent. If the plot of autocorrelation does not display any pattern, it implies that samples drawn maybe independent. Figure B.4 shows Autocorrelation Function (ACF) plots for the artificial data while figure B.5 shows ACF plots for the two parallel chains of the elephant data. The two figures do not show any pattern implying that the samples drawn are independent or simply there is absence of autocorrelation of parameters.

B.3 Trace Plots

Trace plots are the commonly used graphical techniques to monitor convergence. Figure B.2 shows trace plots for the artificial data while figure

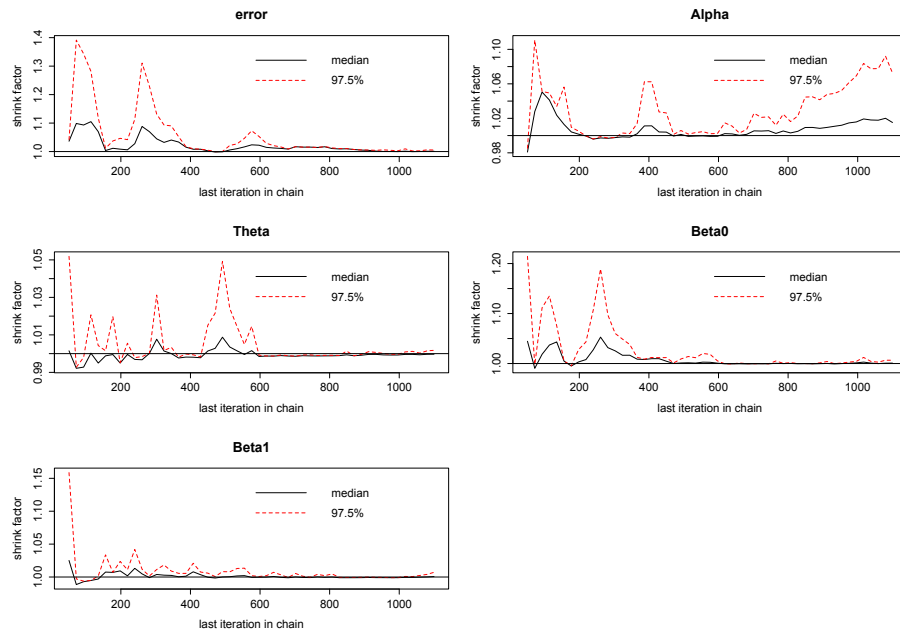


Fig. B.1: Gelman-Rubin scale reduction factor plots.

B.3 shows trace plots for combined chains for the elephant data. The figures shows that the algorithms may have converged and mixes well.

From the above, it can be concluded that samples drawn from the posterior distributions maybe independent and that the chains may have converged and mixed well; inferences can thus be drawn.

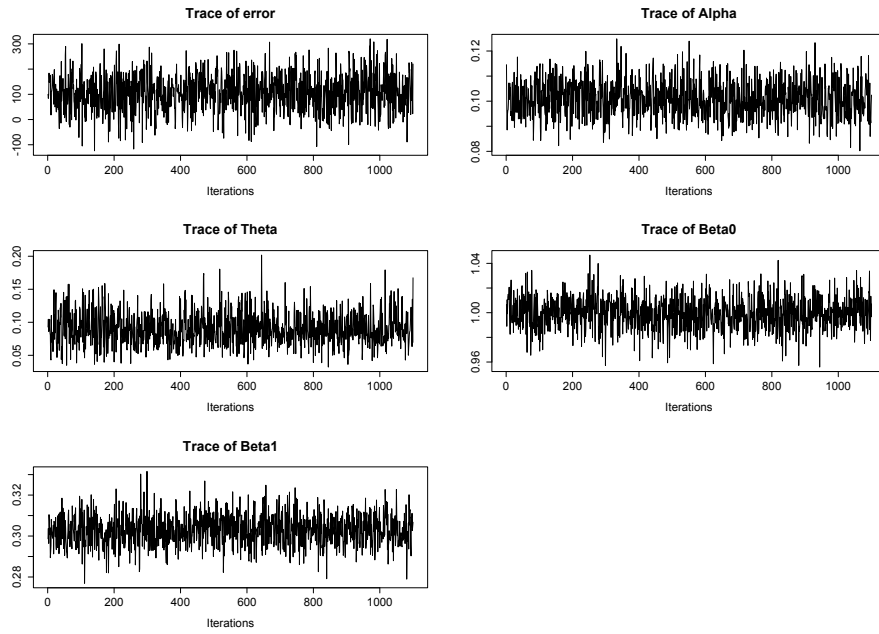


Fig. B.2: Trace plots for the artificial data.

