

Nolte, E. G., J. Bart, B. P. Pauli, G. S. Kaltenecker, and J. A. Heath. 2016. Detectability of migrating raptors and its effect on bias and precision of trend estimates. *Avian Conservation and Ecology* 11(2):9. <http://dx.doi.org/10.5751/ACE-00894-110209>
Copyright © 2016 by the author(s). Published here under license by the Resilience Alliance.

Research Paper

Detectability of migrating raptors and its effect on bias and precision of trend estimates

Eric G. Nolte¹, Jonathan Bart², Benjamin P. Pauli^{1,3}, Gregory S. Kaltenecker⁴ and Julie A. Heath¹

¹Boise State University Department of Biological Sciences and Raptor Research Center, ²Jon Bart Consulting, ³Saint Mary's University of Minnesota Biology Department, ⁴Boise State University Intermountain Bird Observatory

ABSTRACT. Annual counts of migrating raptors at fixed observation points are a widespread practice, and changes in numbers counted over time, adjusted for survey effort, are commonly used as indices of trends in population size. Unmodeled year-to-year variation in detectability may introduce bias, reduce precision of trend estimates, and reduce power to detect trends. We conducted dependent double-observer surveys at the annual fall raptor migration count at Lucky Peak, Idaho, in 2009 and 2010 and applied Huggins closed-capture removal models and information-theoretic model selection to determine the relative importance of factors affecting detectability. The most parsimonious model included effects of observer team identity, distance, species, and day of the season. We then simulated 30 years of counts with heterogeneous individual detectability, a population decline ($\lambda = 0.964$), and unexplained random variation in the number of available birds. Imperfect detectability did not bias trend estimation, and increased the time required to achieve 80% power by less than 11%. Results suggested that availability is a greater source of variance in annual counts than detectability; thus, efforts to account for availability would improve the monitoring value of migration counts. According to our models, long-term trends in observer efficiency or migratory flight distance may introduce substantial bias to trend estimates. Estimating detectability with a novel count protocol like our double-observer method is just one potential means of controlling such effects. The traditional approach of modeling the effects of covariates and adjusting the index may also be effective if ancillary data is collected consistently.

Défectabilité de rapaces en migration et son effet sur les biais et la précision des estimations de tendance

RÉSUMÉ. Les décomptes annuels de rapaces en migration faits à partir de lieux fixes représentent un type de relevé répandu, et la variation des effectifs comptés au fil des ans, ajustée selon l'effort d'inventaire, est utilisée couramment comme indice de l'évolution de la taille des populations. Le fait de ne pas modéliser la variabilité de la détectabilité d'une année à l'autre peut introduire un biais, réduire la précision des estimations de tendance et réduire la capacité à détecter les tendances. Nous avons effectué des décomptes avec double observateurs dépendant lors du suivi annuel des rapaces en migration automnale à Lucky Peak, Idaho, en 2009 et 2010. Nous avons appliqué des modèles de Huggins avec retrait pour une population fermée et sélectionné les modèles basés sur la théorie de l'information afin d'évaluer l'importance relative des facteurs qui affectent la détectabilité. Le modèle le plus parcimonieux incluait les effets de l'identité de l'équipe d'observateurs, de la distance, de l'espèce et du jour durant la saison. Nous avons ensuite simulé 30 années de décomptes comportant une détectabilité individuelle hétérogène, une baisse de population ($\lambda = 0,964$) et une variabilité aléatoire non-expliquée du nombre d'oiseaux disponibles. La détectabilité imparfaite n'a pas biaisé l'estimation de la tendance, mais a haussé par moins de 11 % le temps nécessaire pour atteindre une puissance de 80% à détecter une tendance. Nos résultats indiquent que la disponibilité présente une plus grande source de variance que la détectabilité dans les décomptes annuels, de sorte que les efforts visant à prendre en compte la disponibilité améliorerait la valeur des décomptes réalisés durant la migration. D'après nos modèles, les tendances sur de longues périodes de l'efficacité des observateurs ou de la distance des oiseaux en vol peuvent introduire des biais considérables sur les estimations de tendance. L'estimation de la détectabilité, au moyen d'un protocole de décompte novateur tel que notre méthode de double-observateur, est un exemple de méthodes permettant de contrôler les effets de ce type. L'approche traditionnelle dans laquelle l'effet des covariables est modélisé et l'indice ajusté en conséquence pourrait aussi être efficace si les données complémentaires sont colligées de façon constante.

Key Words: *count; distance; double-observer; hawk; monitoring; perceptibility; sampling; visibility*

INTRODUCTION

Population monitoring is essential to avian conservation (Finch and Martin 1995, Dunn 2002). The North American Breeding Bird Survey (BBS) has proven to be an effective monitoring method for many species, but trend estimates for many raptors (Accipitriformes and Falconiformes) that breed in the remote northern reaches of the continent not covered by the BBS are unreliable (Dunn et al. 2005). Breeding season surveys of North American raptors can be difficult and costly because raptors breed at low densities over large ranges (Fuller and Mosher 1981). Species that breed in forests and do not confront intruders may be easier to observe on migration (Fuller and Mosher 1987).

During migration, concentrations of visible raptors occur at predictable locations (Kerlinger 1989, Zalles and Bildstein 2000, Bildstein 2006). At such locations, termed watch-sites, observers record the numbers of each raptor species that pass overhead hourly or daily during migration (Zalles and Bildstein 2000, Bildstein et al. 2008). HawkCount (<http://hawkcount.org/>), an online data repository for watch-site counts in North and Central America, lists 203 watch-sites reporting in the last two years, or with more than five years of counts (E. G. Nolte, *personal observation* May 2016). Zalles and Bildstein (2000) listed 58 watch-sites outside North America.

A watch-site count is a two-stage sample. In the initial sampling stage, raptors must first fly past the watch-site while an observer is present; only these raptors are available to count. In the second stage, the available raptors may be either counted by an observer, or pass unseen. The count is the product of the available raptors and their probability of detection (Dunn and Hussell 1995, Nichols et al. 2009).

Migration counts are an index and cannot be used to estimate population size (Fuller and Mosher 1981, 1987, Dunn 2005); however, migration counts are thought to change roughly in proportion to change in population size, so population trends may be estimated (Dunn and Hussell 1995, Dunn 2005, Farmer et al. 2007, Farmer and Hussell 2008). If, contrary to our assumptions, the index does not change proportionately to change in the population, perhaps because of trends in availability or detection, then trend analysis may be inconclusive or even misleading (Thompson 2002, Johnson 2008). Field workers and analysts collectively spend thousands of hours every year to produce this index, so the underlying assumption deserves rigorous testing.

Two previous studies have examined the factors affecting probability of detection at raptor migration watch-sites. First, Sattler and Bart (1984), working at the Derby Hill watch-site on the shoreline of Lake Ontario in New York, compared the numbers of hawks counted by one observer tasked with watching the entire sky with six times the number counted by a second observer watching only one-sixth of the sky at a time. They found that detectability varied by observer attentiveness, flight density, flight visibility, and species. Specifically, they found that higher birds were less detectable than lower birds and that the observer was more attentive and detected raptors with greater efficiency in dense flights. Furthermore, raptor species that typically soared were detected at higher rates than species that often did not soar.

Second, Berthiaume et al. (2009), at the *Observatoire d'oiseaux de Tadoussac*, on the shoreline of the Saint-Lawrence estuary in Québec, used a double-observer approach (Nichols et al. 2000) to assess the relative effects of flight behavior and weather. Species affected detectability, with small species having lower detectability than large species. For most species, birds at eye level were most detectable, and detectability decreased with increasing altitude. Cloud cover increased the detectability of high-flying raptors while decreasing the detectability of raptors at lower altitudes. Additionally, the number of raptors migrating in a group had a significant positive effect on detectability. Wind direction and speed, cloud cover, humidity, and hour of the day affected flight altitude, and thereby affected detectability indirectly (Berthiaume et al. 2009).

Some potential factors have not yet been adequately investigated. Detectability may be affected by site-specific factors and the number of observers (Kochenberger and Dunne 1985). The detectability studies of Sattler and Bart (1984) and Berthiaume et al. (2009) were both performed at shoreline watch-sites in the northeast where lone observers made official counts. Interobserver variation in detectability exists in avian point counts (Sauer et al. 1994, Kendall et al. 1996, Cunningham et al. 1999, Nichols et al. 2000, Alldredge et al. 2007, Campbell and Francis 2011), and is likely in raptor migration counts (Dunn and Hussell 1995, Dunn et al. 2008). No previous study at a raptor migration count has tested for interobserver variation in detectability among more than two unique individual observers, nor compared the performance of any number of unique teams of observers.

With the intent to generalize the results of prior raptor detectability research and investigate observer effects, we undertook a new double-observer study of detectability at a Western ridgeline watch-site with teamed observers. Subsequently, we used our model in simulations of long-term trend analysis to estimate the effect of detectability on statistical power. We undertook this study to gain a greater understanding of the relative importance of various sources of statistical error in migration counts and suggest methodological changes and areas of future research to improve the utility of migration counts for population monitoring.

