# Design of a Monitoring Program for Northern Spotted Owls[1]

**Jonathan Bart and Douglas S. Robson[2]**

**Abstract:** This paper discusses methods for estimating population trends of Northern Spotted Owls (*Strix occidentalis caurina*) based on point counts. Although the monitoring program will have five distinct components, attention here is restricted to one of these: roadside surveys of territorial birds. Analyses of Breeding Bird Survey data and computer simulations were used to develop recommendations for design of the roadside surveys. An approach known as "lattice sampling," in which some stations are visited annually and other stations are visited less often, may offer some advantages over the more common practice in wildlife surveys of visiting every station once per year. The analyses suggest that an adequate sample of the roadside surveys could be obtained with less than one person-year of effort per year per state, an expenditure well within current efforts for surveying Northern Spotted Owls, and that a minimum of 8 years, and probably at least 10 years of survey data will be required to obtain reliable estimates of long-term population trends.

The Northern Spotted Owl was listed as a threatened species under the Endangered Species Act on July 23, 1990. During the following months, a Recovery Team was appointed and began meeting to discuss the design of a monitoring program to provide reliable information on population trends of Northern Spotted Owls throughout their range. Northern Spotted Owls live in older forests in western Washington and Oregon and northwestern California. They occur at very low densities (median home range size varies across the range from 2,000 to 12,000 acres) and thus are difficult to survey. Substantial resources will probably be expended on the monitoring program and, because the results it produces will provide the basis for making decisions of considerable economic and biological importance, a detailed study of the design is warranted.

The proposed monitoring program was divided into five parts: (1) roadside surveys; (2) studies of floaters; (3) monitoring of activity sites; (4) transmitter studies; and (5) coordination. Here we discuss design of the roadside surveys. The other segments are discussed in the Recovery Plan (USDI 1992).

Two studies, one based on Breeding Bird Survey data (Robbins and others 1986: Sauer and Droege 1990) and the other based on a computer simulation, were used to investigate how frequently stations should be visited, how many should be visited per year, and how long the survey should be continued before attempting to estimate long-term trends. Particular attention was given to a design in which some of the stations are visited annually and the rest are visited every $t$ years with $t > 1$. Such a design may be useful when the trend is small and when high within-site autocorrelation in successive years exists. An illustration of this approach is shown in *table 1* where some of the sites are visited only every third year while others are visited annually; such a design falls into the class known as "lattice sampling designs" (Yates 1960). Although the results were developed for Northern Spotted Owls, the methods could be applied with little alteration to many other species.

## Methods

### *Analyses of Breeding Bird Survey Data*

Breeding Bird Survey data from a 25-year period (1966-1990) for hawks and owls were used as surrogate pilot data for long-term Spotted Owl data in our analysis of optimal sampling for owl trends. We used the following procedure to select several data sets, each consisting of the count data on all routes for all years for one species within one state or province. We asked the Fish and Wildlife Service (FWS) for up to five data sets per species, each having ≥30 routes surveyed per year. They provided us with 31 data sets. We then discarded routes covered in fewer than 20 of the 25 years (inclusion of poorly covered routes can seriously bias trend estimates), and we discarded data sets in which this reduced the number of routes below 20. This process produced 15 data sets for analysis.

The 15 data sets included 7 species and 9 states or provinces (*table 2*). We calculated the mean number of birds per route recorded each year and plotted these means. Periods during which the changes in mean counts were approximately linear were then delineated by eye, and these intervals were used in the analysis. Most intervals were ≥20 years, but 2 intervals were 17 years. The average number of routes per year, during the intervals used from each data set, varied from 20 to 57; the average number of birds per route varied from 0.7 to 5.2. Average annual trend (referred to below as lambda, $\lambda$) was calculated by fitting an exponential function to the data using regression on logarithms of counts. We expressed the results as percent changes (i.e., a $\lambda$ of -1 percent meant that the population declined at an average annual rate of about 1 percent during the interval). The trends varied from -1.6 percent to 6.9 percent (*table 2*) and had an average value of 2.0 percent. Autocorrelation, as indicated by the Durbin-Watson test, was absent in all but one data set.

