THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

ESTIMATING BALD EAGLE POPULATION SIZE USING DUAL FRAME SAMPLING TECHNIQUES

DAWN E. HAINES AND KENNETH H. POLLOCK

INSTITUTE OF STATISTICS MIMEO SERIES NO. 2287

JUNE 28, 1996

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

Estimating Bald Eagle Population Size Using Dual Frame Sampling Techniques

Dawn E. Haines and Kenneth H. Pollock
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Institute of Statistics Mimeo Series No. 2287

June 28, 1996

Abstract

Monitoring the number of animals in any wildlife population is an important task. Studying the bald eagle population is particularly worthwhile for a variety of reasons. In addition to being our national symbol and having threatened status under the Endangered Species Act of 1973, bald eagles serve as good indicators of environmental health. To estimate the number of bald eagles in a specific area, we recommend the implementation of a dual frame sampling technique widely used in agricultural surveys by the U. S. Department of Agriculture. This methodology combines an incomplete list frame with an area frame that is assumed to be complete. A key assumption in this scenario is that the list and area frame sampling schemes are independent. Hartley's "screening" estimator is used to combine sample information from both frames. The quantities estimated include both the number of active and successful nests along with their standard errors. The properties and advantages of the dual frame method are discussed. Finally, an example illustrating the methodology is presented.

KEY WORDS: Sample, Estimation, Screening Estimator, Multiple Frame Sampling

1 INTRODUCTION

The bald eagle (*Haliaeetus leucocephalus*) is one of the most widely recognized birds in the United States. The eagle, regarded by most Americans as our national symbol, is currently listed as threatened under the Endangered Species Act of 1973. Therefore, populations of bald eagles are widely studied. One reason for continued monitoring of these populations is that bald eagles are good indicators of ecosystem health as well as overall water quality. In particular, eagles nesting around large bodies of water in the United States are being examined for the purpose of measuring several environmental and ecological variables (Bartish 1994).

Deciding how to measure the total well-being of the environment is not a simple task. However, justification for including bald eagles as an indicator of the health of a region is overwhelming. First of all, eagles are at the top of the food chain and depend heavily on a diet consisting of aquatic prey. Toxic substances existing in the aquatic habitat eventually poison the inhabitants. Consequently, bald eagles ingest the poisoned prey and become contaminated themselves. Eagles are known to be very sensitive to many contaminants, especially over an extended period of time. Eagle nestlings, which are young birds not yet ready to leave the nest, are not free from the influence of the toxins. In fact, evidence of toxic compounds has been detected in unhatched bald eagle eggs in addition to nestling plasma samples. Obvious responses to contamination include egg shell thinning and decreased productivity.

Assessment of the health of the overall ecosystem using eagles is necessarily based on several variables. First, eagle productivity and mortality rates are measured. Contaminant burden is another plausible variable to include in our study. Most contamination data are collected on dead eagles, egg shells, and plasma samples. Other data may be collected on contaminant exposure, habitat suitability, and human disturbance pressures but will not play a role in our bald eagle population study. Rather, our study is primarily concerned with estimating the total number of active eagle nests in a predefined region. In addition, we consider estimating the number of successful nests in this population. Note that active nests are defined as nests presently in use by one or more eagles, whereas a successful nest is a nest that produces at least one fledgling. General trends in the population, such as those that may be observed in reproductive performance, are usually evaluated and reported. Although the bald eagle is a desirable species for measuring environmental health, the methodology developed here may be applied to other wildlife populations where the breeding individuals use consistent and highly visible nesting territories. Examples include herring gulls, hawks, osprey, and other raptors. These species also have similar feeding habits to those of the bald eagle and may accumulate toxins in a comparable fashion.

