Journal of Wildlife Management 74(6):1326–1334; 2010; DOI: 10.2193/2009-255

Tools and Technology Article

Estimation of Species Identification Error: Implications for Raptor Migration Counts and Trend Estimation

JOSHUA M. HULL, *Golden Gate Raptor Observatory, Building 1064, Fort Cronkhite, Sausalito, CA 94965, USA*
ALLEN M. FISH, *Golden Gate Raptor Observatory, Building 1064, Fort Cronkhite, Sausalito, CA 94965, USA*
JOHN J. KEANE, *United States Department of Agriculture, Forest Service, Pacific Southwest Research Station, Sierra Nevada Research Center, 1731 Research Park Drive, Davis, CA 95618, USA*
SYLVIA R. MORI, *United States Department of Agriculture, Forest Service, Pacific Southwest Research Station, 800 Buchanan Street, Albany, CA 94710, USA*
BENJAMIN N. SACKS, *Canid Diversity and Conservation Unit, Veterinary Genetics Laboratory, University of California, One Shields Avenue/Old Davis Road, Davis, CA 95616, USA; and Department of Biological Sciences, California State University Sacramento, 6000 J Street, Sacramento, CA 95819, USA*
ANGUS C. HULL,[1] *Golden Gate Raptor Observatory, Building 1064, Fort Cronkhite, Sausalito, CA 94965, USA*

ABSTRACT One of the primary assumptions associated with many wildlife and population trend studies is that target species are correctly identified. This assumption may not always be valid, particularly for species similar in appearance to co-occurring species. We examined size overlap and identification error rates among Cooper's (*Accipiter cooperii*) and sharp-shinned (*A. striatus*) hawks specific to a raptor migration count station along the Pacific Coast of North America. Illustrating the difficulty of distinguishing between these 2 species, we found overlap in 7 metrics among species–sex groups and in 2 metrics between species, and a principal components analysis revealed a continuum of discrete clusters for each species–sex combination in morphospace. Among juvenile hawks ($n = 940$), we found the greatest misidentification rate for male Cooper's hawks (23% of the 156 males were identified as sharp-shinned), lesser error rates for female Cooper's (8%, $n = 339$) and female sharp-shinned (6%, $n = 246$), and the lowest misidentification rate for male sharp-shinned hawks (0%, $n = 199$). We observed a similar pattern of misidentification among adult hawks ($n = 48$). We attempted to use conditional probabilities (identification rates) from calibration data to calculate the true number of adult and juvenile Cooper's hawks and sharp-shinned hawks. Discrepancies between total number of observed accipiters and estimated number using calibration data suggest that daily observer misclassification rates are higher than misclassification rates estimated from calibration data and prevent correction of the raw data. Our results illustrate the importance of testing for and quantifying observer error in species identification in wildlife census and population trend studies particularly when target species may be easily confused with other nontarget species.

KEY WORDS *Accipiter cooperii*, *Accipiter striatus*, California, Cooper's hawk, error, hawk count, identification, migration, monitoring, sharp-shinned hawk.

Raptor populations are often difficult and costly to monitor on their breeding grounds, especially raptors in the genus *Accipiter* due to their low population densities, cryptic colorations, forested habitats, and often secretive behaviors (Fuller and Mosher 1981, Widén 1988, Bildstein and Meyer 2000, Curtis et al. 2006). Among raptor species, a cost-effective alternative to conducting wide-ranging breeding surveys has been to annually monitor numbers at spring and autumn migration concentration points (Farmer et al. 2007). Monitoring using long-term raptor migration counts has played a valuable role in recognition of population trends (e.g., Spofford 1969, Bednarz et al. 1990, Kjellen and Roos 2000). For example, raptor counts initiated in 1934 at Hawk Mountain Sanctuary, Pennsylvania, USA, were vital in documenting population declines of several raptor species in the wake of widespread use of the pesticide dichloro–diphenyl–trichloroethane (DDT), as well as in tracking the increase in migrant raptors following cessation of DDT use in the United States (Bednarz et al. 1990). Raptor migration counts are now commonly conducted worldwide in part to monitor raptor population trends (Zalles and Bildstein 2000).

As with any monitoring method, accuracy and precision of raptor migration counts can be biased by many factors that can introduce both sampling and process variation. For example, raptor migration counts are known to be affected by the number of human counters and visitors at hawk-count sites, local and annual weather patterns, variation in bird flight altitudes, annual variation in reproduction, and annual variation in migration routes among other factors (Fuller and Mosher 1981, Bildstein et al. 2007). One source of potential bias that has not received research attention concerns species identification error rates. Whereas many species are clearly distinguishable based on morphology, plumage, and behavior, other more similar species may be more difficult to correctly identify. Systematic errors in species identification could lead to biased counts that may influence interpretations of population trends and abundance (Robbins and Stallcup 1981). Therefore, determining identification error rates for similar species will improve the understanding of species-specific migration patterns, as well as the usefulness of migrant counts for monitoring long-term raptor population trends.

Cooper's hawks (*Accipiter cooperii*) and sharp-shinned hawks (*A. striatus*) are counted in large numbers at many raptor migration sites across North America (Zalles and

[1] *E-mail: bhull@parksconservancy.org*

Bildstein 2000). Cooper's and sharp-shinned hawks differ in diet and some habitat preferences and their respective population trends may be associated with habitat- and region-specific effects. Thus, correct species identification and estimation of trends may play a role in understanding the ecosystem processes reflected in migration counts of these hawks. Cooper's and sharp-shinned hawks are similar in shape, plumage, flight pattern, and size, complicating identification (Clark 1984, Bildstein and Meyer 2000, Curtis et al. 2006). Further, potential for misidentification may be even more acute in western North America where male Cooper's hawks and female sharp-shinned hawks overlap in size for some morphological measurements (Smith et al. 1990). Due to this overlap, male Cooper's hawks and female sharp-shinned hawks may be more often misidentified for one another, than would female Cooper's hawks and male sharp-shinned hawks. This discrepancy results in species identification error rates that are specific to each of the 4 possible sex–species combinations (e.g., M Cooper's vs. F sharp-shinned hawks).