To determine how many years were needed to obtain reliable estimates of the long-term trends, we selected all possible sets of $k$ sequential years ($k$ = 3-15) from each period in each data set, calculated the estimated percent change per year, and stored the error (estimated trend-true trend). The data were summarized by determining the minimum interval length such that 80 percent of the errors were <0.02, <0.03, and <0.04. The rationale for this procedure was that the main source of concern in using this survey method is that the true trends not be overestimated. Our analyses give

[1] This paper was not presented at the Workshop on Monitoring Bird Populations by Point Counts but is included in this volume because of its interest and value.

[2] Associate Professor of Zoology, Ohio Cooperative Fish and Wildlife Research Unit, Department of Zoology, The Ohio State University, Columbus, OH 43210; and Statistician, 150 MacLaren Street, PH6 Ottawa Ontario, K2P OL2, Canada

**Table 1--*Example of a lattice design in which four sets of routes are visited each year, one annually and the rest every third year***

| | Year | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Route | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | X | X | X | X | X | X | X | X | X | X | X | X |
| 2 | X | | | X | | | X | | | X | | |
| 3 | | X | | | X | | | X | | | X | |
| 4 | | | X | | | X | | | X | | | X |

**Table 2--*Description of Breeding Bird Survey data sets used to estimate number of years required to obtain reliable estimates of long-term trends.***

| Species | State | Years | Average number of routes | Average count/ route | Estimate percent change |
|---|---|---|---|---|---|
| Turkey Vulture | Florida | 1966-90 | 22 | 5.2 | -1.4 |
| | Maryland | 1966-90 | 43 | 4,7 | 3.3 |
| | Ohio | 1966-89 | 24 | 1.5 | 3.1 |
| | Oklahoma | 1967-90 | 23 | 4.0 | [a]0.2 |
| Black Vulture | Alabama | 1966-90 | 28 | 2.1 | 2.1 |
| | Florida | 1966-90 | 21 | 4.7 | 0.1 |
| Red-tailed Hawk | Kansas | 1967-90 | 29 | 2.1 | 1.1 |
| | Oklahoma | 1970-90 | 23 | 1.6 | 3.1 |
| | Wisconsin | 1966-90 | 57 | 0.7 | 4.5 |
| Red-shouldered Hawk | Florida | 1971-90 | 23 | 2.0 | 0.9 |
| American Kestrel | New York | 1974-90 | 46 | 1.4 | -1.6 |
| | Ohio | 1974-90 | 24 | 1.3 | 1.4 |
| | Ontario | 1968-90 | 25 | 1.0 | 3.0 |
| Osprey | Florida | 1966-85 | 20 | 1.0 | 6.9 |
| Great Horned Owl | Kansas | 1967-90 | 29 | 0.7 | 2.2 |

[a] auto-correlation present, based on Durbin-Watson test with $\alpha = 0.05$.

estimates of the sample size requirements when "overestimated" is defined as error of 0.02, 0.03, or 0.04, and the probability of avoiding this error is 80 percent. The rationale for selecting the threshold values 0.02, 0.03, and 0.04 is given below (magnitude of trend that should be detectable).

The procedure is illustrated with data from Red-tailed Hawks (*Buteo jamaicensis*) in Wisconsin. This data set showed an increasing trend throughout the 25-year period (*fig. 1*). The average annual change was 4.5 percent or 0.045. In the 25-year interval, there are 23 different intervals of 3 years each. An estimate of the "true," long-term trend (0.045) was calculated from each of these samples (*table 3*). Row 1 of *table 3* indicates that 57 percent of these estimates were <0.02 higher than the true value (i.e., 57 percent were <0.065); 57 percent were <0.03 higher than the true value, and 61 percent were less than 0.04 higher than the true value. At the opposite extreme, 91 percent of the estimates based on 15-year intervals were <0.02 higher than the true value, and all of these estimates were <0.04 higher than the true value.

### *Computer Simulations*

Computer simulations based on models of population change can also be used to investigate sampling efficiency. Our investigation required specification of an autocorrelation model describing the process by which the survey data would be obtained. We used an analytically tractable Markov chain model.

### *Model Details*

The model in simplest form is defined by the probability ø that a site is initially occupied, the probability $p$ that a site which is occupied in one year will be occupied the next year, and the probability $r$ that a site which is not occupied in one year will be occupied the next year. These define a Markov chain on the two states of nature, "1" for occupied and "0" (*table 4*).