Developing a sampling frame is the first step in any probability sampling scheme. Two types of sampling frames exist and are commonly known as list and area frames. The most significant information contained on a list frame includes a written description of eagle nest locations. Other quantities of interest may be ascertained and, hence, included on the list frame. A list of nest locations from the previous year is typically available although it will never be complete. In contrast to a list frame, an area frame is used to describe the geographic boundaries of a population of eagle nests. Eagle populations are known to physically occupy territories or regions whose exact surface areas are known. To successfully implement an area frame, the population of bald eagles is partitioned into sampling units, or subareas. A simple random sample of subareas is then selected from the population. In addition to simple random sampling, it is common to perform stratified random sampling. Stratification of the bald eagle population is based on habitat type, geography, or another variable of interest. Once stratification is complete, each stratum is divided into subareas. Finally, a simple random sample of subareas is drawn from each stratum. Each subarea in the stratified random sample is then personally examined and characteristics relevant to the study are recorded. Of primary interest is the number and location of active nests. For each active nest discovered via enumeration, the enumerator records whether it was an element of this year's list frame. Further, observed active nests are categorized as successful or not successful.

The statistical literature discusses several advantages and disadvantages of both list and area frames. The list frame is a physical listing of sampling units in the target population. Lists routinely collected for some purpose are usually found to be incomplete, partially illegible, or to contain an unknown amount of duplication. List frames are relatively inexpensive to construct and are more efficient to sample than area frames. In contrast, an area frame is a collection of geographical areas defined by easily identifiable boundaries. Area frames are often utilized because they provide complete coverage of the target population. As a result, every sampling unit has positive probability of selection. Area frames cost more to develop than list frames. The reason for this pertains to the fact that area frames require personal enumeration whereas list frame enumeration can be conducted via mail, telephone, or electronically. One desirable trait of area frames is that they require very little maintenance and, hence, are relatively inexpensive to maintain. Major frame revisions are needed only when geographic characteristics substantially change, which is not common in practice. Additionally, once an area frame is constructed, it lasts many years. These favorable qualities along with advanced developments in aerial photography and satellite imagery necessitate the implementation of area frames.

Instead of favoring one type of frame over the other, our approach is to use a joint list-area sampling design to achieve full coverage of the target population. Multiple frame survey methodology combines one or more incomplete list frames with an area frame that is assumed to be complete. The purpose of this combination of frames is to offset the weaknesses of each individual frame with the strengths of the other frames (Kott and Vogel 1995). Since sampling from the incomplete list frame is conducted at such low cost, a large percentage of sampling units are selected from the list. As a result, the list frame portion of the dual frame estimator provides an efficient method of obtaining precise estimates. A much smaller proportion of sampling units are selected from the area frame due to increased enumeration costs. However, the area frame offers a means to measure the incompleteness of the list frame. In this paper, we illustrate how to combine the list and area frame information to efficiently estimate the number of active eagle nests and the total number of successful nests in the target population. In Section 2, we describe the dual frame method and discuss different estimation methods. Section 3 presents an example that illustrates the methodology. In conclusion, Section 4 offers a brief, general discussion of the paper.

2 THE DUAL FRAME METHOD

2.1 Background: The Screening Estimator

To apply appropriate list frame methodology to the population of eagle nest locations, the experimenter utilizes the list frame from last year. This list frame is usually available but is inevitably incomplete. The list frame of nests includes a detailed description of both active and inactive eagle nest sites. Since we are interested in locating the nests annually, it is imperative to carefully document the exact location of each nest.

The list frame for this year is constructed by taking last year's list frame and adding to it any active nests discovered during last year's area frame sample. There are several ways for an enumerator to discover nests that are not recorded on last year's list frame. First, the nest discovered may be brand new to the population. This is regarded as a birth and indicates that the nest originated since the last eagle study. Another feasible explanation for the emergence of new nests relates to the incompleteness of last year's list frame. That is, not all nests existing in the study area last year were represented on last year's list frame. Finally, eagle nests are often abandoned. Formerly abandoned nest sites may now be occupied by different pairs of eagles. As a result, a nest may be discovered in last year's area sample but not be present on last year's list frame. Once this year's list frame is constructed, we are ready to select our sample.

One method of sampling is initiated by stratifying this year's list frame on a chosen variable. For example, stratification may be based on geography, habitat features, or other physical characteristics. A simple random sample of nest sites is then drawn from each stratum. Next, each sampled site is personally visited by an enumerator. At this point, the enumerator's job is to classify sampled nest sites as active or inactive.

