

RAPTORS PRESENT BUT UNOBSERVED:
DETECTABILITY AT A WESTERN MIGRATION WATCH-SITE
AND ITS EFFECT ON TREND ANALYSIS

By

Eric G. Nolte

A thesis

submitted in partial fulfillment
of the requirements for the degree of
Master of Sciences in Raptor Biology
Boise State University

May 2012

© 2012

Eric G. Nolte

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Eric G. Nolte

Thesis Title: Raptors Present but Unobserved: Detectability at a Western Migration Watch-Site and Its Effect on Trend Analysis

Date of Final Oral Examination: 15 February 2012

The following individuals read and discussed the thesis submitted by student Eric G. Nolte, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Julie A. Heath, Ph.D. Chair, Supervisory Committee

Jonathan Bart, Ph.D. Member, Supervisory Committee

Alfred M. Dufty Jr., Ph.D. Member, Supervisory Committee

The final reading approval of the thesis was granted by Julie A. Heath, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

DEDICATION

To Brianne

ACKNOWLEDGEMENTS

My committee has been unfailingly helpful in the course of the design, implementation, analysis, and presentation of my research. Dr. Julie Heath, my graduate advisor and committee chair, has assisted with every stage, and has been my most valuable ally, without fail, for four years. I feel truly honored to have been granted this opportunity. The line of research presented in Chapter 1 was her idea in the beginning, but she graciously surrendered its design entirely to my judgment. Her resources provided me with a field vehicle, computer, and partial funding for field research. Dr. Jonathan Bart, of the US Geological Survey's Forest and Rangeland Ecosystem Science Center, has provided me with countless invaluable insights into the nuances of sampling theory and the use of simulation. His brief comments and scratch-paper notes are directly responsible for many aspects of the experiment in Chapter 2. Dr. Alfred Dufty Jr. provided valuable input that helped focus the scope of the research, and was a most helpful reviewer.

This thesis was made possible by the cooperation of Gregory Kaltenecker, Director of the Idaho Bird Observatory, who permitted the temporary addition of extra observers at the Lucky Peak Hawk-Watch. I sincerely hope this thesis is only the best sort of publicity for IBO.

A grant for field work and free publicity in *Hawk Migration Studies* was provided by the Hawk Migration Association of North America. Additional funding for field research was provided by an award from the Boise State University Raptor Research

Center. Many thanks to all the workers at IBO in 2009 and 2010, particularly hawk-watchers Jon Kauffman, Preston Alden, David Rankin, Sara Cendejas-Zarelli, Ian Dolly, Jesús Gómez Esteban, Garrett MacDonald, and David Kramer.

The simulations in Chapter 2 were made possible by a generous donation of computer processor-hours by Fulcrum Inquiry, Los Angeles, California. Thank you, Dan and David Nolte. My longtime friend J. Bradley Altenau also provided assistance with computation.

A teaching assistantship provided by the Department of Biological Sciences at Boise State University provided for living expenses while I completed this thesis. I also have been helped immensely with developing and practicing the presentation of my research by fellow graduate students, especially my fellow denizens of Heath lab: Erin H. Strasser, Alyson Webber, Neil Paprocki, Alexandra Anderson, and Dana Owen, and writing circle devotees Amy Ulappa, Jamie Utz, Alex Urquhart, and John O'Keefe among many others, too numerous to mention by name. Last, but certainly not least, the support of my family has been vital to my success; not only my beloved parents in California, but also my wife Brianne and her family in Idaho, the Boesigers and O'Learys.

ABSTRACT

Annual counts of migrating raptors (*Accipitriformes*, *Falconiformes*) are used as indices of population size. Variation in the proportion of the raptor population counted may decrease precision of trend estimates, thereby reducing power of inference. The proportion counted is the product of sample coverage and probability of detection. It is possible to improve the power of trend analysis by the adoption of techniques, such as double-observer or distance sampling, which estimate the probability of detection. I used a dependent double-observer method to estimate detectability at the annual fall raptor migration count at Lucky Peak, Idaho, in 2009 and 2010. I used Huggins closed-capture removal models and information-theoretic multi-model inference to describe important factors affecting detectability. The most parsimonious model included effects of observer identity, distance, wingspan, genus, and day of the season. Competitive models also included wind-speed, cloud cover, and hour of the day. These results demonstrate the importance of controlling observer effort and training at watch-sites, and the potential utility of adjusting daily counts to account for differences in flight distance. I used model-averaging to account for selection-uncertainty in estimating coefficients, and used the resulting equation to simulate 30 years of counts of Sharp-shinned Hawks (*Accipiter striatus*) and Northern Harriers (*Circus cyaneus*) with heterogeneous detectability, a known population trend, and a degree of unexplained random variation in the number of available birds. Imperfect detection did not substantially bias trend estimation, but did increase variance in counts, decreasing power. Correcting for detectability did little to

improve power to detect long-term declines when there was a realistically high variation in the number of available raptors ($CV \geq 0.26$). Detectability-correction by means of double-observer or distance sampling may, in the case of raptor migration counts, not be warranted for the purpose of long-term population monitoring. Efforts may be better focused on improving our understanding of mechanisms that cause changes in the number of migrants available to count.

TABLE OF CONTENTS

DEDICATION	iv	
ACKNOWLEDGEMENTS.....	v	
ABSTRACT	vii	
LIST OF TABLES	xi	
LIST OF FIGURES	xii	
INTRODUCTION.....	1	
Literature Cited	5	
CHAPTER 1: DETECTABILITY OF MIGRATING RAPTORS AT A WESTERN		
RIDGELINE WATCH-SITE		8
Abstract	8	
Introduction.....	8	
Methods.....	11	
Study Site.....	11	
Experiment.....	12	
Statistical Analyses	14	
Results	18	
Discussion	19	
Literature Cited	23	
CHAPTER 2: DOES IMPERFECT DETECTION OF MIGRATING RAPTORS		
AFFECT THE POWER OF POPULATION TREND ANALYSES?.....		33

Abstract	33
Introduction.....	34
Methods	37
Results	40
Discussion	41
Management Implications	43
Literature Cited	44
CONCLUSION	51
Literature Cited	56
APPENDIX	59
Detectability Models Used in Simulations.....	60

LIST OF TABLES

Table 1a	Ordinal Scale Used in Estimating Effects of Distance and Altitude on Detectability	27
Table 1b	Ordinal Scale Used in Estimating the Effect of Cloud Cover on Detectability.....	27
Table 2	Covariates Used in Models of Individual Heterogeneity in Detectability.....	28
Table 3	Model Comparison From a Set of 406 Candidate Models Estimating the Detectability of Migrating Raptors in Double-Observer Counts Conducted at Lucky Peak in 2009 and 2010.....	29
Table 4	Model-averaged Estimates of Coefficients with Standard Errors and Odds Ratios.....	30
Table 5	Estimating the Coefficient of Variation of Annual Numbers of Available Raptors at Lucky Peak	47
Table 6	Number of Years of Counts Necessary to Achieve 80% Power to Detect a -3.5% Annual Population Trend.....	48

LIST OF FIGURES

Figure 1	Effect of Relative Distance and Altitude on Detectability	31
Figure 2	Estimated Mean Detectability of Selected Species	32
Figure 3	Sharp-Shinned Hawk Trend Analysis Simulation Results	49
Figure 4	Northern Harrier Trend Analysis Simulation Results	50

INTRODUCTION

What does it mean to monitor a population? Ideally, we want to be able to estimate the number of individuals at a point in time, or estimate demographic rates such as fecundity or survival, so that we may predict the population size at some point in the future. Representative samples are necessary to guarantee unbiased population estimates, but determining whether a sample is representative requires knowledge of the full extent of the population in space and time. This is difficult for birds and other highly mobile species. In the interest of reducing bias in estimation, wildlife biologists are strongly encouraged to consider probability of detection (Nichols et al. 2000, Buckland et al. 2001, Thompson 2002, Alldredge et al. 2006, 2007a, 2007b). In traditional survey design and analysis of monitoring data, detectability is assumed to be perfect ($= 1$), or at least perfectly consistent. If such methods are applied when detectability is highly variable, estimates may be biased, even when the sampling design is sound (Thompson 2002).

Alternatively, we may decide to only estimate the population trend. To do this, a sample is treated as an index of abundance, an abstract number that changes proportionally to real change in the population (Johnson 2008). This approach relaxes the requirement of a representative sample. Nonetheless, change in detectability over time may violate the assumption of proportionality (Thompson 2002).

Many continental-scale, multi-species monitoring efforts use an index approach. The North American Breeding Bird Survey (BBS) and the Audubon Christmas Bird Counts both attempt to monitor long-term trends in landbirds in the United States and Canada, and both have persisted for over thirty years thanks to an effective utilization of

a corps of skilled volunteers. The BBS is often used in setting management priorities (e.g., Dunn 2002, Dunn et al. 2005), thanks to its more systematic survey design and relatively sophisticated analyses, which have been designed to account for some predictable sources of change in detectability (Sauer et al. 1994, Link and Sauer 1998). The BBS does, however, have some limitations. To make effective use of a volunteer effort, the BBS is confined to latitudes with an extensive road network, leaving much of the boreal and arctic regions of Canada and Alaska uncovered (Dunn et al. 2005). The BBS sampling scheme consists of numerous short-duration counts performed at widely-spaced points (Sauer et al. 1994). Because detectability declines with increasing distance from the observer, these “point counts” have the highest possible ratio of area with low detectability to area with high detectability (Buckland et al. 2001). Therefore detectability can be presumed to be most consistent for species that tend to reside on relatively small, fixed home ranges, and provide abundant cues to the observer. For this reason, it is not surprising that point-counting is the predominant survey method for monitoring breeding songbirds (*Passeriformes*) (Ralph et al. 1995).

Many raptors (*Accipitriformes* and *Falconiformes*), however, are not well suited for BBS trend analysis (Dunn et al. 2005). Being large-bodied and predatory, most species of raptors in North America have relatively large home-ranges in the breeding season (Fuller and Mosher 1981, 1987). Many have large populations breeding in the boreal forest and tundra north of the limit of the surveyed region (Dunn et al. 2005). The raptors problematic for the BBS tend to be long-distance migrants (Kerlinger 1989).

The energetic demand of migration and the vagaries of weather cause migratory flights of many raptors to become concentrated at certain geographic features, known as

leading lines (where lift is improved) or diversion lines (where paths are redirected by neighboring regions of poor lift) (Bildstein 2006). Because these lines are often predictable, raptor-watchers began (first in the Northeastern US) to annually attend fixed sites to count the numbers of raptors that pass. Realizing the value of such counts for monitoring these species, later generations have improved the quality of data at existing sites and began many new watch-sites in the western and southern portions of North America to build a continental monitoring network (Zalles and Bildstein 2000, Bildstein 2006, Bildstein et al. 2008). Diverse origins, priorities, and protocols of watch-site managing entities have made progress from a loose collective of nonprofit organizations to a unified continental monitoring network difficult. Building such a network requires first a widespread agreement on a satisfactory data-collection protocol, followed by the development of a sound method of trend analysis (Titus et al. 1989, Lewis and Gould 2000, Farmer et al. 2007, Bildstein et al. 2008).