METHODS

Study site

The Lucky Peak raptor count is performed each fall by the Intermountain Bird Observatory (formerly Idaho Bird Observatory), a nonprofit research and public education program of Boise State University. At a single point, at least two observers cooperate to count migrating raptors each day, from 25 August to 31 October, as weather permits. Lucky Peak is situated at the southern end of the Boise Ridge, on the western front of the Rocky Mountains, overlooking the Snake River Plain and Boise, Idaho (43° 36'18.7" N, 116° 3'40.6" W; Zalles and Bildstein 2000, Ruelas Inzunza 2008). Owing to the elevation of the site (~1000 m above the plain), visible raptors are distributed both laterally and vertically. The watch-site also includes a raptor banding station on the west slope of the mountain, in sight of the observation point. Captured raptors are reported to the migration observers

by two-way radio. The watch-site is open to the public, and observers provide interpretation for visitors.

Data collection

We conducted a double-observer survey (Nichols et al. 2000) during the autumn raptor migration count at Lucky Peak in 2009 and 2010. Sampled days were 1–4 days apart (mean = 1.8, SD = 1.0) on 29 weekend days and 36 weekdays. Four observers were grouped in teams of two. One team, designated primary, was positioned at the traditional lookout point and attempted to count all raptors passing the lookout. The primary observers called out the identification and location of raptors they observed to avoid double-counting. The other team, designated secondary, was positioned approximately three meters behind the primary team. The secondary observers worked together to record, on a separate sheet, only those additional raptors that were not counted by the primary team. Secondary observers could ask the primary observers questions to clarify which bird had been counted, but had to be silent when identifying any birds the primary observers had missed. Therefore, detection by the primary observers was assumed to be unaffected by the activities of the secondary observers, while detection by the secondary observers was conditional on nondetection by the primary observers. Birds captured by the banding station were reported to the observers via radio and were removed from the data. We randomly assigned observers to teams each day. The observation teams remained consistent over the course of each day, except on four days in 2010 when an observer was substituted mid-day. The teams rotated between the primary and secondary roles at the end of each hour.

For individual raptors, observers recorded species and, when possible, age, sex, and color morph, as well as a distance and altitude category (*Alt.*). Observers scored birds by altitude only when within the range of unaided vision (where differences in background color and viewing angle are greatest when altitude varies), and scored more distant birds strictly by visibility with 10X magnification (Table 1). We chose this system because lateral distance affected apparent size in the same way altitude did, so distance and altitude were difficult to measure separately, and their effects on detectability were likely to be similar enough to complicate inference if modeled independently. Observers classified each bird based on its closest approach to the watch-site, even if it was initially detected further away.

At the midpoint of each hour, observers recorded weather conditions with a hand-held weather station (Kestrel 4000[®], Nielsen-Kellerman, Boothwyn, PA). Observers measured wind velocity in kilometers-per-hour, wind direction in degrees, ambient temperature in degrees Celsius, and visually estimated a cloud cover category. Cloud cover categories (*Cloud*) were: (1) Clear (0% – 15%), (2) Partly Cloudy (16% – 50%), (3) Mostly Cloudy (51% – 75%), and (4) Overcast (76% – 100%).

Statistical analyses

Detectability was estimated by fitting a closed-population mark-recapture model (closed-capture model; Otis et al. 1978). A closed-capture model, unlike simpler logistic-regression approaches, accounts for the presence of animals that were undetected. Closed-capture models are based on three key assumptions: (1) each “capture” attempt, in this case the attempt

of an observer team to detect migrant raptors, has access to the same pool of animals (a closed population), (2) animals are independent in their detection probabilities, and (3) there is no heterogeneity in detection probabilities among individual animals. To relax assumption 3, we used the conditional likelihood approach developed by Huggins (1989, 1991) to account for heterogeneity. Individually-varying detectability was modeled as a linear function of covariates related to the observer, flight, weather, and species of each bird. Observer-specific detectability is only estimable for the primary observer role in the dependent double-observer survey design, so our models require one additional assumption: (4) the detection probability for an observer team was not affected by its role (Nichols et al. 2000). The available migrant raptors were considered a closed population because observer teams were positioned closely enough to view the same extent of sky and the two counts occurred simultaneously. We seldom observed hawks migrating in groups of more than four (approximately 3% of observations), so detection of individuals could be considered independent. We excluded Turkey Vultures (*Cathartes aura*) from our analysis because they migrated in much larger groups. We likewise removed raptors not identified to genus ($n = 100$) from the analysis.

Table 1. Ordinal scale used in estimating effects of distance and altitude on detectability. Migratory flights at Lucky Peak are distributed laterally, with relatively few raptors flying high overhead. Thus, altitude was only noted at close distances, where potential differences in background color and viewing angle were substantial, and altitude could be estimated with confidence. We excluded all birds assigned to category 6 from analysis because they represent birds not within the standard search radius at this site, and not available to all observers because only one spotting scope was present. The distance classification scheme is adapted from flight altitude codes on the data form published by the Hawk Migration Association of North America (2009).

Distance code (Alt.)	Definition
0	Below level of the observers, within range of unaided vision.
1	0 – 30m above observers, within range of unaided vision.
2	> 30 m above observers, within range of unaided vision.
3	Difficult, but possible to see without binoculars.
4	Visible only with aid of 10X binoculars (but clearly seen).
5	Raptor sometimes fades out while viewing with 10X binoculars.
6	Visible only with a $\geq 20X$ spotting scope.

We used an information-theoretic model-selection approach with Akaike’s information criterion corrected for small sample size (AIC_c) as the selection criterion to assess the relative effects of these factors on detection probability (Burnham and Anderson 2002). Model-fitting was performed using the Huggins closed-capture data type in Program MARK (White and Burnham 1999). We coded raptors recorded by the primary observers with encounter history “11,” and raptors recorded only by the

Table 2. Best candidate models estimating the detectability of migrating raptors in double-observer counts conducted at Lucky Peak in 2009 and 2010. ΔAIC_c is the difference in AIC_c between the model and the model with the lowest AIC_c . L is the model likelihood, and w is the AIC_c weight of evidence. K is the number of parameters in the model. For all models, see Appendix 2.

Model	ΔAIC_c^\dagger	w	L	K
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	0	0.257	1	22
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Day} + \text{Day}^2\}$	0.39	0.212	0.823	21
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	1.13	0.146	0.568	23
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	1.92	0.099	0.384	22
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus}\}$	2.28	0.082	0.320	19
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed}\}$	2.34	0.080	0.311	20
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed}\}$	2.63	0.069	0.268	21
$p\{\text{Team} + \text{Alt.} + \text{Alt.}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud}\}$	3.10	0.055	0.212	20

[†] Lowest $AIC_c = 7098.77$

secondary observers with encounter history “01.” We fixed the value of the probability of recapture (c) equal to one because birds detected by the primary observers could not fail to be detected by the secondary observers.

We measured several covariates related to each of the four hypothesized sources of variation in detectability: observers, migratory flight, weather, and species. We examined independent measurable covariates for correlation and any with coefficients $> \pm 0.4$ we did not use in combination during model-building. Initially, we fit all possible models separately for each of the four hypothetical sources of variation, along with a null model with no covariates, and a model with only the effect of year (42 models). At this stage, we removed any variables that reduced model deviance by < 3 from further consideration, and built a general model. From subsets of variables in the general model, we built an all-combinations candidate model set (64 models). In doing so, we kept together any sets of model parameters describing a single covariate (e.g., five genus variables and wingspan, all describing the species).

We modeled the effects of observer teams (combinations of two individual observers) as dichotomous (dummy) variables. Ten teams, representing pair-wise combinations of seven regular observers, participated under a representative range of conditions (> 7 days). We pooled the 17 other observer teams with insufficient samples. The seven regular observers (Tables 2 and 3) were all recent (2004 – 2010, median = 2009) university graduates with B. Sc. degrees from wildlife and natural resource programs. All had prior professional experience assisting with field studies of wildlife (6 – 40 months, median = 15), but only one had prior experience observing bird migration (5 months). We used the number of days since the beginning of the season and the hour of the day as covariates to account for possible effects of practice or fatigue. We also modeled a second-order effect to allow for a nonlinear effect of practice (e.g., diminishing returns).

Two variables described the migratory flight. We used the number of birds observed per hour, including vultures (BPH), a naïve estimate of flight density, as a covariate for all birds observed in that hour. We used the distance category (see Table 1) as an individual covariate to model the effect of flight-line. Because we had decided (for the sake of efficient use of model parameters) to model the ordinal categories as a linear covariate, we considered

it necessary to include a second-order effect in model-selection to allow for unequal units and nonlinear effects.