The expected proportion of occupied sites converges to a limiting value, $r/(1\ p+r)$, independent of the initial fraction ø of occupied sites, so if the initial occupancy rate is less than this equilibrium value, then an upward trend occurs. As an illustration of this phenomenon, the trend over an 8-year peri-

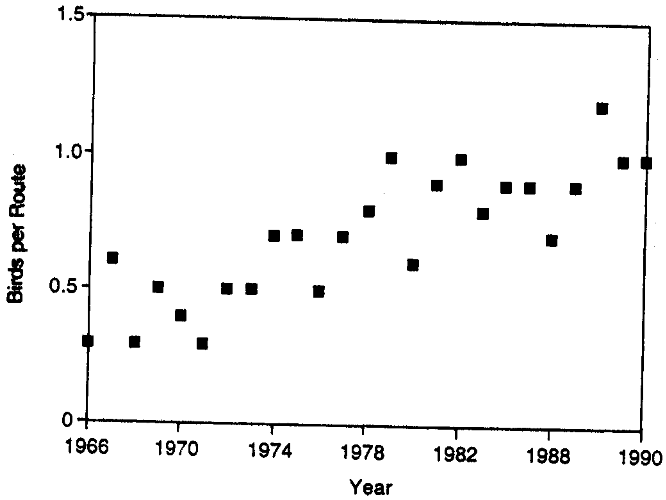


**Figure 1**--Mean population counts by year for Red-tailed Hawks in Wisconsin as indicated by Breeding Bird Survey data.

**Table** ***3-Reliability of trend estimates $\lambda$ for Red-tailed Hawks in Wisconsin as a function of interval length.***

| Number of years in sample | Percent of samples in which the error $\hat{\lambda} - \lambda$ was: | | |
|---|---|---|---|
| | <0.02 | <0.03 | <0.04 |
| 3 | 57 | 57 | 61 |
| 4 | 64 | 68 | 68 |
| 5 | 67 | 81 | 86 |
| 6 | 65 | 75 | 85 |
| 7 | 63 | 68 | 74 |
| 8 | 67 | 78 | 89 |
| 9 | 76 | 88 | 88 |
| 10 | 75 | 81 | 94 |
| 11 | 73 | 73 | 93 |
| 12 | 71 | 79 | 93 |
| 13 | 62 | 92 | 100 |
| 14 | 67 | 100 | 100 |
| 15 | 91 | 91 | 100 |

**Table 4--*Transition probability matrix for Markov chain model***

| Initial probability of site occupancy | Condition | Condition[1] | |
|---|---|---|---|
| | | 1 | 0 |
| ø | 1 | p[2] | 1-p |
| 1-ø | 0 | r[3] | 1-r |

[1] 1: site occupied, 2: site not occupied

[2] p: probability that site remains occupied in second year

[3] r: probability that a site that is not occupied in first year will be occupied in second year

od was calculated for the case $p$ = 0.9 and $r$ = 0.01, giving a limiting value of .01/.11 = .090909 for the fraction occupied after a large number of years (*table 5*). The initial fraction was chosen to be 30 percent smaller than this; namely, ø = .070. During the 8-year interval, this fraction increased 17 percent to 0.082, which is equivalent to an annual, proportional change of 1.019626, or 1.96 percent (i.e., $1.019626^8$ = 1.17, or 17 percent). In the analyses below, we would describe such a change by stating that the population grew at an average annual rate of 1.96 percent.

Detectability bias and noise were introduced into the model by assuming that an occupant is detected with probability *d,* and "detection" is independent from site to site and year to year. Only "false negatives" are allowed; i.e., a true 0 is always recorded as 0, while a true 1 is sometimes (with probability 1-$d$) recorded as 0. This has the effect of (1) reducing the expected slope by the factor $d$, (2) reducing the variance by the factor $d^2$, but (3) adding a noise variance component.

Stochasticity in the transition probabilities $p$ and $r$ was introduced as random multiplicative effects. A year-specific, site-specific $p$ became a product of a random year factor $\alpha$, say, and a random site component $\tau$, say; thus, $p = \alpha\tau$ at where $\alpha$ is common to all sites that year and $\tau$ is common to all years at that site. Randomness in $r$ was introduced by assuming that $r/p$ is a random variable $\theta$, $0 < \theta < 1$ , for $\alpha$ fixed $p = \alpha\tau$; on; the distribution of $\alpha$ and $\tau$ were selected (from the beta family), for convenience of simulation, to be that of the largest order statistic of a sample from a uniform distribution on the unit interval. The uniform sample size was chosen separately for $\alpha$ and $\tau$ to force the expected value of $p$ to be a specified value (namely, the same value as before when $p$ was a constant). Similarly, the distribution of $\theta$ was taken to be that of the smallest order statistic from a uniform sample of a size determined by the previously assumed constant value of $r$. Such distribution choices enable the ready use of probability transforms in simulating values for $\alpha$, $\tau$ and $\theta$ while also permitting some analytic calculations to be readily performed; e.g., the calculation of the vector ($a$) of expected annual occupancies.