A number of changes in the nests may have taken place since last year's survey. For instance, nests that were active last year may now be inactive. This is referred to as a death. The nests deemed active are then designated as successful or not in order to estimate the total number of successful nests in the population.

The area frame procedure is applied to a region whose exact surface area is known. As with the list frame, the area frame is stratified on a designated variable. The stratification variables for the list and area frames are not required to be the same. However, stratification may be based on the same variable for both frames. The analysis is greatly simplified in this case. For the bald eagle area frame, it is reasonable to define $3 - 5$ strata based on a chosen variable. Furthermore, each stratum contains explicit sampling units or subareas. Once stratification is performed, subareas are randomly selected from each stratum. The enumerator visits each subarea in the stratified random sample and thoroughly searches for nest sites. Note that we assume enumeration within subareas is conducted without error. That is, we assume all active nests contained within selected subareas are discovered. Moreover, we assume inactive nests existing within subareas are not recorded in the area sample.

Multiple frame sampling is a popular sampling scheme that combines list and area frame information. The technique of multiple frame estimation has been utilized for more than 40 years. Hartley (1962) was the first to theoretically derive the multiple frame methodology. Considerable extensions to this work are due to Cochran (1965), Lund (1968), Fuller and Burmeister (1972), Hartley (1974), and Bosecker and Ford (1976).

Hartley's approach considers two sampling frames, denoted A and B, along with the domains generated by their intersection. In addition, Hartley assumes simple random sampling within each frame and ignores finite population correction factors. In the case of two frames A and B, we define domains a, b, and ab. Domain a consists of those elements belonging only to frame A, and likewise, domain b consists of those elements belonging only to frame B. Domain ab, however, contains units that belong to both frames A and B.

Multiple frame sampling requires two assumptions, namely *completeness* and *identifiability*. *Completeness* requires that every unit in the target population belong to at least one of the frames in the study. This requirement is satisfied simply by using an area frame, which is complete by itself. As a result, for any k frames it is possible to partition the population elements into $2^k - 1$ mutually exclusive domains. The completeness assumption prohibits the existence of a domain with no member elements; hence, the 1 is subtracted in the above formula. In the same manner, we partition sample units into $2^k - 1$ domains. *Identifiability*, on the other hand, is the ability to discern whether or not a sampled unit from one frame could possibly belong to any other frame in the study. Validating the identifiability assumption determines frame overlap. In our context, a process called *unduplication* is used to remove any sampling units from the area sample that were accessed by way of the list frame sample.

To illustrate multiple frame estimation techniques for the bald eagle population, we first introduce appropriate terminology and later define relevant notation. Unduplication is the process by which sampling units in the area sample are separated into two distinct domains, commonly known as the *nonoverlap* and *overlap* domains. This means active nests discovered in subareas by way of personal enumeration are categorized into one of these two domains. In this context, the portion of the area frame referred to as the nonoverlap domain includes nest sites sampled from the area frame that are *not* included in this year's list frame. Consequently, only area sample nests not on this year's list frame contribute to the area portion of the multiple frame estimate. By contrast, the overlap domain contains nests sampled from the area frame that are also included in this year's list frame.

Unduplication is a complex process. Although the method seems simple to employ, it is typically the leading source of nonsampling error in multiple frame sampling (Nealon 1984). To unduplicate a sample, enumerators carry a copy of this year's list frame with them in the field. After contact with each active nest in the sampled subareas, the enumerator records whether or not this nest was included on this year's list. If area sample nests which appear on the list frame are not identified as such, then resulting multiple frame estimates are inflated. Also, difficulties may arise if a reporting unit is not defined and identified in a consistent manner. For the population of eagle nest sites, this particular issue does not appear to be a concern.

In this framework, an appropriate estimator of the population total, Y_S, of a variable of interest is Hartley's "screening" estimator, denoted

$$\widehat{Y}_S = \widehat{Y}_A + \widehat{Y}_L,$$

where $\widehat{Y}_A$ is the area frame estimate of the population total, Y_A, of the variable of interest in the nonoverlap domain, and $\widehat{Y}_L$ denotes the list frame estimate of the population total, Y_L, of the variable of interest in the overlap domain. As a result, we estimate the population total Y_S by combining the individual estimates of Y_A and Y_L. The name screening estimator follows from the fact that elements of the area frame sample are "screened" and included in the estimation process only if they are not found on this year's list frame.

