Evaluation of Wintering Waterfowl Population Aerial Survey Methodology

by

Stephanie Anne Braswell

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
December 14, 2024

Keywords: aerial survey, wintering waterfowl, detection probability,
count variability, availability, population

Copyright 2024 by Stephanie Anne Braswell

Approved by

Robert A. Gitzen, Co-Chair, Associate Professor of Wildlife Biology and Conservation
William D. Gulsby, Co-Chair, Associate Professor of Wildlife Management
Heath M. Hagy, Project Leader of United States Fish and Wildlife Service Habitat And Population Evaluation Team
Stephen S. Ditchkoff, Alumni & William R. & Fay Ireland Distinguished Professor of Wildlife Ecology, Management, and Nutrition

Abstract

Suitably designed wildlife population surveys account for methodological biases to produce defensible estimates. Although North American waterfowl populations have been aerially surveyed for nearly a century, accounting for biases remains challenging. We evaluated wintering waterfowl aerial population surveys conducted by the Alabama Wildlife and Freshwater Fisheries Division in the Tenneesee River Valley and Mobile coast, bay, and delta during November-February 2021–2024. We quantified detection probability and count variability within aerial surveys and used those estimates with an availability index to assess the relative influence of sources of error and survey design in aerial surveys. Detection probability was highest (0.71-0.95) for geese (*Branta* spp.) and lowest (0.24-0.70) for dabbling ducks (tribe *Anatini*) but was also influenced by seat position, observer experience, and group size. Detectability had lower influence on simulated estimates than availability or count variability. While assessing availability is challenging, managers can address counting variation by conducting replicate surveys.

Acknowledgements

I would like to begin by thanking my co-advisors Dr. Robert Gitzen and Dr. William Gulsby for your continued support throughout this project. Bob, thank you for the countless stats and analyses meetings, teaching us wildlife population analysis using populations of mythical creatures, and keeping me from spiraling further into insanity. Will, thank you for always providing great advice, keeping my tangents relevant, and for showing me how to ask the right questions. I thank you both greatly for taking me on as a graduate student and for being a balanced advisory team. I have enjoyed working with you, and hope you are able to continue teaching your students as much as you taught me.

Next, thank you to my committee members Dr. Heath Hagy and Dr. Steven Ditchkoff for your inputs which have only improved my research and thesis. Thank you both for always being available, for contributing positively, for your time, and for being expert leaders in our field. I thank you both for being dedicated to improving student's research and the field overall.

I would like to thank all of the people who aided in the data collection for this project, for without you it never would have been possible. I would like to especially thank Seth Maddox, Austin LeCroy, and Thomas Harms of ADCNR for acting as aerial observers in our surveys. It was a genuine pleasure to work with each of you. I would also like to thank Seth Maddox and Amy Silvano of ADCNR for their technical support and advice throughout the duration of this research. Also from ADCNR, I would like to thank Ziggy Jirik, for having a contagious love for your job and for being a great pilot who always kept us safe. Lastly, I would like to thank ADCNR overall for funding this research.

Thank you to my lab mates, office mates, and the rotating door of graduate students in the College of Forestry, Wildlife and Environment and beyond. For being a great support system, for

making me a better researcher, and for being great friends. I will truly miss our birding and fishing trips, dog park adventures, and office chats. I wish you endless success in leaving our field and resources better than you found them.

Most importantly, I would like to thank Jesus Christ and my family without whom this would not have been possible. To mama, daddy, Jonathan, and Hank: thank you for your unconditional love which you have given me my whole life, for being my biggest supporters, and for showing me my full potential, I cannot thank you enough for how amazing y'all are. To Ed, thank you for sharing my love of ducks, for making me a better person, for your endless encouragement, for being my best friend; thank you for everything always and all that you are, please never change.

Lastly, I would like to thank the ducks, geese, and other waterbirds for being the subjects of our surveys, for being extremely difficult to count, and for giving me purpose in the field of wildlife biology.

My only hope is that everyone mentioned here knows this list only scratches the surface of things for which I have to thank you for.

Table of Contents

Abstract 2

Acknowledgements 3

List of Tables 7

List of Figures 8

Chapter I (Incorporating Double-Observer Methods Into Wintering Waterfowl Aerial Surveys) ... 9

Introduction 9

Methods 11

Results 18

Discussion 21

Management Implications 28

Chapter II (Relative Importance of Detectability, Availability, and Counting Error in Wintering Waterfowl Aerial Population Surveys) 29

Introduction 29

Methods 32

Results 45

Discussion 50

Management Implications 58

Literature Cited 60

Appendix A (Condensed Results From Full Data Set Analyses) 93

Appendix B (Logistic Regression Results) 103

Appendix C (Supplementary Analyses of Subset Data Results) 105

Appendix D (Supplemental Error Metric Results) 114

Appendix E (Supplementary Availability Results) 115

List of Tables

Table 1 (Aerial Survey Dates).....71

Table 2 (Observer Seat Position and Recording Device Rotation Used in Aerial Surveys) 72

Table 3 (Model Comparison Results from AIC Analyses of Huggins Models).....73

Table 4 (Top Model Beta and Real Coefficients of Huggins Models).....74

Table 5 (Detection Estimates for Varying Combinations of Observer, Seat, and Guild).....75

Table 6 (Linear Model Coefficients of Error Metrics from Simulations).....77

Table 7 (R^2 Values of Error Metrics from Simulations) 78

Table 8 (Overall Average Rankings of Permutations) 79

List of Figures

Figure 1 (Map of Aerial Survey Locations and Flight Paths).....80

Figure 2 (Guild Level Aerial Survey Detectability) 81

Figure 3 (Observer Level Aerial Survey Detectability).....82

Figure 4 (Seat Level Aerial Survey Detectability).....83

Figure 5 (Overall Aerial Survey Detection Probability Estimates) 84

Figure 6 (Land Cover of Gadwall Tracking Sites).....85

Figure 7 (Observed Gadwall Locations) 87

Figure 8 (Overall Average Detection Probability Estimates).....88

Figure 9 (Variability in Coefficient of Variation Among Permutations).....89

Figure 10 (Variability in Percent Relative Bias Among Permutations) 90

Figure 11 (Simulation Flow Chart).....91

Chapter I – Incorporating Double-Observer Methods Into Wintering Waterfowl Aerial Surveys

Biologists have monitored North American waterfowl populations at landscape-level scales for nearly a century, longer than any other taxa (Prenzlow and Lovvorn 1997, USFWS 2004, Whitaker et al. 2019). Estimates of population size gathered from these surveys historically have been used to inform annual harvest regulations (Pearse et al. 2008*a*, Whitaker et al. 2019). However, as methodological shortcomings were discovered, it became apparent that standardized sampling approaches and unified goals were necessary to develop more robust population estimates (Sharp et al. 2002). Further, in the 1994 North American Waterfowl Management Plant (NAWMP) Update, one of the population recommendations stated that governmental agencies must maintain and improve current monitoring to measure the status of North American waterfowl populations (U.S. Department of the Interior et al. 1994). To sustain and bolster datasets from regional to local scales and meet NAWMP objectives, federal and state wildlife agencies conduct annual waterfowl population surveys of both wintering and breeding waterfowl populations (USFWS 2022). The resulting data are used to make informed decisions regarding harvest regulations and formulate management strategies (Pearse et al. 2008*a*, Whitaker et al. 2019). Additionally, monitoring provides insight into changes in population structure and dynamics (Reinecke et al. 1992, Smith et al. 1995), apprises biologists on migration routes and chronology, helps track disease outbreaks (Andersson et al. 2015), and indicates species-level distribution changes (Gilbert et al. 2020). Each of these metrics can be further analyzed in relation to time or environmental factors when population surveys are conducted using rigorous methodologies (Andersson et al. 2015).

State and federal agencies in the North American flyways typically conduct large-scale annual surveys during the wintering portion of the annual waterfowl cycle (Sharp et al. 2002).

Survey results are aggregated by flyway or other geographies to inform biologists of species composition, proportion, and distribution on non-breeding grounds (Sharp et al. 2002, Hagy et al. 2025). Outputs from population surveys appear in their simplest form as a rate, or number of species counted per unit area, which is also known as a count statistic (Nichols et al. 2000). When estimates of detection probabilities are also obtained during surveys, count statistics can be turned into estimates of abundance (Nichols et al. 2000). Although regulatory decisions are substantiated by population estimates and count statistics, this information is subject to limitations (Baldassarre and Bolen 2006). For example, imperfect detection probabilities, variable survey methods (Eggeman and Johnson 1989), and anthropogenically introduced biases (Ross 1985, Smith et al. 1995) are often sources of error in estimates of species presence and abundance (Pagano and Arnold 2009*a*), with many surveys also lacking precision estimates (Reinecke at al. 1992, Eggeman et al. 1997). However, methods incorporating correction factors and defining detection probabilities continue to improve, concurrent with development of novel field approaches (e.g., drone surveys; Chabot and Bird 2015, Clare et al. 2017).

One common approach to quantifying detection probability in aerial surveys is implementing double-observer methods. Double-observer methods have been successfully integrated into aerial waterfowl surveys to evaluate species-level detection probabilities and seat-position effects (Koneff et al. 2008) to improve the overall robustness of non-breeding waterfowl population abundance estimates in the Southeast (Masto 2019), and to generate correction factors and evaluate species-level detection (Gilbert et al. 2021). Data associated with this type of sampling (e.g., flock size, time of observation, etc.) can be used as covariates in double-observer models to better estimate detection probabilities and thus waterfowl abundance (Shirkey et al. 2014, Alisauskas and Conn 2019).

Historically, many state agencies, like the Alabama Wildlife and Freshwater Fisheries Division (AWFF), have relied on count statistics when monitoring non-breeding waterfowl populations. More recently, the AWFF and other state wildlife agencies are becoming increasingly interested in addressing methodological shortcomings of their aerial surveys, with the overall goal of increasing the robustness of their wintering waterfowl population estimates (Pearse et al. 2008*a*). Accordingly, we incorporated double-observer sampling methods into AWFF's wintering waterfowl population surveys to quantify detection probability. This work is intended to give AWFF and other entities flexible options for producing defensible waterfowl population abundance estimates with explicit measures of detectability. Incorporating robust population estimates into managerial and regulatory decision-making processes will ultimately improve understanding of waterfowl populations and sustainable conservation of waterfowl.