Wind speed, wind direction, ambient temperature, and cloud cover category were used to describe the effects of weather. Circular variables cannot be used in linear models so we used the cosine of wind direction as a linear covariate. This number ranges from -1 (wind from the south, a headwind) to 1 (wind from the north, a tailwind). We also used the product of the cosine of wind direction and the wind speed as a covariate. This number was highest for strong tailwinds, and lowest for strong headwinds, with lighter winds and crosswinds having intermediate values. We chose these transformations because the resulting variables were likely to be correlated with the speed of migrating raptors. We chose to limit the number of wind variable interactions to avoid colinearity and make the effect of migration volume and flight line distinguishable from more proximate effects of wind.

We hypothesized that detectability might vary among species because species were of different visible size or flew with different styles. We used an approximate average wingspan for each species (from Sibley 2000) to account for differences in visible size. Raptors of unknown species ($n = 417$) were assigned a roughly approximate size, based on the information available (e.g., genus, large or small). The second-order effect of wingspan was also considered, in case detectability might increase nonlinearly with size. We used a dichotomous variable for each common genus of raptors observed (*Accipiter* $n = 3088$, *Buteo* $n = 1680$, *Circus* $n = 331$, *Falco* $n = 1525$, and *Pandion* $n = 119$) to account for differences in flight style among raptors of similar wingspan. Golden Eagles (*Aquila chrysaetos* $n = 27$) and Bald Eagles (*Haliaeetus leucocephalus* $n = 4$) composed the reference (null) category.

Tests of differences in covariates between years were performed with Fisher Exact Tests for dummy variables and Welch t Tests for quantitative variables ($H_0: \bar{x}_1 = \bar{x}_2$). Means of detectability estimates were calculated with weights of $1 / [\hat{p}_1 + ((1-\hat{p}_1) \hat{p}_2)]$, where \hat{p}_1 is the individual raptor’s estimated detectability for the primary observers and \hat{p}_2 is the individual raptor’s estimated detectability for the secondary observers. The denominator is an estimate of the total probability of the individual being detected by either of the observer teams. Weighting observations by the inverse of the detection probability is necessary to correct for the

Table 3. Estimates of coefficients from the most parsimonious model (β) with standard errors (SE), odds ratios (e^β), and AICC model-selection importance weights (Σw). All covariates were scaled to range from 0 to 1 to show relative magnitudes of effects. Asterisks indicate informative variables ($H_0: \beta \neq 0, \alpha = 0.15$). Unique observer letters represent individuals. Reference categories are: Observer team “Other” (17 teams that participated on < 7 days each) and Genus “Eagles” (*Haliaeetus* and *Aquila* pooled).

Source	Σw	Variable	β	SE	e^β	
Distance	1	<i>Alt.</i>	1.18	0.48	3.26	*
		<i>Alt.</i> ²	-2.48	0.51	0.08	*
Species	1	Wingspan	2.05	0.78	7.77	*
		<i>Falco</i>	1.17	0.98	3.22	
		<i>Accipiter</i>	1.07	0.93	2.91	
		<i>Circus</i>	0.80	0.80	2.22	
		<i>Buteo</i>	0.82	0.74	2.23	
		<i>Pandion</i>	-0.21	0.72	0.81	
		Windspeed	-0.47	0.30	0.63	*
Observer team	1	B & C	0.52	0.13	1.69	*
		B & F	0.40	0.15	1.50	*
		E & G	0.10	0.16	1.10	
		E & A	0.01	0.11	1.01	
		D & G	-0.07	0.16	0.93	
		F & C	-0.25	0.14	0.78	*
		D & C	-0.31	0.13	0.74	*
		D & E	-0.39	0.12	0.68	*
		D & F	-0.45	0.12	0.64	*
		D & B	-0.59	0.13	0.56	*
		Day	0.71	Day	0.07	0.89
		Day ²	0.47	0.89	1.60	
Model	1	Intercept	-0.15	1.00	0.86	

sample bias caused by heterogeneous detectability, i.e., more detectable birds get sampled disproportionately often (Horvitz and Thompson 1952). Descriptive values are presented as means \pm SD.

Simulation

A simulation of the number of birds passing the watch site and the number that were detected (Appendix 1), was written in R (Revolution R Community 4.3 build of R 2.12.2, Revolution Analytics, Palo Alto, California). The number of birds available to detect in year k (for $k > 1$) was $N_k = N_1 \lambda^{k-1} + \epsilon_k$. We made $\lambda = 0.964$, so N would be expected to decline by 50% in 20 years. The Raptor Population Index project adopted a similar benchmark trend ($\lambda = 0.965$) to evaluate power (Farmer and Hussell 2008, Smith et al. 2008). The ϵ_k were normally distributed with mean 0 and $\sigma_k = N_k \cdot CV(N_k)$. $CV(N_k)$ represents the square root of the variance in annual number of birds available, as a proportion of the expected N .

We used our data to define the attributes of a statistical population of possible individuals. We weighted the probability of selecting any record (the attributes of an observed bird) to be simulated by $1/p_i$ to make less detectable attribute combinations occur with realistic frequency. We randomly sampled the population (with replacement) in each year to simulate N_k available individuals. We used the attributes of the available individuals (inherited from the original records) in the model to determine their individual probabilities of detection.

Unlike all the other attributes, the observer team coefficient was modified for the purpose of simulating a realistic sequence of years. In most years at Lucky Peak the counts are predominantly

made by a single pair of observers, and a pair seldom stays the same from one year to the next. Therefore, we applied a single observer team coefficient to all individuals in a year, and changed the coefficient between years. The distribution of observer team effects was a normal distribution with a mean taken from the estimated values for the 10 observer teams and a variance inflated by 20% because the observers in two years were probably more similar than a 30-year sample of observers would be.

We determined whether each bird was detected using a random number generator and the detection probability for the bird. If the random decimal from a uniform distribution between zero and one was greater than the detection probability, the bird was not detected (otherwise, it was). The sum of birds detected in year k was the annual count C_k .

Trends were estimated by fitting an exponential curve to the counts. These regressions were performed for each of 3000 trials for every unique combination of simulation parameters tested. We estimated detectability bias as the absolute difference between the means of trend coefficients from the regressions of C_k and N_k . We assessed precision with a 90% confidence interval (the fifth to 95th percentiles) of trend estimates. Power to detect the trend was estimated as the proportion of trials in which the upper bound of the 90% confidence interval for the trend estimate parameter (β) was < 0 . The effect of variable detection on power was determined using two methods. First, we compared the mean number of years required to achieve 80% power with perfect detection (by regression of N_k) and variable detection (by regression of C_k). Second, we plotted power with perfect and variable detection as a function of survey duration with durations of five to 30 years.

Analyses were carried out for the Sharp-shinned Hawk (*Accipiter striatus*) and Northern Harrier (*Circus cyaneus*). Both species were common at Boise Ridge, rarely traveled in groups, and are high priorities for alternative range-wide surveys by Partners in Flight, i.e., are not satisfactorily monitored by the BBS (Dunn et al. 2005). To avoid potentially drastically underestimating the number of distant Sharp-shinned Hawks, we included all birds identified only as *Accipiter* in the sample. One set of simulations presumed no trend in observer skill over time, included three starting population sizes (N_1), and three values for $CV(N_k)$. Historical annual counts at Lucky Peak suggest an N_1 of Sharp-shinned Hawks of roughly 2000, with a $CV(N_k)$ of 0.26. Counts of Northern Harriers suggested an N_1 of roughly 450, with a $CV(N_k)$ of 0.38 (pers. obs.). The third N_1 of 100 was simulated to show what could be expected when the species is locally rare. The third $CV(N_k)$ of 0.18 approximates the minimum interannual variation in available birds to be expected for any raptor species at any watch-site (Fuller and Mosher 1981, Lewis and Gould 2000). To examine a possible scenario where trend estimates would be biased in the opposite direction of the true trend, thereby dramatically reducing power, we ran a second pair of simulations in which the mean observer effect trended from low (-0.59) in year 1 to high (0.52) in year 30. For these, we only used the realistic population parameters for Sharp-shinned Hawks ($N_1 = 2000$, $CV(N_k) = 0.26$), and Northern Harriers ($N_1 = 450$, $CV(N_k) = 0.38$).