**Table 5--*Illustration of a Markov chain approach toward stochastic equilibrium when r/(1-p+r) is 30 percent higher than the initial fraction ø.***

| Year | Expected proportion of sites occupied |
|---|---|
| 1 | 0.069930 |
| 2 | 0.072238 |
| 3 | 0.074292 |
| 4 | 0.076120 |
| 5 | 0.077746 |
| 6 | 0.079194 |
| 7 | 0.080483 |
| 8 | 0.081630 |

### *Simulation Analysis*

A computer program based on the model permitted power calculation for any combination of values for Ø, p and *r* and for alternative lattice designs. It was used to generate 34. data sets. In each, a population changing in size at a specified rate was monitored for either 8 or 12 years, the probability of detecting the trend (i.e., power) was determined with different sampling designs and sample sizes. The level of significance was set at 10 percent in all tests. Rates of change varied from a decline of 4.8 percent per year to an increase of 4.5 percent per year; numbers of owls recorded per station varied from 0.03 to 0.14; detection rates varied from 0.5 to 1.0; and stochastic year and site effects were present in some analyses and absent in others. The number of stations visited per year varied from 200 to 1000, and from none to all of them were on a 4-year cycle. Power varied from 0.21 to 1.00.

### *Magnitude of Trend That Should Be Detectable*

A decision must be made about how large a trend the survey should be capable of detecting. One step in making this decision is estimating the magnitude of fluctuations that night be expected in Spotted Owl populations that were stable and "healthy." No data for such calculations are available for Spotted Owls, but the estimates are important because we would not expect trends to be exactly zero, even if a population were fully recovered. In any given period it would probably be increasing or decreasing slightly and would thus have roughly a 50-50 chance of declining slightly. Thus some effort must be made to understand the magnitude of trend that might be considered normal and to incorporate this information into sample size guidelines for the monitoring program.

We examined the Breeding Bird Survey data sets described above to help determine natural levels of variation in populations that are stable or close to stable. Four of the 15 populations showed both positive and negative trends during the 25-year period. We estimated both trends in these cases, obtaining a total of 19 trends. Five were negative and 14 were positive. About one-half (42 percent) of the absolute trends exceeded 3 percent per year and two-thirds exceeded 2 percent per year. The preponderance of positive values may have been caused by a slight overall increase in these populations at the regional or national level (Droege, S., telephone conversation) or perhaps by a general increase in surveyor skill (Peterjohn, B., telephone conversation). We can shift the distribution so that it is approximately centered on zero by subtracting 2 percent from all values. Nine of the 19 trends (i.e., about one-half) are then negative and 10 are positive. In this case, 37 percent of the absolute trends exceed 3 percent per year, and 42 percent exceed 2 percent per year. These results suggest that average annual changes, over periods of up to 25 years, in state-wide populations of raptors, are commonly as large as 2.5 percent or 3 percent. Smaller populations probably exhibit somewhat larger fluctuations, so annual changes in a single physiographic province of 3.5-4 percent may be common.

Another factor to consider in deciding how large a trend should be detectable is how the estimate of trend will be combined with other information in determining whether populations are recovering. We believe that conclusions about the long-term stability of the population should not depend solely, or even primarily, on empirical estimates of trend. On the contrary, these data should probably play a minor role, compared to efforts based more on *understanding the causes* of trends (i.e., population modeling). We believe the latter efforts (which are described in the Recovery Plan [USDI 1992]) will provide a more reliable and cost-effective way to estimate or predict trends.

The points above suggest that the roadside surveys should have adequate (i.e., 80 percent) power to detect annual changes of 2.5 percent at the statewide level or 3.5 percent at the province level. Changes of smaller magnitude would probably be hard to interpret, even if they were detected, since such changes may occur commonly in healthy populations. The cost of obtaining higher power would also be hard to justify given that other measures will play at least as important a role in the overall estimation of population trends as will the roadside surveys. We therefore calculated the sample sizes required for 80 percent power of detecting annual changes in the 2-4 percent range.