No matter the type of sampling employed in the multiple frame approach, it is vital that we maintain independence of the area and list frame sample selections. When this assumption holds true, the variance of the screening estimator $\widehat{Y}_S$ is given by

$$\text{Var}(\widehat{Y}_S) = \text{Var}(\widehat{Y}_A) + \text{Var}(\widehat{Y}_L).$$

If independence of sampling procedures between the two frames is guaranteed, it is possible for the area frame to estimate the incompleteness of the list frame. Independence, or guarding one frame from the influences of the other, implies that list frame information cannot be used to aid in the discovery of nests in area frame samples. Additionally, it is necessary to ensure that nests found through area frame sampling are not added to the list frame until we construct the subsequent year's list frame. Adding nests to the list frame prematurely can bias the estimation process by effectively changing selection probabilities (Kott and Vogel 1995). Additions to the list frame should only be allowed when an entirely new sample is selected from both frames.

2.2 Estimation of Active Nests: The Screening Estimator

To relate the screening estimator directly to the bald eagle population, we completely enumerate area frame subareas selected in the stratified random sample. As stated earlier, all area frame nests recorded are assumed to be active. Once a nest is found, it is first classified as to whether or not it is on this year's list frame. If the nest is on this year's list frame, then information collected from this nest is not included when estimating Y_A. Hence, entire nests are subject to withdrawal from the nonoverlap domain estimation process. If a given nest was not a member of this year's list frame, we retain that nest in our sample and record the relevant data.

Throughout the eagle survey, it is assumed that this year's list frame is readily available to the practitioner. Thus, the list frame size is a known quantity. Further, it is recognized that the list frame is not complete and that yearly maintenance is essential. As a result, the screening estimator previously defined must be modified to account for these list frame deficiencies.

In classical sampling theory, a population total estimator is defined as the product of the population size and the sample mean of measurements taken on a variable of interest from elements belonging to that population. In our example, *we are initially interested in estimating the total number of active nests on both the area and list frames.* This is akin to estimating population totals where the nonoverlap and overlap domains constitute the two populations, respectively. We first consider the case where stratification is not used in either the area or list frame sampling schemes. An estimator of the total number of active nests in the nonoverlap domain of the area frame is

$$\widehat{Y}_A = N_A \cdot \bar{y}_A,$$

where

N_A = number of area sampling units in the area frame,

n_A = number of area sampling units selected,

and

$$\bar{y}_A = \text{sample mean number of active nests in the nonoverlap domain} = \frac{\text{total number of active nests in the nonoverlap domain}}{n_A}.$$

An estimator of the total number of active nests in the list frame is given by

$$\widehat{Y}_L = N_L \cdot \bar{y}_L,$$

where

$$\begin{aligned} N_L &= \text{number of nests on this year's list frame,} \\ n_L &= \text{number of nests selected from this year's list frame,} \end{aligned}$$

and

$$\begin{aligned} \bar{y}_L &= \text{sample mean number of active nests on this year's list frame} \\ &= \frac{\text{total number of active nests in sample from this year's list frame}}{n_L}. \end{aligned}$$

To obtain $\bar{y}_L$, define the indicator variable y_i as

$$y_i = \begin{cases} 1 & i^{th} \text{ nest is active} \\ 0 & i^{th} \text{ nest is inactive} \end{cases},$$

$$i = 1, \ldots, N_L.$$

Since y_i is an indicator variable, note that $\bar{y}_L$ also represents the sample proportion of active nests on this year's list frame.

Combining the area and list frame estimators yields the screening estimator

$$\widehat{Y}_S = N_A \cdot \bar{y}_A + N_L \cdot \bar{y}_L,$$

which offers an estimate of the total number of active nests in the population. We express the variance of the screening estimator as

$$\text{Var}(\widehat{Y}_S) = N_A^2 \text{ Var}(\bar{y}_A) + N_L^2 \text{ Var}(\bar{y}_L).$$

The screening estimator is easily generalized to include multiple strata, and is described in detail in the next subsection.