Our objectives were to describe the extent of morphological similarity and to estimate species identification error rates between male and female Cooper's and sharp-shinned hawks counted during autumn migration at the Marin Headlands near San Francisco, California, USA. The Marin Headlands site is an ideal location to investigate accipiter identification errors because both species are counted in large numbers each autumn and the degree of morphological overlap between these species is more extensive at western North American raptor counts than elsewhere on the continent (Smith et al. 1990). After quantifying identification error rates for each species, sex, and age class we then attempted to develop correct identification classification probabilities to estimate how these error rates may affect the raw migration counts.

STUDY AREA

The Golden Gate Raptor Observatory (GGRO) study site was located at the tip of the Marin peninsula between the San Francisco Bay and Pacific Ocean (37°49′49″N, −122°29′59″W). Raptor migration was first described along the Marin peninsula by Binford (1979). Full-season raptor banding efforts (beginning in 1984) and group-based hawk counts (beginning in 1986) were subsequently established on Hawk Hill in the Marin Headlands. Following 3 years of preliminary counts, the count methodology was standardized in 1989 using a site-specific quadrant system (McDermott and Fish 1991).

METHODS

We conducted hawk counts using rotating teams of 7–17 volunteer hawk counters of various degrees of identification skill. Each team was led by an experienced hawk counter with ≥5 years of experience counting hawks in the Marin Headlands. During daily counts it was the responsibility of the leader to verify and make final determinations about individual hawk identifications. From 1995 through 2006, an average of 4,050 sharp-shinned and 2,377 Cooper's hawk sightings were made each autumn at the Marin Headlands (Kauka 2006).

We trapped Cooper's and sharp-shinned hawks from 1984 through 2006 during autumn migration (15 Aug through 15 Dec). Raptor trapping was authorized by the United States Geological Survey (Federal Bird Banding Permit no. 21827), State of California (Department of Fish and Game Scientific Collecting Permit no. SC-001259), and United States National Park Service (GOGA-2001-SCI-0009) permits. We captured each hawk, using standard techniques, in a mist net, dho-ghazza, or bow net (Bloom 1987) and banded captured hawks with a federal aluminum leg band. We identified each individual to species and sex by morphometric measurements and to age by plumage characteristics (Clark and Wheeler 2001, Pitzer et al. 2008). We recorded 7 metrics for each individual: tarsus depth (greatest width at the narrowest point anterior to posterior of tarsus), hallux chord (from tip of hallux to flesh), tail length (from feather insertion to tip of longest tail feather), delta tail (difference in length between no. 1 and no. 6 rectrices), weight, exposed culmen length (from cere to tip of culmen), and unflattened wing chord (from tip of longest primary to wrist). To determine degree of overlap among migrant Cooper's and sharp-shinned hawks, we calculated mean, standard deviation, and range for each metric and then standardized those metrics according to variance. We used a principal components analysis (PCA) in SYSTAT version 11.0 (Systat Software, Inc., Richmond, CA) to visualize the degree of overlap in morphospace between male and female Cooper's and sharp-shinned hawks. We retained 2 components and rotated eigenvalues using varimax (Krzanowski 2000) and we used Tukey tests to compare PCA scores among groups to test for significant differences among species–sex groups.

We conducted accipiter identification tests evenly across years and across time within a given migration season. The hawk-counting team, located 0.30 km from the banding station, was notified by radio when an accipiter was in-hand and available for an identification study. The hawk release was coordinated by the banding team by 2-way radio from behind a prearranged landmark on a nearby ridge. During release, the raptor handler remained out of sight of the hawk counters, preventing them from making relative size comparisons that might assist in identification. Following the hawk's release and subsequent flight, observers each filled out an individual data sheet and chose among 3 alternatives for species: sharp-shinned hawk, Cooper's hawk, or unidentified species. Similarly, a choice was indicated for age: juvenile, adult, or undocumented age. The team leader then used a similar data sheet to indicate a final team decision on each released hawk. The leader was instructed to use the same methods for identifying accipiters as those used for identifying all other raptors counted during the regular hawk count.

We estimated misclassification probabilities using the 1995–2004 accipiter identification data of the captured, banded, and released birds. Data consisted of the number of captured, banded, and released Cooper's (COHA) and sharp-shinned (SSHA) hawks of known age (ad or juv) and known sex (F or M). We classified released banded birds into 8 categories, now called known categories, as follows: 1) COHA adult female, 2) COHA adult male, 3) COHA juvenile female, 4) COHA juvenile male, 5) SSHA adult female, 6) SSHA adult male, 7) SSHA juvenile female, and 8) SSHA juvenile male. This set of data was associated with the hawk-specific identification responses from the observer group. Hawks identified by the observer group were classified into 9 categories, now called observed categories, as follows: 1) COHA adult, 2) COHA juvenile, 3) COHA unknown age, 4) SSHA adult, 5) SSHA juvenile, 6) SSHA unknown age, 7) unidentified adult, 8) unidentified juvenile, and 9) unidentified unknown age.

We included unidentified categories in our estimate because we had no metric that allowed us to allocate to either the Cooper's or sharp-shinned hawk totals individuals called as unidentified during the count. Using simple formulae to allocate unknowns according to proportions of positively identified individuals does not work well for allocating unknowns because we have no way of establishing the known proportion of Cooper's and sharp-shinned hawks in migration without first determining identification error rates and correcting the raw data. Our pool of positively identified individuals consisted of the proportions observed among trapped hawks. Due to trapping biases, this proportion was unlikely to reflect species composition in migration. Relative proportions of Cooper's and sharp-shinned hawks in the raw hawk-count data were unlikely to be correct because of the likely asymmetric misclassifications that our study was designed to measure. Consequently, the unknown category included an unknown proportion of the 2 species. Even if the true proportion of the 2 hawks in the count data, sans unknowns, were known, there is no reason to expect that each species was equally likely to be called unknown by observers; that is, there is a sex–species effect in the determination process. Estimating the species proportion in the unknown category was of great interest because it not only allowed us to estimate the true proportion of the 2 species in the raw data, it also provided an opportunity to understand which sex–species categories were more likely to be called as unknown regardless of the relative proportion in migration.