## Results

### *Analyses of Breeding Bird Survey Data*

The average amount of time (and range) required for 80 percent probability that errors in estimating trend were <0.02 was 11.9 years (range: 8 to >15 (*table 6*)). The corresponding figures for errors of <0.03 and <0.04 were 9.5

**Table 6--*Number of years required to obtain reliable estimates of long term trends from sample Breeding Bird Survey data sets.***

| | | Minimum Number of years for 80 pct Probability that $\hat{\lambda} - \lambda$ was | | |
|---|---|---|---|---|
| Species | State | <0.02 | <0.03 | <0.04 |
| **Turkey Vulture** | Florida | >15 | 9 | 80 |
| | Maryland | 12 | 7 | 6 |
| | Ohio | 15 | 15 | 15 |
| | Oklahoma | 11 | 10 | 9 |
| **Black Vulture** | Alabama | >15 | 13 | 10 |
| | Florida | 10 | 10 | 10 |
| **Red-tailed Hawk** | Kansas | 8 | 7 | 4 |
| | Oklahoma | 14 | 8 | 7 |
| | Wisconsin | 15 | 13 | 6 |
| **Red-shouldered Hawk** | Florida | 8 | **7** | 6 |
| **American Kestrel** | New York | 8 | 6 | 5 |
| | Ohio | 8 | 8 | 8 |
| | Ontario | 13 | 10 | 8 |
| **Osprey** | Florida | 9 | 8 | 7 |
| **Great Horned Owl** | Kansas | 14 | 11 | 9 |
| Average | -- | 11.9 | 9.5 | 7.9 |

years (range: 6-15) and 7.9 years (range: 4-15), respectively. (In these calculations, estimates from table 6 of ">15" were counted as 17 years.)

### *Computer Simulations*

Power increased with increasing value of the following five variables: absolute trends, numbers of birds recorded per 100 stations, number of stations per year, number of years, and fraction of the station on a 4-year cycle. Detection rate and presence or absence of random factors in the simulation had little effect on power. The large number of factors having a substantial influence on power made it difficult to specify conditions required to achieve a specified power. In general, however, few simulations produced power of 80 percent when only 8 years of monitoring data were available. With annual population changes of <3 percent and number of owls recorded per 100 stations of <10, power never reached 80 percent even when 1000 stations were visited per year and all were on a 4-year cycle (power was about 79 percent in this case). Higher absolute trends, or numbers of birds reported per 100 stations, increase power to levels above 0.80. For example, surveying 1000 stations per year and recording 6.5 birds per 100 stations, when the population was declining 4.1 percent per year, produced power equal to 80 percent when 30 percent of the stations were replaced annually. It produced power of 85 percent when all the stations were replaced annually. Despite these examples, in general, it was expensive, and sometimes virtually impossible, to achieve power of 80 percent with only 8 years of monitoring data.

With 12 years of data, many more situations were found in which power was above 80 percent, sometimes by a substantial margin. For example, with a 2.5 percent decline per year, and 6.2 birds recorded per 100 stations, power was 0.82 with 1000 stations, all replaced annually. With 10 birds recorded per 100 stations, power exceeded 80 percent if 70 percent of 800 stations, or all of 600 stations, were replaced annually. With larger trends, smaller samples were sufficient. For example, with a 3.6 percent decline per year and 7.7 birds recorded per 100 stations, power exceeded 80 percent if 60 percent of 600 stations, or all of just 400 stations, were replaced annually. By contrast, if none of the stations were replaced, then 1000 per year were necessary to achieve power of 80 percent.

Results from the simulations were analyzed with a general linear models program in which power was the dependent variable. The simplest equation with high explanatory power was:

$$power = -1.18 + 0.04stns + 0.18chg + 0.04recs + 0.05yrs + 0.16repl$$

where *stns* = number of stations surveyed per year in 100's (e.g., 800 stations per year was coded as 8)
*chg* =annual absolute percent change per year (e.g., if the population decreased 3.2 per cent each year, then *chg* was 3.2)
*recs* =average number of birds recorded per 100 stations
*yrs* =length of monitoring period in years
*repl* =fraction of the stations on a 4-year cycle

The $r^2$ for this equation was 0.87. All variables were highly significant. Slight improvement was obtained by applying a square root transformation to several of the variables and including a few interaction terms, but the gain ($r^2$ equalled 0.92 in the best model) did not seem worth the increased difficulty in interpreting the model. Adding detection rate and presence of random effects did not improve the fit of the model. The standard deviation and coefficient of variation of the residuals were 0.08 and 16 percent, indicating that the model revealed general trends, but did not make individual predictions very well.