METHODS

Study Area

This study focused on two regions covered by AWFF's ongoing crewed aerial, non-breeding waterfowl surveys: the Tennessee River Valley (TRV) in north Alabama and the Mobile coast, bay, and delta (Mobile CBD) in south Alabama (Figure 1d.). The TRV surveys were split into eastern and western portions. Aerial surveys in the TRV east covered the Guntersville Reservoir and both banks of the Tennessee River north from the reservoir to the Alabama-Tennessee border (Figure 1a). Additional state lands connected to the river that were surveyed included Crow Creek Refuge, special opportunity area (SOA), and wildlife management area (WMA); Mud Creek dewatering unit (DWU) and WMA; North Sauty Refuge; and Raccoon Creek DWU and WMA (Figure 1c). The TRV west aerial survey covered Wheeler Reservoir west of U.S. Route 31, Wilson Reservoir, Pickwick Reservoir, Swan Creek DWU and WMA,

and Mallard Fox Creek WMA (Figure 1a,c). Mobile CBD surveys started at the Alabama Port and followed the western coastline (west of Alabama State Route 193) to the Mississippi-Alabama border, continued along the north shore of Dauphin Island, circled and bisected Gilliard Island from west to east, then flew 3.2 km (2-mile) spaced transects laid out east to west from Gilliard Island north to the south side of Interstate 10. Surveys were continued north of I-10 by following 1.6 km (1-mile) spaced transects laid out east to west and concluded at the industrial powerlines on the north side of Grand Bay (Figure 1b).

Within the TRV aerial survey area the landcover composition was ~82% open water, ~5% deciduous forest, ~3% woody wetlands, and 0.5% emergent herbaceous wetlands (Dewitz and USGS 2021). Additional land cover types that are beneficial for waterfowl species occurring in the TRV survey area include cropland, pasture/hay, and grassland/herbaceous but all were <2.5% coverage. All other cover types present occurred at <1.5% coverage (Dewitz and USGS 2021). Within the Mobile CBD aerial survey area the landcover composition was ~81% open water, ~10% emergent herbaceous wetlands, and 5.9% woody wetlands. All other cover types present, including those beneficial for waterfowl, occurred at ≤1% coverage (Dewitz and USGS 2021).

Survey Protocols

Aerial surveys were conducted in a Cessna 182T Skylane fixed-wing aircraft during designated time periods from November-February 2021–2024 and were flown in sets. Each set of surveys was defined spatially and temporally. Spatially, a survey set consisted of 3 surveys, 1 in each of the survey areas (Mobile CBD, TRV East, and TRV West) conducted as close in time as weather and scheduling permitted. Temporally, 4 sets of surveys were conducted each winter, 1 in the middle of November ("Preseason"), 1 in the middle of December ("December"), 1 in the first week of January ("Midwinter"), and 1 in the last week of January ("Late Season"), as

weather allowed (Table 1). Flight speed and altitude varied due to terrain, weather conditions, and landscape obstacles (e.g., power lines, cell phone towers, etc.), but were flown as close to 61-76 m (200-250 ft) above ground level and 148-185 kph (92-115 mph) as possible. All surveys were flown by the same pilot.

Observers sat on the right-hand side of the plane for all surveys and only recorded observations seen out of the right-hand side of the plane. The pilot did not act as an aerial observer. Water level (ft), surface temperature (°F), and start and end times were recorded for surveys. Taxa of interest included all waterfowl species (*Anatidae*), American coots (*Fulica americana*), sandhill cranes (*Grus canadensis*), double-crested cormorants (*Phalacrocorax auritus*), American white pelicans (*Pelecanus erythrorhynchos*), and bald eagles (*Haliaeetus leucocephalus*). All taxa of interest observed during the survey were identified to species level and enumerated following USFWS guidelines as published by Bowman (2014).

We recorded observations using multiple devices and platforms: a SONY ICD-B300 IC Recorder, a WearHome V1.0 voice recorder, the Hodges H2 recording device with the Scribe voice recording software (USFWS and SWYM, LLC.; Scribe version 3.0.11), and the Hodges H1 USFWS-GPS-voice recording software (J. Hodges, USFWS). Additionally, five total surveyors acted as observers throughout the duration of this project. Observers rotated seat positions in the plane in the second and third winters to test seat position as a covariate within our survey system and to account for observer experience differences (Pagano and Arnold 2009*b*). The observer sitting in the front seat was always listed as the primary observer and during the preseason and midwinter flights was always the most experienced surveyor (Table 2).

During all aerial surveys, two observers independently recorded observations during flight while following the general survey protocol above (Alisauskas and Conn 2019). Time

stamps of each observation were made to the minute during aerial surveys. The timing devices used by surveyors were synchronized at the beginning of each survey. The seat positioning of both observers during each survey was also noted. Observers did not speak to or provide visual cues to each other to maximize the independence of observations (Vrtiska and Powell 2011). We did not hang a curtain (e.g., visibility barrier) between observers within the plane cabin which may have allowed the back seat observer to accidentally take on cues from the front seat observer. However, we felt that this occurrence was rare as the back seat observer maintained their focus directly out of the right-hand window within the field of view for the surveys. Additionally, observers were outfitted with helmets which limited their peripheral sight. We feel that this further decreased the chances of either observer accidently taking on cues from the other observer.

Data Processing and Analyses

All double-observer aerial observations were paired by time, species recorded, and estimated group size. Observations of the same or biologically similar species (e.g., Goldeneye [*Bucephala clangula*] and Bufflehead [*B. albeola*]) within the same minute bins were paired as long as they were the only observed group of that species designation within that minute bin or, in the case of multiple groups of the same species designation, if it was obvious that those groups could be paired between observers. Paired groups that were detected by both observers received a "1,1" encounter history to be used in analyses, indicating that each observer detected that group. Each group represented one recorded observation or detection. If one observer detected a group(s) in minute bins where the other observer did not, those observations could be definitively designated a "1,0" if that observation was made by the front seat observer and a "0,1" if made by the back seat observer. Examples of observations that could not be paired

include instances where groups were not well defined (e.g., refuges with large congregations of birds), or where there were multiple (~<3) groups of the same species within a 1-minute bin. A data subset was created from the paired double-observer data by including pairings from species where >67% of detections of that species within 1 survey could be paired. Additionally, we conducted a supplementary analysis using the fully paired data set where all paired detections were included (Appendix A).

We ran closed-population mark-recapture Huggins models in the RMark package (Laake 2013) in R statistical software (v4.4.1, R Core Team 2021) to assess detectability levels within our surveys, using the subset of data with all detection pairings >67% as explained above. The model set included all possible additive combinations of the following covariates: seat position (e.g., front or back), season (e.g., Preseason, December, Midwinter, Late Season), observer, log-transformed observed group size, and either a guild or fine-scale taxonomic grouping variable.

The observer variable was a grouped variable where all 5 participating observers fell into 1 of three categories based on when they participated and experience level. Observer group A was the most experienced observer with >10 years' experience and participated in the first two years of the study. Observer group B included observers with 1-3 years of aerial waterfowl survey experience and were those observers who participated in all three years of surveys or filled in for observers who participated in all three years of surveys. Observer group C was made up of observers with the least amount of experience or who had not participated in aerial surveys for 3+ years and only participated during year three.

Although observed groups were identified to the species level, for analyses we examined two levels of taxonomic grouping, combining species across guild or genera ("fine-scale") to obtain sufficient sample sizes for using either level as detectability covariates. The guild-level

variable grouped observed species into one of four guilds: dabbling ducks (tribe *Anatini*), diving ducks (*Aythya*, and *Bucephala*), geese (*Branta*), and non-waterfowl (*Fulica*, *Grus*, *Haliaeetus*, *Pelecanus*, and *Phalacrocorax*). The fine-scale grouping variable is slightly more specific than the guild level but not quite as specific as a genus level grouping variable. We ran into issues of small sample sizes when we attempted to use a genus-level taxonomic grouping variable; therefore, we combined some genera together to create the fine-scale grouping. The fine-scale variable grouped observed species into the following groups: *Anas*, Other dabbling ducks (*Aix*, *Mareca*, and *Spatula*), *Aythya*, *Bucephala*, *Branta* + *Grus*, *Fulica*, *Haliaeetus*, *Pelecanus*, and *Phalacrocorax*. Any genus not listed in either the guild or fine-scale group was not observed during our aerial surveys or was removed from analysis due to low sample size ($n < 3$ observations) and non-convergence in models.

The model set included all additive combinations of the covariates listed above and every model included the seat covariate. The guild and fine-scale covariates never occurred in the same model. Additionally, 1 model including an interaction between seat and observer was considered. A top model and competing model set was identified using Akaike Information Criterion (AIC; Akaike 1974, Arnold 2010). Results from the covariations within the top model are presented below (Table 4) and covariates appearing in competing models but not the top model were analyzed supplementarily (Appendix C). Final detection probabilities and plots were generated using the RMark (Laake 2013) and ggplot (v3.5.1; Wickham 2016) packages in R.

In addition to those models listed above, we analyzed a set of negative binomial regressions using the glmmTMB function in the TMB (v1.9.14; Kristensen et al. 2016) package in R. These negative binomial models were supplementary to investigate the relationship between survey factors (e.g., observers, seat position, etc.) and observed group size. As our

detection models produced estimates of group-level detection, we wanted to ensure that estimates of group size were not influenced by survey factors as this might indicate that individual detection within a group is different than our group-level detection. For example, if front seat observations were consistently and significantly higher than back-seat observations, that might indicate that within group detection is influenced by seat position. Negative binomials modeled group size by the same covariates listed above, plus time of observation in decimal hours, survey location, a species taxonomic level, and by individual survey. All single covariate models, two covariate additive models where seat was one of those covariates, and a full model were considered for AIC analyses. We simulated residuals and conducted residual-based goodness of fit tests (Dharma package v0.4.6; Hartig 2022) to assess the fit of the top and full models.

Lastly, we calculated overall detection probability estimates from aerial surveys using the equation:

$$1 - ((1 - \mathrm{p1})(1 - \mathrm{p2}))$$

where p1 is the estimated detection for the primary observer in the front seat, and p2 is the estimated detection for the secondary observer in the back seat (Nichols et al. 2000). We calculated overall detection probability estimates representing surveys where both observers are equal in experience, but as one observer cannot occupy both seats simultaneously, these estimates are hypothetical. We did so by using the equation:

$$1 - ((1 - \mathrm{p1})\text{^}2)$$

where p1 represents the estimated detection for the same observer occupying both seats during double-observer sampling methods as detection probability cannot be estimated using single-observer methods (Nichols et al. 2000).

RESULTS

During November – February 2021-2024, 31 aerial surveys were conducted using double-observer methods. There were 5,734 group level detections made by 5 observers in all study years and 1,441 were definitively paired observations. There were 25 total species observed in aerial surveys aggregated in analyses across 4 guilds: 10 dabbling duck species, 8 diving duck species, 2 goose species, and 5 non-waterfowl species. The most frequently observed species in each of the above groups were gadwall (*Mereca strepera*), bufflehead, Canada geese (*Branta canadensis*), and American coot, respectively, for both the front and back seat. The least frequently observed species in each guild were wood ducks (*Aix sponsa*, n = 4), hooded merganser (*Lophodytes cuculatus*, n = 2), ruddy ducks (*Oxyura jamaicensis*, n = 2), snow geese (*Answer caerulescens,* n = 1), and bald eagles respectively. All 25 observed species were detected from the front seat, but back seat observers never detected redhead (*Aythya americana*), ruddy duck, or snow geese. Of all groups used for analyses, 90% were flocks of 100 birds or fewer in size. The smallest flock size observed was 1 individual (multiple species and occurrences), and the largest flock of birds used in analyses was 1,300 ring-necked ducks (*Aythya collaris*).

