VARIANCE ESTIMATION FOR THE FEDERAL WATERFOWL HARVEST SURVEYS

Paul H. Geissler, U. S. Fish and Wildlife Service, Patuxent Wildlife Research Center
Route 197, Laurel, Maryland 20708

KEY WORDS: wildlife, survey, variance, bootstrap

The Federal Waterfowl Harvest Surveys provide estimates of waterfowl harvest by species for flyways and states, harvests of most other migratory game bird species (by waterfowl hunters), crippling losses for ducks, geese, and coots, days hunted, and bag per hunter. The Waterfowl Hunter Questionnaire Survey separately estimates the harvest of ducks and geese using cluster samples of hunters who buy duck stamps at sample post offices. The Waterfowl Parts Collection estimates species, age, and sex ratios from parts solicited from successful hunters who responded to the Waterfowl Hunter Questionnaire Survey in previous years. These ratios are used to partition the duck and goose harvest into species, age, and sex specific harvest estimates. Annual estimates are correlated because successful hunters who respond to the Questionnaire Survey in one year may be asked to contribute to the Parts Collection for the next three years. Bootstrap variance estimates are used because covariances among years are difficult to estimate.

E. M. Martin provided invaluable advise on the procedures for conducting and analyzing the current Waterfowl Harvest Survey. C. M. Bunck, J. E. Hines, R. E. Trost, and B. K. Williams made helpful suggestions, and C. B. Thrash and E. Brown, Jr. provided computing assistance.

DESCRIPTION OF WATERFOWL HARVEST SURVEYS

Information comes from the Postal Service's report of duck stamp sales, the Waterfowl Hunter Questionnaire Survey, and the Waterfowl Parts Collection Survey (Voelzer, et al. 1982, Couling et al. 1982). The Questionnaire Survey estimates days hunted, harvest and unretrieved kill for duck, goose and coot hunting, and the harvest of most other migratory game birds by duck stamp purchasers. In the spring, the Branch of Surveys, Migratory Bird Management Office randomly selects a sample of over 3,000 out of the 16,000 post offices that sell duck stamps. The survey is stratified by geographic zones within states and by 3 post office size categories based on the number of duck stamps sold (<100, 100-999, >999). Numbers of sample post offices are allocated to strata so that the same proportion of the duck stamps are sampled in every stratum. Each post office is a cluster sample of duck stamp buyers.

The stamps and contact cards for the Waterfowl Hunter Questionnaire Survey are sent to sample post offices before sales begin on 1 July. Two reminders to hand out contact cards are sent to sample post offices, one shortly before the season opens and another about mid-season. When a person buys a duck stamp at a selected post office, the clerk gives the buyer a contact card. The buyer is asked to return the card with his name and address, and an indication of whether or not he purchased the stamp for hunting or for collecting. The card informs the buyer that he will receive a questionnaire about his duck harvest and provides a brief diary to record his hunting activity. Potential hunters who return contact cards before a cut off date just prior to the close of the hunting season in their states are included in the Questionnaire Survey. About 70,000 questionnaires are mailed shortly after the close of the season, and after 3 or 4 weeks, a follow up questionnaire is sent to those who did not respond. About 50,000 questionnaires are returned.

The Waterfowl Parts Collection Survey estimates species, age, and sex ratios for partitioning the duck and goose harvests (estimated by the Questionnaire Survey) into those categories. About 30,000 successful hunters are sent envelopes each year and are requested to return duck wings and goose tail feathers from each bird they shoot. About 80,000 birds are sampled each year. The names are selected from last year's Parts Collection and Hunter Questionnaire Surveys. Cooperating hunters may remain in the sample for 3 years before being replaced.