We estimated conditional probability of classifying the accipiter into 1 of the 9 observed categories given that the individual was in a known category using the following ratio:

$$\text{Probability}(\text{observed}/\text{known}) = \text{(no. of birds in the observed category that are also in the known category)}/\text{(total no. of birds in the known category)}$$

We used a logistic regression model as a method to estimate the yearly (10-yr) conditional probabilities for the 9 observer groups applied to each of the 8 banded groups by applying a computer routine that fitted the logistic model (Linsey 1995). This estimation resulted in the same kind of estimates illustrated above; the difference was that we estimated yearly confidence intervals for those probabilities as well. We performed this estimation using Program SAS GENMOD procedure (SAS 9.2; SAS Institute, Cary, NC).

We estimated conditional probabilities (or identification rates) for each year within the 10-year sample. We fitted 8 Logit models (1/known category) to 9 categorical explanatory variables, 1/observed category as shown below. For each i (i = 1, 2, …, 8),

$$\text{Log}\left(p_{\text{ijk}}/\left(1-p_{\text{ijk}}\right)\right) = \text{observed}_{\text{ijk}} + \text{overdispersion\_error},$$

where p_{ijk} = conditional probability for each observed category j given a known category i for year k, j = 1, 2, …, 9 and k = 1, 2, …, 10, $\text{observed}_{\text{ijk}}$ = observed group j for banded group i in year k. The estimation of conditional probabilities should have included 72 (8 × 9) conditional probabilities per year for each of the 10 study years. However, with the exception of 1995, not all observed categories had data every year; thus, we obtained 576 year-banded–observed combinations instead of 8 × 9 × 10 = 720 conditional probabilities. These estimated probabilities are available by request.

We assumed responses to have an overdispersed binomial distribution (McCulloch and Searle 2001). The data set was not large (988 hawks counts in 10 yr) or complete (560 conditional probabilities instead of 720) enough to fit explanatory variables (e.g., wind speed, month) other than the ones resulting from the observed effect per each of the 8 known categories. We used chi-square to test equality among yearly identification rates for each group or banded category. Because chi-square tests were not significant, with the exception of the 1995 conditional probability of being identified by the group as Unidentified species and unknown age given that it was banded as juvenile Cooper's hawk female, we used the average of yearly estimated misclassification probabilities as our correction rates.

We used raw 1992 through 2006 hawk counts to attempt using identification rates (conditional probabilities from calibration data) to correct raw counts (Appendix). The 1992 through 2006 raw data set consisted of 7 observed categories: 1) COHA adult, 2) COHA juvenile, 3) COHA unknown age, 4) SSHA adult, 5) SSHA juvenile, 6) SSHA unknown age, and 7) unidentified accipiter of unknown age. This last category could potentially include hawks other than COHA or SSHA. Therefore, only the first 6 categories match the first 6 observed categories from the calibration set. We used the SAS IML procedure to solve the system of equations (as described in Appendix). We applied the procedure to each of the 15 years of raw migration data (1992–2006) and to the sum over the 15 years.

RESULTS

The only morphometric variables with overlap between species were wing chord and tarsus depth, overlapping between female sharp-shinned hawks and male Cooper's hawks (Table 1). There was no overlap among the 4

Table 1. Mean (±SD), range, and sample size (*n*) for 7 morphometric characters in male and female sharp-shinned and Cooper's hawks in the Marin Headlands, California, USA, 1984–2006. Overlap between female sharp-shinned and male Cooper's hawks is seen in tarsus depth and wing chord.

Morphometric characters	Mean, SD, range, sample size	Sharp-shinned M	Sharp-shinned F	Cooper's M	Cooper's F
Tarsus depth (mm)	$\bar{x}$	3.6	4.6	5.9	7.4
	SD	0.2	0.3	0.4	0.4
	range	3.1–4.0	3.9–5.2	5.1–6.6	6.2–8.0
	n	1,527	2,700	1,596	3,381
Hallux chord (mm)	$\bar{x}$	11.4	14.4	19.3	22.9
	SD	0.6	0.7	0.9	1.0
	range	10.2–12.5	13.0–15.9	17.6–21.0	20.9–25.0
	n	1,503	2,639	1,552	3,325
Tail length (mm)	$\bar{x}$	135	160	189	214
	SD	4	5	6	7
	range	127–144	150–171	176–202	199–228
	n	1,479	2,633	1,540	3,281
Delta tail (mm)	$\bar{x}$	3.4	7.2	21.3	25.5
	SD	3.6	3.2	3.2	3.8
	range	−3.8–10.6	0.8–13.7	14.9–27.7	17.9–33.1
	n	1,520	2,673	1,488	3,222
Wt (g)	$\bar{x}$	98	164	277	411
	SD	6	11	20	33
	range	86–110	142–186	238–316	344–477
	n	1,150	2,058	1,268	2,973
Exposed culmen (mm)	$\bar{x}$	9.5	11.7	14.8	17.5
	SD	0.6	0.7	0.8	0.9
	range	8.3–10.8	10.3–13.1	13.3–16.4	15.8–19.2
	n	1,496	2,619	1,515	3,253
Wing chord (mm)	$\bar{x}$	166	198	216	247
	SD	5	5	6	6
	range	157–175	188–209	205–227	236–259
	n	1,524	2,697	1,588	3,301

species–sex groups for weight. There was only a small degree of overlap in the other individual morphometric variables, and only for a limited set of the species–sex groups. The greatest overlap occurred for the delta tail variable, which overlapped broadly within species between males and females, but did not overlap between species. The PCA resulted in 2 useful components. Principal component 1 (PC1) accounted for 61.4% of variance in the data set and principal component 2 (PC2) accounted for 35.7% of variance. Although individual metrics showed overlap between sex–species groups, a graphical plot of PC1 versus PC2 revealed a continuum of discrete clusters in morphospace for each of the 4 sex–species combinations (Fig. 1). We found significant differences in PC1 and PC2 values among each of the 4 groups (Tukey tests, $P < 0.001$ in all cases).