RESULTS

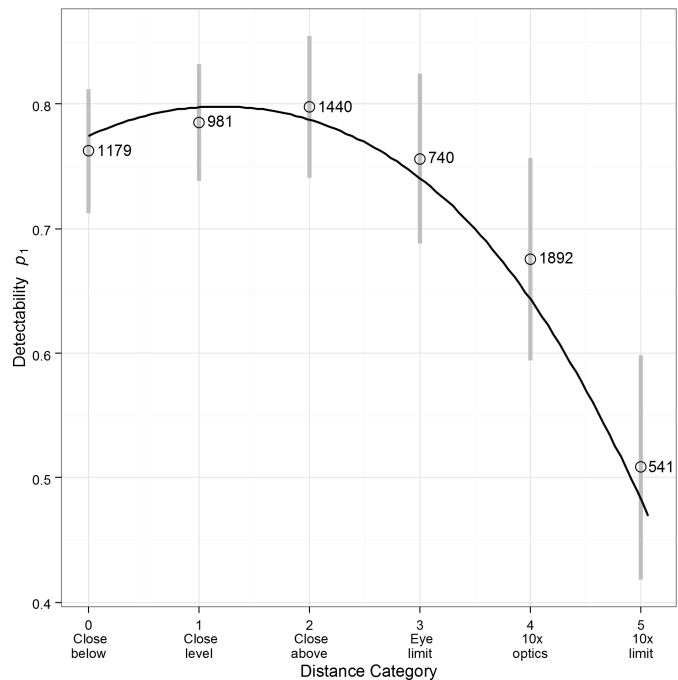
Double-observer trials

Observers detected 6773 raptors in 390 hours on 65 days. Primary observers detected 77% of raptors observed, and secondary observers made 23% of detections (effective sample size = 1571). Many variables differed significantly between years (Appendix 2). Comparison of AIC_c between the year-effect model and models representing other hypotheses suggested that the other covariates had superior explanatory value, so we did not consider year in any additional model-selection to avoid co-linearity (Appendix 3).

The general model ($K = 23$) included variables for observer, distance, species, cloud cover, wind speed, and day. The most parsimonious model ($AIC_c = 7098.77$, selection weight = 0.26, $K = 22$) included all the terms of the general model except cloud cover (Table 2). Eight models of the 64 candidates were reasonably competitive in model selection ($\Delta AIC_c < 4$) and every one included all the variables for observer, distance, and species (Table 2). The rest of the models had $\Delta AIC_c > 20$ (Appendix 2).

Coefficients from the most parsimonious model suggest detection probabilities differed by observer team (Table 3). Detectability greatly decreased with distance beyond the range of unaided vision (Fig. 1). Species with longer wingspans were generally more detectable, but Ospreys (*Pandion haliaetus*) were unusually difficult to detect for their size (Fig. 2). Weather variables and day had relatively little effect on detectability independent of species, flight, and observers (Table 3), but in four of the top eight models (Table 2) wind speed was estimated to have a negative effect on detectability (Table 3).

Fig. 1. Effect of relative distance and altitude on detectability. For definitions of distance categories, see Table 1. Points are weighted mean detectability with bars of ± 1 SD, labeled with the number of individuals detected. The curve shows the model prediction for a hypothetical individual with average covariates. The effect of ordinal distance category was modeled as a quadratic function (Table 3).



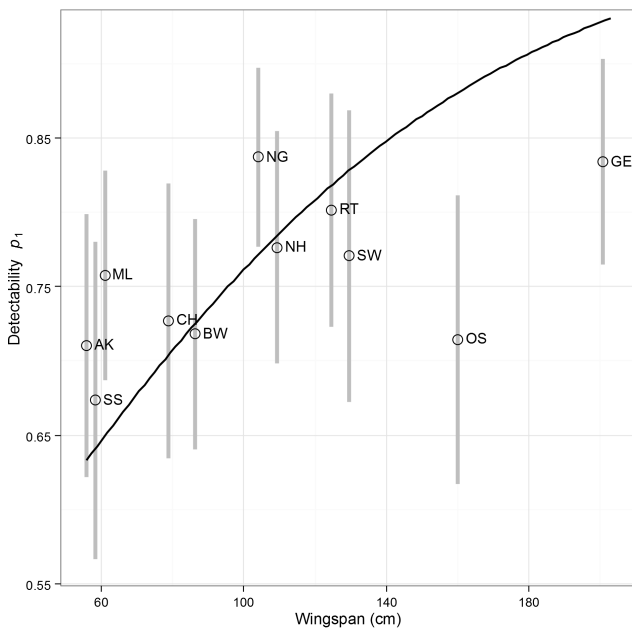
Estimated detectability of individual raptors observed ranged from 0.36 to 0.94 for the two primary observers and 0.59 to 0.99 for all four observers. Weighted mean detectability was 0.72 ± 0.11 with the two primary observers and 0.92 ± 0.07 with all four observers.

Simulation

Imperfect detection and heterogeneous detectability affected power mainly by increasing count variance. The bias in trend estimation introduced by imperfect detection was minimal in simulations with no predefined trend in detectability ($|\beta_C - \beta_N| \leq 2.6 \cdot 10^{-4}$). Bias was greater when observer ability was simulated to improve over time ($|\beta_C - \beta_N| \approx 0.01$ Fig. 3).

The relative effect of limited detectability on power is inversely related to variation in the number of raptors available $CV(N_k)$ (Figs. 4 and 5). The effect of heterogeneous detectability on power was minimal and accounted for ≤ 1 year difference in time to attain the 80% power benchmark with realistic parameters and no trend in observer skill. The effect of imperfect detectability on precision and power in both species was more negative when variation in availability was lower and population size was smaller. When the two species were compared with equal N_1 and $CV(N_k)$ values, the decline in power resulting from imperfect detectability was greater in Sharp-shinned Hawks, the smaller and less perceptible species.

Fig. 2. Estimated mean detectability of selected species, in order of increasing wingspan. Points are weighted means with bars of ± 1 SD. Species effects on detectability were modeled by additive effects of wingspan and genus (Table 3). The curve illustrates the modeled effect of varying wingspan, with genus as “Circus” and all other covariates fixed at median values. Species are: AK = American Kestrel (*Falco sparverius*), SS = Sharp-shinned Hawk (*Accipiter striatus*), ML = Merlin (*Falco columbarius*), CH = Cooper’s Hawk (*Accipiter cooperii*), BW = Broad-winged Hawk (*Buteo platypterus*), NG = Northern Goshawk (*Accipiter gentilis*), NH = Northern Harrier (*Circus cyaneus*), RT = Red-tailed Hawk (*Buteo jamaicensis*), SW = Swainson’s Hawk (*Buteo swainsoni*), OS = Osprey (*Pandion haliaetus*), and GE = Golden Eagle (*Aquila chrysaetos*).

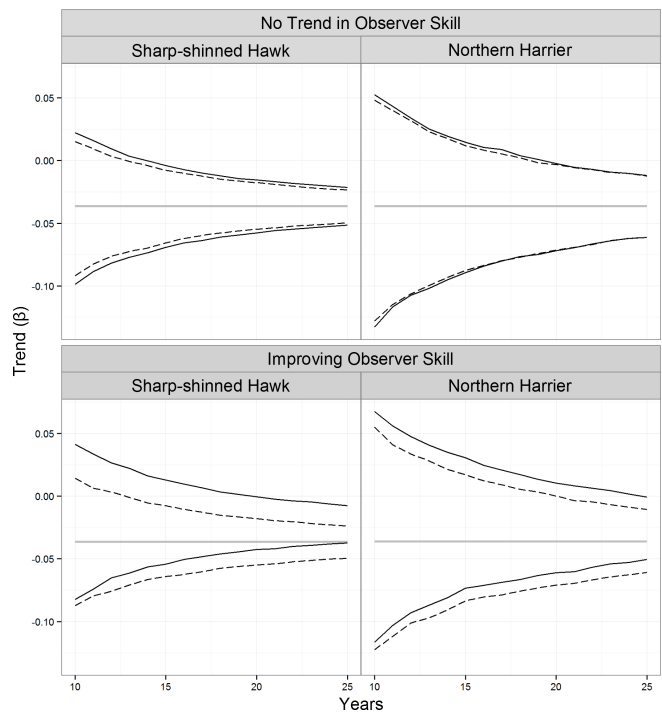


We estimate that Lucky Peak Hawk-Watch would require 19 years to achieve 80% power to detect a decline in Sharp-shinned Hawks with 90% confidence when the true trend is -3.5% annually. Twenty-five years of counts would be required in the case of Northern Harriers. Harriers’ greater probability of detection did not compensate much for the loss of precision from small population size and highly variable availability. When mean observer skill was simulated to improve over the 30 years, the average number of years necessary to achieve 80% power to detect a decline increased to 26 years for Sharp-shinned Hawks and exceeded 30 years for Northern Harriers. Moreover, the decline detected can be expected to be of lesser magnitude than the true trend (Fig. 3).