2.3 Estimation of Other Related Variables

In practice, it may be of interest to estimate variables such as the total number of successful nests, number of young fledged, or contaminant burden among the active nests. Let x_i, $i = 1, \ldots, N_L$, denote the value of the variable of interest and N_L denote the total number of nests on this year's list frame. For example, if we are interested in estimating the total number of successful nests, x_i takes the value 1 if the i^{th} nest is successful, and 0 otherwise. That is,

$$x_i = \begin{cases} 1 & i^{th} \text{ nest is successful} \\ 0 & i^{th} \text{ nest is not successful} \end{cases},$$

$$i = 1, \ldots, N_L.$$

However, if we are interested in estimating the number of young fledged, x_i gives the number of young fledged in the i^{th} nest. As before, let y_i denote 1 or 0, depending on whether the i^{th} nest on the list frame is active or inactive, respectively. Define the random variable

$$z_i = x_i y_i.$$

If x_i is zero for all inactive nests, as in the case of the number of successful nests or young fledged, then z_i is the same as x_i. For the sake of concreteness, let us assume that we are interested in estimating the total number of successful nests.

As previously stated, the total number of successful nests in the population, Z_S, consists of two parts:

$$Z_A = \text{total number of successful nests in the nonoverlap domain,}$$

and

$$Z_L \;=\; \text{total number of successful nests in this year's list frame.}$$

To estimate Z_S, a random sample of size n_A area sampling units is selected from the N_A subareas in the area frame. In addition, an independent random sample of size n_L nests from the list frame is selected. Let $\bar{z}_A$ denote the sample mean number of successful nests that are in the nonoverlap domain of the area samples, and let $\bar{z}_L$ denote the proportion of successful nests among n_L nests selected. That is,

$$\bar{z}_A = \frac{\text{total number of successful nests in the nonoverlap domain}}{n_A},$$

and

$$\bar{z}_L = \frac{\text{total number of successful nests in sample from this year's list frame}}{n_L}.$$

A screening estimator for the total number of successful nests is

$$\begin{aligned} \widehat{Z}_S &= \widehat{Z}_A + \widehat{Z}_L \\ &= N_A \cdot \bar{z}_A + N_L \cdot \bar{z}_L. \end{aligned}$$

The variance of the screening estimator is

$$\begin{aligned} \text{Var}(\widehat{Z}_S) &= N_A^2 \,\text{Var}(\bar{z}_A) + N_L^2 \,\text{Var}(\bar{z}_L) \\ &= N_A^2 \,\frac{\sigma_{zA}^2}{n_A} + \frac{N_L}{n_L}\,(N_L - n_L)\,\sigma_{zL}^2, \end{aligned}$$

where σ_{zA}^2 and σ_{zL}^2 are the population variances of the number of successful nests in the nonoverlap domain and the list frame, respectively. Note that the finite population correction factor (fpc) is utilized only in the list frame portion of the screening estimator. This is because a moderately-sized sample is selected from the list frame, whereas only a small percentage of area sampling units are selected. An unbiased estimator of the variance of $\widehat{Z}_S$ is given by

$$\widehat{\text{Var}}(\widehat{Z}_S) = N_A^2 \frac{s_{zA}^2}{n_A} + \frac{N_L}{n_L}(N_L - n_L)\, s_{zL}^2,$$

where

$$s_{zA}^2 = \frac{1}{n_A - 1}\left[\sum_{i=1}^{n_A} z_{A_i}^2 - n_A \bar{z}_A^2\right],$$

and

$$s_{zL}^2 = \frac{1}{n_L - 1}\left[\sum_{i=1}^{n_L} z_{L_i}^2 - n_L \bar{z}_L^2\right]$$

$$= \frac{n_L}{n_L - 1}\left[\bar{z}_L(1 - \bar{z}_L)\right].$$

Note that z_{A_i} is the number of successful nests in the nonoverlap (NOL) domain of the i^{th} subarea unit selected, whereas z_{L_i} is an indicator variable for the success of the list frame nests. That is,