Raptor migration counts are rightly considered indices rather than estimates of population size because the location of raptor migration watch-sites is neither systematic nor random, and the observed flight does not represent a complete coverage of the population (Kerlinger 1989, Farmer et al. 2007). The *a priori* assumptions of traditional survey design do not apply. However, sound statistical analysis of raptor migration counts may still be possible. A conceptual framework for inference from a sample drawn from a previously selected sub-population is known in the statistical literature as a superpopulation model (Hartley and Sielken 1975).

Raptor migration counts are an example of such a two-stage sampling procedure. First the raptors must migrate near a watch-site while the observers are present. Second,

the observers must see, identify, and record those raptors. The first stage is limited by sample coverage, and the second stage is limited by probability of detection (Nichols et al. 2009). At each stage, sample bias is possible. To improve the confidence with which managers might make decisions based on raptor migration counts, researchers should seek to quantify these biases, identify their causes, and mitigate their statistical consequences. In this thesis, I present my research examining the causes and consequences of detection bias in raptor migration counts.

In Chapter 1, I present an empirical study conducted at the Idaho Bird Observatory's Lucky Peak watch-site, near Boise, Idaho during the fall counts in 2009 and 2010. The goal of this study was to quantify the magnitude and variance of detectability of migrating raptors at an inland leading line watch-site, using a double-observer survey design (Nichols et al. 2000). This was the first study of detectability at an elevated site far from a coastline, or outside of the Atlantic Flyway. I modeled the relative effects of factors related to observers, flight line, species, and weather, with the goal of identifying the most important factors to consider in designing improved trend analyses or survey protocols for raptor watch-sites.

In Chapter 2, I present computer simulations that utilized the empirical data and models from Chapter 1 to estimate the effect of heterogeneous, imperfect detectability on trend analyses of standardized raptor migration counts. I estimated the expected variance, bias, and resulting loss of statistical power attributable to detectability. I estimated the relative effect of varying sample coverage on power by comparing the simulated detectability-related variance with the total variance in 15 years of historical counts.

Finally, I assess the strengths and weaknesses of raptor migration counts as a tool for monitoring and conservation, and suggest some directions for research into mitigating extraneous variation in sample coverage and detectability. By empirically verifying the theoretical basis for raptor migration counts as an index of population change, hypothesis-based research may improve the value of raptor migration counts as a technique for population monitoring.

Literature Cited

- Allredge, M. W., K. H. Pollock, and T. R. Simons. 2006. Estimating detection probabilities from multiple-observer point counts. *Auk* 123:1172-1182.
- Allredge, M. W., K. H. Pollock, T. R. Simons, J. A. Collazo, and S. A. Shriner. 2007a. Time-of-detection method for estimating abundance from point-count surveys. *Auk* 124:653-664.
- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007b. Factors affecting aural detections of songbirds. *Ecological Applications* 17:948-955.
- Bildstein, K. L. 2006. *Migrating raptors of the world : their ecology & conservation*. Comstock Pub. Associates, 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 Ornithologist's Union, Washington, D.C., USA.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling*. Oxford University Press, Oxford, United Kingdom.
- Dunn, E. H. 2002. Using decline in bird populations to identify needs for conservation action. *Conservation Biology* 16:1632-1637.
- 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. *In* *Partners in Flight Technical Series. Partners in Flight*, [Online] www.partnersinflight.org/pubs/ts/02-monitoringneeds.pdf.
- 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. *Studies in Avian Biology* 6:235-246.
- _____. 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.
- Hartley, H. O., and R. L. Sielken. 1975. Super-population viewpoint for finite population sampling. *Biometrics* 31:411-422.
- Johnson, D. H. 2008. In defense of indices: The case of bird surveys. *Journal of Wildlife Management* 72:857-868.
- Kerlinger, P. 1989. *Flight strategies of migrating hawks*. University of Chicago Press, Chicago, Illinois, 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.
- Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. *Ecological Applications* 8:258-268.
- 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.
- 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.
- Ralph, C. J., J. R. Sauer, and S. Droege. 1995. *Monitoring bird populations by point counts*. Forest Service General Technical Report PSW-149, Albany, California, USA.
- Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* 111:50-62.
- Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. *Auk* 119:18-25.
- Titus, K., M. R. Fuller, and J. L. Ruos. 1989. Considerations for monitoring raptor population trends based on counts of migrants. In B.-U. Meyburg, and R. D. Chancellor, editors. *Raptors in the modern world : proceedings of the III World Conference on Birds of Prey and Owls*. International Council for Bird Preservation, Berlin, Germany.

Zalles, J. I., and K. L. Bildstein. 2000. Raptor watch : a global directory of raptor migration sites. Birdlife International, Cambridge, United Kingdom.

CHAPTER 1: DETECTABILITY OF MIGRATING RAPTORS AT A WESTERN RIDGELINE WATCH-SITE

Abstract

Annual counts of migrating raptors are used as indices of population size. Heterogeneous detectability may cause the counted proportion of raptors to vary. This variation may reduce the precision of population trend estimates. I used a dependent double-observer method to estimate detectability at the annual fall raptor migration count at Lucky Peak, Idaho, in 2009 and 2010. I used Huggins closed-capture removal models and information-theoretic multi-model inference to determine factors affecting detectability. The most parsimonious model included effects of observer identity, distance, wingspan, genus, and day of the season. Competitive models also included wind-speed, cloud cover, and hour of the day. These results demonstrate the importance of controlling observer skill and effort and the potential utility of adjusting daily counts to account for differences in flight distance. By employing methods that address the factors that affect detectability, raptor-observatory organizations may be able to produce trend assessments with greater statistical power, thereby better informing timely management decisions.

Introduction

Population monitoring is essential to avian conservation (Finch and Martin 1995, Dunn 2002). The North American Breeding Bird Survey has proven to be an effective monitoring method for many species, but trend estimates for many raptors

(*Accipitriformes*, *Falconiformes*) 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 and many breed in the remote northern reaches of the continent not covered by the BBS (Fuller and Mosher 1981, Dunn et al. 2005). Some species, such as those that breed in forests and do not confront intruders, are easier to observe on migration (Fuller and Mosher 1987).

During migration, wind drift, leading lines, and diversion lines create concentrations of visible migrants at predictable locations (Zalles and Bildstein 2000, Bildstein 2006). At such locations, termed watch-sites, observers record the numbers of each raptor species that pass (Zalles and Bildstein 2000, Bildstein et al. 2008). In North America, over 117 watch-sites have engaged in long-term monitoring of raptor migration. Additionally, at least 58 monitoring watch-sites have been established elsewhere in the world (Zalles and Bildstein 2000).

However, the relationships between raptor migration counts and biological populations are complicated and poorly understood, making inference difficult (Kerlinger 1989, Dunn and Hussell 1995). Migration counts are not a representative sampling of biological populations; however, changes in migration counts over time may be considered an index of change in population size (Farmer et al. 2007, Farmer and Hussell 2008). Precision of trend estimation is reduced by variation in the proportion of the population counted (Thompson 2002, Johnson 2008). The proportion counted depends on the sample coverage and the probability of detection (Nichols et al. 2009).

The probability of a raptor being available to be counted is the product of three constituent probabilities (Nichols et al. 2009): 1) The watch-site is on the raptor's

migratory path (p_s), 2) the raptor is present during the hours observers are present (p_p), and 3) the raptor behaves in such a way as to not be invisible (p_a) (Dunn and Hussell 1995).

The count of available raptors is limited by the probability of detection (p_d) (Nichols et al. 2009). Imperfect detection results in the count of available birds being lower than the actual value. Variation in detectability contributes to count variance, reducing statistical power to detect trends (Thompson 2002). A trend in detectability over time may bias estimates of trends in the number of available birds (Thompson 2002, Johnson 2008).

Two previous studies have examined the factors affecting detectability 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, found that detectability varied by observer attentiveness, flight density, flight visibility, and species. Specifically, they found that higher birds were less visible and detectable than lower birds and that the observer was more attentive and detected raptors with greater efficiency during times of high flight density. 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 St. Lawrence estuary in Quebec, used a double-observer approach 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 thus affected detectability indirectly (Berthiaume et al. 2009).

The detectability studies of Sattler and Bart (1984) and Berthiaume et al. (2009) were performed at watch-sites in the Northeast on shorelines and at sites where observers worked alone rather than in a team. Neither study identified any differences in detectability between observers. However, observer effects exist in avian point counts (Campbell and Francis 2011, Alldredge et al. 2007, Nichols et al. 2000, Cunningham et al. 1999, Kendall et al. 1996, Sauer et al. 1994), and are likely in raptor migration counts (Dunn and Hussell 1995, Dunn et al. 2008). Furthermore, detectability may be affected by site-specific factors and the number of observers (Kochenberger and Dunne 1985).

I used a double-observer sampling design to estimate the detectability of migrating raptors at a mountain-ridge site in the Western interior with paired observers. I investigated the relative effects of observers, characteristics of the migratory flight, weather, and species in determining detectability. My objective was to improve our understanding of statistical error in migration counts and suggest methodological and statistical applications that may improve the utility of migration counts for population monitoring.

Methods

Study Site

The Lucky Peak Hawk-Watch is performed each fall by the Idaho Bird Observatory, a nonprofit research program of Boise State University. At least two

observers count migrating raptors each day, from 25 August to 31 October, as weather permits. Counts are suspended only in the event of electrical storms, or precipitation which reduces visibility substantially. 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 (approx. 1000 m above the plain), visible migrant raptors are distributed both laterally and vertically. Counts at this site from 1994 to 2005 were analyzed by Smith et al. (2008). 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 via two-way radio. The watch-site is open to the public, and observers provide interpretation for visitors.

Experiment

I conducted a double-observer sample (Nichols et al. 2000) during the autumn raptor migration count on 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 located at the traditional lookout positions and attempted to count all raptors passing the lookout. The other team, designated secondary, was positioned approximately three meters behind the primary team. The secondary observers recorded, on a separate sheet, only additional raptors that were not counted by the primary team. The primary observers called out the identification and location of raptors they observed so the secondary observers could avoid double-recording raptors. Secondary observers could ask the primary observers

questions to clarify which bird had been counted, but were quiet 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 non-detection by the primary observers. Birds captured in nets and reported to the observers via radio were removed from the data. I randomly assigned observers to teams for 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 switched 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 visibility-based distance and altitude category. Observers assigned birds to one of three categories by altitude only when within the range of unaided vision (where differences in background color and viewing angle are greatest when altitude varies), and assigned birds to visibility-based distance categories without regard for altitude when they were more distant (definitions in Table 1a). I chose this system because lateral distance affected apparent size in the same way as difference in altitude, so distance and altitude were difficult to measure separately, and their effects on detectability were likely to be similar enough to complicate model-fitting if they were to be considered independently. Observers classified each bird based on its closest approach to the watch-site, even if it was detected farther 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 (Table 1b).

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 undetected in the survey. 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 capture probabilities, and 3) there is no heterogeneity in capture and recapture probabilities among individual animals.

One additional assumption is unavoidable with the dependent double-observer survey design, because observer-specific detectability is only estimable for the primary observers (Nichols et al. 2000): The detection probability for an observer team is not affected by whether it is in the primary or secondary role.

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. Predatory raptors at Lucky Peak were very seldom seen migrating in groups of > 4 birds (approximately 3% of observations), so detection of individuals was generally independent. Turkey Vultures (*Cathartes aura*) typically were counted in large groups, so this species was excluded from analysis.