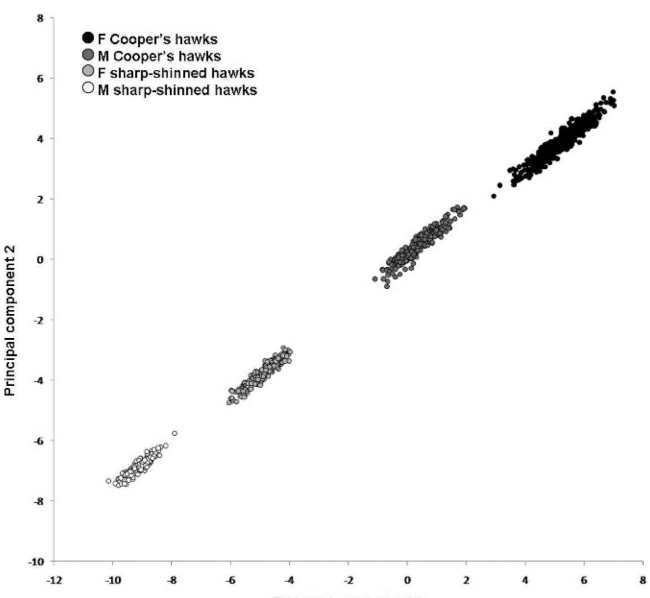

Figure 1. Principal components analysis clusters for each species–sex category showing discrete clusters in morphospace for principal component 1 and principal component 2 in male sharp-shinned, female sharp-shinned, male Cooper's, and female Cooper's hawks sampled in the Marin Headlands, California, USA, 1984–2006.

We released 988 individual accipiters (sharp-shinned hawks or Cooper's hawks) for identification studies during autumn 1995 through 2004. Identification studies were made on 417 days during that same 10-year period. We released 469 sharp-shinned hawks, including 208 males and 261 females. We released 519 Cooper's hawks, including 166 males and 353 females (Table 2).

Identification error rates were lower when observers only classified birds to species and differed between Cooper's and sharp-shinned hawks (Tables 3, 4; Fig. 2). Identification error rates were higher and more variable when observers attempted to classify birds to both species and age class. Overall, species identification error rates were higher among Cooper's than sharp-shinned hawks. Male Cooper's hawks were particularly likely to be recorded as the wrong species, with 23% of juvenile and 22% of adult males misidentified as sharp-shinned hawks. Male Cooper's hawks were most likely to be assigned to the unidentified category by observers (16% of juv, 31% of ad). Female Cooper's hawks

Table 2. Number of Cooper's (COHA) and sharp-shinned (SSHA) hawks included in the identification study and identification success rates among sex–species categories across the 10-year period of the study in the Marin Headlands, California, USA, 1995–2004.

	Known (banded birds)								
	COHA ad		COHA juv		SSHA ad		SSHA juv		
Identified as	F	M	F	M	F	M	F	M	Total
COHA ad	10	4	1	0	0	0	0	0	15
COHA juv	0	0	259	84	0	0	17	0	360
COHA unknown	2	0	17	3	0	0	1	0	23
SSHA ad	1	3	0	0	11	9	4	3	31
SSHA juv	0	0	23	34	1	0	172	153	383
SSHA unknown	0	0	3	6	1	0	17	22	49
Unidentified ad	0	2	1	2	1	0	1	0	7
Unidentified juv	0	0	20	18	1	0	21	11	71
Unidentified unknown	1	1	15	9	0	0	13	10	49
Total	14	10	339	156	15	9	246	199	988

were less likely to be recorded as the wrong species, with 7% of juveniles and 3% of adults called sharp-shinned hawks by observers. Female Cooper's hawks were also less likely to be called unidentified (10% juv, 10% ad). We found similar error rates for female sharp-shinned hawks, where 6% of juveniles and 0% of adults were identified as Cooper's hawks, and 14% of juveniles and 10% of adults were called unidentified. We found the lowest error rates among male sharp-shinned hawks, which were misidentified at a rate of 0% for juveniles and 0% for adults. Finally, male sharp-shinned hawks were least likely to be called unidentified (11% juv, 0% ad).

Using the system of equations with the assumption of a 1:1 sex ratio, we estimated true number of Cooper's and sharp-shinned hawks observed at the Marin Headlands station (Tables 4 and 5). Our correction methodology accounted for misidentified individuals and assigned unidentified individuals into species–age categories. Observed total number of Cooper's hawk adults increased from 1,890 to 3,124 and number of Cooper's hawk juveniles increased from 25,527 to 34,822. Similarly, the total number of sharp-shinned hawk juveniles increased from 41,702 to 49,995. However, the total number of sharp-shinned hawk adults decreased from 3,595 to 2,549. The effect of misidentification and correction can be seen in the change in percentage of Cooper's hawks in the total count. In the uncorrected count, total number of observed accipiters was 116,808, with 14.3% classified as unidentified accipiters (Cooper's hawks or sharp-shinned hawks), but in the corrected count total number of birds was estimated as 90,490, with a calculated 12.3% as unidentified Cooper's hawks or sharp-shinned hawks. The difference between uncorrected total and corrected total count was 22.5% of the uncorrected total. Changing the sex ratio did not reduce the difference. For example, for a sex ratio of 2:3 the percent difference was 22.3% and for a ratio of 3:2 the difference was 22.7%. The discrepancy between the uncorrected and corrected totals is a result of inherent differences between the raw data and calibration data. When hawk-count teams observed known banded hawks they classified fewer hawks as unidentified Cooper's and unidentified sharp-shinned hawks than when they observed hawks during the normal migration count (8,633 of 116,808 vs. 23 of 988 unidentified Cooper's hawks and 18,702 of 116,808 vs. 49 of 988 unidentified sharp-shinned hawks).