HARVEST ESTIMATION

Quantities such as the harvest of all duck species combined, harvest of all goose species combined, harvest of coots, crippling losses, days hunted, and the harvest of other migratory game bird species can be estimated from the Waterfowl Hunter Questionnaire Survey. That harvest K_g is estimated as

$$\hat{K}_g = \sum_z \left[\left(\frac{\sum_p k_{zgp}}{\sum_p h_{zgp}} \right) \left(\frac{\sum_y^m f_{zy}}{m} \right) A_z \, H_z \right] \qquad (1)$$

where

- z = zone and post office size strata of stamp purchase,
- g = species group such as ducks,
- p = post office where stamp was purchased,
- y = year,
- k_{zgp} = kill (or days hunted) reported by hunters returning questionnaires,
- h_{zgp} = number of hunters returning questionnaires,
- f_{zy} = proportion of stamps sold for hunting, a pooled estimate for all zones in a state,
- m = number of annual proportions f_{zy} (usually 10),
- A_z = adjustment factor for memory and prestige bias, and junior hunters, and
- H_z = number of duck stamps sold.

The first factor $\left(\frac{\sum_p k_{zgp}}{\sum_p h_{zgp}} \right)$ is the success (kill/hunter or days hunted/hunter) reported by the hunters that respond to the Waterfowl Hunter Questionnaire Survey in geographic zone and post office size stratum z for estimate group g (e.g. ducks). The number of hunters is estimated by multiplying the stamp sales H_z and by the mean proportion of stamps sold for hunting $\left(\frac{\sum_y^m f_{zy}}{m} \right)$. Harvest is estimated by multiplying success by the number of hunters and by A_z to adjust for memory and prestige bias and for junior hunters. Memory and prestige adjustment A_z is treated as a known constant although it is actually an estimate with an unknown variance. Estimates for the harvest in a particular state are made by restricting the kill k_{zgp} to that state in (1).

Other estimates require species, age, and/or sex ratios. Groups of species such as ducks, geese or coots will be referred to as a group. Estimates for groups such as the harvest of ducks from the Questionnaire Survey are partitioned into subgroups such as adult mallards using ratios estimated by the Parts Collection. Including the ratio $\left(\frac{\sum_p w_{zgsp}}{\sum_p \sum_i w_{zgip}} \right)$ in (1) provides an estimate for a subgroup

$$\hat{K}_{gs} = \sum_z \left[\left(\frac{\sum_p k_{zgp}}{\sum_p h_{zgp}} \right) \left(\frac{\sum_p w_{zgsp}}{\sum_p \sum_i w_{zgip}} \right) \left(\frac{\sum_y^m f_{zy}}{m} \right) A_z \, H_z \right] \qquad (2)$$

where

s = subgroup such as adult mallards,

w_{zgsp} = number of wings or other parts, and

i = summation index.

The post office p for the parts refers to the post office where the hunter purchased the duck stamp that resulted in his being included in the Hunter Questionnaire Survey. If the hunter is included in the Parts Collection in subsequent years, the hunter is still identified by the post office where his name was first selected.

Bootstrap estimates (Efron 1982, Diaconis and Efron 1983, Efron and Tibshirani 1986) are used because they reduce the bias associated with the estimation of ratios from order n^{-1} to n^{-2} (n is number of post offices) and because bootstrap variance estimates include the effects of correlations among years. Bootstrap estimates of variance start by drawing a sample of n_z post offices with replacement out of the n_z post offices in zone and post office size stratum z. Post offices are the primary sampling units and must be used to estimate variances. Data on the proportion of stamps sold for hunting are not available at the post office level. Consequently the variance due to that proportion is calculated among years instead among post offices. A sample of m years are selected with replacement out of the m years used to estimate the proportion of stamps sold for hunting. A particular bootstrap sample b consists of a sample of n_z post offices and m years, each selected with replacement from the original sample. A bootstrap sample estimate of harvest $\hat{K}'_{g(b)}$ or $\hat{K}'_{gs(b)}$ is calculated from this bootstrap sample using Equations 1 and 2 respectively. A large number (B = 200) of bootstrap samples are selected, and the harvest $\hat{K}'_g$ or $\hat{K}'_{gs}$ is estimated as the median of the bootstrap sample estimates. The mean is usually used but the median may be superior for nonsymmetric distributions (Geissler 1987). The variance is estimated as

$$\hat{v}(\hat{K}'_g) = \sum_{b=1}^{B} (\hat{K}'_{g(b)} - \hat{K}'_g)^2/(B\text{-}1) \; . \qquad (3)$$

The 100 α % confidence interval is then estimated as

$\hat{K}'_g \pm \underline{t}_{p\alpha} \sqrt{\hat{v}(\hat{K}'_g)}$ where $\underline{t}_{p\alpha}$ is the Student's $\underline{t}$ value with $p = \sum_z (n_z\text{-}1)$ degrees of freedom. The variance and confidence interval for $\hat{K}'_{gs}$ are similar.