Heterogeneity in detection probability among individual raptors has been shown in previous studies (Sattler and Bart 1984, Berthiaume et al. 2009). To account for

individual differences I used the conditional likelihood approach developed by Huggins (1989, 1991). Heterogeneity in detection probability was incorporated as a linear function of multiple covariates related to the observer, flight, weather, and species of each bird.

I 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. I model-averaged models with $\Delta AIC_c < 2$ to account for model-selection uncertainty in estimating effect sizes (Burnham and Anderson 2002). Model-fitting was performed using the Huggins closed-capture data type in Program MARK (White and Burnham 1999). I coded raptors recorded by the primary observers with encounter history "11," and raptors recorded only by the secondary observers with encounter history "01." I 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.

I measured several covariates related to each of the four hypothesized sources of variation in detectability: observers, migratory flight, weather, and species. I examined independent measurable covariates for correlation and any with coefficients $> \pm 0.4$ were not used the same model. Initially, I fit all possible models representing 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). For each source, I selected the model with the lowest AIC_c as representative of the working hypothesis. I used the variables from these four models to construct a general model. I then built, from subsets of

variables in the general model, a set of candidate models with all possible combinations (364 models). In doing so, I kept sets of variables describing a single covariate together.

I 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). I pooled the 17 other observer teams with insufficient samples. The seven regular observers (symbolized by A – G in Tables 2 and 4) 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 any prior experience observing bird migration (5 months). I 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. I also modeled a second-order effect of number of days since the beginning of the season to account for a non-linear effect of practice.

I used the number of birds observed per hour (BPH), representing a naïve estimate of flight density, as a covariate for all birds observed in that hour. I used the distance category (see Table 1a) as an individual covariate to model the effect of flight-line. I also included a second-order effect of distance on detectability to account for non-linearity. Non-linearity was strongly suspected for two reasons: 1) Distance category was an ordinal variable, and units were likely to be unequal, and 2) non-linear distance-detectability functions are common (Buckland et al. 2001).

I included wind speed, ambient temperature, and cloud cover category as covariates. As circular variables cannot be used in linear models, I used the cosine of

wind direction as a linear covariate. This number ranged from -1 (wind from the south, a headwind) to 1 (wind from the north, a tailwind). I 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. I chose these transformations because the resulting variables were likely to be correlated with the speed of migrating raptors. I chose to limit the number of wind variable interactions to avoid co-linearity and make the effect of migration volume and flight line distinguishable from more proximate effects of wind.

I hypothesized that detectability might vary among species because species were of different visible size or flew with different styles. I used an approximate average wingspan for each species (Sibley 2000) as a variable to account for visible size. The second-order effect of wingspan was also considered, in case detectability might increase non-linearly with size. To account for differences in flight style among raptors of similar wingspan, I used a dichotomous variable for each genus of raptors observed, with the exception of *Aquila* and *Haliaeetus* (Eagles), which were pooled because of similarity of flight style and small sample sizes.

Distributions of covariates were described with arithmetic means and standard deviations. Tests of differences in covariates between years were performed with Pearson χ^2 tests for dummy variables and Welch t tests for quantitative variables ($H_0: \bar{x}_1 = \bar{x}_2, \alpha = 0.05$). 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 sample bias caused by heterogeneous detectability (i.e., more-detectable birds get sampled disproportionately often) (Horvitz and Thompson 1952).

Results

Observers detected 6873 raptors in 390 hours on 65 days. Secondary observers made 23% of detections (effective sample size = 1595). Observer teams that participated on fewer than seven days made a far greater proportion of observations in 2010, and different observer teams participated in each year (Table 2). We began double-observer data collection 12 days later in the season in 2010 than in 2009 (Table 2). The distance category for observed raptors was higher on average in 2009 (Table 2). Mean ambient temperature, wind-speed, and cosine of wind direction differed between years, but cloud cover did not (Table 2). Comparison of AIC_c between the year-effect model and models representing other hypotheses suggested that the other covariates had superior explanatory value, and I did not consider year in any additional model-selection to avoid co-linearity. Therefore, caution is necessary when interpreting model selection results (Table 3) and estimates of effect sizes (Table 4) for the covariates that differed between years.

The most parsimonious model (evidence ratio to second model = 1.5) included the effects of observers, flight distance, species, and day of the season (Table 3). Nine of the 406 models in the candidate set had a $\Delta AIC_c < 2.0$, all of which included every parameter in the top model (Table 3).

Detection probabilities differed among observer teams (Table 4). Detectability increased with the number of days since the beginning of the season (Table 4), suggesting a positive effect of practice on detection probability (odds ratio of last day to first = 1.76). Detectability greatly decreased with distance beyond the range of unaided vision (Figure 1). Species with longer wingspans were more detectable, with the exception of Ospreys (*Pandion haliaetus*), which were unusually difficult to detect for their size (Figure 2). Otherwise, genus did not have a significant conditional effect on detectability. Weather had little effect on detectability independent of species, flight, and observers (Cloud cover importance weight = 0.39, wind speed importance weight = 0.46).

Estimated detectability of individual raptors observed ranged from 0.23 to 0.99 for the two primary observers. The weighted mean detectability with two observers was 0.66 (SD = 0.14). The weighted mean detectability with all four observers present was 0.86 (SD = 0.10).

Discussion

Detectability of migrant raptors at Lucky Peak varied depending on the identities of the observers, the distance of the migratory flight, and species characteristics. These results emphasize the importance of maintaining consistent levels of observer skill and morale, and the utility of collecting high-quality spatial data. Differences in detectability among species may cause comparison of counts of different species at a watch-site to not accurately reflect their true relative abundance.

Varying observer effects are well-known in point counts (Campbell and Francis 2011, Nichols et al. 2000, Cunningham et al. 1999, Kendall et al. 1996, Sauer et al. 1994), and have been suspected to occur in raptor migration counts (Dunn and Hussell

1995, Dunn et al. 2008). My results confirm that observer effects are important in determining the detectability of migrating raptors, contradicting the conclusion of the previous double-observer study (Berthiaume et al. 2009).

In my opinion, the prior double-observer raptor migration count study (Berthiaume et al. 2009) found no differences in detectability among observers because the experimental design was not adequate for detecting such differences. The models incorporated an assumption that detection probabilities of primary and secondary observers were mutually independent (fixed $c = p_2$). However, the secondary observers were not prevented from viewing the activity of the primary observers (Berthiaume et al. 2009). Unintentional provision of visual cues by the primary observer may have violated the assumption of mutually independent detection (Alldredge et al. 2006). If this occurred, comparison of the estimated detection probabilities of observers in the primary and secondary roles may not be valid. In the same role, the prior study compared only two observers with similar levels of experience (Berthiaume et al. 2009). The design of this study differed in key respects, and followed more closely the methods of Nichols et al. (2000), which may have made observer differences more apparent: I used more observers, rotated observers between roles, treated secondary observers as non-independent, and equalized recording burdens between roles.

Apart from the observer effect, results were consistent with Berthiaume et al. (2009). Detectability was greatest for raptors within the range of unaided 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 (*Circus cyaneus*). The low detectability of Ospreys was more pronounced in this study than in Berthiaume et al. (2009), but was consistent with results from Sattler and Bart (1984). Ospreys at Lucky Peak in 2009 and 2010 were relatively uncommon (< 2% of raptors), 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 of the field of view with few raptors, making uncommon raptors with atypical migration strategies less detectable (Kochenberger and Dunne 1985). Alternatively, the Osprey's plumage may provide particularly effective camouflage against the sky.

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 since Lucky Peak is a mountaintop site where raptors are often detected near the horizon, whereas Tadoussac is a shoreline site close to sea level, and birds are likely detected at higher angles. Sattler and Bart (1984) observed that cloud cover improved visibility at Derby Hill, another weakly-elevated 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 relatively 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 experienced at the Derby Hill watch-site (over 200 raptors in 30 minutes), or, because

only one observer's efficiency was quantified, it may be an observer-specific effect (Sattler and Bart 1984).

Double-observer techniques for estimating detectability may not be appropriate for all raptor species and watch-sites. In particular, those species at watch-sites with flight densities high enough to cause most birds to be detected in "kettles" or "clusters" (Berthiaume et al. 2009) are likely to pose challenges. The method's assumption of independent detection is problematic in such cases. The method may be adapted to treat a cluster as the independently-detectable unit (Cook and Jacobson 1979, Buckland et al. 2001), provided clusters are well-defined and their constituent birds are homogenous in detectability. The latter condition is unlikely to be true for mixed-species assemblages.

At sites where the majority of counted raptors are recorded from estimations of the sizes of very large groups, variance arising from imperfect estimation may be of far greater magnitude than variance arising from imperfect detection (Boyd 2000). At such sites, the best options for assessing the relationship of the count to the number of available birds may be photography and radar, though each has limitations (Boyd 2000, Gauthreaux and Belser 2003).

Berthiaume et al. (2009) and I both used simple visibility-based metrics to model effects of distance and both 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 among 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 is a far superior format for a spatial covariate, which can be collected with little additional effort. Because distance affects detectability, and weather affects distance, collecting high-quality distance data may provide a means to develop more accurate models of weather-related count bias (Berthiaume et al. 2009). Alternatively, distance sampling may be investigated as a means of partially correcting for heterogeneous detectability (Buckland et al. 2001).

Literature Cited

- Allredge, M. W., K. H. Pollock, and T. R. Simons. 2006. Estimating detection probabilities from multiple-observer point counts. *Auk* 123:1172-1182.
- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007. Factors affecting aural detections of songbirds. *Ecological Applications* 17:948-955.
- Berthiaume, E., 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.
- Bildstein, K. L. 2006. *Migrating raptors of the world: their ecology and conservation*. Comstock Pub. Associates, 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 Ornithologist's Union, Washington, D.C., USA.
- Boyd, W. S. 2000. A comparison of photo counts versus visual estimates for determining the size of snow goose flocks. *Journal of Field Ornithology* 71:686-690.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling*. Oxford University Press, Oxford, United Kingdom.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd edition. Springer, New York, New York, USA.
- 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.

- Cook, R. D., and J. O. Jacobson. 1979. Design of estimating visibility bias in aerial surveys. *Biometrics* 35:735-742.
- 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.
- Dunn, E. H. 2002. Using decline in bird populations to identify needs for conservation action. *Conservation Biology* 16:1632-1637.
- 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. *In* Partners in Flight Technical Series. [Online] www.partnersinflight.org/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.
- 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.
- 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.
- _____. 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, DC.
- Gauthreaux, S. A., and C. G. Belser. 2003. Radar ornithology and biological conservation. *Auk* 120:266-277.

- HMANA. 2009. Forms. Hawk Migration Association of North America, Plymouth, New Hampshire, USA. [Online] http://www.hmana.org/data_entry_paper.php.
- 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.
- Huggins, R. M. 1989. On the statistical-analysis of capture experiments. *Biometrika* 76:133-140.
- _____. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* 47:725-732.
- Johnson, D. H. 2008. In defense of indices: The case of bird surveys. *Journal of Wildlife Management* 72:857-868.
- 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.
- Kerlinger, P. 1989. Flight strategies of migrating hawks. University of Chicago Press, Chicago, Illinois, USA.
- 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.
- 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.
- 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.
- 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*:7-135.
- 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, DC, 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.
- Sibley, D. 2000. *The Sibley guide to birds*. 1st 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, DC, USA.
- Thompson, W. L. 2002. Towards reliable bird surveys: Accounting for individuals present but not detected. *Auk* 119:18-25.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120-139.
- Zalles, J. I., and K. L. Bildstein. 2000. *Raptor watch: a global directory of raptor migration sites*. Birdlife International, Cambridge, United Kingdom.

