Utility of pheasant call counts and brood counts for monitoring population density and predicting harvest

Clifford G. Rice

Wildlife Management Program, Washington Department of Fish and Wildlife, Olympia, Washington

Follow this and additional works at: https://scholarsarchive.byu.edu/wnan

Recommended Citation
Available at: https://scholarsarchive.byu.edu/wnan/vol63/iss2/4
UTILITY OF PHEASANT CALL COUNTS AND BROOD COUNTS FOR MONITORING POPULATION DENSITY AND PREDICTING HARVEST

Clifford G. Rice

ABSTRACT.—Call counts and brood counts are frequently used to evaluate Ring-necked Pheasant (Phasianus colchicus) populations and to forecast harvest. Given the variability commonly observed in these counts, I evaluated their utility in performing these functions in the state of Washington. Pheasant harvest, call counts, and brood counts have all declined in Washington State since 1982. Power for detecting trends in call counts was higher than for brood counts, but substantial sampling was required to reliably detect even large changes in the short term (e.g., power = 0.9 for a 40% decline between 2 years with 12 routes). Brood counts predicted harvest with greater precision than did call counts, but predictions were meaningful only at the statewide scale (i.e., not for counties or major river basins). This was true for predicting total harvest and relative harvest (high, medium, or low).

Key words: brood counts, call counts, crowing counts, harvest prediction, pheasant, population monitoring, roadside surveys, upland game.

Call counts and brood counts are 2 techniques used by the Washington Department of Fish and Wildlife (WDFW) for monitoring Ring-necked Pheasant populations. The objectives of these surveys are to monitor population levels and to predict hunter harvest during the following fall hunting season.

Previous papers on these techniques have dealt primarily with refinements for minimizing variability and the influence of confounding variables such as season, time of day, and weather (e.g., Fisher et al. 1947, Kimball 1949, Kozicky 1952, Klonglan 1955, Gates 1966, Luukkonen et al. 1997). While some have advocated the use of call counts as an index of population levels (Robertson 1958, Gates 1966, Warner and David 1982, Snyder 1985, Rotella and Ratti 1986), others have felt that these counts were best used to document presence (Anderson 1983). Statistical power for detecting differences in call or brood counts has not been addressed. Several authors reported high correlations between field surveys and subsequent harvest levels (Suchy et al. 1991, and several unpublished reports), but significant correlations do not necessarily yield precise predictions. To overcome this deficiency, I examined the precision of predictions based on past survey indices and harvest.

STUDY AREA AND METHODS

Trend Analysis

Call count and brood count transects were conducted in the primary management zone for pheasant in Washington: the Snake, Columbia, and Yakima River basins (Fig. 1). The Columbia and Yakima River basins were historically almost entirely shrub-steppe habitat (Johnson and O'Neil 2001). The Snake River basin was historically primarily eastside (interior) grassland habitat, with some areas of herbaceous wetlands and eastside (interior) riparian-wetlands habitat (Johnson and O'Neil 2001). Currently, all 3 basins are predominantly agriculture, pasture, and mixed environs habitat (Johnson and O'Neil 2001) containing only fragments of the original habitats.

WDFW biologists subjectively selected the locations of call and brood count routes to represent high-quality pheasant habitat. Call counts were conducted along fixed transects on 1 or 2 days between 25 April and 15 May, from 50 minutes before sunrise to 10 minutes after sunrise. At 20 stations at ≥1.6-km (1-mile) intervals, the observer recorded all pheasant calls heard for a 2-minute period. Counts were not conducted if wind speed exceeded about 8 km · hr⁻¹ (5 miles · hr⁻¹), and stations were moved...
Fig. 1. Map of southeastern Washington showing counties and the 3 river basins comprising the primary management zone for Ring-necked Pheasant.

91 m (300 feet) if localized noise (e.g., traffic) interfered with hearing pheasant calls. If a transect was surveyed on >1 occasion, the highest count was selected.

Brood counts were conducted along fixed transects 32 km (20 miles) in length on (usually) 3 days. Timing varied according to plant phenology each year (1st half of August in phenologically advanced years, middle 2 weeks of August in average years, and 2nd half of August in phenologically late years). In the morning at, or within 15 minutes after, sunrise or in the evening leading up to sunset, the observer drove the route at 24–32 km · h⁻¹ (15–20 miles · hr⁻¹) and recorded all pheasants seen: adult male, adult female with brood, adult female without brood, or chick (juvenile). If necessary, broods were flushed to obtain a complete count. Counts were not conducted in rain, fog, or windy conditions.