A guild-level grouping variable, seat position, a grouped observer variable, and a log-transformed observed group size variable affected group-level detectability (Table 3). Further, a season variable and a within-guild fine-scale grouping variable affected group-level detectability (Appendix C).

We estimated detection probability estimates for four guilds within our aerial surveys: dabbling ducks, diving ducks, geese, and non-waterfowl. When observer and seat were held constant, groups of geese were estimated to have the highest detection probability of all guilds

starting at 0.46 (CI [0.29, 0.64]) for individuals and reaching 0.79 (CI [0.64, 0.89]) for flocks of 100 birds in size. Flocks of geese were 2.35 (CI [1.14, 4.86]) times as likely to be detected compared to flocks of non-waterfowl. Non-waterfowl flocks of size 500 and 100 birds were estimated to have detection probability estimates of 0.72 (CI [0.56, 0.84]) and 0.61 (CI [0.48, 0.73]) respectively, whereas individual birds within the non-waterfowl guild were estimated to have detection probabilities of 0.27 (CI [0.18, 0.37]). Non-waterfowl flocks were 1.74 (CI [0.92, 3.27]) times as likely to be detected compared to flocks of diving ducks, where individuals within the diving duck guild have an estimated detection probability of 0.17 (CI [0.09, 0.30]). Detectability estimates for diving ducks reached 0.48 (CI [0.33, 0.62]) at a flock size of 100 birds and 0.60 (CI [0.43, 0.75]) at a flock size of 500 birds. Dabbling ducks were the guild with the lowest modeled detection probability estimates. Diving ducks and geese were 1.84 (CI [0.77, 4.44]) and 7.54 (CI [2.84, 20.03]) times as likely to be detected compared to dabbling ducks, respectively. Lone dabbling ducks were estimated to have a detection probability of 0.10 (CI [0.04, 0.22]) but detection of flocks of 100 birds reached 0.33 (CI [0.19, 0.51]). We observed that detection probability estimates increase at a much slower rate after flock size increased over 100 birds, but maximum modeled detection probability of dabbling ducks at flock size 350 was estimated at 0.42 (CI [0.25, 0.62]) (Figure 2).

Observer group A (ObsA, most experienced group) had the highest detectability estimates in both seats, and across all guilds and flock sizes (Table 5). With other covariates held constant (guild = dabbling ducks, seat = front), groups were 0.35 (CI [0.21, 0.59]) and 0.31 (CI [0.18, 0.56]) times as likely to be observed by ObsB (middle experience) or ObsC (least experience) than by ObsA, respectively, and detection was 0.58 (CI [0.43, 0.77]) times as likely for ObsC than ObsB (Figure 3).

Group detection was 0.70 (CI [0.54, 0.91]) times as likely in the back seat compared to the front seat when observer and guild were held constant. However, this changed among guilds and observers. Detection differences between the front and back seat were smallest when flocks were identified as being in the goose guild. ObsA had the least amount of difference in detection probability between front and back seat observations. Detection differences between the front and back seat were greatest when flocks were identified as being in the diving duck guild. ObsC had the greatest amount of difference in detection probability between front and back seat observations. At constant observer (ObsA) and guild (dabbling ducks), front and back seat detection probability estimates were 0. 24 (CI [0.11, 0.45]) and 0.18 (CI [0.08, 0.38]) for individual birds in each seat respectively. These estimates increased to 0.58 (CI [0.37, 0.76]) and 0.49 (CI [0.28, 0.71]) for flocks 100 birds in size, and up to 0.67 (CI [0.46, 0.83]) and 0.59 (CI [0.36, 0.79]) for flocks 350 birds in size (Table 5, Figure 4).

We ran a suite of negative binomial models to investigate the influence of factors such as seat position on within-group detectability of birds. We found no significant difference between seat positions in log group size counts ($P = 0.55$). Additionally, no significant relationships ($P > 0.14$) existed between log group size counts and bird guild. Lastly, there were no significant relationships observed between log group size counts and observer ($P > 0.20$). Residual indicated that the top model (group size as a function of fine-scale group) showed insignificant deviation ($P = 0.83$), dispersion ($P = 0.54$), and outliers (0.84).

Overall detectability estimates were highest with ObsA in the front seat and ObsB in the back seat. With that observer combination and at flock size 100, detection probability estimates were 0.69 for dabbling ducks, 0.83 for diving ducks, 0.98 for geese, and 0.91 for non-waterfowl. The observer combination with the lowest overall detection probability estimates were ObsC in

the front seat and ObsB in the back seat. For that observer combination and at flock size 100, detectability estimates were 0.48 for dabbling ducks, 0.66 for diving ducks, 0.93 for geese, and 0.80 for non-waterfowl (Figure 5).

DISCUSSION

We incorporated double-observer methods into wintering waterfowl population aerial surveys to quantify detection probabilities. Our detectability estimates were lower than expected for duck guilds, especially dabbling ducks. Gilbert et al. (2021) reported dabbling and diving duck detection probabilities in their waterfowl fixed-wing aircraft surveys to be 0.92 and 0.84, respectively. Our detection estimates appear to be lower for dabbling ducks at 0.69 but similar for diving ducks at 0.83; however, they calculated full visibility bias considering count bias as well as detection probability. Additionally, we employed multiple observers with various experience levels which may contribute to lower estimates overall, as observer effects have been found to influence detection (Erwin 1982, Cox et al. 2022). We also may have observed lower detection estimates due to differences in survey design (Pearse 2007), available habitat (Davis et al. 2014), or species compositions (Sharp et al. 2002). Further, average detection probably estimates for individuals across guilds in our study were ~0.10 lower than detection probability estimates of decoys observed during double-observer aerial surveys over open wetlands in Mississippi (Pearse et al. 2008*b*). Maximum detection probability estimates from that same study began leveling off around a flock size of 30 birds in open water, whereas our estimates began leveling off between 50-100 birds across all surveyed areas. It was especially difficult to detect small groups of birds and specifically ducks, in areas of dense, emergent vegetation within our study which may contribute to detection leveling off at larger group sizes compared to Pearse et al. (2008*b*). Additionally, Koneff et al. (2008) reported much higher detection probabilities of

black ducks (*Anas rubripes*) and mallards (*Anas platyrhynchos*) in double-observer breeding waterfowl aerial surveys than we achieved in our non-breeding surveys.

In our study, the most experienced observer in the front seat was estimated to have a detection probability of 0.24 for individual dabbling ducks, whereas the front seat observer from Koneff et al. (2008) was estimated to have a 0.74 detection probability of mallards and a 0.61 detection probability of black ducks of group sizes less than 10 birds. As their work investigated breeding waterfowl population survey methodology, we suspect that duck behavior during the non-breeding season and a lack of well-defined group boundaries may have contributed to our lower detection probability estimates. In our surveys, ducks frequently appeared in large congregations (e.g., on refuges, rafted in reservoirs or large open water), which makes grouping and pairing observations between observers extremely difficult (Martinson and Kaczynski 1967, Musgrove and Holloway 1997). Pairing observations was further complicated as aerial observers often aggregate flocks within voice recordings with aggregation increasing with greater bird abundance. Additionally, the most frequently occurring dabbling duck species in our surveys was gadwall (Henry 2012). It was difficult to pair data between observers for such commonly occurring species; therefore, many gadwall observations were censored from analyses. This contributed to lower detection estimates as some observations censored from analyses were likely true double-detections as shown in the full subset analysis (Appendix A) where dabbling duck detectability was higher than in the subset analysis. The same trend is true for American coots, the most frequently occurring species in our aerial survey. American coot had 0.08 higher detection from full analyses (Appendix A) than from subset analyses (Appendix C) indicating that adequate pairing may not have been achieved in our surveys.

It is not surprising that detection probability was greatest for the goose guild, as this guild encompasses large-bodied birds that are easily recognized in most of the cover types we surveyed. This finding was consistent with Fletcher and Hutto (2006), who also reported high detection probability for Canada geese compared to other species of interest. Additionally, they reported that body size and detection probability were positively correlated, which matches the pattern of detectability among the guilds in our study (Fletcher and Hutto 2006).

Our results also were consistent with our expectation of high detection probability for the non-waterfowl guild, as this group is also principally made up of large-bodied, easily distinguishable species. For example, in crewed helicopter surveys, the only species with greater detection than sandhill cranes were Canada geese, tundra swans (*Cygnus columbianus*), and northern pintail (*Anas acuta*; Alisauskas and Conn 2019). Further, our non-waterfowl detection estimates were similar to those reported in other non-breeding, fixed-wing aerial surveys. Gilbert et al. (2021) reported species-level detection probability estimates of American white pelicans to be 0.67, double-crested cormorants to be 0.27 and American coots to be 0.83. Our group-level detection estimates for the non-waterfowl guild ranged from 0.27 for individuals up to 0.72 for large groups. Of the species which encompass our non-waterfowl guild, double-crested cormorants were most likely to be detected as individuals (anecdotal observation; Appendix A), and our detectability estimate for this species was very similar to that of Gilbert et al. (2021).

Beyond guild-level detectability differences, we expected to observe a difference in detection probability estimates among observer groups, especially given that observers in our surveys had differing levels of experience (Erwin 1982, Pagano and Arnold 2009*b*). The observer group with the highest detectability estimates (ObsA) was our most experienced observer, a result similar to previous research (Roy et al. 2022). The observer group (ObsC) with the lowest

detectability estimates was considered to be our least experienced observer group, again similar to previous research (Sharp et al. 2002). The intermediate observer group's (ObsB) detection probability estimates fell in between the other two. Fortunately, our negative binomial models showed that differences in observed group sizes between different observers were statistically insignificant, so greater detection probability estimates for ObsA did not result from overcounting individuals within groups.

Our results from seat position analyses are also consistent with most literature, as many studies have reported higher detection probabilities of waterfowl and waterbirds in the front seats compared to back seats (Alisauskas and Conn 2019, Conroy et al. 2008, Koneff et al. 2008). However, Roy et al. (2022) reported breeding waterfowl helicopter survey detection estimates were higher in the back seat due to what they believe was extra tasks assigned to front seat observers. In our surveys, front seat observers had a larger field of view than back seat observers and were not assigned additional in-flight tasks. Therefore, it is not surprising that detection estimates were higher for our front seat observers. Additionally, within our surveys, two of the three species that were observed from the front seat but not from the back were the least frequently observed species in their respective guilds (i.e., ruddy ducks for diving ducks and snow geese for geese). We suspect that these groups went undetected by the back seat observer because they were so infrequently encountered.