DISCUSSION

Using a test data set of positively identified Cooper's and sharp-shinned hawks, we quantified the extent of morphological overlap between species and the specific age–sex–species misidentification rates. We were unable to develop a correction factor for the raw hawk-count data. Although accipiters have long been recognized as difficult to identify in the field, to the best of our knowledge this is the first study to attempt to quantify field identification error rates for this taxon (Kaufman 1990, Clark and Wheeler 2001, Liguori 2005).

Our morphological analysis of a large sample of Cooper's and sharp-shinned hawks helps to define the range of overlap and illustrates a primary contributing factor to identification error. The PCA recovered a continuum of discrete clusters in morphospace for each of the groups, consistent with expectations and illustrating the size gradient among the 4 sex–species categories (Bildstein and Meyer 2000, Curtis et al. 2006). In our analyses of individual morphometric characters, we found that weight exhibited no overlap across the species–sex groups. Additional individual metrics, other than delta tail, showed only slight overlap among sex–species combinations, and between species only the tarsus depth and wing chord overlapped between female sharp-shinned hawks and male Cooper's hawks. Overlap in wing chord between species suggests some similarity in appearance of flight profile between male Cooper's and female sharp-shinned hawks and may be a contributing factor to errors in species identification.

Within the 10-year test data set, we found a systematic error in observer identification of Cooper's and sharp-shinned hawks in the Marin Headlands. In particular, Cooper's hawks were more likely to be misidentified as sharp-shinned hawks than visa versa. The result of this pattern suggests a consistent underestimate of the relative abundance of Cooper's hawks during migration and an overestimate of sharp-shinned hawks. This bias impacts estimation of relative abundance of each species as well as estimation of the absolute number of each species observed

Table 3. Identification rates for test Cooper's (COHA) and sharp-shinned hawks (SSHA) of known species, age, and sex groups observed and identified to species and age by hawk counters in the Marin Headlands, California, USA, 1995–2004. Values are means with 95% confidence interval in parenthesis. Error rates were highest among male Cooper's hawks and lowest among male sharp-shinned hawks.

Identified as:	Known (banded) birds															
	COHA ad				COHA juv				SSHA ad				SSHA juv			
	F[a]	F CI	M[b]	M CI	F[c]	F CI	M[d]	M CI	F[e]	F CI	M[f]	M CI	F[g]	F CI	M[h]	M CI
COHA ad	0.74	0.43–1	0.47	0–0.95	0.002	0–0.01	0	0–0	0	0–0	0	0–0	0	0–0	0	0–0.02
COHA juv	0	0–0	0	0–0	0.78	0.7–0.85	0.60	0.45–0.74	0	0–0	0	0–0	0.06	0.03–0.09	0	0–0
COHA unk[i]	0.13	0–0.4	0	0–0	0.05	0.02–0.08	0.01	0–0.09	0	0–0	0	0–0	0.002	0–0.1	0	0–0
SSHA ad	0.03	0–0.11	0.22	0–0.65	0	0–0	0	0–0	0.73	0.37–1	1	1–1	0.01	0–0.03	0.02	0–0.04
SSHA juv	0	0–0	0	0–0	0.07	0.3–0.11	0.19	0.08–0.31	0.14	0–0.49	0	0–0	0.71	0.65–0.77	0.77	0.71–0.82
SSHA unk	0	0–0	0	0–0	0.007	0–0.02	0.04	0–0.07	0.03	0–0.1	0	0–0	0.07	0.04–0.1	0.11	0.05–0.17
Unid[i] ad	0	0–0	0.14	0–0.37	0.002	0–0.01	0.008	0–0.02	0.03	0–0.1	0	0–0	0.002	0–0.01	0	0–0
Unid juv	0	0–0	0	0–0	0.056	0.03–0.08	0.10	0.03–0.17	0.07	0–0.25	0	0–0	0.08	0.04–0.12	0.055	0.03–0.08
Unid unk	0.10	0–0.38	0.17	0–0.60	0.04	0–0.08	0.05	0.01–0.08	0	0–0	0	0–0	0.06	0.03–0.1	0.05	0.02–0.08
Total	1		1		1		1		1		1		1		1	

[a] $n = 14$, birds were caught and banded during 5 yr.
[b] $n = 10$, birds were caught and banded during 6 yr.
[c] $n = 339$, birds were caught and banded during 10 yr.
[d] $n = 156$, birds were caught and banded during 10 yr.
[e] $n = 15$, birds were caught and banded during 7 yr.
[f] $n = 9$, birds were caught and banded during 6 yr.
[g] $n = 246$, birds were caught and banded during 10 yr.
[h] $n = 199$, birds were caught and banded during 10 yr.
[i] unk = unknown; unid = unidentified.

at the Marin site. Although the magnitude of migration for each species may need reevaluation, trend estimations are not necessarily influenced by the pattern of error documented if we assume that misidentification error rates are consistent across years. We did not detect a significant difference in error rates across years; however, sample sizes for some of the sex–species categories were low and between-year variance high. Because the annual Marin Headlands count is performed by rotating teams of volunteers with high retention across years, differences in