$$z_{L_i} = \begin{cases} 1 & i^{th} \text{ list frame nest is successful} \\ 0 & \text{otherwise} \end{cases},$$

$$i = 1, \ldots, N_L.$$

One further extension that is applicable to the screening estimator is the stratification of the list and/or the area frame. The stratification may be performed on different variables in different frames. Suppose that the area frame consists of l' strata and the list frame consists of l strata. Stratified random samples are selected from both frames. A screening estimator of Z_S is given by

$$\widehat{Z}_S = \sum_{h=1}^{l'} N_{A(h)} \bar{z}_{A(h)} + \sum_{h=1}^{l} N_{L(h)} \bar{z}_{L(h)}, \tag{1}$$

where

$$\bar{z}_{A(h)} = \frac{\text{total number of successful nests in } NOL \text{ domain of } h^{th} \text{ area frame stratum}}{n_{A(h)}},$$

$$\bar{z}_{L(h)} = \frac{\text{total number of successful nests in sample from } h^{th} \text{ list frame stratum}}{n_{L(h)}},$$

$n_{A(h)}$ = number of area sampling units selected from h^{th} area frame stratum, and

$n_{L(h)}$ = number of nests selected from h^{th} stratum on this year's list frame.

An unbiased estimate of the variance of $\widehat{Z}_S$ is given by

$$\widehat{\mathrm{Var}}(\widehat{Z}_S) = \sum_{h=1}^{l'} N_{A(h)}^2 \frac{s_{zA(h)}^2}{n_{A(h)}} + \sum_{h=1}^{l} N_{L(h)}(N_{L(h)} - n_{L(h)}) \frac{\bar{z}_{L(h)}(1 - \bar{z}_{L(h)})}{(n_{L(h)} - 1)}, \qquad (2)$$

where

$$s_{zA(h)}^2 = \frac{1}{n_{A(h)} - 1} \left[\sum_{i=1}^{n_{A(h)}} z_{A(h)_i}^2 - n_{A(h)} \bar{z}_{A(h)}^2 \right].$$

Note that $\widehat{Y}_S$, the estimator of the total number of active nests, and $\widehat{\mathrm{Var}}(\widehat{Y}_S)$, an unbiased estimator of the variance of $\widehat{Y}_S$, may be obtained using (1) and (2), respectively, where z_i's are replaced by y_i's. In Section 3, we give an example to illustrate the above computations.

3 EXAMPLE

Consider a population consisting of three strata defined in the upper Michigan peninsula. Assume that in this particular study a single variable is used to define three strata for both the area and list frames. That is, $l' = l = 3$. Hypothetical data from such an experiment are presented in Tables 1 and 2.

Table 1: List Frame Data for Upper Michigan Peninsula

h	$N_{L(h)}$	$n_{L(h)}$	No. Active Nests	$\bar{y}_{L(h)}$	No. Successful Nests	$\bar{z}_{L(h)}$
1	40	32	30	0.9375	27	0.84375
2	41	33	29	0.8788	28	0.84848
3	19	16	15	0.9375	13	0.81250

Table 1 contains typical data from a list frame. These numbers represent possible observed values for the eagle population in the upper Michigan peninsula. To make sense of the data, we describe how the numbers are obtained as well as the implications of their values. For strata 1, 2, and 3, the current list frame has 40, 41, and 19 nests, respectively. For each stratum, we take an approximate 80% sample of nests. A large percentage of nests are selected from the list frame since it is relatively inexpensive to locate list frame nests. In this experiment, we have selected 32, 33, and 16 nests from strata 1, 2, and 3, respectively. Of these nests, we observed that 30, 29, and 15 are active. As a result, the sample proportions, $\bar{y}_{L(h)}$, of active nests for the three strata are

$$\bar{y}_{L(1)} = \frac{30}{32} = 0.9375,$$

$$\bar{y}_{L(2)} = \frac{29}{33} = 0.8788,$$

and

$$\bar{y}_{L(3)} = \frac{15}{16} = 0.9375.$$

In addition to recording whether or not a nest is active, the enumerators also recorded whether sampled nests were successful or not. They observed that in the sample of nests selected, 27, 28, and 13 nests were successful in strata 1, 2, and 3, respectively. Therefore,

$$\bar{z}_{L(1)} = \frac{27}{32} = 0.84375,$$

$$\bar{z}_{L(2)} = \frac{28}{33} = 0.84848,$$

and

$$\bar{z}_{L(3)} = \frac{13}{16} = 0.81250.$$

It is important to notice that $\bar{z}_{L(1)}$ here estimates the proportion of successful nests in stratum 1, *not the proportion of active nests in stratum 1 that were successful.* The same holds true for strata 2 and 3.