A similar procedure is used for estimating the percent change in the harvest between years, harvest trends, and contrasts between the harvest in years with liberal and conservative hunting regulations. A function of the harvest estimates is calculated for each bootstrap sample. For example the function $\hat{F}'_b = 100\ (\hat{K}'_{gy(b)} - \hat{K}'_{gy'(b)}) \ / \ \hat{K}'_{gy'(b)}$ represents the percent change in the harvest between years y and y' where $\hat{K}'_{gy(b)}$ is the harvest estimate from (1) for year y and bootstrap sample b. The function is estimated by $\hat{F}'_g$ the median of the function estimates for the bootstrap samples and the variance is estimated as in (3), substituting F for K. Variance estimates for these quantities which involve more one year must include the correlations among years. Annual estimates are correlated because successful hunters who respond to the Questionnaire Survey in one year may be asked to contribute to the Parts Collection for the next 3 years. Bootstrap variance estimates correctly include the effects of correlations among years and surveys, because estimates for functions are made independently for each bootstrap sample. Correlations (covariances) do not have to be explicitly considered because the variance is estimated from differences among the independent post offices.

EXAMPLE

In the example (Tables 1 and 2), a state has 2 zones. The Hunter Questionnaire Survey has a sample of 6 post offices in zone 1 and 4 post offices in zone 2. The adjustment factor is 0.868 for this state. A total of 21,957 duck stamps were sold in zone 1 and 15,095 in zone 2. The total duck harvest estimate based on the Hunter Questionnaire Survey for the original sample from Equation 1 is

$$\left[\left(\frac{1{,}933}{474}\right)(0.901)(0.868)(21{,}957)\right]$$

$$+\left[\left(\frac{924}{215}\right)(0.901)(0.868)(15{,}095)\right] = 120{,}763.$$

Parts (duck wings) were collected from hunters who bought duck stamps the previous year at 4 different post offices in zone 1 and from 7 different post offices in zone 2. Because the hunter's names for the Parts Collection came from the Questionnaire Survey the previous year, those hunters are identified with the post office where they were originally sampled. Including this information from the Parts Collection Survey, the estimated total mallard harvest for the original sample from Equation 2 is

$$\left[\left(\frac{1{,}933}{474}\right)\left(\frac{13}{41}\right)(0.901)(0.868)(21{,}957)\right]$$

$$+\left[\left(\frac{924}{215}\right)\left(\frac{21}{59}\right)(0.901)(0.868)(15{,}095)\right] = 40{,}262.$$

Variances are estimated among bootstrap samples. These samples are generated from the original sample by randomly selecting post offices and years with replacement (Tables 1 and 2). Thus Apex is included in the bootstrap sample twice but Bowie is excluded. Substituting the values for the bootstrap sample given in Tables 1 and 2 into Equation 1 gives a bootstrap sample estimate of 120,218. Continuing this process with 3 more bootstrap samples gives the estimates 120,763, 120,218, 122,987, 119,606, and 123,666 for the original sample and the first 4 bootstrap sample estimates of the duck harvest. These estimates for the mallard harvest are 40,262, 36,434, 42,158, 38,428, and 38,835. The standard errors estimated from these 2 sets of 5 estimates are 1,779 and 2,136 respectively. The 95% confidence intervals with 18 degrees of freedom are $120{,}763 \pm 2.101(1{,}779) = 117{,}025 - 124{,}501$ and $38{,}835 \pm 2.101(2{,}136) = 34{,}347 - 43{,}323$ respectively.