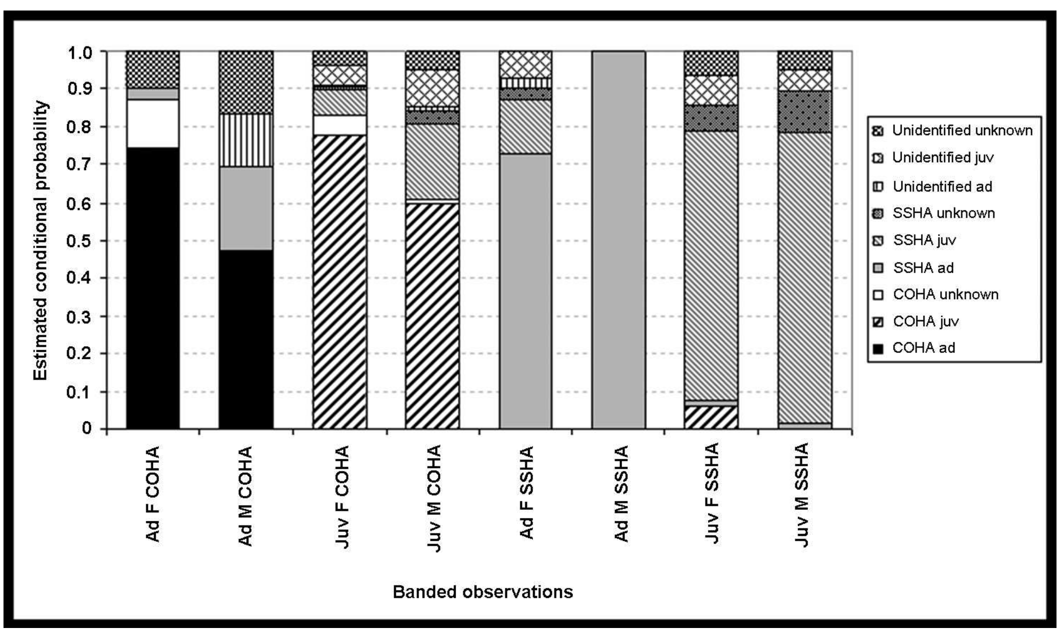

Figure 2. Estimated conditional probabilities of particular observed species–age categories for Cooper's (COHA) and sharp-shinned (SSHA) hawks belonging to a known species–age–sex category in the Marin Headlands, California, USA, 1995–2004.

Table 4. Uncorrected and corrected annual total counts assuming 1:1 sex ratio for migrating sharp-shinned (SSHA) and Cooper's (COHA) hawks using the 1995–2004 calibration data in the Marin Headlands, California, USA, 1994–2006. Uncorrected sums reflect the total numbers from the raw count, corrected sums are totals for each category estimated using a system of equations on the calibration data, and estimated uncorrected sums are a validation set testing the validity of the system of equations using uncorrected data.

Hawk	Uncorrected sum	Corrected sum	Estimated uncorrected sum
COHA ad	1,890	3,124	1,890
COHA juv	25,527	34,822	25,527
COHA unknown	8,633		1,248
Total no. of COHA	36,050	31,698	28,665
SSHA ad	3,595	2,549	3,595
SSHA juv	41,702	49,995	41,702
SSHA unknown	18,702		5,408
Total no. of SSHA	63,999	52,544	50,705
Unidentified COHA or SSHA ad	Not reported		431
Unidentified COHA or SSHA juv	Not reported		5,951
Unidentified COHA or SSHA unknown age	Not reported		4,738
Unidentified accipiter (COHA, SSHA, or others)	16,759		11,120
Identified as COHA or SSHA	100,049	90,490	79,370
Total no. of birds	116,808	90,490	90,490
% unidentified	14.3%		12.3%
% unexplained			22.5%

error rates between years may be less likely than at counts where one or a few individuals perform the annual count and there is a high degree of staff turnover between years, which suggests that the impact of misidentification error may influence trend estimates at some sites more than others depending on staffing practices. Implementation of a correction factor would be most critical at sites where between-year variance in misidentification rates is high.

As might be expected, observers had the greatest difficulty in identifying to species the species–sex combination with overlap in size: male Cooper's and female sharp-shinned hawks. However, because error rates between these 2 sex–species categories are not equal (M Cooper's hawk > F sharp-shinned hawk), size overlap alone may not account for the observed pattern of identification error. Therefore, we suggest that other factors besides similarity in size contributed to observed error rates.

An explanatory factor in this pattern of unequal error rates may be the use of flight behavior in the identification process. Just prior to the moment of each test hawk's release from behind the ridgeline, hawk counters train their binoculars on a flagged shrub. The concealed bander then releases the hawk near the same location. The hawk, having been in captivity for 10–30 minutes prior to this release, is anxious to escape the bander and become aloft. The resulting view from the counters' perspective is that of a rapidly flapping hawk exhibiting flight behaviors for achieving maximal speed and lift, resulting in flight behavior more typical of sharp-shinned hawks than of Cooper's hawks and atypical of hawks of either species observed during the standard hawk count. This altered flight profile may contribute to the pattern of misidentification, which presents a confounding element into our estimation of error rate, because many hawks observed during daily counts are not in rapid, escape flight. Therefore, a modification of our testing protocol may provide a better estimation of field error rates. For example, instead of having the entire group of observers watch the release, the team leader could watch the release and then have the rest of the team join observations after a set delay.

Table 5. Annual mean counts and ranges for sharp-shinned (SSHA) and Cooper's (COHA) hawk raw counts in the Marin Headlands, California, USA, 1992–2006.

Hawk	Yr mean count	Range
COHA ad	126	80–246
COHA juv	1,701	1,330–2,114
COHA unknown	576	384–830
SSHA ad	240	116–463
SSHA juv	2,780	1,586–4,118
SSHA unknown	1,247	608–2,016
Unidentified COHA or SSHA ad	Not reported	
Unidentified COHA or SSHA juv	Not reported	
Unidentified COHA or SSHA unknown age	Not reported	
Unidentified accipiter (COHA, SSHA, or others)	1,117	744–1,594
Overall yr mean (15 yr)	7,787.2	4,848–11,381
Total no. of birds	116,808	

Another factor confounding our estimation of error is that only Cooper's and sharp-shinned hawks were included in the test data set; therefore, observers could assume no other raptor species would be presented. Confusion between these small accipiters and other raptors commonly observed during migration at the Marin Headlands is likely to add negligible error to the count. However, additional low-level error may exist that we did not document. To address this issue, we modified our study for the 2008 migration to include all raptors trapped at the banding station.