Table 2: Area Frame Data for Upper Michigan Peninsula

h	$N_{A(h)}$	$n_{A(h)}$	No. Active Nests Per Subarea NOL	No. Successful Nests Per Subarea NOL	$\bar{y}_{A(h)}$	$s_{yA(h)}$	$\bar{z}_{A(h)}$	$s_{zA(h)}$
1	170	9	1,1,1,2,1,0,2,1,1	1,1,0,2,1,0,2,1,1	1.1111	0.601	1.000	0.7071
2	200	10	0,1,2,1,0,1,2,1,1,2	0,1,2,1,0,1,2,0,0,2	1.100	0.738	0.900	0.8760
3	130	7	2,1,1,2,1,0,0	2,1,1,2,1,0,0	1.000	0.816	1.000	0.8165

Table 2 contains similar hypothetical data from area frame sampling. For strata 1, 2, and 3, there are 170, 200, and 130 area sampling units, respectively. We take an approximate 5% sample of subareas within each stratum and carefully enumerate all active nest sites within each sampled subarea. The set of all active nest sites thus enumerated is compared with the current list frame sites to determine the number of active nests in the nonoverlap domain. Also, the number of successful nests in the nonoverlap domain within each sampled subarea is recorded.

From Table 2, we note that 10 active nests not found on this year's list frame are discovered via enumeration of the nine subareas selected in stratum 1. Similarly, 11 and 7 active nests are found in the nonoverlap domain in the subareas selected in strata 2 and 3, respectively. Such large numbers of active nests in the nonoverlap domain emphasize the incompleteness of the current list frame. As the quality (*i.e.* completeness) of the list frame improves, we expect to see a decrease in the number of nests in the nonoverlap domain.

Based on the above data, the screening estimate of the total number of active nests in the population is

$$\widehat{Y}_S = (170 \cdot 1.1111) + (200 \cdot 1.1000) + (130 \cdot 1.0000) + (40 \cdot 0.9375) + (41 \cdot 0.8788) + (19 \cdot 0.9375)$$

$$= (538.887 + 91.3433) = \widehat{Y}_A + \widehat{Y}_L = 630.23.$$

Therefore, our estimate for the total number of active nests is 631, of which 539 are from the nonoverlap domain and 92 are from the list frame. If one uses only the list frame, then an estimate of the total number of active nests is 92, which is seriously biased downwards. An unbiased estimate of the variance of $\widehat{Y}_S$ is

$$\begin{aligned}
\widehat{\text{Var}}(\widehat{Y}_S) &= \left[170^2 \frac{(0.601)^2}{9} + 200^2 \frac{(0.738)^2}{10} + 130^2 \frac{(0.816)^2}{7}\right] \\
&+ \left[\frac{40}{31}(8)(0.9375)(.0625) + \frac{41}{32}(8)(0.8788)(.1212) + \frac{19}{15}(3)(0.9375)(.0625)\right] \\
&= [1159.857 + 2178.576 + 1607.567] + [0.6048 + 1.0917 + 0.2227] \\
&= (4946 + 1.9192) = \widehat{\text{Var}}(\widehat{Y}_A) + \widehat{\text{Var}}(\widehat{Y}_L) = 4948.
\end{aligned}$$

As a result, our estimate of the active number of nests is 631 with an estimated standard error of 70.342. Notice that a large portion of the estimated variance of $\widehat{Y}_S$ is due to the area frame sample. Both the sampling rate (5%) and the variability in the number of active nests in the nonoverlap domain contribute significantly to the large estimated variance of $\widehat{Y}_A$. Also, the high sampling rate (80%) for the list frame and the high percentage of nests in the list frame being active yield a small estimated variance for $\widehat{Y}_L$.