RESULTS AND DISCUSSION

The 95% confidence intervals of the harvest estimates by flyways (Figs. 1 and 2) show that the harvest estimates are very reliable for mallards (mean confidence interval of ±8% of the harvest estimate), Canada geese (±11%), and black ducks (±16%); and moderately reliable for canvasbacks (±32%), snow geese (±43%), and brant (±46%). Snow goose harvest estimates are very reliable in the Central Flyway (±15%) where they are most numerous. The confidence intervals are relatively larger for smaller harvest estimates because it is more difficult to precisely estimate the harvest of less common species.

Confidence intervals for the estimates of the percent change between the annual harvest estimates indicate how large a change in harvest can be detected by the Federal surveys (Figs. 3 and 4). At a flyway level, small changes can be detected with mallards (10%), canada geese (13%), and black ducks (22%); and larger changes with canvasbacks (42%), snow geese (43%), and brant (58%).

Confidence intervals for state estimates (Fig. 5) are much larger than those for the flyway estimates. The larger state estimates such as the harvest of mallards in Louisiana (±37%) and the corresponding percent change estimates (±32%) are reasonably precise. However, other state estimates such as the harvest of black ducks in Maryland (±89%) and the corresponding percent change estimate (±129%) are only marginally useful.

LITERATURE CITED

Couling, L. M., A. R. Sen, and E. M. Martin. 1982. Reliability of kill and activity estimates in the U. S. waterfowl hunter survey. Spec. Sci. Rep. - Wildl. 240. U. S. Fish and Wildlife Service, Washington, D.C., 14 pp.

Diaconis, P., and B. Efron. 1983. Computer intensive methods in statistics. Sci. Am., May 116-130.

Efron, B. 1982. The Jackknife, the bootstrap and other resampling plans. Soc. for Industrial and Applied Math., Philadelphia. 92 pp.

Efron, B. and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1:54-77.

Geissler, P. H. 1987. Bootstrapping the lognormal distribution. Proc. of Comput. Sci. and Stat.: Nineteenth Symp. on the Interface: 543-545.

Voelzer, J. F., E. Q. Lauxen, S. L. Rhoades, and K. D. Norman. 1982. Waterfowl status report 1979. Spec. Sci. Rep. - Wildl. 246. U. S. Fish and Wildlife Service, Washington, D.C., 96 pp.

Table 1. A sample of post offices and one bootstrap sample used to illustrate the estimation of harvest confidence intervals.

Original sample						Bootstrap sample					
		Questionnaires		Parts				Questionnaires		Parts	
Zone	Post office	Number returned	Ducks reported	Total wings	Mallard wings	Zone	Post office	Number returned	Ducks reported	Total wings	Mallard wings
1	Apex	52	216			1	Apex	52	216		
1	Bowie	90	366			1	Apex	52	216		
1	Cecil			14	4	1	Cecil			14	4
1	Clinton	106	431			1	Clinton	106	431		
1	Jay	109	443			1	Jay	109	443		
1	Knox			8	2	1	Knox			8	2
1	Martin	95	387			1	Knox			8	2
1	Pace	22	90			1	Knox			8	2
1	Waldo			13	5	1	Pace	22	90		
1	Webster			6	2	1	Webster			6	2
Total		474	1,933	41	13	Total		341	1,396	44	12
2	Bangor			7	3	2	Bangor			7	3
2	Cary	33	133			2	Cary	33	133		
2	Clay			9	3	2	Cary	33	133		
2	Ithaca			4	1	2	Ithaca			4	1
2	Kent			18	7	2	Kent			18	7
2	Lyon			5	2	2	Oxford	74	305		
2	Oxford	74	305			2	Oxford	74	305		
2	Polk			10	3	2	Polk			10	3
2	Wayne	63	258			2	Polk			10	3
2	York	45	228	6	2	2	York	45	228	6	2
Total		215	924	59	21	Total		259	1,104	55	19

Table 2. A sample of years and one bootstrap sample used illustrate the estimation of harvest confidence intervals.