DISCUSSION

Detectability of migrant raptors at Lucky Peak depended on the observer team, the distance of the migratory flight, and species characteristics. It is important to note that as much as individual detectability varied, the mean detectability was considerably

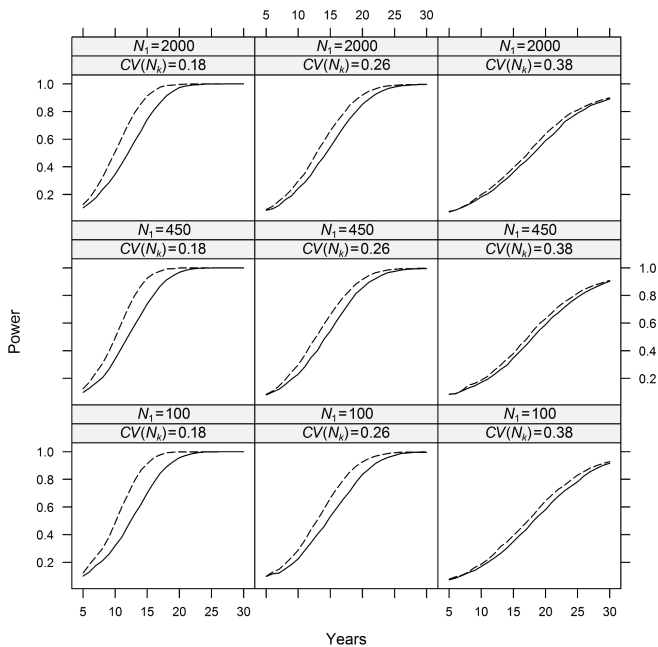
Fig. 3. Precision (90% CI) of trend parameter estimates with increasing study duration from 3000 simulation trials. Solid black lines are from the regressions of simulated counts with imperfect detectability. The dashed black lines are from the regressions of the available population. The gray line is the value for the true trend. When no trend in observer skill was simulated (the observer effect varied in each year but the mean did not change) there was little bias, and only a slight loss of precision from imperfect detection. When the mean observer skill was simulated to improve over time the trend in counts was biased high, relative to the population trend. As a result, increased precision with longer study duration ceases to improve the likelihood of detecting the true magnitude of population decline.



higher than in some other surveys of raptors (McLeod and Andersen 1998, Ayers and Anderson 1999). Interspecific differences in detectability will not bias species-specific estimates of trend for population monitoring. The sizable effects of observers and distance, however, may be detrimental to data quality and should be addressed at raptor watch-sites. Specifically, the number and ability of observers may change over time (Dunn et al. 2008), and changing weather may change flight distance over time (Berthiaume et al. 2009). Longer term trends in these effects would contribute to bias in trend estimates and greater loss of power (Kéry et al. 2009, Paprocki et al. 2014, Crewe et al. 2015).

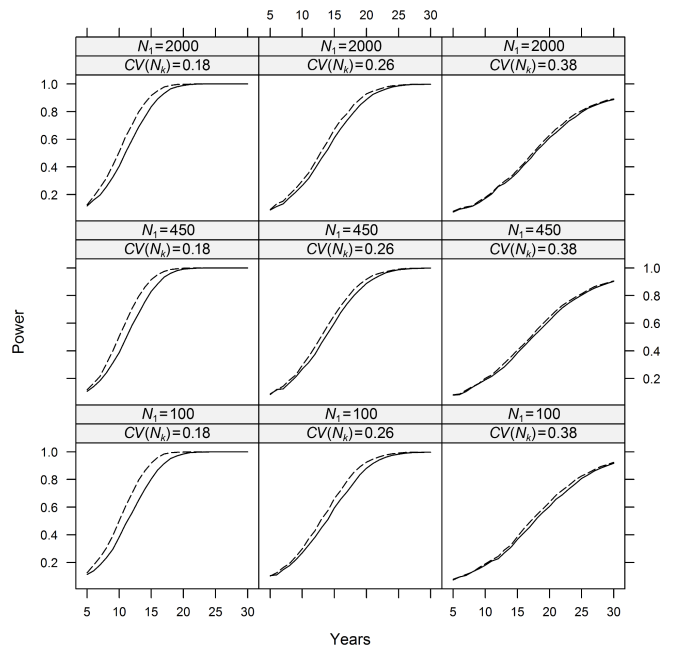
Apart from the observer effect, our results were consistent with the findings from Berthiaume et al. (2009) suggesting double-observer sampling is a robust technique for quantifying relative detectability at many raptor migration watch sites. In both studies detectability was greatest for raptors within the range of unaided

Fig. 4. Sharp-shinned Hawk (*Accipiter striatus*) trend analysis simulation results. Estimated statistical power ($\alpha = 0.1$, two-tailed test) to detect a significant declining trend ($\lambda = 0.964$) by the number of years of study duration. Dashed lines depict power in simulations with detectability = 1. Solid lines depict power in simulations with imperfect detectability. N_1 is the expected available population in the first year and $CV(N_k)$ is the square root of variance in the annual number of birds available as a proportion of N . Each simulated scenario was iterated 3000 times.



vision viewed against sky, lower for raptors viewed against the ground, and declined with increasing distance or altitude. Likewise, smaller species were considerably less detectable than larger species. Ospreys were an exception to this trend and were less detectable than smaller *Buteo* species and Northern Harriers. The low detectability of Ospreys was more pronounced in this study than in Berthiaume et al. (2009), but was consistent with the results of Sattler and Bart (1984). Ospreys at Lucky Peak in 2009 and 2010 were relatively uncommon (< 2% of raptors detected), and often flew along very different flight lines than the majority of migrants. Observers seeking to detect the greatest proportion of migrants may pay more attention to heavily-populated flight lines than regions in the field of view with few raptors, making uncommon raptors with atypical migration strategies less consistently detectable. This hypothesis closely resembles the one suggested by Kochenberger and Dunne (1985), to explain low counts of Peregrine Falcons (*Falco peregrinus*) on busy days at Cape May, New Jersey. If this is true, “specialty” watch-sites with concentrations of species that are rare at high-volume watch-sites may offer excellent monitoring value, e.g., the Florida Keys site for Peregrine Falcons (Lott 2006).

Fig. 5. Northern Harrier (*Circus cyaneus*) trend analysis simulation results. Estimated statistical power ($\alpha = 0.1$, two-tailed test) to detect a significant declining trend ($\lambda = 0.964$) by the number of years of study duration. Dashed lines depict power in simulations with detectability = 1. Solid lines depict power in simulations with heterogeneous detectability < 1. N_1 is the expected available population in the first year and $CV(N_k)$ is the square root of variance in the annual number of raptors available as a proportion of N . Each simulated scenario was iterated 3000 times.



Comparing the results of this study with previously published results (Sattler and Bart 1984, Berthiaume et al. 2009), it appears some factors may predict detectability better at some sites than others. Cloud cover was associated with greater detectability in all three studies, but the effect was of lesser predictive value at Lucky Peak than at Tadoussac (Berthiaume et al. 2009). This might be expected because Lucky Peak is a mountaintop site where raptors are often detected near or below the horizon, whereas Tadoussac is a shoreline site close to sea level, and birds may be detected at higher angles. Sattler and Bart (1984) observed that cloud cover improved visibility at Derby Hill, another low-elevation shoreline watch-site. At Derby Hill, flight density had a significant direct effect on detectability, whereas at Tadoussac and Lucky Peak flight density was of little value in predicting detectability (Sattler and Bart 1984, Berthiaume et al. 2009). This difference may be attributable to the relatively high peak flight densities observed at the Derby Hill watch-site (over 200 raptors in 30 minutes).

Our simulation results support the findings of prior power analyses. Lewis and Gould (2000) estimated the power of trend analysis for seven watch-sites and concluded that an interannual CV of 30% or less was necessary to have 80% power ($\alpha = 0.1$) to

detect a 50% population decline in 25 years, provided the mean number of birds counted per year was at least 20. At their seven watch-sites, among species counted in numbers > 20 per year, only 43% of species-by-site combinations had a CV that met this standard. In our simulations of a slightly faster decline (-50% in 20 years) we estimate that a $CV \leq 38\%$ is necessary to attain 80% power in 25 years (Figs. 4 and 5). Detectability-correction alone is unlikely to increase the number of species or watch-sites from which reliable trend estimates may be obtained because detectability had little effect on CV when $CV \geq 30\%$.

The results of these simulations provide insight into the conditions in which double-observer or distance-sampling detectability correction may be useful. The most important consideration is the potential for a trend in detectability over time, as we simulated with a trend in observer ability. Policies may be adopted to attempt to keep such trends to a minimum. Watch-site managers should consider adopting staffing policies that produce minimal changes in observer ability at longer time scales, i.e., month to month or year to year. We concur with Dunn et al.'s (2008) recommendation to use teams of two or more observers and rotate a pool of equivalently trained observers from day to day, instead of employing only one or two observers each year who may be exceptionally skilled or unexpectedly mediocre. Methodological drift may be avoided by producing written protocols, training new observers carefully, and periodically having more experienced observers check the work of those less experienced. Official observers can sometimes be isolated from visitors. Alternatively, a detectability-estimating survey method can be used.

In our simulations without a trend in detectability and low bias, detectability had a substantial effect on power when the number of available birds was consistent from year-to-year ($CV(N) < 25\%$), the species was uncommon at the watch-site (20 to a few hundred observed each year), and individuals of the species were relatively difficult to detect. Few combinations of species and watch-sites are likely to meet these qualifications. Raptor migration counts have rates of detection of 66% or higher (Berthiaume et al. 2009), utilize an index approach (Dunn and Hussell 1995), and are primarily useful for long-term monitoring (Fuller and Mosher 1981, 1987), making double-observer or distance-sampling detectability correction less potentially beneficial in this method.