To estimate the total number of successful nests in the population, we use equation (1). The screening estimate of the total number of successful nests in the population is

$$\widehat{Z}_S = (170 \cdot 1.000) + (200 \cdot 0.900) + (130 \cdot 1.000) + (40 \cdot 0.84375) + (41 \cdot 0.84848) + (19 \cdot 0.81250)$$

$$= (480 + 83.975) = \widehat{Z}_A + \widehat{Z}_L = 563.975.$$

Therefore, our estimate for the total number of successful nests is 564, of which 480 are from the nonoverlap domain and 84 are from the list frame. An unbiased estimate of the variance of $\widehat{Z}_S$ is

$$\widehat{\text{Var}}(\widehat{Z}_S) = \left[170^2 \frac{(0.7071)^2}{9} + 200^2 \frac{(0.8760)^2}{10} + 130^2 \frac{(0.8165)^2}{7}\right]$$

$$+ \left[\frac{40}{31}(8)(0.84375)(.15625) + \frac{41}{32}(8)(0.84848)(.15152) + \frac{19}{15}(3)(0.81250)(.1875)\right]$$

$$= [1605.0707 + 3069.504 + 1607.567] + [1.3609 + 1.3178 + 0.5789]$$

$$= (6282.1417 + 3.2576) = \widehat{\text{Var}}(\widehat{Z}_A) + \widehat{\text{Var}}(\widehat{Z}_L) = 6285.4.$$

Thus, an estimate of the total number of successful nests in the population is 564 with an estimated standard error of 79.28.

4 DISCUSSION

The methodology proposed in this paper is not restricted to bald eagle populations. Rather, it is useful for monitoring any wildlife population where the breeding individuals have highly visible nesting territories which are stable over the course of many years. As stated, we conduct our estimate of population totals by observing active nests with breeding birds. Note that this measure automatically underestimates the total number of eagles since it does not incorporate adolescents and non-breeding adults. However, as long as we are able to estimate the breeding portion of the population, we are able to effectively monitor the status of the entire population. Another issue to mention is that this paper only addresses *dual* frame sampling techniques. This means only one area frame and one incomplete list frame are employed in the estimation process. One area of future research will focus on incorporating one area frame with multiple incomplete list frames. The resulting estimator will be called a multiple frame estimator. Capture-recapture techniques may also be utilized to obtain improved estimates of list frame size when there is more than one incomplete list frame available.

References

[1] Tim Bartish, *Design Considerations for Trust Species Monitoring: Bald Eagles and Their Habitat in the Great Lakes Region*, Biomonitoring of Environmental Status and Trends (BEST) Program, December 1994.

[2] R. R. Bosecker and B. L. Ford, *Multiple Frame Estimation with Stratified Overlap Domain*, Proceedings of the Social Statistics Section, American Statistical Association (1976), 219–224.

[3] R. S. Cochran, *Theory and Applications of Multiple Frame Surveys*, Ph.D. thesis, Iowa State University, 1965.

[4] W. A. Fuller and L. F. Burmeister, *Estimators for Samples Selected from Two Overlapping Frames*, Proceedings of the Social Statistics Section, American Statistical Association (1972), 245–249.

[5] H. O. Hartley, *Multiple Frame Surveys*, Proceedings of the Social Statistics Section, American Statistical Association (1962), 203–206.

[6] ______, *Multiple Frame Methodology and Selected Applications*, Sankhyá **36** (1974), 99–118.

[7] P. S. Kott and F. A. Vogel, *Multiple-Frame Business Surveys*, Survey Methods for Businesses, Farms, and Institutions (B. G. Cox, ed.), John Wiley & Sons, 1995, pp. 185–203.

[8] R. E. Lund, *Estimators in Multiple Frame Surveys*, Proceedings of the Social Statistics Section, American Statistical Association (1968), 282–288.

[9] J. P. Nealon, *Review of the Multiple and Area Frame Estimators*, Staff Report 80, U. S. Department of Agriculture, Statistical Reporting Service, Washington, D. C., 1984.