Our negative binomial models suggested that once a group is detected, individual detectability within that group does not differ between guild, observer, or seat position. This indicates that group size estimation and count error are consistent among observer groups in our study. Consistency between observers was also found by Pearse et al. (2008*b*). As observers are consistent in their flock size estimation, current observer training protocols and aerial survey

guidelines do not need to be adjusted within our system. This is consistent with the suggestions of Eggeman and Johnson (1989).

Double-observer methods allow us to model both individual and overall detection probability estimates. Per-observer detectability estimates obtained from double-observer methods serve to estimate detection probabilities that would be expected from a single observer. By combining both individual estimates from double-observer surveys, we calculated an overall detection probability estimate using the equation published by Nichols et al. (2000). This overall estimate was greater than any individual detection probability estimate, even those from our most experienced observer. We saw an 0.11, 0.09, 0.07, and 0.09 increase in overall detection in each guild (e.g., geese, non-waterfowl, diving ducks, and dabbling ducks, respectively) from detection estimates of a survey if the most experienced observer was the only participant. Further, we saw an 0.18, 0.21, 0.16, and 0.22 increase in overall detection in each guild (e.g., geese, non-waterfowl, diving ducks, and dabbling ducks, respectively) from detection estimates of a survey if the least experienced observer was the only participant. Additionally, our overall detection probability estimates may provide beneficial information to agencies by showing relative changes in guild-level detection probabilities between different combinations of observers and seat positions. Managers should consider what information they are aiming to gather from their surveys and what resources are available to them when deciding which observers to send on surveys. When considering who to employ as observers within double-observer surveys, managers should keep in mind that minimally experienced individuals can limit overall detection, and substantially experienced observers can maximize overall detection. Managers conducting aerial waterfowl surveys would ideally employ two highly experienced observers; however, that is not always feasible due to limited survey resources (Eggeman and Johnson

1989, Musgrove and Holloway 1997). When two highly experienced observers are not available for surveying, agencies still benefit from utilizing the double-observer method because they are able to estimate detection probabilities using multiple observers and generate correction factors within their survey systems.

We emphasize that we are not generalizing our specific modeled overall detection probabilities to other surveys but stating that the differences in detectability of varying observer combinations provide insight as to what detection levels may be acceptable to agencies based on observer availability and associated experience. Nonetheless, our overall double-observer detectability estimates overlap those published by Fletcher and Hutto (2006), who reported an overall double-observer detection probability of 0.88 across all species of interest, compared to our estimates which ranged from 0.69 to 0.98, depending on guild. More specifically, Fletcher and Hutto (2006) estimated overall double-observer detection probability of non-waterfowl species to be 0.87-0.99, Canada geese to be 0.98, and all ducks to be 0.91. These estimates are only higher than our overall double-observer estimates for duck guilds.

We conducted analyses on a subset of the full data set because we did not want to artificially inflate detection by using observations that were not definitively paired. The most frequent issue with matching observations was when multiple groups of the same species were observed in the same minute bin. This typically occurred with the most common species (e.g., American coots, gadwall, etc.). Larger flocks were also more difficult to pair than smaller flocks (Appendix B). Another issue that arose mostly on areas where dense or large congregations of waterfowl occurred (e.g., refuges) was not having clear definitions of what constituted a group or flock. Pairing of observations occurred post-aerial survey as observations between observers were made independently during the survey. Occasionally, during post-survey observation

pairing, it was obvious that both observers detected a group of birds but identified them as different species. This was not a frequent occurrence and when it did occur, the disagreement was not substantial and rarely affected analyses at the guild or fine-scale group level. For example, one observer identified the group as bufflehead and the other goldeneye, or one observer identified the group as scaup (*Aythya spp.*) and the other ring-necked ducks.

Future research conducted should take additional measures to ensure observations are pairable as detection estimates are influenced by the achieved level of observation pairing within surveys (Appendices A and C). Additionally, future research should further investigate the relative importance of detectability versus counting bias within a system and test these estimates as correction factors for surveys. Lastly, we recognize that creating definitive protocols for defining flocks and flock boundaries within areas of high waterfowl concentrations would be beneficial to increasing the number of paired observations, but is impractical to implement during surveys. One option to address this issue is to define discernable units within aerial surveys to aggregate waterfowl within. Bolstering methods to include these discernable units and aggregating all waterfowl within units would remove the need to define individual flocks (Hagy et al. 2024). Further, these methods would allow agencies to examine count bias (an observer's tendency to under- or overcount group sizes) within their systems. We did not assess count bias here; however, with both defined detection probability estimates and count bias, agencies can achieve full population estimates for their surveyed populations further increasing the robustness of their estimates (Gilbert et al. 2021).

Implementing double-observer strategies begins to address the issue of imperfect detection probability within any aerial survey (Pearse et al. 2008*b*). Addressing issues associated with aerial surveys results in better data and understanding of waterfowl populations and their

dynamics (Reinecke et al. 1992, Smith et al. 1995), aids in the fulfilment of NAWMP goals (U.S. Department of the Interior et al. 2018), and directly affects stakeholders as harvest regulations are informed by aerial survey population estimates (Baldassarre and Bolen 2006, Whitaker et al. 2019). Based on our results, incorporating double-observer methods can help agencies quantify detection probabilities, but without definitive pairing of observations the utility of the method to provide defensible population estimates during winter is limited.

MANAGEMENT IMPLICATIONS

Agencies interested in assessing detection probability within their wintering waterfowl population aerial surveys should consider utilizing double-observer methods as we and others have demonstrated. It is crucial that state wildlife agencies investigate the relationships of detection probability and covariates within aerial survey systems because it allows them to quantify what proportion of the population is being accounted for within aerial surveys. For AWFF, we showed that whole populations are not accounted for by single-observer strategies because detectability is less than 100%. Other agencies interested in evaluating detection probability within their aerial surveys should ensure that observations can be paired. In addition to quantifying detection probability estimates within aerial surveys, agencies should aim to quantify count bias. In combination, estimates of detection and count bias allow agencies to quantify fully robust population estimates. After implementing double-observer methods and quantifying detection using paired observations, agencies can use detection probability and count bias estimates to define correction factors and confidence intervals for their population estimates.

Chapter II – Relative Importance of Detectability, Availability, and Count Variability in Aerial Surveys for Wintering Waterfowl

The objectives of the North American Waterfowl Management Plan (NAWMP) are to conserve waterfowl (*Anatidae*) and their habitat at the continental scale (U.S. Department of the Interior and Environment Canada 1986). One of the population recommendations listed in the NAWMP is to standardize and improve the accuracy of breeding population surveys at the flyway and continental scales (U.S. Department of the Interior and Environment Canada 1986). Further, in the 1994 NAWMP Update, one of the population recommendations stated that governmental agencies must maintain and improve current monitoring to measure the status of North American waterfowl populations (U.S. Department of the Interior et al. 1994). These recommendations are typically implemented at the landscape scale using aerial surveys (USFWS and CWS 1987, Koneff et al. 2008). Biologists use outputs from aerial surveys (e.g., abundance estimates) to help set harvest regulations (Pearse et al. 2008*a*), formulate management strategies (Whitaker et al. 2019), determine priority areas for conservation (Soulliere et al. 2013), track migration chronology (Andersson et al. 2015), assess changes in population structure (Reinecke et al. 1992, Smith et al. 1995), and indicate species-level distribution changes (Gilbert et al. 2020). Given the many uses of these data, they are extremely valuable to waterfowl biologists. However, challenges in the methods used to collect these data may prevent survey outputs from fulfilling NAWMP goals if the methods do not produce defensible estimates for all metrics of interest (Reinecke at al. 1992, Eggeman et al. 1997).

Imperfect detection probabilities are one source of potential bias in aerial waterfowl surveys (Eggeman and Johnson 1989). Detection probability (or 'detectability') is the probability that a bird or group of birds is seen and counted by an observer. Failure of surveyors to detect

birds or flocks is common across studies (e.g., flocks or individuals; Shirkey et al. 2014, Alisauskas and Conn 2019), and flock size is typically underestimated (Pollock and Kendall 1987, Conroy et al. 2008, Davis et al. 2022). Additionally, variability in group size estimates among observers increases with flock size, decreasing accuracy of population counts (Gilbert et al. 2021). Spatial and temporal variability in survey coverage may also prevent biologists from detecting changes in population trends over time (Eggeman and Johnson 1989).

Fortunately, corrective methodology can reduce bias and error in aerial waterfowl surveys (Pearse et al. 2008*b,* Cox et al. 2022). Accordingly, previous research has focused on assessing visibility bias within aerial surveys where visibility bias represents the combined effects of detection probability and count error (Gilbert et al. 2021). In our study, we define detection probability as stated above, and count error as the combined effects of count bias and count variability. Count variability is the measure of variability in group size estimates among observers which varies slightly from count bias referring to an observer's tendency to over- or undercount group sizes. Specifically, methods incorporated into aerial surveys to increase detection probability and account for count variability include distance sampling, double-observer methods, or aerial photography (Koneff et al. 2008, Johnson et al. 1989, Bayliss and Yeomans 1990, etc.). However, some estimates corrected for imperfect detection and count variability still do not consider availability of species of interest (Prenzlow and Lovvorn 1996, Eggeman and Johnson 1989). When availability, or the likelihood that a bird in the target population is within the survey area during the time of the survey, remains undefined, the proportion of the population represented by corrected estimates remains unknown (Diefenbach et al. 2007, O'Donnell et al. 2015). Few studies have investigated the effect of availability on abundance estimates because they use probabilistic sampling frameworks encompassing the

entire area being used by the surveyed population (Pearse et al. 2008*a*) removing the need to quantify availability within survey areas. However, for established surveys that do not employ probabilistic sampling frameworks, in which some birds may be outside the survey area during any single survey, quantifying availability offers mangers an option to obtain full region population estimates through the use of correction factors for availability.

Beyond quantifying sources of error in aerial surveys, few studies have examined the relative importance of various sources of error on abundance estimates (Cox et al. 2022), or waterfowl managers' ability to detect changes in population trends through time. Additionally, although many researchers have implemented corrective methodology to address bias and variability, information on the relationship between survey effort and the resulting accuracy and precision of abundance estimates is lacking (Gilbert et al. 2021, Bayliss and Yeomans 1990, Cordts et al. 2002). Such information would inform agencies of how they can best allocate their available survey resources (e.g., observers, time, flights, etc.) to produce defensible abundance estimates (Kumar and Rice 2021) or detect changes in population trends over time.

State wildlife agencies and managers are becoming increasingly interested in evaluating and improving their wintering waterfowl population aerial surveys (Pearse et al. 2008*a*). Therefore, we implemented a study to further inform the relative influence of sources of error in aerial waterfowl surveys and ways to address those sources of error while optimizing survey options available to managers. Specifically, we quantified detection probability, count variability, and availability for aerial surveys during the non-breeding season in Alabama. We used these estimates to inform simulations assessing the impact of each source of error and determine plausible alternative survey approaches within our study system. Agencies can use this information to better understand the relative influence of each source of error within their survey

systems and how they affect their ability to assess temporal trends in abundance. As a result, surveys can be designed to optimize the relationship between survey effort and accuracy and precision of waterfowl abundance estimates.