The coefficients above describe the general relationship between the variables and power. The general trend was to obtain an increase in power of about 0.04 for each of the following: (1) increasing the number of stations per year by 100; (2) increasing the annual trend in the population by 0.25 percent (e.g., from 3 percent to 3.25 percent); (3) increasing the number of birds recorded per 100 stations by 1; (4) increasing the number of years on which the estimate was based by 1; and (5) putting an additional 25 percent of the stations on a 4-year cycle.

Obviously, these statements hold only for powers well below 1.0 and for appropriate ranges of the variables, and as noted above, the specific predictions of the regression model were often in error by 0.08 to 0.10 or even more. Nonetheless, the results above provide at least a rough guideline to the ways that power is affected by altering the variables.

## Discussion

### *Analyses of Breeding Bird Survey Data*

In 11 of the 15 data sets, 7 to 11 years were required before estimates of the long-term trend were within 3 percent of the true values. There is little basis, at present, for deciding which of our data sets most closely resemble the data that will be collected for Northern Spotted Owls. We studied the analyses to determine effects of density, sample size, outliers, and autocorrelation but were unable, with this small sample size, to reach definitive conclusions. Even if we had, it would probably be difficult to predict the form of the Northern Spotted Owl data set. For example, survey data on diurnal raptors such as the Red-tailed Hawk that inhabits open landscapes might be considerably different from survey data for Northern Spotted Owls.

### *Computer Simulations*

These analyses suggested that a minimum of 8 years will probably be required for 80 percent probability of detecting trends in owl populations unless such trends exceed 3 percent and >10 owls are recorded per 100 stations. Such a program would probably require that >1000 stations be visited per year. If 12 years of data are available to estimate trends, then 600 stations per year, if visited on a 4-year cycle, might be sufficient to detect annual trends in the 2-3 percent range, particularly if ≥8 birds are recorded per 100 stations. Obviously these conclusions are based on the assumptions inherent in the model, and these assumptions can be refined and improved as data from the monitoring program are collected. Furthermore, it must be remembered that these surveys do not detect trends in the total population, but provide information only about the territorial population (nonterritorial birds are almost never detected on these surveys).

The analyses bring out the value of increasing the length of the monitoring period. For example, increasing the period from 8 to 12 years increases power by approximately .16 (4 years x a gain of .04/year). If power were sufficient initially, then the number of stations visited per year could be decreased by 400 (4 years x a loss in power of .04 per year). Thus, even this simple analysis shows the great value of extending the monitoring period.

The analyses also illustrate the possible advantage of the lattice design over a design in which each station is visited once each year. The regression equation suggested that, for each 25 percent of the stations put on a 4-year cycle, 100 fewer stations could be visited per year without losing power, or power would increase by about 0.04 if the same number of stations was visited. Specific analyses using the computer simulation (rather than the linear model) also indicate the potential value of the lattice design. For example, in one simulation, a population declining at an annual rate of 3.4 percent was surveyed for 8 years. With 600 stations visited per year, power was 66 percent when all stations were visited each year and increased to 75 percent when all stations were visited every fourth year. The potential value of the lattice design is also evident by comparing sample sizes required to achieve a given level of power. For example, power was about the same with 800 stations, all visited on a 4-year cycle, as with 1000 stations all visited each year. Thus, the lattice-design, in this case, would permit a 20 percent reduction in the number of stations surveyed per year without any loss in power to detect the trend. Note, however, that the lattice design in this case would require the identification of 3200 stations, rather than the 1000 stations needed if all stations were visited each year. The lattice design thus requires that more routes be selected, but permits a smaller number to be surveyed in each year (to achieve a given power) than a design in which each route is surveyed each year.

The decision on whether to adopt a lattice design can be postponed until the second year of monitoring. At that time, a decision must be made to revisit every route surveyed in the first year or to temporarily drop some routes and introduce a corresponding set of new routes. If the new set were spatially interpenetrating with those that were dropped, then the logistics of the program would not be compromised by this tactic and the geographic dispersion of the sample would remain essentially the same. (Such a spatially interpenetrating design on a 4-year cycle is being implemented by the Environmental Protection Agency in their newly instituted Environmental Monitoring and Assessment Program (EMAP)).