Table 1a. 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. I excluded all birds assigned to category 6 from my analyses, as they represent birds not within the standard search radius at this site, and not available to all observers since 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 category	Definition
0	Below level of the observers, within range of unaided vision.
1	0 - 30m above observers, within range of unaided vision.
2	>30m 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.

Table 1b. Ordinal scale used in estimating effect of cloud cover on detectability.

Categories correspond to a subset of sky condition codes on the HawkWatch

International Flight Information and Weather Data Form.

Cloud cover category	Definition
0	Clear: 0% to 15% cover
1	Partly Cloudy: 16% to 50% cover
2	Mostly Cloudy: 51% to 75% cover
3	Overcast: 76% to 100% cover

Table 2. Covariates used in models of individual heterogeneity in detectability.

Columns show range, mean for each year, and mean for all observations. Variables for Genus and Observers were dichotomous, and a value of 1 indicates the condition was true, so the mean is a ratio. Continuous variables and discrete ordinal variables (which were treated in the same way) are shown with standard deviations. Asterisks indicate significant differences between years ($p < 0.05$).

Variable	Min.	Max.	2009 ($n = 4164$)	2010 ($n = 2709$)	All ($n = 6873$)
Observers B&C	0	1	0.22		0.13
Observers C&F	0	1	0.19		0.11
Observers B&F	0	1	0.16		0.09
Observers B&D	0	1	0.15		0.09
Observers D&F	0	1	0.13		0.08
Observers C&D	0	1	0.12		0.07
Observers A&E	0	1		0.22	0.08
Observers D&E	0	1		0.12	0.05
Observers D&G	0	1		0.10	0.04
Observers E&G	0	1		0.10	0.04
Observers (Other)	0	1	0.05	0.47*	0.22
Genus <i>Accipiter</i>	0	1	0.42	0.49	0.45
Genus <i>Buteo</i>	0	1	0.25	0.24	0.24
Genus <i>Circus</i>	0	1	0.05	0.05	0.05
Genus <i>Falco</i>	0	1	0.25	0.17	0.22
Genus <i>Pandion</i>	0	1	0.02	0.02	0.02
Genus (Eagles)	0	1	0.01	0.04	0.02
Wingspan (cm)	56	203	83.0 (30.9)	84.0 (30.4)	83.4 (30.7)
Wind-speed (kph)	0	38.5	11.4 (5.8)	8.3 (5.0)*	10.2 (5.7)
cos(Wind direction)	-1	1	0.01 (0.61)	-0.18 (0.57)*	-0.06 (0.60)
Temperature (°C)	-3.6	39.1	18.7 (8.4)	21.3 (5.6)*	19.7 (7.5)
Cloud Cover	0	3	0.6 (0.9)	0.9 (1.1)	0.7 (1.0)
Distance	0	5	2.5 (2.3)	1.6 (1.7)*	2.4 (1.6)
BPH	1	216	42.1 (30.8)	35.9 (23.6)	39.7 (28.3)
Day	3	67	27.9 (14.6)	33.1 (12.1)*	30.0 (13.9)
Hour	10	19	14.2 (2.0)	14.0 (1.9)	14.1 (2.0)

Table 3. Comparison of 406 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 . \mathcal{L} is the model likelihood, and w is the AIC_C weight of evidence. K is the number of parameters in the model. The top nine models, with $\Delta AIC_C < 2.0$, were model-averaged to estimate effect sizes for covariates.

Rank	Model	ΔAIC_C^\dagger	w	\mathcal{L}	K
1	p (Team + Distance + Distance ² + Wingspan + Genus + Day)	0.00	0.100	1	0
2	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Hour + Windspeed)	0.82	0.067	0.664	2
3	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Hour)	0.83	0.066	0.661	1
4	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Windspeed)	0.91	0.064	0.634	1
5	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Cloud Cover)	1.10	0.058	0.576	1
6	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Cloud Cover + Windspeed + Hour)	1.27	0.053	0.529	3
7	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Day_2)	1.68	0.043	0.433	1
8	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Cloud Cover + Windspeed)	1.69	0.043	0.430	2
9	p (Team + Distance + Distance ² + Wingspan + Genus + Day + Cloud Cover + Hour)	1.81	0.041	0.404	2
32	General model: Observers + Flight + Species + Weather	5.24	0.007	0.073	5
145	Observers: p (Team+Day+Day ² +Hour)	95.70	0	0	4
284	Flight: p (Distance+Distance ² +BPH)	179.09	0	0	4
334	Species: p (Wingspan+Genus)	215.42	0	0	7
375	Year effect: p (Year)	227.53	0	0	2
378	Weather: p (Cloud Cover + Windspeed)	231.16	0	0	3
395	Constant: p (.)	236.59	0	0	1

† The lowest AIC_C value was 7212.48

Table 4. Model-averaged estimates of coefficients ($\hat{\beta}$) with standard errors (SE) and odds ratios ($e^{\hat{\beta}}$). All covariates have been scaled to range from 0 to 1 to allow comparison of relative magnitudes of effects. Asterisks indicate significant effects ($H_0: \hat{\beta} \neq 0, \alpha = 0.05$). Each unique observer letter represents an individual observer. The letters were randomly assigned. Reference categories are: Observers (Other) (17 teams that participated on < 7 days) and Genus (Eagles) (*Haliaeetus* and *Aquila* pooled).

Parameter	$\hat{\beta}$	SE	$e^{\hat{\beta}}$
Intercept	0.836	0.497	2.310
Observers B & C	0.558	0.136	1.750*
Observers B & F	0.462	0.155	1.590*
Observers E & G	0.139	0.156	1.150
Observers A & E	0.038	0.109	1.040
Observers D & G	-0.162	0.152	0.850
Observers C & F	-0.278	0.138	0.760*
Observers C & D	-0.308	0.131	0.730*
Observers D & E	-0.433	0.125	0.650*
Observers D & F	-0.486	0.120	0.620*
Observers B & D	-0.616	0.141	0.540*
Distance	0.987	0.475	2.680*
Distance ²	-2.337	0.524	0.100*
Wingspan	1.593	0.649	4.920*
Genus <i>Accipiter</i>	-0.025	0.437	0.980
Genus <i>Buteo</i>	-0.067	0.348	0.940
Genus <i>Circus</i>	-0.127	0.396	0.880
Genus <i>Falco</i>	0.014	0.472	1.010
Genus <i>Pandion</i>	-1.033	0.491	0.360*
Cloud Cover	0.195	0.146	1.220
Windspeed	-0.405	0.313	0.670
Day	0.566	0.286	1.760*
Day ²	0.543	0.878	1.720
Hour	0.338	0.220	1.400

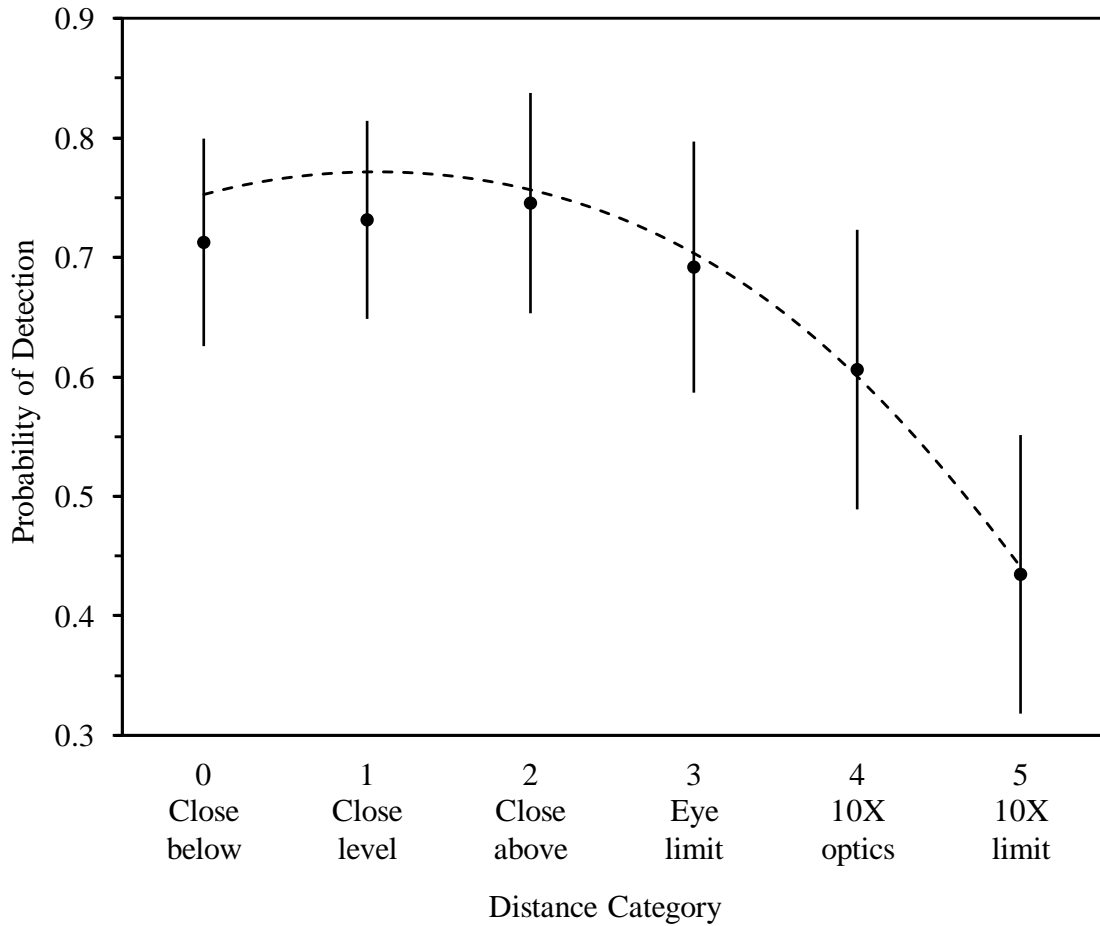


Figure 1. Effect of relative distance and altitude on detectability. For definitions of distance categories, see Table 1a. Points are weighted mean detectability with bars of ± 1 SD. The dashed curve shows the model prediction for a hypothetical individual with average covariates. The effect of ordinal distance category was modeled as a quadratic function (Table 4).