Short-term detectability studies can yield valuable insights into cost-effective ways to improve count protocols. For example, our results show that there may be substantial differences in raptor-counting ability among individuals who are very comparable in terms of bird-watching experience, age, or visual acuity. For this reason, prior screening of observers may not be an effective safeguard against observer variation. Further research into the causes of variation in observer ability may suggest more effective screening criteria, or methods to adjust annual indices for observer ability.

The relative importance of factors affecting availability is in need of further research. Apart from survey effort, the proportion of the population available to count may be affected by changes in migration routes, distances, and timing, as well as rates of

fecundity and survival. Temporal data on the rate of passage of raptors at watch-sites are collected at an hourly scale at most watch-sites in North America, providing a rich source of information for availability compensation in trend analyses (Farmer et al. 2007, Farmer and Hussell 2008). Collecting similarly useful spatial datasets should be a priority.

This study and Berthiaume et al. (2009) both used simple visibility-based metrics to model effects of distance on individual raptors and found similar effects. This suggests that visibility-based distance and altitude codes, already in use at most watch-sites, may be useful covariates for adjusting counts to more accurately reflect the number of raptors present. However, at most sites, the code is recorded hourly, and represents a poorly defined central tendency for all the birds observed in that hour. The hourly measure provides no information on the distribution of distances, or how flight lines differ among species. A visibility-based distance (or altitude) code for each individual raptor or flock can be recorded with very little additional effort (E. G. Nolte, *personal observation*).

Present techniques for trend analysis of raptor migration counts avoid addressing the issues of detectability and availability by making no estimate of total migratory volume at any scale (Farmer and Hussell 2008, Crewe et al. 2013). This approach limits raptor migration counts to detecting long-term, continental trends in populations. Progress in stable hydrogen isotope analysis (Domenech et al. 2015, Hobson et al. 2015, Nelson et al. 2015) may make producing accurate trends at smaller spatial scales especially informative, by linking fluctuations in certain watch-site counts with climate, habitat, or prey base changes in specific regions. Simple innovations in data-collection and analysis to account for detectability and availability bias, such as recording individual flight distances, rotating observers, and empirically studying the relative efficiency of observers, may enable watch-site networks to engage in more hypothesis-based research addressing current conservation issues of interest (e.g., Dennhardt et al. 2015).

Responses to this article can be read online at:
<http://www.ace-eco.org/issues/responses.php/894>

Acknowledgments:

We thank all the workers at IBO who took part in data-collection. The late Dr. Alfred Dufty Jr. provided guidance in the design of this study and preparation of the manuscript. Dr. Erica Dunn and Dr. Matthew Etterson provided helpful critiques. Funding and support for this study was provided by the Hawk Migration Association of North America and the Department of Biological Sciences, Raptor Research Center, and Intermountain Bird Observatory at Boise State University. Computing resources for this material is based in part upon work supported by the National Science Foundation under Grant No. 1229709.

LITERATURE CITED

- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007. Factors affecting aural detections of songbirds. *Ecological Applications* 17:948-955. <http://dx.doi.org/10.1890/06-0685>
- Ayers, L. W., and S. H. Anderson. 1999. An aerial sightability model for estimating Ferruginous Hawk population size. *Journal of Wildlife Management* 63:85-97. <http://dx.doi.org/10.2307/3802489>
- Berthiaume, É., M. Bélisle, and J.-P. Savard. 2009. Incorporating detectability into analyses of population trends based on hawk counts: a double-observer approach. *Condor* 111:43-58. <http://dx.doi.org/10.1525/cond.2009.080081>
- Bildstein, K. L. 2006. *Migrating raptors of the world: their ecology and conservation*. Comstock, Ithaca, New York, USA.
- Bildstein, K. L., J. P. Smith, E. Ruelas Inzunza, and R. R. Veit. 2008. *State of North America's birds of prey*. Nuttall Ornithological Club and American Ornithologists' Union, Washington, D.C., USA.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Second edition. Springer, New York, New York, USA. <http://dx.doi.org/10.1007/b97636>
- Campbell, M., and C. M. Francis. 2011. Using stereo-microphones to evaluate observer-variation in North American Breeding Bird Survey point counts. *Auk* 128:303-312. <http://dx.doi.org/10.1525/auk.2011.10005>
- Crewe, T. L., P. D. Taylor, and D. Lepage. 2015. Modeling systematic change in stopover duration does not improve bias in trends estimated from migration counts. *PLoS ONE* 10:e0130137. <http://dx.doi.org/10.1371/journal.pone.0130137>
- Crewe, T. L., P. D. Taylor, D. Lepage, L. Goodrich, J. Brown, and J. Sodergren. 2013. *2013 Analysis methods and trend results*. The Raptor Population Index. [online] URL: <http://rpi-project.org/2013/>
- Cunningham, R. B., D. B. Lindenmayer, H. A. Nix, and B. D. Lindenmayer. 1999. Quantifying observer heterogeneity in bird counts. *Australian Journal of Ecology* 24:270-277. <http://dx.doi.org/10.1046/j.1442-9993.1999.00971.x>
- Dennhardt, A. J., A. E. Duerr, D. Brandes, and T. E. Katzner. 2015. Integrating citizen-science data with movement models to estimate the size of a migratory Golden Eagle population. *Biological Conservation* 184:68-78. <http://dx.doi.org/10.1016/j.biocon.2015.01.003>
- Domenech, R., T. Pitz, K. Gray, and M. Smith. 2015. Estimating natal origins of migratory juvenile Golden Eagles using stable hydrogen isotopes. *Journal of Raptor Research* 49:308-315. <http://dx.doi.org/10.3356/JRR-14-19.1>
- Dunn, E. H. 2002. Using decline in bird populations to identify needs for conservation action. *Conservation Biology* 16:1632-1637. <http://dx.doi.org/10.1046/j.1523-1739.2002.01250.x>
- Dunn, E. H. 2005. *Counting migrants to monitor bird populations: state of the art*. Pages 712-717 in C. J. Ralph and T. D. Rich, editors. *Bird conservation implementation and integration in the Americas: Proceedings of the Third International Partners in Flight Conference*. 20-24 March 2002, Asilomar, California, Volume 2. Gen. Tech. Rep. PSW-GTR-191. U.S. Forest Service, Pacific Southwest Research Station, Albany, California, USA.
- Dunn, E. H., B. L. Altman, J. Bart, C. J. Beardmore, H. Berlanga, P. J. Blancher, G. S. Butcher, D. W. Demarest, R. Dettmers, W. C. Hunter, E. E. Iñigo-Elias, A. O. Panjabi, D. N. Pashley, C. J. Ralph, T. D. Rich, K. V. Rosenberg, C. M. Rustay, J. M. Ruth, and T. C. Will. 2005. *High priority needs for range-wide monitoring of North American landbirds*. Technical Series No. 2. Partners in Flight, Laurel, Maryland, USA. [online] URL: <http://www.pwrc.usgs.gov/pif/pubs/ts/02-monitoringneeds.pdf>
- Dunn, E. H., and D. J. T. Hussell. 1995. Using migration counts to monitor landbird populations: review and evaluation of current status. *Current Ornithology* 12:43-88. http://dx.doi.org/10.1007/978-1-4615-1835-8_2
- Dunn, E. H., D. J. T. Hussell, and E. Ruelas Inzunza. 2008. Recommended methods for population monitoring at raptor-migration watchsites. Pages 447-459 in K. L. Bildstein, J. P. Smith, E. Ruelas Inzunza, and R. R. Veit, editors. *State of North America's birds of prey*. Nuttall Ornithological Club and American Ornithologists' Union, Washington, D.C., USA.
- Farmer, C. J., and D. J. T. Hussell. 2008. The raptor population index in practice. Pages 165-178 in K. L. Bildstein, J. P. Smith, E. Ruelas Inzunza, and R. R. Veit, editors. *State of North America's birds of prey*. Nuttall Ornithological Club and American Ornithologists' Union, Washington, D.C., USA.
- Farmer, C. J., D. J. T. Hussell, and D. Mizrahi. 2007. Detecting population trends in migratory birds of prey. *Auk* 124:1047-1062. [http://dx.doi.org/10.1642/0004-8038\(2007\)124\[1047:DPTIMB\]2.0.CO;2](http://dx.doi.org/10.1642/0004-8038(2007)124[1047:DPTIMB]2.0.CO;2)
- Finch, D. M., and T. E. Martin. 1995. *Ecology and management of neotropical migratory birds: a synthesis and review of critical issues*. Oxford University Press, New York, New York, USA.
- Fuller, M. R., and J. A. Mosher. 1981. Methods of detecting and counting raptors: a review. *Studies in Avian Biology* 6:235-246.
- Fuller, M. R., and J. A. Mosher. 1987. Raptor survey techniques. Pages 37-65 in B. A. Giron Pendleton, B. A. Millsap, K. W. Kline, and D. M. Bird, editors. *Raptor management techniques manual*. National Wildlife Federation, Washington, D.C., USA.
- Hawk Migration Association of North America (HMANA). 2009. *Data submission*. HMANA, Plymouth, New Hampshire, USA. [online] URL: <http://www.hmana.org/data-submission>
- Hobson, K. A., S. L. Van Wilgenburg, E. H. Dunn, D. J. T. Hussell, P. D. Taylor, and D. M. Collister. 2015. Predicting origins of passerines migrating through Canadian migration monitoring stations using stable-hydrogen isotope analyses of feathers: a new tool for bird conservation. *Avian Conservation and Ecology* 10 (1):3. <http://dx.doi.org/10.5751/ace-00719-100103>
- Horvitz, D. G., and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47:663-685. <http://dx.doi.org/10.1080/01621459.1952.10483446>