Call counts were initiated in some areas in 1961, but they were more widely implemented after 1976 and were initiated in the Snake River basin in 1982. Brood counts were conducted after 1976 in the Columbia River basin, and after 1983 or 1984 in the Yakima and Snake River basins. Analysis was limited to 1982–1998, and since then call and brood counts have not been conducted.

Pheasant harvests were estimated through an annual hunter questionnaire survey, which was sent to a random sample of hunters and received from about 10% of them (yearly average of 6520 questionnaires received). These estimates were at the county, WDFW region, and statewide levels. WDFW Regions 1, 2, and 3 correspond to the Snake, Columbia, and Yakima River basins, respectively. The coefficient of variation for pheasant harvest (1993–1999) in the selected counties (see below) averaged 0.091 (σ = 0.018).

Trends were assessed for the years 1982–1998 because many routes were not run prior to that period. Estimates by county were limited to Adams, Benton, Columbia-Garfield, Franklin, Grant, Walla Walla, Whitman, and Yakima Counties.

For analysis of trend and statistical power, indices of pheasant abundance (calls/station for call counts and pheasants/day for brood counts) were log-transformed and the analysis performed on the transformed values. This was done so that linear regression of transformed values against time evaluated the proportional change in the index rather than its absolute change. For example, a decline from 50 pheasants/day to 25 pheasants/day produces the same regression coefficient or slope as a decline of 10 pheasants/day to 5 pheasants/day. Thus, the relative decline across all routes is evaluated, regardless of their initial densities. The proportional decline was estimated as $e^b - 1$, where
where $e$ is the base of the natural log, and $b$ is the slope of the regression. Percent change in the index equaled $(e^b - 1)(100)$. Index values predicted by the regression can be back-transformed into the original units (predicted index = $e^{predicted}$).

**Power Analysis**

Statistical power (1 - the type II error rate) is the probability of detecting a change of a specified magnitude given the variability in the data set, sample size, and significance level ($\alpha$ or type I error rate) for the statistical test (Cohen 1988). Lower variability increases power; increasing sample size increases power; and choosing a higher significance level (e.g., $\alpha = 0.10$ rather than 0.05) increases power. In monitoring and trend analysis, the statistical test can be compared to an alarm system (Rice et al. 2001). We want the alarm to signal when something we measure reaches a certain value. However, because our system is imperfect, we know that sometimes the alarm will go off when that value has not been reached and sometimes it will remain silent, even when the value is reached. For power analysis, variability in the data is a measure of how imperfect the alarm system is. The value to trigger the alarm is the magnitude of change we want to be able to detect. The likelihood of false alarms is set by the $\alpha$ level for the statistical test. The likelihood of an alarm going off when it "should" is statistical power. Power and $\alpha$ are tradeoffs; if we can live with many false alarms, we can achieve much higher power, and vice versa. In this analysis an $\alpha$ of 0.10 was used and the criterion for power was 0.90. This implies equal importance in avoiding false alarms and ensuring the detection of change with the statistical test.

Like other statistical parameters, power calculations are estimates. Also, in evaluating future sampling requirements, these estimates depend on data gathered in the future being comparable in variability to those upon which the power analysis is performed. The results of power analysis should be interpreted and applied with this in mind.

An important issue is evaluating underlying variability. If there are repeated measures, assessing within-cell variability is one approach to this. However, for call counts there was only one count within each cell (at each route each year). Assessing within-cell variability might be feasible for brood counts (with 3 counts/route/year), but in the available data sets, individual counts were infrequently reported (17% of 321 cells).

I performed an analysis of variance (ANOVA) and estimated variability from the residual error; i.e., the variation which was not explained by the ANOVA was used as an estimate of the inherent variability in the data set. In the ANOVA the factor for route was nested within basin because that is how routes occur. Another factor for year was nested within basin, which estimated a year effect for each basin. The least-squares estimated index was then dependent on which route it was and the values for other routes in that basin during that year compared with other years.

For brood counts I estimated variability for pheasants/day, adults/day, adult females/day, and adult males/day. From these I selected the measure with the lowest variability (which would have the greatest power) for power analysis.

I estimated power using JMP (SAS Institute, Inc. 2000) power analysis options.

**Predicting Harvest**

There are 2 measures of harvest which would be important for managers to be able to predict, total harvest and harvest/day (pheasants harvested/hunter-day). These predictions have utility at the county, basin, and statewide levels and might be based on call or brood counts. For brood counts I considered pheasants/day to be the most likely predictor of harvest. Rather than examine all permutations of these variables and scales, I started with call counts predicting harvest/day at the county level, then sought to improve the prediction by using brood counts. Next, I tested whether total harvest yielded a better prediction, and finally I broadened the scale to basin and statewide levels.