Our attempt to directly use conditional probabilities from the calibration data set derived from the banded birds to correct raw count numbers revealed a final confounding influence. Results from the attempted correction factor suggest that correct identification rates are lower and the proportion of each species classified in 1 of the 3 unidentified accipiter categories is greater in the raw count data than in the calibration set. In other words, it seems that an observer doing routine counts is more uncertain in identifying a free-flying bird than when identifying banded birds. A potential explanation for this difference may be related to the difference between distances over which accipiters are detected and identified during the identification study and actual migration counts. In the identification study, all birds were released at a distance of 0.30 km and about 34-m elevation below the hawk-count location. In contrast, during actual migration counts, the observers are routinely detecting and attempting to identify accipiters at distances up to 1.5 km (based on distances to known landmarks) and at various elevations above and below the hawk-count location.

This hypothesized distance effect may explain why a greater proportion of accipiters are recorded as unidentified in the raw count. Further, we cannot evaluate with the current data whether there is a differential distance effect, that is, whether the relative ability to distinguish Cooper's and sharp-shinned hawks varies with distance or how identification success and distance are related. Thus, an additional step in the process of developing accurate correction factors for our count data will be to address potential distance effects to achieve conditional probabilities from future calibration sets that are more directly representative of conditional probabilities inherent to the count data. Because other trapping stations are located at various distances from the count station, distance effects could be evaluated by expanding our identification study to release study birds across the gradient of distances that hawk counters experience during actual counts rather than the single distance we used to assess identification error rates. A possible mechanism to mimic the potential range of altitudes observed during the regular count would be to have the team of observers delay their observations until notified by the team leader when the hawk achieves a particular altitude. Success of this approach would depend on whether individual hawks remained in the area long enough to reach the desired altitude.

In addition to providing valuable information about identification error rates of Cooper's and sharp-shinned hawks, the accipiter identification study played an unanticipated role in changing the behavior of counters collecting raw data. Prior to initiation of our study, a lower proportion of Cooper's and sharp-shinned hawks were classified as unidentified. Once the trial testing procedure began, prior to initiation of our formal study and errors in identification became apparent, the proportion of accipiters in raw counts classified as unidentified increased. We hypothesize that this trend reflected more careful decision-making by counters with a new understanding of their identification errors.

MANAGEMENT IMPLICATIONS

Although further study is required to validate our estimates of accipiter identification error rates and develop a correction factor, our results indicate that accipiter identification errors can be a source of bias in migration counts, and understanding inherent errors provides an opportunity to more accurately study raptor migration. Due to the geographic variation in size overlap between Cooper's and sharp-shinned hawks, identification error rates we presented may only be applicable to Pacific-region hawk counts. Because of known geographic variation among Cooper's hawks (Smith et al. 1990), we expect that identification error rates may be greatest at Pacific Coast sites, such as the Marin Headlands, where the size overlap with sharp-shinned hawks is most pronounced. However, the importance of identifying systematic errors in identification applies to all raptor counts, as well as to counts of other taxa, where co-target and nontarget species with similarities in size, shape, or plumage occur. We hope that our example will encourage other avian researchers to investigate the potential for species identification errors and to further develop methods for assessing how this type of bias may affect monitoring counts and estimates of population trends.

ACKNOWLEDGMENTS

The GGRO's Accipiter Identification Study was a 13-year cooperative effort among hundreds of volunteer banders and hawk counters. We are grateful to these volunteers for putting their interest in collecting the best possible data ahead of being correct all the time. We owe particular thanks to the count and banding leaders who took time from their full days to engage their teams in our study. K. Swenson, K. Rippens, and K. Ostrom contributed to initiation of our study in 1994. M. B. McEachern, J. Petersen, R. Simmons, and M. Stephens provided valuable comments on an earlier version of this manuscript. This analysis was partially supported by GGRO's DAP Fund. The GGRO is sponsored by the Golden Gate National Parks Conservancy and the National Park Service. This is contribution number 67 from GGRO.

LITERATURE CITED

Bednarz, J. C., D. Klem, L. J. Goodrich, and S. E. Senner. 1990. Migration counts of raptors at Hawk Mountain, Pennsylvania, as indicators of populations trends, 1934–1986. Auk 107:96–109.

Bildstein, K. L., and K. Meyer. 2000. Sharp-shinned hawk (*Accipiter striatus*). Account 482 *in* A. Poole, editor. The birds of North America online. Cornell Lab of Ornithology, Ithaca, New York, USA. <http://bna.birds.cornell.edu/bna/species/482>. Accessed 15 June 2009.

Bildstein, K. L., J. P. Smith, and R. Yosef. 2007. Migration counts and monitoring. Pages 101–115 *in* D. M. Bird and K. L. Bildstein, editors. Raptor research and monitoring techniques. Hancock House, Blaine, Washington, USA.

Binford, L. C. 1979. Fall migration of diurnal raptors at Pt. Diablo, California. Western Birds 10:1–16.

Bloom, P. H. 1987. Capturing and handling raptors. Pages 99–123 *in* B. A. G. Pendleton, B. A. Millsap, K. W. Cline, and D. M. Bird, editors. Raptor management techniques manual. National Wildlife Federation, Washington, D.C., USA.

Clark, W. S. 1984. Field identification of accipiters in North America. Birding 16:251–263.

Clark, W. S., and B. K. Wheeler. 2001. A field guide to hawks of North America. Second edition. Houghton Mifflin, Boston, Massachusetts, USA.

Curtis, O. E., R. N. Rosenfield, and J. Bielefeldt. 2006. Cooper's hawk (*Accipiter cooperii*). Account 75 *in* A. Poole, editor. The birds of North America online. Cornell Lab of Ornithology, Ithaca, New York, USA. <http://bna.birds.cornell.edu/bna/species/075>. Accessed 15 June 2009.

Farmer, C. J., D. J. T. Hussell, and D. Mizrahi. 2007. Detecting population trends in migratory birds of prey. Auk 124:1047–1062.