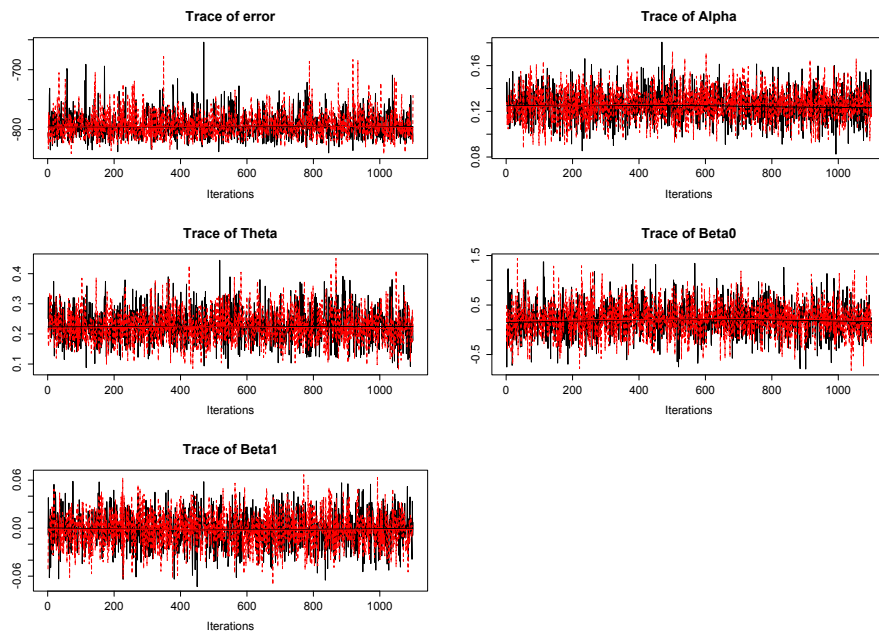


Fig. B.3: Trace plots for combined chains for the elephant data.

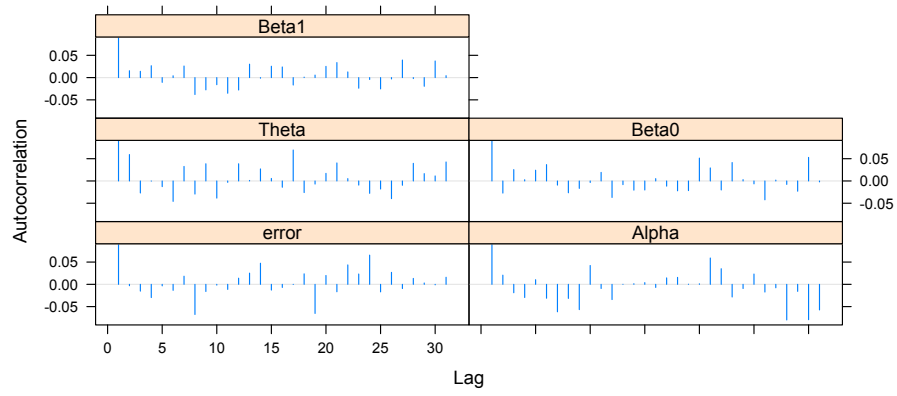


Fig. B.4: Plots of the autocorrelation function.

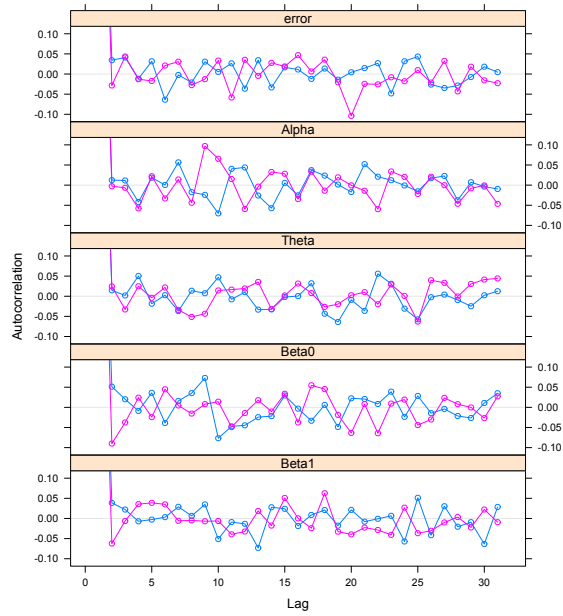


Fig. B.5: Autocorrelation function for the elephant data.