Both call counts and brood counts showed significant relationships with harvest (see below), but this does not mean that they function well in predicting harvest. Because of the large sample size and the large decline in harvest since 1984, a relationship between harvest and survey indices does not mean it is a close relationship, but merely a consequence of large changes in both measures. I portrayed this relationship by estimating the 90% confidence interval around predicted harvest for the survey
indices given various sampling intensities. Given the sampling level and variability in the relationship, the band encompasses 90% of the variation in the predicted harvest given the index value. An indication of the utility of the surveys is the minimum difference in 2 index values which would yield nonoverlapping confidence intervals for the 2 estimates.

Upland game managers in Washington acknowledge that predicting harvest exactly may not be important, but predicting high, medium, or low harvest levels would be sufficient. Since this prediction would be relative to the average for a given area, I used the weighted average of pheasants/day for each year and applied ordinal logistic regression to assess the prediction probabilities of high, medium, and low harvest. Because counties differed greatly in the harvest level, assigning harvest estimates to high, medium, and low was done separately for each county. The distribution of basin-wide harvests was comparable across basins, so they were pooled for this assignment.

Ordinal logistic regression calculates the probability for each outcome (high, medium, low), given the observed index values (e.g., pheasants/day). The best estimate is the outcome with the highest probability for a given index value. The error of the outcome estimate is \(1 - \text{the highest outcome probability (or the sum of the probabilities for the other outcomes)}\).

**RESULTS**

**Trend Analysis**

The number of call and brood counts conducted varied considerably among years (Table 1). The 61 call count transects were conducted an average of 6.4 times during 1982–1998 (median = 6, range 1–16) and the 23 brood count transects an average of 12.6 times (median = 12, range 3–19).

Estimated total harvest declined over the entire primary management zone and in each of the basins in 1984–1997 (Table 2). All counties but one showed a significant decline in total harvest (Table 3) at rates of 3.9–9.8% per year. Pheasants harvested/day did not show a consistent trend. Statistically significant but moderate declines were found in 2 basins (Table 2) and 3 counties (Table 3). Calls/station declined over the entire primary management zone and in all 3 basins from 1982 to 1998, and the declines were of similar magnitude (Table 2). In the Snake River basin, the decline in calls/station was evident in Columbia-Garfield and Walla Walla Counties but not in Whitman County (Table 3). In the Columbia basin, the decline was evident in all 3 counties (Adams, Franklin, and Grant), as was the case in the Yakima River basin, although the trend was not significant in Yakima County (Table 3).

The total number of pheasants seen/day (pheasants/day) in brood counts declined over the entire primary management zone, which was reflected in significant declines in the Columbia and Yakima River basins but not in the Snake River basin (Table 2). Each of the counties in each basin matched this pattern (Table 3).

**Power Analysis**

Variability (root mean square error) for call counts was estimated at 0.4004077. Brood count variability was 0.7684459 for pheasants/day, 0.5333018 for adults/day, 0.5201298 for adult males/day, and 0.5806677 for hens/day. I used males/day in power analysis for brood counts.

For call counts a power of 0.9 was attained for detecting a decline between 2 years of about 40% with 11 routes, 60% with 4 routes, or 80% with 2 routes (Fig. 2). Similarly, power of 0.9 was attained for detecting increases of 60% with 13 routes and 80% with 9 routes.

For detecting trends in adult males/day in brood counts (assuming that 3 days were recorded at each brood route), a power of 0.9 was attained for detecting a decline of about 40% with 18 routes, 60% with 6 routes, or 80% with 3 routes (Fig. 3). Similarly, a power of 0.9 was attained for detecting increases of 60% with 22 routes and 80% with 14 routes. With the exception of large declines (>50%) where differences in samples sizes were less, brood counts required 60–70% more routes than call counts to achieve equivalent power.

**Predicting Harvest**

The relationship between calls/station and harvest/day at the county level was weak (Fig. 4). Even for counties with 20 routes, nonoverlapping confidence intervals occur for very large differences in calls/station (e.g., 0–25 calls/
### Table 1. Number of call count and brood count routes conducted in Washington State yearly, 1982-1998, by river basin.

<table>
<thead>
<tr>
<th>Year</th>
<th>Call counts</th>
<th>Brood counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Snake</td>
<td>Columbia</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1984</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1985</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1986</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1987</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1988</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1989</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1990</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>105</td>
</tr>
</tbody>
</table>

As values greater than 20 calls/station have been rare, this magnitude of distinction is unlikely. So, although we can predict harvest/day based on calls/station, the estimate is unlikely to be statistically different from an estimate for any other value of calls/station.