METHODS

Overview

We calculated availability, detectability, and count variability as sources of error within aerial surveys in Alabama. Availability was calculated using an unpublished data set from a previous study evaluating gadwall (*Mareca strepera*) movement ecology in the Tennessee River Valley (TRV). Detectability and count variability were calculated from aerial surveys conducted in the TRV and in the Mobile coast, bay, and delta (Mobile CBD). We describe where and how the data were collected to calculate these sources of error in the Study Area and Field Data Collection sections below. Further, we describe specifically how each source of error was calculated in the Sources of Error section below.

We then used the results from calculating our aerial survey sources of error to inform simulations evaluating the relative influence of each source of error on wintering waterfowl population aerial survey estimates and manager's ability to detect changes in population trends over time (Figure 11). Details describing how simulations were conducted are in the Survey Evaluations section below. Finally, concluding the Survey Evaluations section are details describing specific error metrics we used to compare simulated survey estimates to true populations.

Study Area

This study focused on the TRV in north Alabama and the Mobile CBD in south Alabama, where the Alabama Wildlife and Freshwater Fisheries Division (AWFF) conducted ongoing

crewed aerial, non-breeding waterfowl surveys. During November-February 2017-2019 we tracked gadwall to quantify availability within North Sauty and Crow Creek Refuges in Jackson County, and within Wheeler National Wildlife Refuge (Wheeler NWR) in Morgan, Limestone, and Maddison Counties in north Alabama (Figure 6). North Sauty Refuge (2,027 ha; NSR) and Crow Creek Refuge (1,354 ha; CCR) were managed by the AWFF and were open to the public for recreational purposes such as fishing, boating, deer and small game hunting, bird watching, etc., but were closed to waterfowl hunting. Landcover composition in NSR was 35% open water, 17% deciduous forest, 29% woody wetlands, and 2% emergent herbaceous wetlands (Dewitz and USGS 2021, Figure 6). Additional land cover types in NSR that are beneficial to waterfowl included ~3% cropland and ~4% pasture/hay; all other cover types present occurred at ≤4% coverage (Dewitz and USGS 2021). Landcover composition in CCR was 42% open water, 9% deciduous forest, 17% woody wetlands, and 2% emergent herbaceous wetlands (Dewitz and USGS 2021, Figure 6). Additional land cover types in CCR that are beneficial to waterfowl included 11% cropland and 9% pasture/hay; all other cover types present occurred at ≤5% coverage (Dewitz and USGS 2021).

Wheeler NWR (14,164 ha) was managed by the United States Fish and Wildlife Service (USFWS) and was open to the public for the same recreational purposes as NSR and CCR but was also closed to all waterfowl hunting (USFWS 2007). Wheeler NWR was ~30% open water, ~38% forested wetlands, 5.4% swamp, 5.4% wetland impoundments, ~10% managed cropland, ~8% upland hardwoods, ~3% pine plantation, with the rest being old fields, native warm-season grasses, mudflats, and other land cover types (USFWS 2007, Figure 6). Each of these refuges was extremely important to wintering waterfowl and similar species (e.g., other waterbirds such as American coots [*Fulica americana*], sandhill cranes [*Grus canadensis*], double-crested

cormorants [*Phalacrocorax auratus*], and American white pelicans [*Pelecanus erythrorhynchos*]) as they provided feeding, roosting, refuge, and stop-over areas. Large congregations of waterfowl utilized these areas throughout the entire wintering portion of their annual cycle and these areas were where the largest concentrations of wintering waterfowl occurred in Alabama (USFWS 2007).

During November-February 2021-2024 we conducted aerial surveys to quantify detectability and count variability. Aerial surveys were flown over a large portion of the TRV of north Alabama and in the Mobile CBD in south Alabama. The TRV surveys were split into an eastern and western portion (Figure 1). The eastern portion included the Guntersville Reservoir and both banks of the Tennessee River north from the reservoir to the Alabama-Tennessee border. Additional state lands surveyed included Crow Creek Refuge, special opportunity area (SOA), and wildlife management area (WMA); Mud Creek dewatering unit (DWU) and WMA; North Sauty Refuge; and Raccoon Creek DWU and WMA. The western portion covered Wheeler Reservoir west of U.S. Route 31, Wilson Reservoir, Pickwick Reservoir, Swan Creek DWU and WMA, and Mallard Fox Creek WMA. Within the TRV aerial survey area the landcover composition was: ~82% open water, ~5% deciduous forest, ~3% woody wetlands, and 0.5% emergent herbaceous wetlands (Dewitz and USGS 2021). Additional land cover types that are beneficial for waterfowl species in the TRV survey area included cropland, pasture/hay, and grassland/herbaceous but all were <2.5% coverage. All other cover types present occurred at <1.5% coverage (Dewitz and USGS 2021).

Mobile CBD surveys started at the Alabama Port and followed the western coastline (west of Alabama State Route 193) to the Mississippi-Alabama border, continued along the north shore of Dauphin Island, circled and bisected Gilliard Island from west to east, then flew 3.2 km

(2-mile) spaced transects laid out east to west from Gilliard Island north to the south side of Interstate 10. Surveys were continued north of I-10 by following 1.6 km (1-mile) spaced transects laid out east to west and concluded at the industrial powerlines on the north side of Grand Bay (Figure 1). Within the Mobile CBD aerial survey area, the landcover composition was: ~81% open water, ~10% emergent herbaceous wetlands, and 5.9% woody wetlands. All other cover types present, including those beneficial for waterfowl, occurred at ≤1% coverage (Dewitz and USGS 2021).

Field Data Collection

We assessed gadwall availability within aerial survey areas to quantify what proportion of the population was available for detection within aerial survey areas during typical aerial survey dates. We used unpublished data from a previous study of gadwall survival and movement in the TRV. Gadwall were captured using swim-in traps and rocket nets baited with corn from November-February 2017-2019. From November-February 2017-2018 (hence, "year 1"), captured gadwall were outfitted with dorsally mounted GPS transmitters (Cellular Tracking Technologies, Rio Grande, NJ) equipped with a solar panel, and from November-February 2018-2019 (hence, "year 2"), captured gadwall were outfitted with very high frequency (VHF) prong-and-suture transmitters (model A440; Advanced Telemetry Systems, Isanti, MN). All animal capture and handling procedures were approved by the Auburn University Institutional Animal Care and Use Committee (Protocol 2017-3140), along with the United States Geological Survey Bird Banding Permit (#06587). The first transmitters were deployed on December 4, 2017, and December 20, 2018 in year 1 and 2, respectively, and the last transmitters were deployed on January 5, 2018, and January 31, 2019 in year 1 and 2, respectively. The GPS transmitters in year 1 were programmed to collect waypoints at 4-hr intervals and data were downloaded from

transmitters on Mondays, Wednesdays, and Fridays every week there was a transmitter active. The VHF radiomarked birds in year 2 were tracked diurnally and nocturnally with transmitters set to 40 pulses/min. Radiomarked gadwall were triangulated using trucks equipped with roof-mounted 4-element antennas in a null peak array with an electric compass. Trapping and transmitter deployment were staggered throughout the duration of the wintering period in the event that a radiomarked bird migrated out of the area.

We note that the years of data collection for assessing availability do not overlap those in which our aerial surveys were conducted. However, the data set opportunistically used for evaluating availability was the best available local information accessible. Moreover, the plausible general level of availability faced in annual winter surveys in our study area was used to inform subsequent multi-decade simulations. For this use, we saw mismatch in years used to calculate presence vs. detectability and count variability as a negligible concern.

Aerial surveys were conducted in a Cessna 182T Skylane fixed-wing aircraft during designated time periods during November–February 2021–2024 and were flown in sets. Each set of surveys was defined spatially and temporally. Spatially, a survey set consisted of 3 surveys, 1 in each of the survey areas (TRV East, TRV West, and Mobile CBD) conducted as close in time as weather and scheduling permitted. Temporally, four sets of surveys were conducted each winter, 1 in the middle of November ("Preseason"), 1 in the middle of December ("December"), 1 in the first week of January ("Midwinter"), and 1 in the last week of January ("Late Season"), as weather allowed. Flight speed and altitude varied due to terrain, weather conditions, and landscape obstacles (e.g., power lines, cell phone towers, etc.), but were flown as close to 61-76 m (200-250 ft) above ground level and 148-185 kph (92-115 mph) as possible. All surveys were flown by the same pilot.

Observers sat on the right-hand side of the plane for all surveys and only recorded observations seen out of the right-hand side of the plane. The pilot did not act as an aerial observer. Water level (ft), surface temperature (°F), and start and end times were recorded for surveys. Species of interest included all waterfowl species (*Anatidae*), American coots (*Fulica americana*), sandhill cranes (*Grus canadensis*), double-crested cormorants (*Phalacrocorax auritus*), American white pelicans (*Pelecanus erythrorhynchos*), and bald eagles (*Haliaeetus leucocephalus*). All species of interest observed during the survey were identified to species level and enumerated following USFWS guidelines as published by Bowman (2014).

We recorded observations using multiple devices and platforms: a SONY ICD-B300 IC Recorder, a WearHome V1.0 voice recorder, the Hodges H2 recording device with the Scribe voice recording software (USFWS and SWYM, LLC.; Scribe version 3.0.11), and the Hodges H1 USFWS-GPS-voice recording software (J. Hodges, USFWS). Overall, five surveyors acted as observers over the duration of this project. Observers rotated seat positions in the plane in the second and third seasons to account for observer experience differences (Pagano and Arnold 2009*b*). The observer sitting in the front seat was always listed as the primary observer and during the preseason and midwinter flights was always the most experienced surveyor (Table 2).

During all aerial surveys, two observers independently recorded observations during flight while following the general survey protocol above (Alisauskas and Conn 2019). Time stamps of each observation were made to the minute during aerial surveys. The timing devices used by surveyors were synchronized at the beginning of each survey. The seat positioning of both observers during each survey was also noted. Observers did not speak to or provide visual cues to each other to maximize the independence of observations (Vrtiska and Powell 2011). We did not hang a curtain (e.g., visibility barrier) between observers within the plane cabin which

may have allowed the back seat observer to accidentally take on cues from the front seat observer. However, we felt that this occurrence was rare as the back seat observer maintained their focus directly out of the right-hand window within the field of view for the surveys. Additionally, observers were outfitted with helmets which limited their peripheral sight. We feel that this further decreased the chances of either observer accidently taking on cues from the other observer.

Sources of Error

We quantified a metric of availability representing the proportion of the population on survey areas during typical times aerial surveys were conducted to correct the surveyed population for the birds utilizing non-surveyed areas. To do so, we filtered all gadwall GPS locations from 2017-2019 by diel period to exclude nocturnal points because aerial surveys are conducted diurnally. We mapped duck locations in R (v4.4.1, R Core Team 2021) and determined the proportion of gadwall locations within the TRV aerial survey flight paths. Aerial survey flight path polygons were created in ArcGIS (v2.5.0, ESRI 2020) to include a 250 m buffer accounting for the viewing area of the aerial surveys beyond the flight path the plane traveled. Locations inside the polygon were assigned a value of 1, whereas those outside of the polygon were assigned a value of 0.