Fuller, M. R., and J. A. Mosher. 1981. Methods of detecting and counting raptors: a review. Pages 235–246 *in* C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in Avian Biology 6. Cooper Ornithological Society, Allen Press, Lawrence, Kansas, USA.

Kaufman, K. 1990. Advanced birding. Houghton-Mifflin, Boston, Massachusetts, USA.

Kauka, K. 2006. Hawkwatch 2005. Pacific Raptor Report 27:8–11.
Kjellen, N., and G. Roos. 2000. Population trends in Swedish raptors demonstrated by migration counts at Falsterbo, Sweden, 1942–97. Bird Study 47:195–211.
Krzanowski, W. J. 2000. Principles of multivariate analysis: a user's perspective. Revised edition. Oxford University Press, Oxford, United Kingdom.
Liguori, J. 2005. Hawks from every angle. Princeton University Press, Princeton, New Jersey, USA.
Linsey, J. K. 1995. Modelling frequency and count data. Clarendon Press, Oxford, United Kingdom.
McCulloch, C. E., and S. R. Searle. 2001. Generalized, linear, and mixed models. John Wiley & Sons, New York, New York, USA.
McDermott, F., and A. Fish. 1991. How do you know you haven't counted that hawk before? Hawk Migration Studies 16:20–21.
Newton, I., and M. Marquiss. 1979. Sex ratio among nestlings of the European sparrowhawk. American Naturalist 113:309–315.
Pitzer, S., J. Hull, H. B. Ernest, and A. C. Hull. 2008. Gender determination of three raptor species using morphology and molecular techniques. Journal of Field Ornithology 79:71–79.
Rao, C. R. 1973. Linear statistical inference and its applications. John Wiley & Sons, Hoboken, New Jersey, USA.
Robbins C. S., and R. W. Stallcup. 1981. Problems in separating species with similar habits and vocalizations. Pages 360–365 *in* C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Cooper Ornithological Society, Studies in Avian Biology 6. Allen Press, Lawrence, Kansas, USA.
Smith, J. P., S. W. Hoffman, and J. W. Gessaman. 1990. Regional size differences among fall-migrant accipiters in North America. Journal of Field Ornithology 61:192–200.
Spofford, W. R. 1969. Hawk Mountain counts as population indices in northeastern America. Pages 323–332 *in* J. J. Hickey, editor. Peregrine falcon populations—their biology and decline. University of Wisconsin Press, Madison, USA.
Widén, P. 1988. The hunting habitats of goshawks *Accipiter gentilis* in boreal forests of central Sweden. Ibis 131:205–231.
Zalles, J. I., and K. L. Bildstein. 2000. Raptor watch: a global directory of raptor migration sites. Birdlife International, Cambridge, United Kingdom; and Hawk Mountain Sanctuary, Kempton, Pennsylvania, USA.

APPENDIX: METHODOLOGY FOR CALCULATION OF THE TRUE NUMBER OF INDIVIDUALS FOR EACH KNOWN CATEGORY

Let the following set denote the names of the banded categories:

True = {(1)COHA [Cooper's hawk] adult females,
(2)COHA adult males,
(3)COHA juvenile females,
(4)COHA juvenile males,
(5)SSHA adult females,
(6)SSHA [sharp-shinned hawk] adult males,
(7)SSHA juvenile females, and
(8)SSHA juvenile males}.

Let T_{ijk} denote the number of birds in the true category (i, j, k), i = species (1 = COHA, 2 = SSHA), j = age (1 = ad, 2 = juv), and k = sex (1 = F, 2 = M). Let G_{nm} denote the number of birds observed by the group as category (n, m), n = species index (1 = COHA, 2 = SSHA, 3 = unidentified), and m = age (1 = ad, 2 = juv, 3 = unknown age). Let the conditional probability P_{nmijk} be defined as the conditional probability of identifying (by the group) a bird as species n of age m given that the bird was banded as species i, age j, and sex k:

$$P_{nmijk} = \text{Probability}[(n,m)/(i,j,k)].$$

Therefore, using probabilistic theorems (Rao 1973) we have that

$$G_{nm} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} P_{mnijk} * T_{ijk} \qquad (1)$$

where m = 1, 2, 3, n = 1, 2, 3, and $\sum_{n=1}^{3}\sum_{m=1}^{3} P_{nmijk} = 1$, for all i, j and k.

For the example provided earlier: Probability (observed/known) = 10/14 = 0.7143 is equivalent to P_{11111} = 10/14 = 0.7143.

We can calculate the value T_{ijk} (true count) as solution of the system of equations (1) when ≥8 from the 9 G_{nm} (observer raw counts) and 8 P_{nmijk} (calibration conditional probabilities) are available. Unfortunately, even though we can estimate all 8 P_{ijnms} using the 1995–2004 Banded data (Table 1), we can estimate only 6 G_{nm} for each of the explanatory levels from the 1990–2005 observed counts. That the observer can assume that the released bird was either COHA or SSHA makes the category "Unidentified species of unknown age" from the calibration set different than the same category from the raw data set. One possible solution to this problem was to make an assumption about the sex ratio for each species at each age class. With a given sex ratio, let R_{ij} = female:male ratio within each species i of age j group. Therefore,

$$T_{111} = R_{11} \times T_{112},$$

$$T_{121} = R_{12} \times T_{122},$$

$$T_{211} = R_{21} \times T_{212},$$

$$T_{221} = R_{22} \times T_{222}.$$

If we assume the sex ratio, R_{ij} known, then the system (1) is reduced to 4 equations with 4 unknowns: T_{112}, T_{122}, T_{212}, and T_{222}. As suggested by the primary sex ratio (Newton and Marquis 1979), and for the purpose of illustration, we could assume that the sex ratio, female:male, of the migrating population of both species was 1:1. Then:

$$T_{111} = T_{112},$$

$$T_{121} = T_{122},$$

$$T_{111} = T_{112}, \text{ and}$$

$$T_{121} = T_{122}.$$

Associate Editor: Bechard.