- Huggins, R. M. 1989. On the statistical-analysis of capture experiments. *Biometrika* 76:133-140. <http://dx.doi.org/10.1093/biomet/76.1.133>
- Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* 47:725-732. <http://dx.doi.org/10.2307/2532158>
- Johnson, D. H. 2008. In defense of indices: the case of bird surveys. *Journal of Wildlife Management* 72:857-868. <http://dx.doi.org/10.2193/2007-294>
- Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. *Auk* 113:823-829. <http://dx.doi.org/10.2307/4088860>
- Kerlinger, P. 1989. *Flight strategies of migrating hawks*. University of Chicago Press, Chicago, Illinois, USA.
- Kéry, M., R. M., Dorazio, L. Soldaat, A. Van Strien, A. Zuiderwijk, and J. A. Royle. 2009. Trend estimation in populations with imperfect detection. *Journal of Applied Ecology* 46:1163-1172. <http://dx.doi.org/10.1111/j.1365-2664.2009.01724.x>
- Kochenberger, R., and P. J. Dunne. 1985. The effect of varying observer numbers on raptor count totals at Cape May, New Jersey. Pages 281-293 in M. Harwood, editor. *Hawk Migration Conference IV*. Hawk Migration Association of North America, Rochester, New York, USA.
- Lewis, S. A., and W. R. Gould. 2000. Survey effort effects on power to detect trends in raptor migration counts. *Wildlife Society Bulletin* 28:317-329.
- Lott, C. A. 2006. A new raptor migration monitoring site in the Florida Keys: counts from 1999-2004. *Journal of Raptor Research* 40:200-209. [http://dx.doi.org/10.3356/0892-1016\(2006\)40\[200:ANRMMS\]2.0.CO;2](http://dx.doi.org/10.3356/0892-1016(2006)40[200:ANRMMS]2.0.CO;2)
- McLeod, M. A., and D. E. Andersen. 1998. Red-shouldered Hawk broadcast surveys: factors affecting detection of responses and population trends. *Journal of Wildlife Management* 62:1385-1397. <http://dx.doi.org/10.2307/3802005>
- Nelson, D. M., M. Braham, T. A. Miller, A. E. Duerr, J. Cooper, M. Lanzone, J. Lemaître, and T. Katzner. 2015. Stable hydrogen isotopes identify leapfrog migration, degree of connectivity, and summer distribution of Golden Eagles in eastern North America. *Condor* 117:414-429. <http://dx.doi.org/10.1650/CONDOR-14-209.1>
- Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* 117:393-408. [http://dx.doi.org/10.1642/0004-8038\(2000\)117\[0393:ADOAFE\]2.0.CO;2](http://dx.doi.org/10.1642/0004-8038(2000)117[0393:ADOAFE]2.0.CO;2)
- Nichols, J. D., L. Thomas, and P. B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201-235 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling demographic processes in marked populations*. Springer, New York, New York, USA. http://dx.doi.org/10.1007/978-0-387-78151-8_9
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62:3-135.
- Paprocki, N., J. A. Heath, and S. J. Novak. 2014. Regional distribution shifts help explain local changes in wintering raptor abundance: implications for interpreting population trends. *PloS ONE* 9:e86814. <http://dx.doi.org/10.1371/journal.pone.0086814>
- Ruelas Inzunza, E. 2008. Raptor-migration watchsite descriptions. Pages 279-302 in K. L. Bildstein, J. P. Smith, E. Ruelas Inzunza, and R. R. Veit, editors. *State of North America's birds of prey*. Nuttall Ornithological Club and American Ornithologists' Union, Washington, D.C., USA.
- Sattler, G., and J. Bart. 1984. Reliability of counts of migrating raptors: an experimental analysis. *Journal of Field Ornithology* 55:415-423.
- Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* 111:50-62. <http://dx.doi.org/10.2307/4088504>
- Sibley, D. 2000. *The Sibley guide to birds*. First edition. Alfred A. Knopf, New York, New York, USA.
- Smith, J. P., C. J. Farmer, S. F. Hoffman, G. S. Kaltenecker, K. Z. Woodruff, and P. F. Sherrington. 2008. Trends in autumn counts of migratory raptors in western North America. Pages 217-252 in K. L. Bildstein, J. P. Smith, E. Ruelas Inzunza, and R. R. Veit, editors. *State of North America's birds of prey*. Nuttall Ornithological Club and American Ornithologists' Union, Washington, D.C., USA.
- Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. *Auk* 119:18-25. [http://dx.doi.org/10.1642/0004-8038\(2002\)119\[0018:TRBSAF\]2.0.CO;2](http://dx.doi.org/10.1642/0004-8038(2002)119[0018:TRBSAF]2.0.CO;2)
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120-139. <http://dx.doi.org/10.1080/00063659909477239>
- Zalles, J. I., and K. L. Bildstein. 2000. *Raptor watch: a global directory of raptor migration sites*. Birdlife International, Cambridge, UK.



Appendix 1. Descriptive statistics for all covariates by year.

Please click [here](#) to download file 'appendix1.r'.

Appendix 2 Descriptive statistics for all covariates by year. Continuous and ordinal variables are presented as mean (SD), and compared with Welch t-tests. Dummy (Boolean) variables are presented as ratios and are compared with Fisher's Exact tests.

Variable	Min.	Max.	2009 (<i>n</i> = 4150)	2010 (<i>n</i> = 2623)	<i>P</i> -value
Observers B&C	0	1	0.22		
Observers F&C	0	1	0.19		
Observers B&F	0	1	0.16		
Observers D&B	0	1	0.14		
Observers D&F	0	1	0.13		
Observers D&C	0	1	0.11		
Observers E&A	0	1		0.21	
Observers D&E	0	1		0.12	
Observers D&G	0	1		0.11	
Observers E&G	0	1		0.10	
Observers (Other)	0	1	0.05	0.47	<0.001
Genus <i>Accipiter</i>	0	1	0.43	0.50	<0.001
Genus <i>Buteo</i>	0	1	0.25	0.25	0.75
Genus <i>Falco</i>	0	1	0.25	0.18	<0.001
Genus <i>Circus</i>	0	1	0.05	0.05	0.73
Genus <i>Pandion</i>	0	1	0.02	0.02	0.78
Genus (Eagles)	0	1	0.004	0.006	0.27
Wingspan (cm)	56	203	83.0 (31.0)	83.4 (30.8)	0.55
Wind-speed (kph)	0	38.5	11.4 (5.8)	8.3 (5.0)	<0.001
cos(Wind dir.)	-1	1	0.01 (0.61)	-0.18 (0.57)	<0.001
Temperature (°C)	-3.6	39.1	18.7 (8.4)	21.4 (5.6)	<0.001
Cloud	0	3	0.6 (0.9)	0.9 (1.1)	<0.001
Alt.	0	5	2.5 (1.6)	2.2 (1.7)	<0.001
BPH	0.5	216	42.1 (30.8)	35.9 (23.7)	<0.001
Day	3	67	27.9 (14.6)	33.1 (12.1)	<0.001
Hour	10	19	14.2 (2.0)	14.0 (1.9)	<0.001

Appendix 3 All models grouped by stage of selection, then ordered by increasing AIC_c. Lowest AIC_c was 7098.769. *K* is the number of model parameters. For relative model likelihoods and AIC_c weights, see Table 2 (All but the top eight models have zero relative likelihood).