Proportion of stamps sold for hunting											
Original sample						Bootstrap sample					
Year 1	Year 2	Year 3	Year 4	Year 5	Mean	Year 2	Year 2	Year 3	Year 4	Year 4	Mean
0.943	0.896	0.840	0.930	0.896	0.901	0.896	0.896	0.840	0.930	0.930	0.898

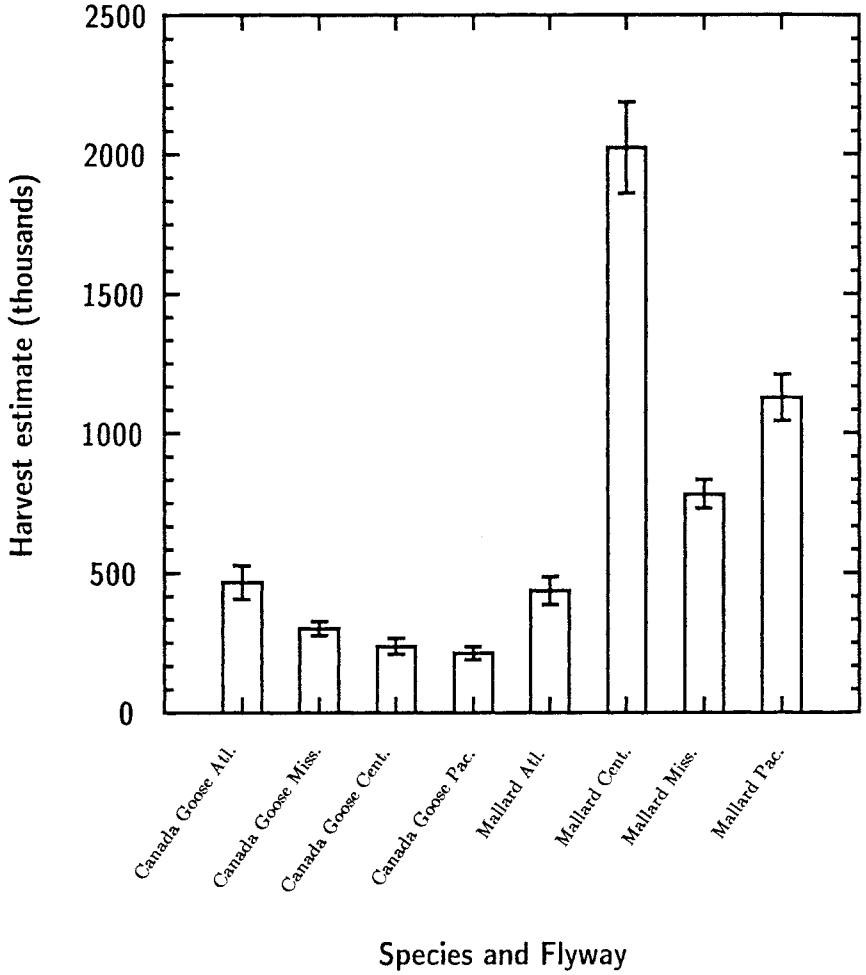

Fig. 1. Ninety-five percent confidence intervals for the annual harvest of Canada geese and mallards in the Atlantic, Mississippi, Central, and Pacific Flyways (mean of 1983 and 1984 intervals). The bar specifies the harvest estimate and the vertical line indicates the confidence interval.