The relationship between total number of pheasants seen/day for brood surveys and harvest/day at the county level was stronger than with calls/station, so the precision of the prediction was improved, but not by much. Large differences in pheasants/day (e.g., 0–50) would be required to result in nonoverlapping confidence intervals.

Although predicting harvest/day at the basin level improved the fit, precise prediction was still problematic. Predicting total harvest from brood counts on a basin level produced a better fit that might have some utility (Fig. 5). Nevertheless, substantial differences in pheasants/day are needed to produce statistically different estimates (e.g., differences of 34 pheasants/day with 90% confidence with 20 routes/year).

At the statewide level (all 3 basins together), harvest can be predicted with reasonable confidence (Fig. 6) with 20 routes, differences of about 28 pheasants/day produce statistically different estimates (90% confidence).

On the county level, low harvest was most likely when fewer than 10 pheasants/day were observed on brood routes; medium harvest was most likely for 10–30 pheasants/day (Fig. 7). However, the probability of error was around 50% except for high values of pheasants/day (>40). On the basin level, definition improved somewhat in that mean pheasants/day below about 10 gave a high likelihood of having low harvest, with error probabilities near 50% otherwise except at very high values of pheasants/day. Statewide, predicting low, medium, or high harvest became more feasible and error probabilities were near 50% only at the transition areas (Fig. 7). However, if one desires 90% precision, this level of certainty is achieved only at fewer than 11 or more than 33 pheasants/day.

### DISCUSSION

#### Trend Analysis

The location of survey routes was determined subjectively. Thus, trends observed in survey results apply to high-quality areas. This was not felt to be a drawback because management concern focuses on high-quality areas, and the same can be expected of hunting effort and harvest. Determining trends over the entire range of the pheasant, including poor and marginal areas, would be difficult, costly, and of little utility in management. Because of these considerations, the subjective selection of route location in high-quality habitat is the
most suitable use of limited resources for monitoring for management purposes.

The selection of the single highest call count as the estimate implies a highly skewed distribution of values about the best estimate (i.e., there are many "random" factors that might lower a count, but few that would raise it above the best estimate). Whether or not this was appropriate, this procedure introduced a bias into the data set in cases when only 1 count was made. In recent years (1990 and later), whether the index value was based on 1 count or 2 counts was not recorded for 30 (24%) of 123 counts. Of the remaining 93 counts, 16 (17%) were single counts. Overall, the mean (a more conventional estimate for a given survey) was 11% below the highest value. At a minimum, because the 2 values were mixed together in a single index, the resulting analysis can be expected to have higher variance and lower power.

Variation in the number of counts conducted each year (Table 1) reflects changes in land use, personnel, and administrative priorities over time. Although this was undesirable from an analytic standpoint, this data set reflects reality and thus represents the basis upon which managers must make decisions about the resources they manage.

Estimated trends in survey indices were not in agreement. While close agreement would probably not be expected, this disparity was particularly true in the Snake River basin, where significant declines occurred in call counts but not in brood counts (Table 2), which appears to be primarily due to this effect being prominent in Walla Walla County (Table 3). Of 24 locations where both types of routes were run, 42% were in agreement in that they detected significant declines in both surveys. At 33% of the locations, a significant change was detected in only 1 survey technique. Seventy-five percent (6) of these were significant changes in call counts not reflected in brood counts, while 25% (2) were significant changes in brood counts not reflected in call counts.

**Power Analysis**

For both call counts and brood counts, power was higher for detecting decreases than it was for comparable increases. This was a consequence of changing the reference value used
to calculate the percent change. In comparisons between 2 years, the statistical test compares 2 values regardless of their order. Thus, the difference between 7 and 14 is the same as the difference between 14 and 7. However, the percentage changes for these 2 examples are +100% and -50%, respectively. Thus, the power to detect a 100% increase is the same as that for detecting a 50% decrease (Figs. 2, 3).

For trend analysis over several years, power might be expected to increase due to the greater number of samples. However, as the specified level of change is distributed over additional years, the slope (steepness) of the change decreases, which concomitantly reduces power. What remains constant is the power to detect a given magnitude of change for a given number of total samples. For example, we can expect to detect a 40% decline in call counts between 2 years with 12 routes/year, or 24 route-years altogether. Likewise, we can expect to detect a 40% decline over 3 years with 8 routes/year, or over 8 years with 3 routes/year (total of 24 route-years). This relationship between power and total sample size was also found for bear bait stations (Rice et al. 2001).