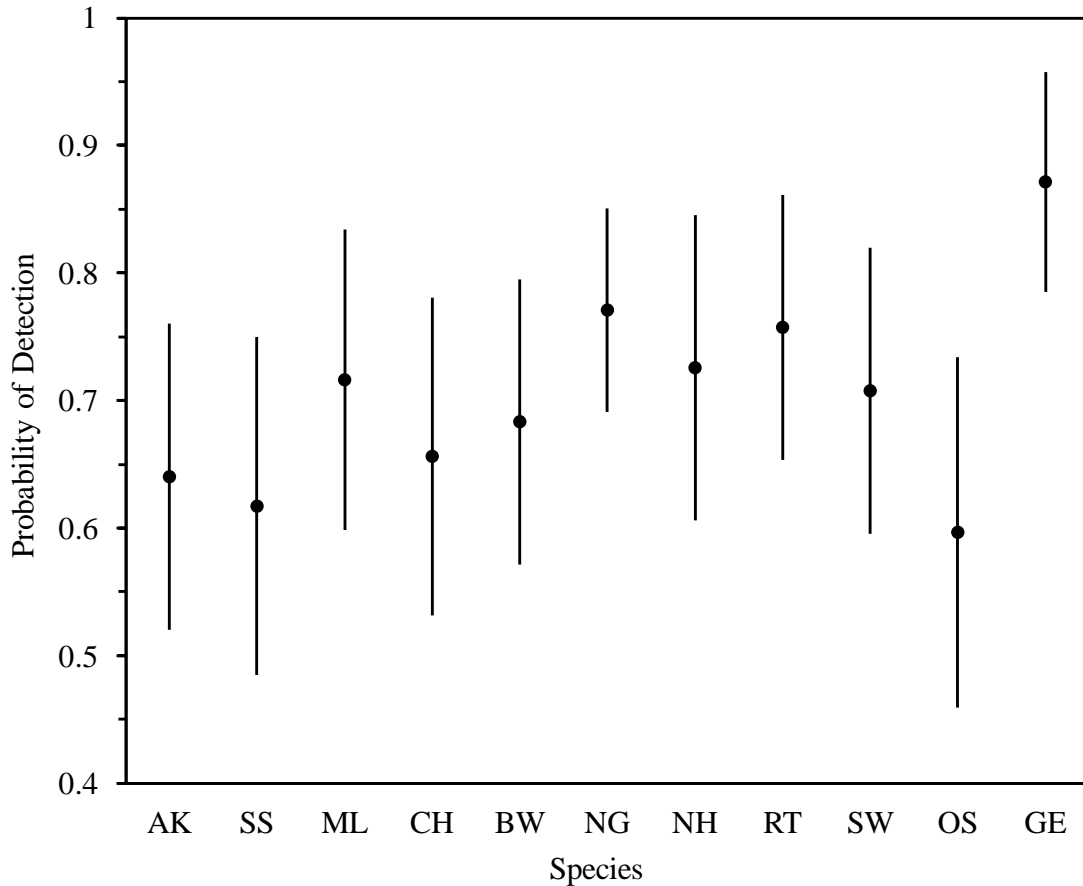


Figure 2. Estimated mean detectability of selected species, ordered by increasing wingspan. Points are weighted means with bars of ± 1 SD. Species effects on detectability were modeled by additive effects of wingspan and genus (Table 4). Species are: AK = American Kestrel (*Falco sparverius*), SS = Sharp-shinned 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*).

CHAPTER 2: DOES IMPERFECT DETECTION OF MIGRATING RAPTORS
AFFECT THE POWER OF POPULATION TREND ANALYSES?

Abstract

Power to detect trends may be decreased by unexplained variation in raptor migration count data. Techniques such as double-observer or distance sampling, which estimate the probability of detecting birds that pass a watch-site, may reduce unexplained variation. I conducted double-observer sampling at Idaho Bird Observatory's Lucky Peak watch-site during the fall migration in 2009 and 2010, and estimated the effects of observers, species, flight distance, and weather on detectability. I used the model in simulations of 30 years of Sharp-shinned Hawk (*Accipiter striatus*) and Northern Harrier (*Circus cyaneus*) counts with heterogeneous individual detectability, a population decline ($\lambda = 0.964$), and a degree of unexplained random variation in the number of available birds. I ran ≥ 1000 iterations of each model parameterization. I estimated the power of a regression to detect the true trend as the proportion of iterations where the null hypothesis $\lambda < 1$ was rejected. Imperfect detection caused minimal bias in trend estimates ($\leq 10^{-4}$), but heterogeneity in detectability did increase variance in counts and reduced power. Simulated counts with perfect detectability required $< 11\%$ fewer years to detect the decline with 80% power ($H_0: \lambda = 0, \alpha = 0.1$) when variance in annual numbers of raptors available to count was realistically high ($CV \geq 0.26$). Detectability correction by means of double-observer or distance sampling, in the case of raptor migration counts, may not be warranted for the purpose of long-term population monitoring. Efforts may

be better focused on improving our understanding of mechanisms that cause changes in the number of migrants available to count.

Introduction

For many reasons, migration counts at fixed locations (watch-sites) are an appealing method for monitoring populations of many raptor species (*Accipitriformes* and *Falconiformes*) (Chapter 1). However, the relationship of a count of migrating raptors to breeding or wintering populations is complicated by both the dynamic nature of migration and technological limitations (Fuller and Mosher 1981, 1987, Kerlinger 1989, Dunn and Hussell 1995). Migration monitoring may be ineffective for monitoring populations because the relationship between the counted population and the total population is unknown, and because migration counts have more unexplained variability than breeding season counts (Svensson 1978, Fuller and Mosher 1981, Titus et al. 1989). Additionally, counts at watch-sites may be biased representations of the volume of the migratory flight, much of which may be at higher altitudes than can be detected visually, and may not follow leading lines (Kerlinger and Gauthreaux 1984, 1985, Kerlinger et al. 1985, Kerlinger 1989).

For these reasons, estimating the total number of migrating raptors at any particular watch-site is not a practical objective. Instead, a raptor migration count is considered an index of abundance, based on the assumption that changes in the count are proportional to changes in the population (Farmer et al. 2007, Farmer and Hussell 2008). For the index to be useful for monitoring, analysis of trends should ideally be both accurate and powerful. To be accurate, trends in the index must match the population trend in direction and magnitude. To be powerful, the results of analyses must be

statistically significant at a desirable level of confidence, given a limited number of years of counts (Bart et al. 2004).

Power in trend analysis is particularly important as a goal for conservation, as monitoring is typically undertaken to detect declines that require management action, and in such cases acting quickly may improve the probability of success (Bart et al. 2004). For trend analysis of index counts to be accurate and powerful, the proportion of the population of interest counted must not change over time (Thompson 2002). If there is a progressive change in the proportion counted, accuracy is reduced, but even random variation can reduce power.

The relationship of a count n to the super-population N^* from which it is sampled is determined by sample coverage and the probability of detection (Nichols et al. 2009). To be available to count, migrants must pass the watch-site within visible range while an observer is present. Let a_i equal the probability of an individual raptor meeting this condition. Only raptors available to count comprise the sample population N , that is:

$$N = \sum_{i=1}^{N^*} a_i$$

Secondly, the available raptors must be observed, identified, and recorded (i.e., detected). Let p_i equal the probability of detection for an available individual [p_d in Nichols et al. (2009)]. So:

$$n = \sum_{i=1}^N p_i$$

The mean probability of being available $\bar{a} = N/N^*$, and the mean detectability $\bar{p} = n/N$, therefore $n = N^* \times \bar{a} \times \bar{p}$.

I have previously developed a model of detectability incorporating individual covariates based on double-observer data collected at the Lucky Peak watch-site near Boise, Idaho in 2009 and 2010 (Chapter 1). The model parameters include effects of observers, species, distance, weather, and day of the season (Table 4). A distance category was recorded at the individual level, contrary to standard practice where data are recorded in hourly tables (HMANA 2009). With individual estimates of detectability (\hat{p}_i), the detectability-corrected count (\hat{N}) may be calculated with a Horvitz-Thompson (1952) estimator:

$$\hat{N} = \sum_{i=1}^n \frac{1}{\hat{p}_i}$$

Provided the number of present raptors accurately reflects the population trend, detectability-correction should cause trends in the count to more accurately reflect population trends. Since detectability presumably varies independently of abundance, it may also reduce unexplained variance and improve the power of trend analysis.

However, performing raptor migration counts in such a way as to make detectability estimable is non-traditional and may require more observers (Sattler and Bart 1984, Dunn et al. 2008, Berthiaume et al. 2009, Chapter 1). The potential for improvement in accuracy and power should be weighed against the costs of adopting such methods.

In this study, I used the migration count data and model to perform stochastic simulations of analyses of annual counts with known trends confounded by variation in sample coverage and detectability. My objective was to investigate the potential for detectability-correction to improve statistical power in trend-analysis. I chose a 50% decline in numbers of migrating raptors occurring over 20 years ($\lambda = 0.964$) as the benchmark trend, recommended by Bart et al. (2004) as a pragmatic balance of

conservation priorities and statistical precision. The Raptor Population Index project adopted a similar benchmark trend ($\lambda = 0.965$) to evaluate power [“moderate precision threshold” in Farmer et al. (2008) and Smith et al. (2008)].

Methods

I wrote a script for the R statistical environment (Revolution R Community 4.3 build of R 2.12.2, Revolution Analytics, Palo Alto, California) to simulate a series of yearly counts of a sub-population with a given starting value and a logarithmic trend, random variation in the number of available migrants and individual-based imperfect detectability.

I specified $E(N_1)$ and the population trend was determined by the geometric growth function $E(N_y) = E(N_1) \cdot (\lambda)^{y-1}$ with $\lambda = 0.964$. I simulated variation in the number of available raptors by generating a normally distributed pseudo-random deviance from the trend in each year (d_y) with $\mu = 0$, and $\sigma = E(N_y) \cdot CV_N$ with CV_N being the simulation parameter determining the coefficient of variation of available birds. I simulated the number of available birds in each year $N_y = E(N_y) + d_y$. Therefore, in a linear regression of $\log(N)$ by year, the slope is an unbiased estimator of the true trend ($\lambda - 1$), but power is limited by variation in sampling (CV_N).

I used a modified version of the individual-based detectability model from Chapter 1 (Table 4) to generate \hat{p}_i . The simulation model included an annual observer effect, fixed species effects, and individual covariate effects for distance category, wind speed, cloud cover category, and day of the season (Appendix A). I omitted the second-order effect of day of the season and the effect of hour of the day to simplify the computation, because neither effect was significant.

I assumed that two trained observers would be present at all times and that observers would change every year, but remain throughout each season. Hiring practices and established protocol at Lucky Peak generally support these assumptions. I generated normally-distributed pseudo-random annual observer effects. The mean of this distribution was the mean of the observer team effects estimated by the detectability model, and the standard deviation was 20% greater than the standard deviation of the estimated effects of the ten unique observer combinations. I increased the simulated observer effect variance because the observers in the experiment in 2009 and 2010 were likely to be more similar in performance than all observers who might participate at Lucky Peak, on account of the atypical degree of supervision during the experiment (generally speaking, observers at Lucky Peak are not actively supervised).

I chose a single species for each simulation, which determined the wingspan and genus effects. I randomly sampled the data with replacement and assigned each bird in the available population a set of distance, weather, and day covariate values matching those of an observed bird of the species. I weighted the covariate sampling by the inverse of the estimated detectability of each bird (under the actual conditions of data-collection, see Chapter 1) to avoid sample bias. I calculated \hat{p}_i by the detectability model (Appendix A), generated a pseudo-random number x_i from a uniform distribution in the range (0, 1), and determined available bird i to be detected when $x_i \leq \hat{p}_i$. The sum of detections was the count for the year (n_y).