Stage	Model	<i>K</i>	Deviance	ΔAIC _c
Final	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	22	7054.12	0
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Day} + \text{Day}^2\}$	21	7056.56	0.39
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	23	7053.19	1.13
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	22	7056.03	1.92
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus}\}$	19	7062.56	2.28
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed}\}$	20	7060.56	2.34
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed}\}$	21	7058.81	2.63
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud}\}$	20	7061.33	3.10
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	16	7088.30	21.88
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Day} + \text{Day}^2\}$	15	7091.24	22.78
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	17	7087.60	23.23
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Day} + \text{Day}^2\}$	16	7090.89	24.47
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Windspeed}\}$	15	7103.98	35.52
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Cloud}\}$	14	7107.06	36.56
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2 + \text{Windspeed}\}$	14	7107.41	36.91
	$p\{\text{Team} + \text{Alt} + \text{Alt}^2\}$	13	7109.74	37.20
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Day} + \text{Day}^2\}$	19	7146.29	86.01
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	20	7144.24	86.01
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Cloud}\}$	18	7148.86	86.53
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	21	7143.08	86.91
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	20	7145.59	87.36
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed}\}$	19	7147.94	87.66
	$p\{\text{Team} + \text{Wingspan} + \text{Genus}\}$	17	7152.19	87.82
	$p\{\text{Team} + \text{Wingspan} + \text{Genus} + \text{Windspeed}\}$	18	7151.75	89.42

Stage	Model	K	Deviance	ΔAIC_c
Final, cont.	$p\{\text{Team} + \text{Day} + \text{Day}^2\}$	13	7172.98	100.44
	$p\{\text{Team} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	14	7171.43	100.93
	$p\{\text{Team} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	14	7172.46	101.96
	$p\{\text{Team} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	15	7170.58	102.12
	$p\{\text{Team} + \text{Cloud}\}$	12	7184.19	109.63
	$p\{\text{Team} + \text{Cloud} + \text{Windspeed}\}$	13	7183.45	110.91
	$p\{\text{Team}\}$	11	7188.99	112.39
	$p\{\text{Team} + \text{Windspeed}\}$	12	7188.70	114.13
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	13	7211.95	139.42
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	12	7214.28	139.72
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed}\}$	11	7217.58	140.98
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Windspeed}\}$	10	7221.03	142.41
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Day} + \text{Day}^2\}$	11	7219.59	142.99
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	12	7218.17	143.60
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus} + \text{Cloud}\}$	10	7223.87	145.24
	$p\{\text{Alt} + \text{Alt}^2 + \text{Wingspan} + \text{Genus}\}$	9	7226.21	145.56
	$p\{\text{Alt} + \text{Alt}^2 + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	6	7246.57	159.85
	$p\{\text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	7	7244.63	159.93
	$p\{\text{Alt} + \text{Alt}^2 + \text{Day} + \text{Day}^2\}$	5	7252.26	163.53
	$p\{\text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Day} + \text{Day}^2\}$	6	7251.11	164.40
	$p\{\text{Alt} + \text{Alt}^2 + \text{Cloud} + \text{Windspeed}\}$	5	7260.70	171.97
	$p\{\text{Alt} + \text{Alt}^2 + \text{Windspeed}\}$	4	7265.84	175.10
	$p\{\text{Alt} + \text{Alt}^2 + \text{Cloud}\}$	4	7267.68	176.94
	$p\{\text{Alt} + \text{Alt}^2\}$	3	7271.52	178.77
	$p\{\text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	11	7288.52	211.92
	$p\{\text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Windspeed}\}$	9	7293.73	213.07
	$p\{\text{Wingspan} + \text{Genus} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	10	7293.31	214.69
	$p\{\text{Wingspan} + \text{Genus} + \text{Cloud} + \text{Day} + \text{Day}^2\}$	10	7293.52	214.89

Stage	Model	K	Deviance	ΔAIC_c
Final, cont.	$p\{\text{Wingspan} + \text{Genus} + \text{Cloud}\}$	8	7298.58	215.91
	$p\{\text{Wingspan} + \text{Genus} + \text{Day} + \text{Day}^2\}$	9	7297.08	216.42
	$p\{\text{Wingspan} + \text{Genus} + \text{Windspeed}\}$	8	7299.70	217.03
	$p\{\text{Wingspan} + \text{Genus}\}$	7	7303.22	218.52
	$p\{\text{Cloud} + \text{Windspeed} + \text{Day} + \text{Day}^2\}$	5	7314.28	225.55
	$p\{\text{Windspeed} + \text{Day} + \text{Day}^2\}$	4	7318.12	227.38
	$p\{\text{Cloud} + \text{Day} + \text{Day}^2\}$	4	7318.42	227.68
	$p\{\text{Day} + \text{Day}^2\}$	3	7321.31	228.55
	$p\{\text{Cloud} + \text{Windspeed}\}$	3	7326.62	233.87
	$p\{\text{Cloud}\}$	2	7330.94	236.18
Initial: Observer	$p\{\text{Windspeed}\}$	2	7333.76	239.00
	$p\{\text{Team} + \text{Day} + \text{Day}^2 + \text{Hour}\}$	14	7170.55	99.82
	$p\{\text{Team} + \text{Day} + \text{Day}^2\}$	13	7172.98	100.24
	$p\{\text{Team} + \text{Day} + \text{Hour}\}$	13	7173.99	101.25
	$p\{\text{Team} + \text{Day}\}$	12	7176.34	101.60
	$p\{\text{Team}\}$	11	7188.99	112.24
	$p\{\text{Team} + \text{Hour}\}$	12	7188.26	113.51
	$p\{\text{Day} + \text{Day}^2 + \text{Hour}\}$	4	7317.80	227.03
	$p\{\text{Day} + \text{Hour}\}$	3	7320.58	227.81
	$p\{\text{Day} + \text{Day}^2\}$	3	7321.31	228.54
	$p\{\text{Year}\}$	2	7323.95	229.19
	$p\{\text{Day}\}$	2	7323.97	229.20
	$p\{.\}$	1	7336.82	240.05
	$p\{\text{Hour}\}$	2	7335.50	240.74
Initial: Flight	$p\{\text{Alt} + \text{Alt}^2 + \text{BPH}\}$	4	7269.00	178.24
	$p\{\text{Alt} + \text{Alt}^2\}$	3	7271.52	178.75
	$p\{\text{Alt} + \text{BPH}\}$	3	7291.06	198.29
	$p\{\text{Alt}\}$	2	7293.64	198.87

Stage	Model	K	Deviance	ΔAIC_c
Initial: Flight (cont.)	$p\{\text{Year}\}$	2	7323.95	229.19
	$p\{BPH\}$	2	7333.16	238.39
Initial: Species	$p\{\text{Wingspan} + \text{Genus}\}$	7	7303.22	218.46
	$p\{\text{Wingspan} + \text{Wingspan}^2\}$	3	7311.71	218.94
	$p\{\text{Wingspan}\}$	2	7314.54	219.77
	$p\{\text{Wingspan} + \text{Wingspan}^2 + \text{Genus}\}$	8	7302.91	220.15
	$p\{\text{Genus}\}$	6	7308.41	221.65
Initial: Weather	$p\{\text{Year}\}$	2	7323.95	229.19
	$p\{\text{Cloud} + \text{Windspeed}\}$	3	7326.62	233.85
	$p\{\text{Cloud} + \text{Windspeed} + \cos(\text{Wind dir.})\}$	4	7325.68	234.91
	$p\{\text{Cloud} + \text{Windspeed} + \text{Temperature}\}$	4	7326.32	235.55
	$p\{\text{Cloud} + \text{Windspeed} + \cos(\text{Wind dir.}) + \text{Temperature}\}$	5	7324.65	235.88
	$p\{\text{Cloud} + [\text{Windspeed} \times \cos(\text{Wind dir.})]\}$	3	7328.85	236.08
	$p\{\text{Cloud}\}$	2	7330.94	236.18
	$p\{\text{Cloud} + \cos(\text{Wind dir.})\}$	3	7329.17	236.41
	$p\{\text{Cloud} + [\text{Windspeed} \times \cos(\text{Wind dir.})] + \text{Temperature}\}$	4	7328.41	237.65
	$p\{\text{Cloud} + \cos(\text{Wind dir.}) + \text{Temperature}\}$	4	7328.77	238.00
	$p\{\text{Cloud} + \text{Temperature}\}$	3	7330.94	238.18
	$p\{\text{Windspeed}\}$	2	7333.76	239.00
	$p\{\text{Windspeed} + \text{Temperature}\}$	3	7332.54	239.77
	$p\{\text{Windspeed} + \cos(\text{Wind dir.}) + \text{Temperature}\}$	4	7330.78	240.01
	$p\{.\}$	1	7336.82	240.05
	$p\{\text{Windspeed} + \cos(\text{Wind dir.})\}$	3	7333.17	240.41
	$p\{[\text{Windspeed} \times \cos(\text{Wind dir.})]\}$	2	7335.53	240.76
	$p\{\cos(\text{Wind dir.})\}$	2	7335.64	240.87
	$p\{[\text{Windspeed} \times \cos(\text{Wind dir.})] + \text{Temperature}\}$	3	7334.11	241.34
	$p\{\cos(\text{Wind dir.}) + \text{Temperature}\}$	3	7334.22	241.46
$p\{\text{Temperature}\}$	2	7336.45	241.68	