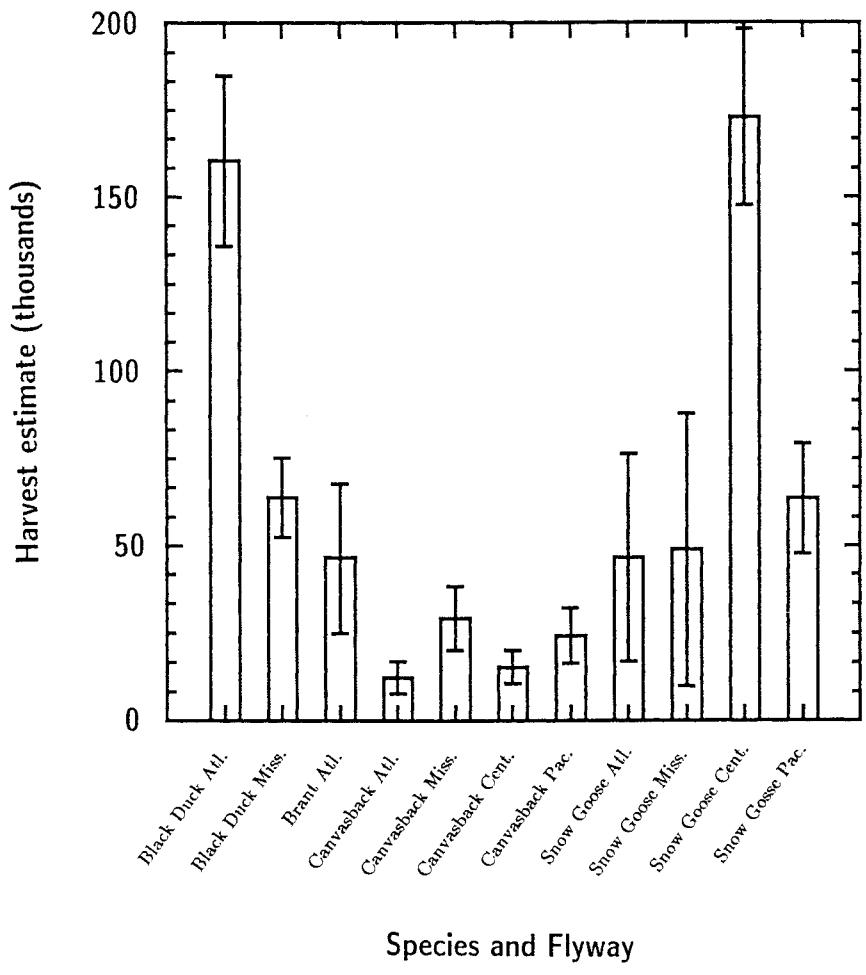

Fig. 2. Ninety-five percent confidence intervals for the annual harvest of black ducks, brant canvasbacks, and snow geese in the Atlantic, Mississippi, Central, and Pacific Flyways (mean of 1983 and 1984 intervals). The bar specifies the harvest estimate and the vertical line indicates the confidence interval.