I computed at least 1000 iterations of every simulation. At the completion of each iteration of the simulation, I fit an ordinary least-squares linear regression of $\ln(1 + n)$ by year, and calculated a 90% confidence interval for the slope. The decline was detected

if the upper limit of the confidence interval was < 0 , and undetected otherwise. I estimated statistical power as the number of iterations in which the decline was detected as a proportion of the total number of iterations. This analysis was performed with each sample size (number of years of counts) from 5 to 30, to determine the rate at which power increases. I repeated the analysis with a linear regression of $\ln(1 + N)$ by year. I estimated the loss of power caused by imperfect detection as the difference in power between the regressions of n and N .

I chose the Sharp-shinned Hawk (*Accipiter striatus*) and the Northern Harrier (*Circus cyaneus*) as model species. Both species were well-represented in my data set ($n = 1,919$ and $n=332$, respectively) and showed no tendency to flock at Lucky Peak. Both species have been identified as high-priority species for alternative range-wide surveys by Partners in Flight (Dunn et al. 2005). The two species differed substantially in estimated detectability in my dataset. The average detectability of Sharp-shinned Hawks was 0.62 (SD = 0.13), and the average detectability of Northern Harriers was 0.73 (SD = 0.12) (Chapter 1). The two species also differed in abundance at Lucky Peak. In annual Lucky Peak counts from 1994 to 2010 (IBO unpublished data), Sharp-shinned Hawks were relatively abundant (573 – 1,962 per year) compared with Northern Harriers (128 – 438 per year). Year-to-year variability in counts also differed between the two species (Sharp-shinned Hawk CV= 0.31, Northern Harrier CV= 0.38). I calculated CV as simply the standard deviation divided by the mean, because neither species showed a significant linear trend in counts (trend $< 1.4\%$, $p > 0.3$). At long distances, Sharp-shinned Hawks may sometimes only be identified to genus, because of the species' visual similarity with the Cooper's Hawk (*Accipiter cooperi*) (Hull et al. 2010). In simulations of counts of

Sharp-shinned Hawks, I included all records of “unidentified *Accipiter sp.*, small” and “unidentified *Accipiter sp.*, size undetermined” in the covariate dataset, to avoid under-representing the frequency of distant Sharp-shinned Hawks.

I ran simulations of each species with no trend and a range of values for CV_N from 0.1 to 0.4 to estimate the total CV of counts (\widehat{CV}_n) according to the level of variation in the number of available raptors. By comparing the \widehat{CV}_n from these simulations with the observed CV, I chose a values for CV_N in later simulations that may be realistic, assuming there was no population trend observed at Lucky Peak.

I chose an $E(N_1) = 2,000$ for Sharp-shinned Hawks and an $E(N_1) = 450$ for Northern Harriers, which represent the species’ maximum annual counts at Lucky Peak, rounded up <5% to a multiple of 50. Using a high number relative to the sample mean is reasonable since probability of detection is < 1 , and thus the count can be presumed to be lower than the number of available birds.

I estimated the historical CV_N for Sharp-shinned Hawks to be 0.26, and the historical CV_N for Northern Harriers to be 0.37 (Table 5). I selected an additional level of $CV_N = 0.18$ and an additional level of $E(N_1) = 100$. These values represent proportionally equal decreases, relative to the differences in the selected levels for the two species.

I simulated all combinations of species and parameter levels to evaluate the relative effects of species, rarity, and extraneous variation in sample population on the effect of detectability on power.

Results

The bias in trend estimation introduced by imperfect detection was minimal in all simulations ($|\hat{\lambda} - \lambda| \leq 10^{-4}$). Imperfect detection and heterogeneous detectability

affected power mainly by increasing count variance. The relative effect of limited detectability on power is inversely related to variation in the number of raptors available (CV_N) (Figures 3 and 4).

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 (Table 6). The effect of detectability on power in both species increased with decreasing variation in availability and decreasing population size (Table 6, Figures 3 and 4). When species were compared in terms of the effect of detectability on power to detect trends with equal $E(N_1)$ and CV_N values, detectability had a greater effect on power to detect trends in Sharp-shinned Hawks, the smaller and generally less perceptible species (Table 6, Figures 3 and 4).

I estimated that Lucky Peak Hawk-Watch would require 19 consecutive annual counts to achieve 80% power to detect a -3.5% annual trend in Sharp-shinned Hawks with 90% confidence (Table 6). Achieving 80% power to detect the same trend with the same analysis in Northern Harriers would require 25 years, because of the greater count variance and lower abundance of Northern Harriers, relative to Sharp-shinned Hawks (Table 6).

Discussion

Lewis and Gould (2000) estimated the power of trend analysis for seven watch-sites and concluded that a 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. This study appears to support the results of their power analysis. Results from simulation of a more severe decline (-50% in 20 years) estimated that a $CV \leq 38\%$ is necessary to attain 80% power in 25 years. Detectability correction alone appears 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 $CV \geq 30\%$ (Table 6).

The results of these simulations provide insight into the conditions in which detectability correction may be useful. 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 each year), and individuals of the species were relatively difficult to detect. Unfortunately, few combinations of species and watch-sites are likely to meet these qualifications. Raptor migration counts may have rates of detection of 66% or higher (Berthiaume et al. 2009, Chapter 1), utilize an index approach (Dunn and Hussell 1995), and are primarily useful for long-term monitoring (Fuller and Mosher 1981, 1987), making detectability correction less potentially beneficial in this method than those methods used in shorter duration studies or situations where animals are less detectable.

The accuracy of any simulation result is contingent on the realism of the model. These simulations depended on the assumption that observer ability varies randomly from year to year in a normal distribution. If the observer combinations in 2009 and 2010 were not a representative sample of all the observer teams who may be employed at Lucky Peak, or if the true observer skill distribution is skewed, these simulations may have underestimated the importance of detectability by overestimating the proportion of

variance attributable to availability. These simulations did not allow trends in average detection probability over time. If, for example, observers became more adept over time, as has occurred in the case of the North American Breeding Bird Survey (Sauer et al. 1994, Link and Sauer 1998), bias in estimation of declines would be greater, and correcting for detectability would improve power more than shown here (Bart et al. 2004). There may be a tendency for the average number of observers to increase as watch-sites become more widely known as birdwatching destinations. This would likewise affect the accuracy of trend estimates.

Management Implications

Watch-site managers should consider adopting staffing policies that ensure a symmetrical distribution of observer effects with low year-to-year variation to ensure detectability has little effect on power. I 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.

The relative importance of factors affecting the number of raptors available to count is in need of further research, in light of the potential power of inference to be gained by accounting for variation in sample coverage. 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 high priority.

Literature Cited

- Bart, J., K. P. Burnham, E. H. Dunn, C. M. Francis, and C. J. Ralph. 2004. Goals and strategies for estimating trends in landbird abundance. *Journal of Wildlife Management* 68:611-626.
- Berthiaume, E., 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.
- 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.
- 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. *In* *Partners in Flight Technical Series*. [Online] www.partnersinflight.org/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.
- Farmer, C. J., R. J. Bell, B. Drolet, L. J. Goodrich, E. Greenstone, D. Grove, D. J. T. Hussell, D. Mizrahi, F. J. Nicoletti, and J. Sodergren. 2008. Trends in autumn counts of migratory raptors in Northeastern North America, 1974-2004. Pages 179-215 *in* K. L. Bildstein, J. P. Smith, E. Ruelas Inzunza, and R. R. Veit, editors. *State of North America's Birds of Prey*. Nuttall's Ornithological Club and American Ornithologist's Union, Washington, D. C.
- 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.

- Fuller, M. R., and J. A. Mosher. 1981. Methods of detecting and counting raptors: a review. *Studies in Avian Biology* 6:235-246.
- _____. 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.
- HMANA. 2009. Forms. *In* Hawk Migration Association of North America, Plymouth, NH. [Online] http://www.hmana.org/data_entry_paper.php.
- 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.
- Hull, J. M., A. M. Fish, J. J. Keane, S. R. Mori, B. N. Sacks, and A. C. Hull. 2010. Estimation of Species Identification Error: Implications for Raptor Migration Counts and Trend Estimation. *Journal of Wildlife Management* 74:1326-1334.
- Kerlinger, P. 1989. *Flight strategies of migrating hawks*. University of Chicago Press, Chicago, Illinois, USA.
- Kerlinger, P., V. P. Bingman, and K. P. Able. 1985. Comparative flight behavior of migrating hawks studied with tracking radar during autumn in central New York. *Canadian Journal of Zoology* 63:755-761.
- Kerlinger, P., and S. A. Gauthreaux. 1984. Flight behavior of Sharp-shinned Hawks during migration: 1. Over land. *Animal Behaviour* 32:1021-1028.
- _____. 1985. Seasonal timing, geographic-distribution, and flight behavior of broad-winged hawks during spring migration in south Texas - a radar and visual study. *Auk* 102:735-743.
- 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.
- Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. *Ecological Applications* 8:258-268.
- 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.
- 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.
- 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.
- Svensson, S. E. 1978. Efficiency of 2 methods for monitoring bird population levels - breeding bird censuses *contra* counts of migrating birds. *Oikos* 30:373-386.
- Titus, K., M. R. Fuller, and J. L. Ruos. 1989. Considerations for monitoring raptor population trends based on counts of migrants. In B.-U. Meyburg, and R. D. Chancellor, editors. *Raptors in the modern world : proceedings of the III World Conference on Birds of Prey and Owls*, Eilat, Israel, 22-27 March 1987. International Council for Bird Preservation, Berlin, Germany.

Table 5. Estimating the coefficient of variation of annual numbers of available raptors at Lucky Peak. \widehat{CV}_n is the estimated coefficient of variation of simulated annual counts. In these simulations, the expected available population $E(N)$ remained constant, detectability varied stochastically for each species according to an individual-based model, and simulation parameter CV_N (the square root of variance in the available population as a proportion of $E(N)$) was manipulated to determine a level that approximated the observed coefficient of variation at Lucky Peak (IBO unpublished data).

Species	$E(N)$	historical CV	CV_N	\widehat{CV}_n	SE
Sharp-shinned Hawk	2000	0.31	0.18	0.23	0.0007
			0.26	0.30	0.0009
			0.37	0.40	0.0013
Northern Harrier	450	0.38	0.18	0.21	0.0006
			0.26	0.28	0.0009
			0.37	0.38	0.0013

Table 6. Simulation results: Number of years of counts necessary to achieve 80% power to detect a -3.5% annual population trend (least-squares linear regression, two-tailed test, $\alpha = 0.1$) with heterogeneous ($p < 1$) and perfect ($p = 1$) detectability. $E(N_1)$ specifies the expected number of available migrants in the first year. CV_N (coefficient of variation in the available population) specifies the square root of variance in the number of available migrants as a proportion of $E(N_y)$.