Next, we used binomial models to incorporate the potential influence of covariates on duck use of flight path areas and used the top model to estimate an index of availability. The covariates we analyzed included the capture location for each bird, study year, and a continuous date covariate that was scaled such that day 1 was November 15th, which coincided with the approximate first day aerial surveys were conducted each winter. We ran an intercept only model and a series of single covariate models with fixed effects of trap site, study year, or date. We also

considered a polynomial date model, a polynomial date + trap site model, and a polynomial date + study year model, which were all scaled as described above. All models included a random effect of individual bird. Models were assessed using Akaike Information Criterion adjusted for small sample size (AICc) scores (Akaike 1974, Arnold 2010) and compared using ΔAICc values with the top and competing models having ΔAICc values of 0 and ≤2, respectively (Akaike 1974, Arnold 2010). We used the predict function to model probability of availability of individual birds over time by trap site (Appendix E) and we modeled an index of overall availability of birds on aerial survey dates using a variation of the top model with trap-site nested with individual bird as a random effect. The index of overall availability of birds was used within simulations, as explained below, as the availability level.

We produced model-fitted detectability estimates for group observations using an RMark Huggins closed-population mark-recapture model including seat position, observer, guild, and log-transformed group size covariates, as these covariates have been shown to best explain variation in our system (Chapter 1). Detectability was estimated at the group-level and it was assumed that individual detectability was the same as group detectability for every bird in a group. Using the covariate.predictions function in RMark (Laake 2013), we generated separate detectability estimates for each group detected for each guild (dabbling duck, diving duck, goose, or non-waterfowl) and seat (front or back) combination possible (n = 8). We then subset the data frames so that each one included only estimates for 1 designated guild, seat, and observer (A, B, or C) combination, resulting in 24 total data frames with detection probability estimates.

For this study we were primarily interested in overall average survey-level detection estimates. Therefore, for each data frame specified above, we calculated weighted estimates of overall detectability for each survey by multiplying the group-level detection probability

estimate by the associated observed group size. Then, we divided the sum of the weighted estimates by the sum of the observed group sizes to calculate an overall weighted estimate for each guild, seat, and observer combination. We calculated the variance of the weighted average using the wtd.var function in the Hmisc package (v5.1.3, Harrell 2024) in R. We used the average of the weighted estimates among combinations to inform the detectability levels within our simulations as explained below.

We estimated count variability by taking the summed group size counts from aerial surveys to obtain total counts by guild and surveyor for each survey, thus resulting in a front-seat count and back-seat count for each survey. For each guild, we ran a linear mixed effects model ("lmer"; lme4 package v1.1.35.5, Bates et al. 2015) modeling log-transformed total count vs. fixed effects of seat position and observer and a random effect of survey. With this model, there are two variance components estimated: 1) the survey random effect variance component, which was not of interest but necessary to account for relative (log-scale) variation in average counts among surveys, and 2) the residual variance component, which estimates average within-flight relative variability between the two counts per flight. Both variance component estimates were model adjusted for the fixed effects accounting for average seat position and observer effects. Therefore, the residual variance estimate for each guild analysis is an estimate of within-flight relative repeatability of counts (i.e., relative counting variability - after accounting for seat and observer identities, Stoffel et al. [2017]). To check whether relative count variability estimates obtained from this modeling were reasonable, we also assessed raw coefficients of variation values within the data. We calculated the mean and standard deviation between total counts by observers for each guild in each survey then divided the standard deviation by the mean and multiplied by 100. Further, we averaged the squared calculated coefficient of variation values by

guild then took the square root to obtain an overall count variability estimate from surveys by guild. We used this overall count variability estimate to inform the count variability levels in our simulations as described below.

Survey Evaluations

We used simulations to compare alternative aerial survey approaches potentially feasible for AWFF, including multiple options for numbers of observers present during the survey, whether or not surveys are conducted in a mark-recapture framework, and the number of repeated surveys conducted. We ran permutations including all possible combinations of 1-2 observers, mark-recapture or non-mark-recapture (e.g., mark-recapture represented true double-observer methods and non-mark-recapture represented two unmatched independent counts from each flight), and 1-3 survey flights within a survey window representing 9 total survey situations evaluated. We conducted simulations such that those incorporating multiple survey flights within a survey window assumed subsequent flights occurred within a timeframe to adequately assume population closure. Each of the 9 survey situations was conducted with crossed combinations of two levels of availability, two levels of count variability, and two levels of detectability representing 8 total error source combinations, each level having been quantified as explained above. Therefore, 72 total permutations were analyzed (9 survey situations × 8 error levels = 72). Levels of availability and count variability were high and low, where the high level was set to the mean level of each respective error we found. The low level of availability and count variability was the value half-way between the calculated mean and the absence of that source of error (e.g., 0).

Levels of detectability were also high and low. In permutations with 1 observer, high detection level was equal to the mean level of detectability found for a highly experienced

observer and the low detection level was equal to the mean level of detectability found across all observers. Including all participating observers is relevant to managers as it accounts for observer differences (Erwin 1982, Pagano and Arnold 2009*a*). In permutations with 2 observers, the front seat, high detection level was equal to the mean level of detectability found for a highly experienced observer in the front seat and the back seat, high detection level was equal to the mean level of detectability found for a highly experienced observer in the back seat (Erwin 1982, Musgrove and Holloway 1997). The front seat, low detection level was equal to the mean level of detectability found for a highly experienced observer in the front seat and the back seat, low detection level was equal to the mean level of detectability found for a mid-level experienced observer in the back seat.

Within each permutation, we ran 250 simulations of estimated survey counts (unpaired data) or population estimates (paired data) over 50 years per simulation. We set "true" population size equal to 50,000 in year 1 and projected the population size over 50 years with a non-varying lambda of 0.99 (1% decline per year). We defined the number of available birds in our simulated surveys as those from a binomial draw from the true population size each year, with the probability of availability set as a mean equal to each level of availability as defined above. We defined the number of detected birds in our simulated surveys as a binomial draw from the number of present birds, with detection probability equal to each of the high and low levels as specified above. In permutations with two participating observers with mark-recapture frameworks, we simulated 1 estimated survey count representing the overall detection achieved by double-observer methods calculated using the equation published by Nichols et al. (2000):

$$1 - ((1 - \mathrm{p1})(1 - \mathrm{p2}))$$

where p1 is the mean overall weighted detection estimate among all guilds for the front seat observer and p2 is the mean overall weighted detection estimate among all guilds for the back seat observer. We then calculated the variance associated with this overall detection rate by using the Delta method (Powell and Gale 2015) and simulated a detection estimate to be used in simulations by drawing a random value from a normal distribution with a mean of the calculated overall detection and a variance of the calculated variation from the overall detection estimate.

We incorporated count variability by setting our simulated survey count variability equal to a high and low level as specified above. To simulate counting variability, we used a truncated normal distribution, with mean equal to the number of detected birds and standard deviation equal to each of the count variability levels. Lastly, for permutations with 0 repeated surveys (1 survey total) no additional factors were incorporated. For permutations with 1-2 repeated surveys, we treated each additional survey as an additional observer without mark-recapture frameworks and without modeled variation in availability, detectability, or count variability. For example, in a permutation with 1 observer and 1 repeated survey, we generated two unmatched separate estimated population counts, with each holding availability, detectability, and count variability at their respective levels.

To assess statistical power to detect a trend in wintering waterfowl abundance, for each simulation replicate in each permutation we ran log-linear models with a fixed effect of year as a numeric variable, starting in year 5. With permutations with 1 survey per year and either only 1 observer or two paired observers producing a single mark-recapture estimate, we used 'lm' to run simple log-linear regression models. For all other permutations, which produced at least 2 counts or mark-recapture estimates, we used 'lmer' models (lme4 package v1.1.35.5, Bates et al. 2015) with the fixed effect of year as a numeric variable and a random effect of year as a categorical

variable (Urquhart et al. 1998). The ‘lm’ or ‘lmer’ model was constructed in a for loop in R such that a model was run for the first year through every survey year starting in year 5 for all simulations. For example, the first ‘lmer’ would model years 1:5 of data for the first simulation, the second model 1:6, etc. through the end of the simulation then repeat through all repetitions. The P-value associated with each ‘lmer’ run was extracted from model summary outputs in R and stored in a data matrix with columns representing simulations and rows representing the simulated analysis year. The number of significant P-values occurring in each row was tallied and divided by the number of columns to define power. The first year in which power was 0.8 or greater represented the number of years required to conduct aerial surveys with that permutation’s bias specifications to correctly detect a change in population trend.

Lastly, we evaluated relative influence of each source of error by comparing survey bias and precision results among all combinations. Specifically, we compared mean squared error values, mean count error values, percent relative bias, and a coefficient of variation among permutations. We calculated mean squared error values for each simulation within each permutation by subtracting the true population value from the simulated population value for every year within a simulation, squaring each value, taking the sum of all the squared values, then divided the sum of squared values by the total number of years projected. This resulted in a mean squared estimate value for each simulation run per permutation. Mean count error was calculated by taking the square root of mean squared error. We calculated percent relative bias by averaging all simulated estimates per year projected, subtracting the true population size from the average simulated population size for each year, then divided each difference by the true population size and multiplied by 100. This resulted in a percent relative bias value for each year projected per permutation. We calculated the standard deviation between all simulated survey

estimates within years using the "sd" function in R. Then, we calculated a coefficient of variation (CV) for each year of simulations per permutation by dividing the calculated standard deviation by the calculated average estimate per year and multiplying by 100.

We averaged mean squared error, CV, and percent relative bias across simulations and years to assign each permutation 1 value per error metric. We used linear regression for each error metric to examine the relative contribution of the sources of error and survey options in explaining the variability among permutations. Specifically, we fit a model with main effects and all two-way interactions between sources of error and permutations, then calculated the R^2 explained by each effect. For each error metric, we conducted a second regression analysis to estimate the average main effect of each error source and survey design factor (number of surveys, number of surveyors, and whether or not two-observer surveys used a double-observer method [paired vs. unpaired]). Lastly, we ranked all 72 permutations by their values of power, mean squared error, percent relative bias, and CV where values closest to 0 were ranked highest. An overall, average rank was assigned to each permutation by averaging all 4 ranks listed above.