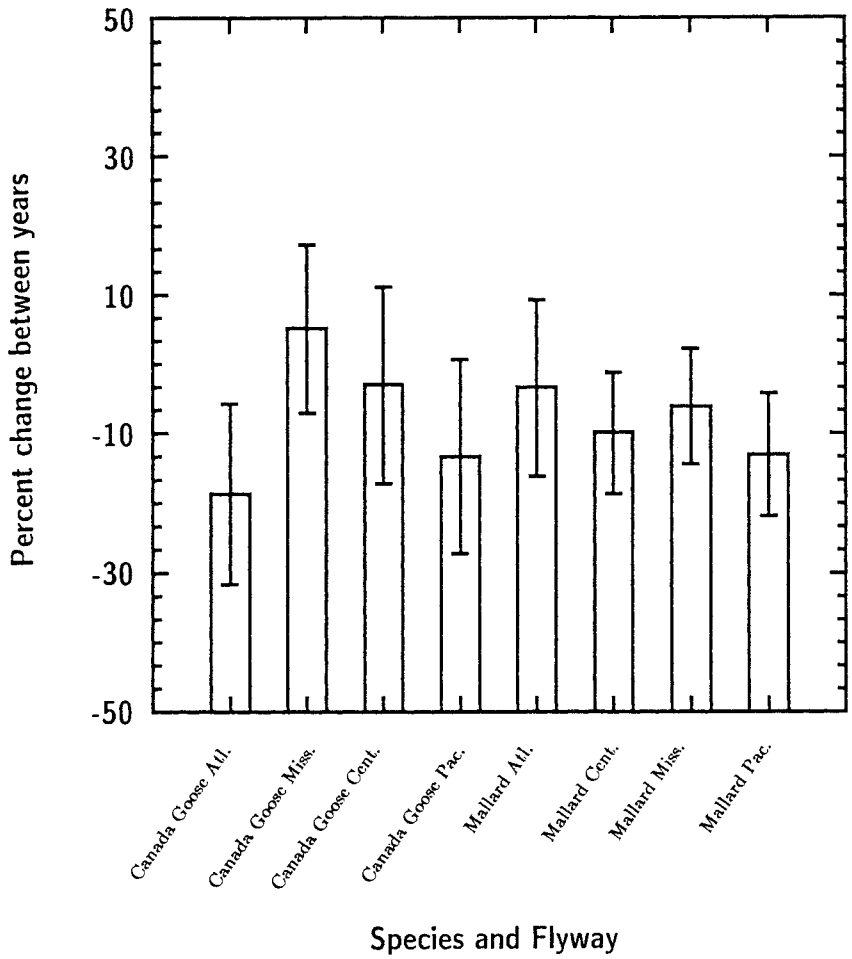

Fig. 3. Ninety-five percent confidence intervals for the percent change between estimates of the 1983 and 1984 annual harvest of Canada geese and mallards in the Atlantic, Mississippi, Central, and Pacific Flyways. The bar specifies the harvest estimate and the vertical line indicates the confidence interval.

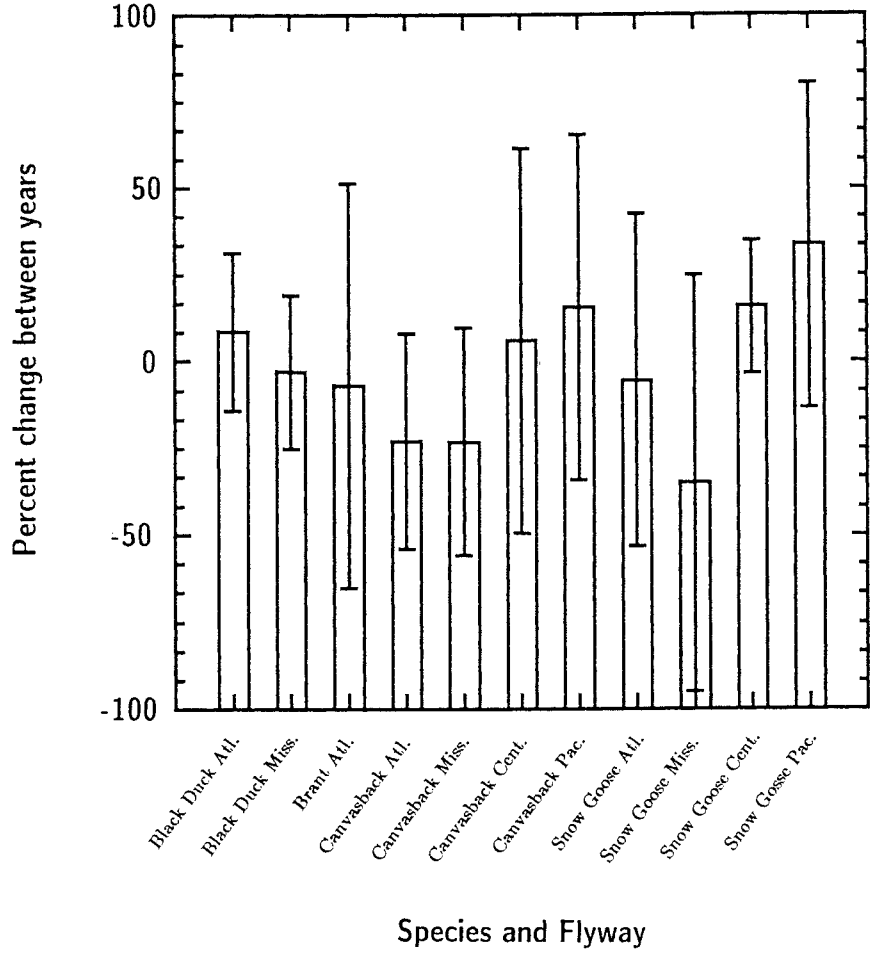

Fig. 4. Ninety-five percent confidence intervals for the percent change between estimates of the 1983 and 1984 annual harvest of black ducks, brant canvasbacks, and snow geese in the Atlantic, Mississippi, Central, and Pacific Flyways. The bar specifies the harvest estimate and the vertical line indicates the confidence interval.