Species	CV_N	$E(N_1)$	years ($p < 1$)	years ($p = 1$)	difference
Sharp-shinned Hawk	0.37	2000	26	25	1
		450	25	24	1
		100	25	24	1
	0.26	2000	19	18	1
		450	19	17	2
		100	19	17	2
	0.18	2000	16	14	2
		450	16	13	3
		100	17	13	4
Northern Harrier	0.37	2000	25	24	1
		450	25	25	0
		100	25	24	1
	0.26	2000	19	18	1
		450	19	18	1
		100	19	18	1
	0.18	2000	15	14	1
		450	15	14	1
		100	15	14	1

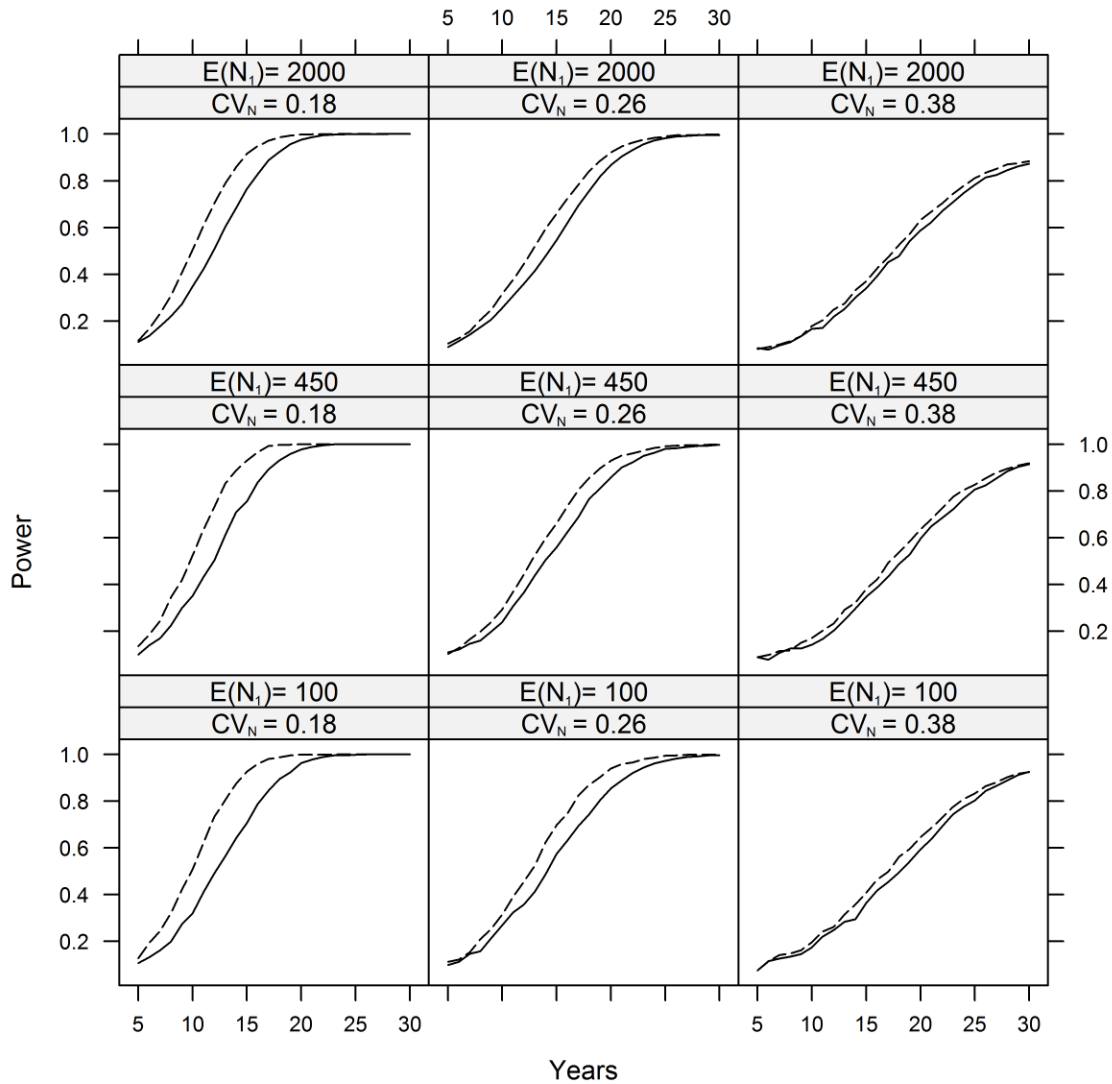


Figure 3. Sharp-shinned Hawk trend analysis simulation results. Estimated statistical power ($\alpha = 0.1$, two-tailed test) to detect a significant declining trend ($\lambda = 0.964$) by the number of years of study duration. Dashed lines depict power in simulations with detectability = 1. Solid lines depict power in simulations with heterogeneous detectability < 1. $E(N_1)$ is the expected available population in the first year, and CV_N is the square root of variance in the annual number of birds available as a proportion of N . The number of iterations was ≥ 1000 for each scenario.

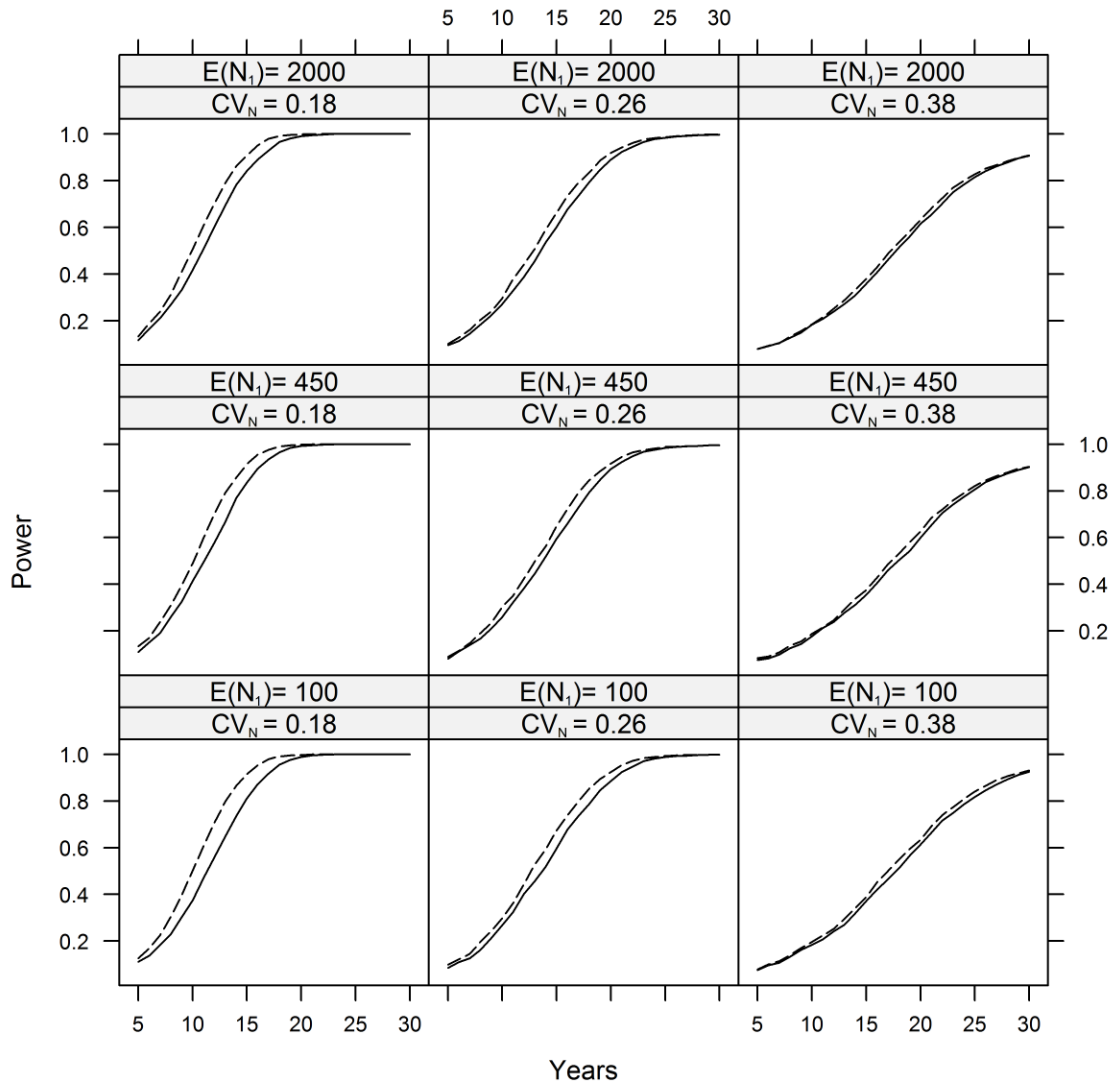


Figure 4. Northern Harrier trend analysis simulation results. Estimated statistical power ($\alpha=0.1$, two-tailed test) to detect a significant declining trend ($\lambda = 0.964$) by the number of years of study duration. Dashed lines depict power in simulations with detectability = 1. Solid lines depict power in simulations with heterogeneous detectability < 1. Parameters are: $E(N_1)$, the expected available population in the first year, and CV_N , the square root of variance in the annual number of raptors available as a proportion of N . Each simulated scenario was iterated 3000 times.

CONCLUSION

While the difficulties of valid inference in non-representative sampling and analysis of indices are easy to see (Ellingson and Lukacs 2003), there are good reasons why indices are still used for most long-term monitoring programs (Link and Sauer 1998, Johnson 2008), and why raptor migration counts have continued in spite of this criticism.

Raptor migration counts have a long history, and may potentially be continued for a long time, conditional on the sustained enthusiasm of a continually renewed corps of observers to collect data at low cost, and a consistent minimal level of funding (to maintain oversight, training, and coordination, thus ensuring relatively consistent, high-quality data). Long-term data sets are a rare commodity, and are very important for effective population monitoring.

While much is made of new analytic techniques that estimate and correct imperfect detection, distance sampling and double-observer methods were developed as a solution to a problem inherent in surveys of animals present at low densities, or that are likely to be present but unobserved (Nichols et al. 2000, Buckland et al. 2001). This problem is: How may we improve the likelihood of gathering enough information (from detections) to make reasonably precise estimates when our survey effort is constrained (Buckland et al. 2001)? Raptor migration counts solve this problem with *a priori* knowledge of the locations and timing of high animal density and detectability (Bildstein 2006). The results of Chapter 1 and those of previous studies (Sattler and Bart 1984, Berthiaume et al. 2009) confirm that detectability is generally high at raptor migration

watch-sites. The fact that detectability at raptor migration watch-sites is unusually high and not representative of the migrating population (Kerlinger et al. 1985) simply means that watch-site counts are likely an inaccurate representation of the migratory behavior of species. From a monitoring standpoint, the high detectability and density of migrating raptors at watch-sites are strengths of the method, not justification for abandoning it.

These two strengths of raptor migration counts for monitoring are partially offset by two corresponding weaknesses. First, the continuation of raptor migration counts for the long, continuous spans of time necessary to make valuable inferences (generally ≥ 20 years) frequently fails (Zalles and Bildstein 2000, HWI 2010). Efforts to build public enthusiasm for raptor migration counts are advisable, in the interest of recruiting volunteer and student labor, thereby keeping costs low.