### *Calibration of Roadside Surveys*

The discussion above assumes that the roadside surveys will be a typical index in which results would be expressed as birds recorded per station or some other measure of effort, and investigators would assume that this measure had an approximately constant relationship to true density in the surveyed area. Under this assumption, changes in the index reflect changes in the population being surveyed. This assumption, however, might be incorrect for many reasons as noted by numerous researchers (e.g., Robbins and others 1986; Sauer and Droege 1990). Furthermore, many areas, including several of considerable size, will probably be searched thoroughly for territorial birds each year regardless of whether index or plot-type methods are employed in the monitoring program. These data provide a basis for "calibrating" the index data by using the technique known as double-sampling. With this approach, the index data may be adjusted to provide estimates of density (of territorial birds). We describe this method below.

A double-sampling approach for monitoring territorial Northern Spotted Owls might proceed as follows. First, areas to be searched thoroughly would be selected and delineated on maps. The main criterion for selecting these areas would be availability of surveyors willing to search the areas thoroughly (defined, for example, as searching according to protocols that have been developed by the USDA Forest Service). The areas could be large (e.g, demographic study areas) or small (e.g., single patches of old-growth) and would not have to be selected randomly, though random selection might be advantageous in some cases. Presence of a bird within the area would be defined as occurring when the bird's activity center was within the thoroughly searched area. The assumption would be made that such searches constituted censuses of the territorial birds in the areas. The results would thus provide "true densities" for each area. Of course, in reality some birds would be missed, but such errors would probably have little effect on the estimate of trend if fewer than, say, 10 percent of the birds were missed, and this figure did not vary greatly between years. The thoroughly searched areas would be regarded as one stratum.

The next step in developing the program would be the delineation of additional strata. Strata could be defined to include owl conservation areas and other high-interest areas, areas in which density is anticipated to be high, areas in which density is anticipated to be low, etc. The strata would not have to be contiguous; thus one stratum might consist of all the areas in a given region dominated by old-growth, another stratum might consist of all the areas with moderate amounts of old-growth, and so on. Each year, randomly selected routes in all of the strata would be visited. Sampling intensity could vary between these strata so that more stations were located within areas of high interest, easy access, or high density. Sampling intensity in these strata could be determined subjectively or by using formulas for maximizing statistical efficiency.

Intensive work within the thoroughly studied stratum would reveal the actual densities in these areas. Two estimates would thus be available: the results from the roadside routes and the actual density. The ratio of these two results would be used to "calibrate" the index in other strata. For example, if the mean number of birds per roadside route was 2, and the true density of owls per 100 $km^2$ was 3, then the results for roadside routes in other strata would be multiplied by 1.5 to obtain an estimate of actual density per 100 $km^2$. The multiplier might differ between years and areas.

One problem with this approach is that in double-sampling methods, the sample of sites searched thoroughly is usually a random sample from the entire population. This permits unbiased estimation of actual densities. In the case of Northern Spotted Owls, the thoroughly searched areas have not been randomly selected. This may not cause serious problems and, even if problems do arise, various ways can be imagined for resolving them. The issue, however, needs to be addressed before the final design is determined.

## Conclusion

The analysis of Breeding Bird Survey data and the computer simulations suggest that the roadside survey should include visits to approximately 750 stations per year in each of the three states, and that data collection will have to continue for at least 8-10 years before reliable estimates of long-term trends can be obtained. An average of about 15 stations are usually visited per person-night, so the fieldwork would require 50 person-days per state, or about 15 days per physiographic province, a modest expenditure of effort that could easily be continued for many years.

Although these are preliminary estimates and will need revision after the first few years of data have been collected, the analyses above identify some of the most important design considerations, suggest methods that appear feasible and efficient, and indicate that the needed data can be collected at reasonable cost. In combination with the demographic and population modeling studies recommended in the Recovery Plan (USDI 1992), these methods will provide comprehensive information about trends in different areas and habitats and should provide a sound basis for refining the recovery program and ultimately for delisting the subspecies.

## Acknowledgments

We appreciate the help of the FWS Breeding Bird Survey program in providing data on short notice. Earlier drafts of this paper were reviewed by Robert Anthony, Sam Droege, Rocky Gutierrez, Josefa O'Malley, C.J. Ralph, John Sauer, Ed Starkey, and Jerry Verner.