Kozicky et al. (1952) estimated sample size requirements for detecting differences in brood counts between areas. Power was not explicitly identified in their analysis, but given that they apparently used \( \alpha \) values of 0.05 and 0.20 (and possibly equivalent type II error rates),
and adjusting for different measures of change, my interpretation is that their findings were comparable to mine.

Call count routes exhibit higher power than brood routes for detecting changes in respective index values (Figs. 2, 3). Disregarding factors such as convenience, timing during the year, costs, and correspondence to the actual population density, call counts are better than brood counts for evaluating changes in population levels. Nevertheless, there is high likelihood of detecting only ≤−50% or ≥+100% change between 2 years. Smaller changes in calls/station are not as likely to yield a statistically significant result. Detecting long-term trends is more feasible as is evident in the trend analysis.

**Predicting Harvest**

Total harvest was predicted more precisely than harvest/day. This is not surprising given the definitive trends in total harvest, call counts, and brood counts, but not in harvest/day. Brood counts were better predictors of total harvest than were call counts. Nevertheless, predicting harvest at the 90% confidence level was realistic only at the statewide level.

This analysis was influenced by 2 factors which might indicate that predictions are actually better than was estimated. First, harvest...
estimates were treated in the analysis as being without error. This was not the case, and although variability of these estimates was not calculated except at the statewide level, it likely was considerable. This introduced variability in the analysis, which likely reduced precision of the prediction. True harvest, although we do not know what it is, may be predicted considerably better than estimated harvest was. Second, total harvest is likely to depend on the amount of pheasant habitat in a given area. Thus, a high density of pheasants in limited
habitat (which would have high index values) may yield a lower total harvest than lower densities in more extensive areas of habitat.

The prediction of relative harvest level (high, medium, low) should not suffer from this last drawback because these levels were referenced relative to harvest in each county. However, observations of >40 pheasants/day comprised only about 30% of our observations, so predictions at the county level are of limited utility and those at the basin level are only marginally better. As for linear prediction, prediction based on harvest level functioned reasonably at the state level.

CONCLUSIONS

Indices for both call counts and brood counts as well as harvest of pheasants have declined in the eastern Washington study area over the last 15 years, but accurately quantifying this decline was difficult due to varying sensitivities of the measures, variation among sites, and short-term fluctuations in environmental effects. Overall, declines were probably in the range of 5–10% per year. This may be the result of declines in both habitat quantity and quality as a consequence of intensive agricultural operations utilizing suboptimal and marginal areas.

Pheasants have a high reproductive potential, and populations can fluctuate considerably in response to annual differences in environmental conditions (Dale 1956). In particular, fair weather during the nesting season promotes chick survival, and adequate rainfall for vegetative growth increases reproduction through enhanced nutrition and improved vegetative cover for nesting and rearing of broods (Hill and Robertson 1988). Brood counts evaluate this reproductive level, and because most of the harvest is of pheasants <1 year of age (Robertson 1958), a brood count serves as the best predictor of harvest. However, this prediction appears to function satisfactorily only at broad scales. This predictive capability, nevertheless, can be improved in several ways: (1) measure the environmental variables and include them in the predictive model; (2) model the environmental variables with the previous year’s harvest; (3) incorporate site-specific effects in the analysis. In the present analysis the functional relationship between brood counts and total harvest was considered to be the same in all areas pooled for a given scale of analysis. This was done to maintain a simple predictive formula with general application. Incorporating site-specific effects would likely improve the prediction but would result in a more complex formula for estimation or perhaps separate formulas for each area.

The recent increase in the annual release of pen-reared pheasants is likely to reduce our ability to predict harvest in the future. This is not solely a result of additional birds affecting each year’s predictions. In addition, the refinement of the predictive equation based on future surveys would be confounded by the harvest of birds not present during brood counts. It would therefore be desirable to include the number of birds released as a variable in future analyses.

Pheasant chick survival varies substantially among years (Riley et al. 1998), hence the higher variability in pheasant/day in brood counts. Nevertheless, variability in brood counts for adults/day, adult females/day, and adult males/day was higher than that for crows/station in call counts and had lower power. I conclude that call counts are most effective for population monitoring, while brood counts are better for predicting harvest. However, even with substantial samples for call counts, only large changes are likely to be detected in the short term (2–3 years), and the precision of harvest prediction based on brood counts is low. For these reasons, these surveys have not been judged as cost-effective in Washington and have been discontinued.

ACKNOWLEDGMENTS

This study was supported by Federal Aid in Wildlife Restoration. I am grateful to G. Koehler, D. Martorello, and M. Schroeder for suggesting improvements to the manuscript.

LITERATURE CITED


Received 2 January 2002
Accepted 14 May 2002