Second, raptor migration counts often vary tremendously from day to day and year to year (Titus et al 1989, Lewis and Gould 2000). Variance in counts and statistically derived indices remains high at many watch-sites for many species, in spite of continued attempts to develop appropriate methods (Bildstein et al. 2008). This variation may not be the result of varying effort, but the result of variation in migration patterns, demographic rates, or detectability (Titus et al. 1989). The results of the simulations in Chapter 2 and those of Berthiaume et al. (2009) suggest that detection probability is not the primary source of this extraneous variation. However, the potential importance of observer effects is still uncertain, because I was not able to accurately assess the individual performance of observers, and the study's sample of observer teams may not have represented the distribution of all possible observer effects. Given these

results, I will suggest a few priorities for research into the improvement of count protocols and methods for analysis.

Although the duration of standardized watch-site counts is one of the method's greatest assets, fear of invalidating past data is not a satisfactory reason to forgo all opportunities to improve data quality now or in the future. If a permanent improvement in methodology reduces variance in annual counts, the value of older data is undiminished, and power of analyses will increase more rapidly. Should the mean annual count be affected, the older data will need to be transformed, resulting in a partial loss of information (Dunn et al. 2008). However, because recent trends are of greater conservation importance than older ones, and statistical power increases logarithmically with increasing study duration (e.g. Figures 3 and 4), the importance of the information-loss declines over time. Therefore, when managers can be reasonably certain a permanent change in protocol will substantially reduce variance in annual indices, the change should be made, provided such changes are made infrequently (Dunn et al. 2008).

The models of detectability developed in Chapter 1 show individual observers may have a considerable influence on counts. Estimating an effect size for each individual observer will probably not be feasible. An alternative approach is to tolerate a certain amount of variation in observer effects, but mitigate the effect of any single observer on trends by using more observers, rotated at intervals of days, not years (Dunn et al. 2008). Employing observers with particularly poor skills may skew the distribution of observer effects, reducing power (Chapter 2). Managers should attempt to confirm the basic competency of every observer.

Distance of the flight from the watch-site affects both detectability (at the visible scale) and spatial survey coverage (at the regional scale) (Johnson 2008). At present, the distance and direction of migrating raptors usually are estimated as central tendencies for the entire flight in each hour (HMANA 2009). These data provide no information on the dispersion or skew of the distribution of the migratory flight, and no information regarding differences in flight line between species. With this approach, the lack of an identifiable dominant flight line in an hour results in loss of information.

The alternative is to record distance data for individual raptors (Berthiaume et al. 2009, Chapter 1). In practice, recording distance estimations for individuals is not prohibitively difficult, even with dense flights, because a single datum can be recorded for a large group. In light of the advantages of individual level data for modeling and inference, I recommend discontinuing the use of a tabular tally form for raptor migration counts and adopting a form with a line for each individual raptor or homogenous group. Recording of hourly variables should continue to be implemented in a separate table.

At the regional scale, spatial variation in availability could potentially be differentiated from temporal variation by use of the extant public network of Doppler weather radar stations (Gauthreaux and Belser 2003, Gauthreaux et al. 2008, Van Gasteren et al. 2008, Buler and Diehl 2009) to estimate relative densities of the migratory flight across the landscape (Bildstein et al. 2008). Accurate models of spatial and temporal variation in migration according to regional weather patterns are likely to provide insights into the autocorrelation of daily counts. Accounting for autocorrelation may improve the power of analyses (Legendre et al. 2002).

Lastly, the role of demographic change on raptor migration counts is poorly understood and potentially problematic for the interpretation of trends. Raptor migration counts performed during spring migration can show more precise long-term trends than autumn counts in the same region (Farmer and Smith 2010), as expected when annual reproduction varies and over-winter survival is density dependent. Presently, the great majority of raptor migration counts are only performed in autumn. By increasing the number of spring counts, researchers may be able to gain a better understanding of demographic effects on index trends (Bildstein et al. 2008).

In finding that detectability is high at Lucky Peak, and failing to find any substantial bias in simulations of trend analysis, I believe my study supports the continuation of raptor migration counts, following the protocols of Dunn et al. (2008). The index calculation and trend analysis procedure currently in use (The Raptor Population Index (RPI); Farmer et al. 2007, Bildstein et al. 2008) accounts for varying survey effort, but not heterogeneous availability or detectability, by the use of ancillary data. This thesis is at least the third study to show that detectability at hawk migration counts may be predicted by such covariates as distance, altitude, and species (Sattler and Bart 1984, Berthiaume et al 2009, Chapter 1). Numerous other studies have shown that weather variables affect daily counts (Richardson 1978, Hall et al. 1992, Allen et al. 1996, Maransky et al. 1997, Yates et al. 2001, Miller et al. 2002, Panuccio et al. 2010, Miller et al. 2011). In the design of the present indices, the use of covariate-corrected annual totals was rejected out of hand in favor of using an adjusted annual mean of daily counts (Farmer et al. 2007). The RPI project ought to reconsider this decision. The enhanced performance of a mean as an index (compared to an annual total) is a result of

limiting the influence of uncommonly high and low-count days. Using covariates to correct daily counts for known effects would have a similar effect on variance, but would be superior in terms of conserving information. Days with very high counts provide more information about abundance than those days with very low counts, on account of the many potential factors unrelated to low abundance which may contribute to low counts (i.e. availability and detectability). Thus, while the use of a geometric mean as an index appeared advantageous from a standpoint of variance reduction at some sites (Farmer et al 2007), it performed poorly at some high volume watch-sites where the distribution of daily counts was especially skewed (Bildstein et al. 2008).

Literature Cited

- Allen, P. E., L. J. Goodrich, and K. L. Bildstein. 1996. Within- and among-year effects of cold fronts on migrating raptors at Hawk Mountain, Pennsylvania, 1934-1991. *Auk* 113:329-338.
- Berthiaume, E., 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.
- Bildstein, K. L. 2006. *Migrating raptors of the world: their ecology & conservation*. Comstock Pub. Associates, 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 Ornithologist's Union, Washington, D.C., USA.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling*. Oxford University Press, Oxford, United Kingdom.
- Buler, J. J., and R. H. Diehl. 2009. Quantifying bird density during migratory stopover using weather surveillance radar. *IEEE Transactions on Geoscience and Remote Sensing* 47:2741-2751.
- Dunn, E., 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.

- Ellingson, A. R., and P. M. Lukacs. 2003. Improving methods for regional landbird monitoring: a reply to Hutto and Young. *Wildlife Society Bulletin* 31:896-902.
- Farmer, C. J., D. J. T. Hussell, and D. Mizrahi. 2007. Detecting population trends in migratory birds of prey. *Auk* 124:1047-1062.
- Farmer, C. J., and J. P. Smith. 2010. Seasonal differences in migration counts of raptors: utility of spring counts for population monitoring. *Journal of Raptor Research* 44:101-112.
- Gauthreaux, S. A., and C. G. Belser. 2003. Radar ornithology and biological conservation. *Auk* 120:266-277.
- Gauthreaux, S. A., J. W. Livingston, and C. G. Belser. 2008. Detection and discrimination of fauna in the aerosphere using Doppler weather surveillance radar. *Integrative and Comparative Biology* 48:12-23.
- Hall, L. S., A. M. Fish, and M. L. Morrison. 1992. The influence of weather on hawk movements in coastal Northern California. *Wilson Bulletin* 104:447-461.
- HMANA. 2009. Forms. Hawk Migration Association of North America, Plymouth, New Hampshire, USA. [Online] http://www.hmana.org/data_entry_paper.php.
- HWI. 2010. Important message about migration site closures. Hawk Watch International, Salt Lake City, Utah, USA. [online] www.hawkwatch.org.
- Johnson, D. H. 2008. In defense of indices: the case of bird surveys. *Journal of Wildlife Management* 72:857-868.
- Kerlinger, P., V. P. Bingman, and K. P. Able. 1985. Comparative flight behavior of migrating hawks studied with tracking radar during autumn in central New York. *Canadian Journal of Zoology* 63:755-761.
- Legendre, P., M. R. T. Dale, M. J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25:601-615.
- 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.
- Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: Application to the North American Breeding Bird Survey. *Ecological Applications* 8:258-268.

- Maransky, B., L. Goodrich, and K. Bildstein. 1997. Seasonal shifts in the effects of weather on the visible migration of Red-tailed Hawks at Hawk Mountain, Pennsylvania, 1992-1994. *Wilson Bulletin* 109:246-252.
- Miller, M. W., E. M. Greenstone, W. Greenstone, and K. L. Bildstein. 2002. Timing and magnitude of Broad-winged hawk migration at Montclair Hawk Lookout, New Jersey, and Hawk Mountain Sanctuary, Pennsylvania. *Wilson Bulletin* 114:479-484.
- Miller, R. A., J. D. Carlisle, and G. S. Kaltenecker. 2011. Effects of regional cold fronts and localized weather phenomena on autumn migration of raptors and landbirds in southwest Idaho. *Condor* 113:274-283.
- 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.
- Panuccio, M., N. Agostini, G. Lucia, U. Mellone, S. Wilson, J. Ashton-Booth, G. Chiatante, and S. Todisco. 2010. Local weather conditions affect migration strategies of adult western honey buzzards *Pernis apivorus* through an isthmus area. *Zoological Studies* 49:651-656.
- Richardson, W. J. 1978. Timing and amount of bird migration in relation to weather - review. *Oikos* 30:224-272.
- Sattler, G., and J. Bart. 1984. Reliability of counts of migrating raptors: an experimental analysis. *Journal of Field Ornithology* 55:415-423.
- Titus, K., M. R. Fuller, and J. L. Ruos. 1989. Considerations for monitoring raptor population trends based on counts of migrants. *In* B.-U. Meyburg, and R. D. Chancellor, editors. *Raptors in the modern world : proceedings of the III World Conference on Birds of Prey and Owls*. International Council for Bird Preservation, Berlin, Germany.
- Van Gasteren, H., I. Holleman, W. Bouten, E. Van Loon, and J. Shamoun-Baranes. 2008. Extracting bird migration information from C-band Doppler weather radars. *Ibis* 150:674-686.
- Yates, R. E., B. R. McClelland, P. T. McClelland, C. H. Key, and R. E. Bennetts. 2001. The influence of weather on Golden Eagle migration in northwestern Montana. *Journal of Raptor Research* 35:81-90.
- Zalles, J. I., K. L. Bildstein. 2000. *Raptor watch: a global directory of raptor migration sites*. Birdlife International, Cambridge, United Kingdom.

APPENDIX

Detectability Models Used in Simulations

Detectability Models Used in Simulations

Observer Effect:

$$Obs \sim \mathcal{N}(\mu = 0.10241, \sigma = 0.44294)$$

Detectability of Sharp-shinned Hawks:

$$\begin{aligned} \text{logit}(\hat{p}_i) = & 0.671 + Obs_y + 0.197 dist_i - 0.093 dist_i^2 + 0.065 cloud_i - 0.011 wind_i \\ & + 0.009 day_i \end{aligned}$$

Detectability of Northern Harriers:

$$\begin{aligned} \text{logit}(\hat{p}_i) = & 1.118 + Obs_y + 0.197 dist_i - 0.093 dist_i^2 + 0.065 cloud_i - 0.011 wind_i \\ & + 0.009 day_i \end{aligned}$$

Where noted above, *dist* is the altitude/distance category (Table 1a, Table 2), *cloud* is the cloud cover category (Table 1b, Table 2), *wind* is the wind velocity in kilometers per hour (Table 2), and *day* is the number of days after August 24 (Table 2).