RESULTS

Sources of Error

During November-February 2017-2019, 51 gadwall were captured and outfitted with GSM backpack transmitters (n = 25; Males = 11, Females = 14; CCR = 7, NSR = 9, Wheeler NWR = 9) or VHF backpack transmitters (n = 26; Males = 13, Females = 13; NSR = 15, Wheeler NWR = 11). Of the transmitters deployed, 11 were harvested while 7 succumbed to non-hunting mortality, 11 were censored due to premature transmitter failure or unconfirmed death, 21 survived to expected transmitter failure, and 1 transmitter fell off the bird. From December 8, 2017 – March 12, 2018, 2,594 locations of individuals were recorded including

1,228 diurnal locations and 1,366 nocturnal locations. From December 27, 2018 – February 28, 2019, 776 locations of individuals were recorded, including 385 diurnal locations and 391 nocturnal locations. Of the 1,881 total daily locations of birds we observed, 1,265 (67%) were within surveyed areas (Figure 7).

The top model (ΔAICc = 0) for factors influencing selection of surveyed areas by birds included the fixed effects of trap site and the scaled polynomial date variable with a random effect of individual (Appendix E). This model carried 90% of the weight and there were no competing models (ΔAICc ≤ 2). We modeled an availability index for each relative day an aerial survey was conducted throughout the duration of the wintering period for birds. The model included a polynomial date scaled fixed effect with a random effect of individual nested within trap site. Availability of birds for December aerial surveys (~ Dec 15) was 0.93 (SE = +/- 0.10), Midwinter (~Jan 5) aerial survey availability was 0.82 (SE +/- 0.24), and availability in the Late Season aerial survey was 0.70 (SE = +/- 0.34). The variance of the individual and trap site interaction random effect was 2.32 (SE = +/- 1.52) and the variance of the trap site random effect was 0.66 (SE = +/- 0.81).

Overall average detectability including all observers, seats, and guilds was 0.64. On average, front seat weighted detectability estimates were 0.08 higher than back seat weighted detectability estimates with overall average detectability estimates being 0.68 in the front seat and 0.60 in the back seat. The most experienced observer had the highest overall average detectability estimate at 0.77 while the mid-level experienced observer and low-level experienced observer each had overall average detectability estimates of 0.57. The non-waterfowl guild had the highest overall average detectability estimate at 0.76, with the goose guild following at 0.73, then diving ducks at 0.53 and dabbling ducks at 0.52 (Figure 8).

Log-scale count variability estimates followed opposite patterns to those of detectability. Dabbling and diving duck guilds had the highest variability at 0.80 (0.58, 1.24; 95% CL) and 0.85 (0.56, 1.30; 95% CL), respectively. Counting variability was much lower for geese at 0.39 (0.40, 0.89; 95% CL), and was slightly higher in non-waterfowl at 0.48 (0.46, 0.92; 95% CL). Coefficient of variation estimates calculated from the raw data (not accounting for additional covariates as in lmer models) were consistent to slightly higher than those obtained from lmer modeling. Dabbling duck count variability was 0.78 (78%), diving ducks variability was 0.97 (97%), goose variability was 0.78 (78%), and non-waterfowl variability was 0.56 (56%).

Survey Evaluations

We calculated the mean availability index to be 0.82 and used that value to represent a high level of availability in simulations and 0.41 to represent a low level of availability in simulations. We used a coefficient of variation value of 0.78 (0-1 scale) to represent the high level of count variability in simulations and 0.39 as the low level of count variability in simulations. We calculated the mean detectability to be 0.64 among all observers and 0.80 for a highly experienced observer; we used those values to represent low and high levels of detectability, respectively, for single observer permutations. We calculated the front seat mean detectability of a highly experienced observer to be 0.80, the back seat mean detectability of a mid-level experienced observer to be 0.52, and the back seat mean detectability of a highly experienced observer to be 0.74. For simulations with two paired observer counts, we calculated the overall mean detection as 0.90 and the overall high detection as 0.95.

Power analyses showed that most survey designs we examined could not accurately detect a 1% annual population decline within a 50-year survey period. Survey designs only including 1 survey flight never reached suitable power by year 50, with the exception that 1

survey flight employing 2 observers with unpaired data, at high availability, high detectability, and low count variability reached suitable power at year 50. Survey designs employing 1 observer in 2-3 flights achieved suitable power when count variability was low (power for 2 observers = 47-50 years, 3 observers = 45-46 years), except with high availability and detectability or with low availability and high detectability in which case suitable power was not achieved for those survey designs. Suitable power was achieved by year 50 in scenarios employing 2 observers with paired or unpaired data flying in 2-3 flights with low count variability. The years in which power was obtained flying with 2 observers were 49-50 years for paired data in 2 surveys, 43-46 years for paired data in 3 surveys, 39-42 years for unpaired data in 2 surveys, and 35-38 years for unpaired data in 3 surveys. The only instance where power was not achieved for 2 observers with paired data flying in 2 surveys at low count variability was when availability and detectability were also low.

Mean count error of simulated survey estimates compared to true values was significantly lower with high levels of detectability and availability, multiple flights (2-3), and with paired data (Table 6). Mean count error was 2,252 birds (SE = +/- 416) lower with high detectability compared to low detectability ($P < 0.001$) and 10,050 birds (SE = +/- 416) lower with high availability compared to low availability ($P < 0.001$). Availability accounted for 79.6% of the variation in mean count error (Table 7). Mean count error was 2,867 birds (SE = +/- 509) lower with paired data compared to unpaired data ($P < 0.001$), 1,250 birds (SE = +/- 509) lower with 2 flights compared to 1 ($P = 0.017$), and 1,807 birds (SE = +/- 509) lower with 3 flights compared to 1 ($P < 0.001$). Lastly, permutation (or survey situation) explained 8.8% of the variability in mean count error (Table 7).

Coefficient of variation (CV) of simulated survey estimates compared to true values was significantly lower with low count variability, 2 observers, and multiple flights (2-3), and higher for paired data (Table 6). CV was 13.77% (SE = +/- 0.55) lower with low count variability compared to high count variability ($P < 0.001$) and 10.48% (SE = +/- 0.68) lower with 2 observers compared to 1 ($P < 0.001$). Count variability explained 33.2% of the variability in CV (Table 7). CV was 12.6% (SE = +/- 0.68) and 18.23% (SE = +/- 0.68) for 2 or 3 flights compared to 1, respectively. CV was 10.43% (SE = +/- 0.68) higher for paired data than unpaired data ($P < 0.001$). Additionally, permutation (or survey situation) explained 64.1% of the variability in CV (Table 7, Figure 9).

Percent relative bias ("relative bias") of simulated survey estimates compared to true values was significantly more negative with low count variability and closer to zero with high detectability, high availability, and paired data (Table 6). Relative bias was 6.77% (SE = +/- 1.01) more negative with low count variability compared to high count variability ($P < 0.001$). Relative bias was 7.15% (SE = +/- 1.01) closer to zero for high detectability than low detectability ($P < 0.001$), and 34.08% (SE = +/- 1.01) closer to zero for high availability than low availability ($P < 0.001$). Availability explained 78.5% of the variability in relative bias. Relative bias was 13.43% (SE = +/- 1.23) higher for paired data compared to unpaired data ($P < 0.001$). Additionally, permutation (or survey situation) explained 10.6% of the variability in relative bias (Table 7, Figure 10).

Permutations with high availability, high detectability, and low count variability had the highest overall average ranks (best performance) while permutations with low availability, low detectability, and high count variability had the lowest overall average ranks (worst performance) (Table 8). Most of the time, permutations with high availability were ranked above those with

low availability. Patterns of levels of detectability and count variability were not as obvious, but generally permutations with high detectability were ranked higher than those with low detectability and permutations with low count variability were ranked higher than those with high count variability (Table 8). The best permutations overall were those employing 2 surveyors with paired data on 3 surveys, and the worst permutations overall were those employing only 1 surveyor on 1 survey (Table 8, Appendix D).

DISCUSSION

Overview

We quantified availability, detectability, and count variability as sources of error in wintering waterfowl population aerial surveys. We discuss the results from calculating these levels of error in the Patterns of Bias and Error section below. Further, we used these levels of sources of error to inform simulations generating wintering waterfowl population aerial survey estimates and evaluated error metrics from the simulations as discussed in the Patterns of Survey Performance section below. Lastly, we discuss the broad takeaways, and some direct management recommendations managers should consider in the Survey Implications section below.

Patterns of Bias and Error

Our estimates of detectability and count variability from aerial surveys were not dramatically different than what has been reported in other studies. Therefore, we felt comfortable incorporating our estimates of detectability and count variability into subsequent simulations. Specifically, our average estimates of overall detectability were 0.64-0.95, compared to other studies of aerial waterfowl surveys which have reported detection values between 0.27-0.96 (depending on species of interest, Gilbert et al. 2021) for surveys employing a

consistent observer and between 0.33-0.99 (Pearse et al. 2008*b*) for multiple observers. Results support our incorporation of multiple observers in our detection estimates, and variability among observers has been shown in other literature (Erwin 1982, Musgrove and Holloway 1997). Further, our estimate of counting variability, which averaged 0.78 (78%), was also comparable to values reported in other studies. For example, Davis et al. (2022) reported that observer counts ranged from 40%–150% of true flock size for large groups. Moreover, Fredrick et al. (2003) reported high counting variation among observers with the average absolute value of error reaching 50%.

Quantifying availability for aerial surveys is not as common in the literature as quantifying detectability or count variability due to some aerial surveys using probabilistic sampling frameworks, but given the literature that does exist, we felt comfortable with the levels of availability we calculated and incorporated into simulations. For example, Conkin and Alisauskas (2013) modeled waterfowl encounter probability based on satellite imagery and reported values for Canada Geese between 0.51 and 0.92. We feel our calculated availability (0.82) was comparable to their encounter probability estimates. Additionally, Schummer et al. (2018) assessed waterfowl presence within survey areas during aerial breeding population surveys based on locations of telemetry marked birds. They reported that predicted probability of female lesser scaup being on surveyed areas during a survey was 0.35-0.64 depending on survey year. Our estimates may be slightly higher than Schummer et al. (2018) due to wintering behavior of birds compared to migrating and breeding behaviors, as well as their study area being substantially larger than ours.

We found that gadwall presence decreased throughout the winter, perhaps because waterfowl congregate in large concentrations on refuges in the TRV of north Alabama during the

hunting season and may deplete food resources quickly (Henry 2012). For example, vegetation mass on North Sauty Refuge and Guntersville Reservoir in north Alabama decreased from 250 g/m^2 in October to <20 g/m^2 in February (McKnight and Hepp 1998). Further, Hagy et al. (2014) reported that wintering dabbling duck densities were negatively correlated with cumulative precipitation throughout the wintering period. They posit that, in addition to food resources being depleted, waterfowl opportunistically forage on newly available food resources following precipitation (e.g., flooded agricultural fields [Hagy et al. 2014]). Additionally, wintering dabbling ducks in Oklahoma decreased use of large reservoirs in favor of small wetlands during late winter (Heitmeyer and Vohs 1984). We believe that the decrease in availability on surveyed areas is due to resource depletion and new resources being made available by an increase in water on the landscape. We do not believe that the decrease in availability we observed is due to spring migration chronology as wintering gadwall in the Mississippi flyway do not migrate north until mid-March (Paulus 1984). The variability and changes of availability we observed throughout the wintering season only adds to the challenge of considering availability bias in aerial surveys where probabilistic sampling frameworks are not employed.