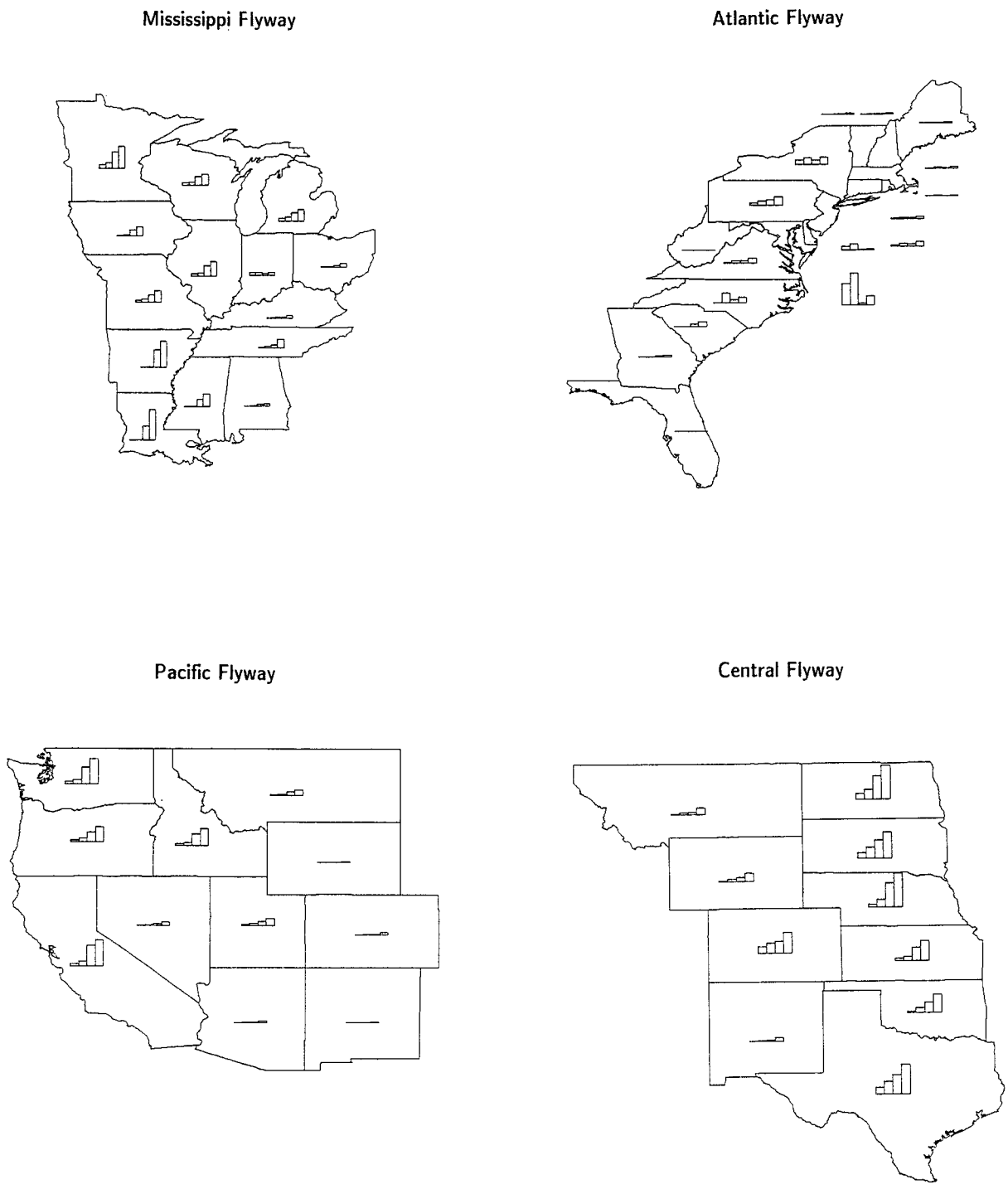

Fig. 5. State 95% confidence intervals for the harvest of Canada Geese and Mallards (mean of 1983 and 1984 intervals). The histogram bars for from left to right are lower Canada goose confidence limit, upper Canada goose confidence limit, lower mallard confidence limit, upper mallard confidence limit. Portions of Colorado, Montana, New Mexico, and Wyoming are in the Central Flyway and portions of those states are in the Pacific Flyway.