Gadwall were the most commonly occurring dabbling ducks in our survey area, and large concentrations of gadwall winter in the TRV annually (Henry 2012). However, dabbling duck cover type selection varies between species (McClanahan 2015). This indicates that some species of dabbling ducks within our aerial survey areas may be using our study areas differently than other species of dabbling ducks within our aerial survey areas. Therefore, our availability index may be limited in scope as we only used gadwall tracking data to quantify availability in our aerial surveys. Tracking additional species would have resulted in a more broadly applicable availability index across guilds, as inter- and intraguild habitat selection varies (Pearse et al.

2012, Davis et al. 2014). Additionally, we speculate that due to the linear nature of our aerial surveys, waterfowl may make small movements within the broader geography of the TRV, utilizing areas that are just outside our aerial survey areas. Dabbling ducks, and specifically mallards, have been shown to not leave established wintering home ranges, but may make daily movements large enough to leave our aerial survey areas (Yetter et al. 2018, Dittmer et al. 2023, Osborne et al. 2024).

Patterns of Survey Performance

We evaluated "mean count error" as an error metric in our study which measured the average difference between the number of birds in the true population and simulated estimates. It was clear that levels of high detectability and increased survey effort in the form of multiple surveys and paired data reduced mean count error in simulated estimates. Additionally, we found that mean count error was mostly influenced by availability in our study, as availability accounted for almost 80% of the variability in mean count error across 72 permutations. This is logical as mean count error is a combination of bias and noise within the system and indicates that quantifying levels of availability in agency surveys is crucial where relevant, and maybe more so than quantifying detectability levels. However, this pattern is specific to our system, and even in other Southeastern non-breeding season waterfowl surveys, the relative importance of the two sources of bias may vary.

In addition to mean count error, which incorporated both bias and noise in simulated estimates, we assessed these two components of error separately with percent relative bias and coefficient of variation metrics. We observed a decrease in absolute percent relative bias with both high levels of detectability and availability and with paired data versus unpaired data. Further, we found that availability again accounted for almost 80% of the variability in percent

relative bias across 72 permutations. However, we found that absolute percent relative bias increased with low count variability compared to high count variability. We speculate that increased count variability might result in simulated overcounting of group sizes, therefore negating the effect of missed groups of birds and reducing the observed absolute relative bias. Lastly, our overall average percent relative bias value (-48%) was comparable to other reported percent bias. Specifically, Fredrick et al. (2003) reported percent bias ranging from approximately -60% to 10%.

We observed highly variable CV values among individual permutations (e.g., survey situations). Permutations with 1 or 2 observers (with paired data) participating in 1 survey showed the highest CV regardless of the levels and sources of error associated with that permutation. We believe this was due to survey effort as we observed lower CV values with an increase in survey effort. Specifically, our results show that CV was significantly lower when 2 or 3 surveys were conducted, which is consistent with Pearse (2007) who reported an increase in survey number decreased CV in their study estimating wintering waterfowl abundance in Mississippi. Moreover, not surprisingly, we found that CV decreased with low count variability. Others have reported similar findings (Erwin 1982, Gilbert et al. 2021). Lastly, we observed an increase in CV when data were paired vs. unpaired. This is because accounting for detectability serves to reduce bias, but at the expense of increased variance (Skalski and Robson 1992).

In our simulations, we found that surveys reached adequate power to correctly detect changes in populations trends in 35 years or more. These results are consistent with White (2019), who reported monitoring efforts for species in the taxonomic class *Aves* with population trends of 1% could take a minimum of ~40 years to achieve power at the 0.8 level. We found the time required to detect a population trend was most influenced by count variability as surveys

with high levels of count variability never achieved adequate power. Lewis and Gould (2000) also reported an inverse relationship between power and count variability, showing that very low count variability (<0.30 CV) was required to detect a 3% change in annual raptor population estimates over a 25-year period. Our simulations took a greater number of years to achieve power than Lewis and Gould (2000) because we, by contrast, modeled a 1% decline in annual population change. Power would have been achieved earlier in our study for larger changes in population size. Additionally, we did not incorporate natural variation in population rate of change, instead assuming deterministic 1% declines in annual population size. Incorporating process variation into simulated population change to more closely mimic natural population fluctuation would act to lower power.

Moreover, time to achieve adequate power was influenced by survey effort with an increased number of observers and replicate flights reducing the number of years required to achieve adequate power. Generally, employing 2 observers achieves power in a fewer number of years than 1 observer, and 3 replicate flights achieves power in a fewer number of years than 2 replicate flights. Power was never achieved when conducting only 1 flight. Our results are consistent with Lougheed et al. (1999) who showed that power to detect changes in waterfowl populations increases with an increasing number of survey replicates. Similarly, our simulations assumed simple regression trend analyses incorporating independent estimates of population size each year. More sophisticated analyses (e.g., hierarchical double-observer models) in which the entire analysis is updated periodically using all data collected to that point could improve power and precision of abundance estimates.

Overall, permutations with low sources of error performed best, and those with high sources of error performed worst. Generally, even though the influence of availability on

estimates is not well studied, it has been shown that levels of detection probability and count variability can affect survey estimates negatively (Eggeman and Johnson 1989, Shirkey et al. 2014, Davis et al. 2022, etc.). Further, permutations with greatest survey effort performed best, and surveys with the least survey effort performed the worst. Other studies have also shown that increased waterfowl population survey efforts produce estimates with lower variability and greater ability to accurately detect trends over time (Lougheed et al. 1999, Pearse 2007).

Survey Implications

By understanding the relative magnitudes of sources of potential noise (counting variability, daily availability variation) and bias (incomplete detectability and availability, counting bias), managers designing surveys may make informed decisions about which sources of error are most important to address and ways to address them. In our simulations, reducing count variability was the most important factor to address in achieving suitable power to detect trends, along with increasing precision in count indices or abundance estimates. Managers have two options for addressing count variability: reducing its magnitude in each survey flight and increasing the number of surveys within a short-term survey window (Lougheed et al. 1999, Pearse 2007). Reducing count variability is more cost-effective, and could be done by incorporating visual correction factors from ground surveys (e.g., Waterfowl Breeding Population and Habitat Survey), increasing training of aerial observers, or assessing group size estimation using photographs with known numbers of individuals (especially for inexperienced observers [Erwin 1982]). Although replicating flights or having multiple observers may cost more, increasing the number of distinct counts is commonly recommended in count-based monitoring (Dennis 2010). Specifically, our results showed adding 1 or 2 replicate flights or an additional observer significantly reduced CV, mean count error, and power. Adding 1 replicate

flight increases precision more than an extra observer for CV and mean count error but adding a second observer increases power more than an extra flight. Further, adding 2 replicate flights improves all precision estimates more than an additional observer and 1.5 times more than 1 replicate flight. Overall, conducting replicate flights and employing multiple observers drastically improved precision within estimates, resulting in a 27.8% decrease in CV, producing a 1,857 bird decrease in mean count error, and achieving adequate power 5.5 years sooner.

Our simulations did not include systematic changes in detectability or availability over time, which, if present but not addressed, could lead to inaccurate conclusions about trends. Availability and detectability biases also would lead to significant undercounting of birds, producing survey counts that represent some fraction of birds using the study area. In many situations, survey approaches can allow simultaneous estimation of detectability and availability (Marsh and Sinclair 1989, Johnson 2008). However, in our study system the linear nature of the TRV system and degree of counting variability limit the utility of some approaches of accounting for detectability on its own, let alone both detectability and availability. While the double-observer method or aerial photography may have utility for quantifying detectability, accounting for both sources of bias simultaneously remains a challenge in this system.

In our system, at least based on gadwall movements, availability bias appears to be a more significant source of error than detectability bias, and the most important single source of error overall. This is not surprising, as our simulations incorporated availability estimates (0.41-0.82) which were lower than the detectability estimates of 0.64-0.95 we incorporated. Our results indicate availability biases are likely greater than detectability biases in our system, but this difference should be considered when interpreting these results. Overall, we believe this conclusion is understandable given that the survey area follows a relatively narrow belt along the

TRV where birds could easily relocate to nearby areas off the river (e.g., private lands, agricultural fields, etc.) which are not surveyed. However, this is unfortunate, because estimating availability from single-flight surveys in a highly dynamic system is difficult. Options for addressing availability are not straightforward but may include supplementary telemetry studies (Schummer et al. 2018), ensuring survey areas adequately cover variable population distributions (Prenzlow and Lovvorn 1996), or predicting species distributions (Conkin and Alisauskas 2013). Addressing availability is important in the context of population trends. For example, if availability changes but the real population trend does not, a survey could erroneously detect a trend that does not exist. Further, if a population trend declines from one year to the next while availability simultaneously increases, a survey could fail to detect a real trend. Specific considerations for assessing availability within our system and others should include quantifying a correction factor applicable across all species of interest and doing so across the entire wintering or survey period.

MANAGEMENT IMPLICATIONS

Agencies interested in increasing the robustness of their waterfowl aerial surveys should consider sources of error within their systems and practical methods for reducing those sources relative to their objectives. Agencies should attempt to quantify availability, detectability, and count variability within their waterfowl aerial surveys if their objective is to estimate true population size within a region. Surveys with high availability, high detectability, and low count variability will require less survey effort to achieve more robust estimates than surveys with low detectability, low availability, and high count variability. Overall, managers should carefully consider the goals of their surveys and the levels of precision they hope to achieve when considering waterfowl population survey design. Agencies should take these and other results

into consideration when allocating resources such as personnel and flight hours for waterfowl aerial surveys.

LITERATURE CITED

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716-723.

Alisauskas, R. T., and P. B. Conn. 2019. Effects of distance on detectability of Arctic waterfowl using double-observer sampling during helicopter surveys. Ecology and Evolution 9:859-867.

Andersson, K., C. A. Davis, G. Harris, and D. A. Haukos. 2015. An assessment of non-breeding waterfowl surveys on National Wildlife Refuges in the Central Flyway. Wildlife Society Bulletin 39:79-86.

Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's Information Criterion. The Journal of Wildlife Management 74:1175-1178.

Baldassarre, G. A., and E. G. Bolen. 2006. Waterfowl Ecology and Management. Krieger, Malabar, FL, USA.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bayliss, P., and K. M. Yeomans. 1990. Use of low-level aerial photography to correct bias in aerial survey estimates of magpie goose and whistling duck density in the Northern Territory. Wildlife Research 17:1-10.

Bowman, T. D. 2014. Aerial observer's guide to North American waterfowl. U.S. Fish and Wildlife Service, Division of Migratory Bird Management, Anchorage, AK, USA.

Chabot, D., and D. M. Bird. 2015. Wildlife research management methods in the 21st century: Where do unmanned aircraft fit in? Journal of Unmanned Vehicle Systems 3:137-155.

Clare, J., S. T. McKinney, J. E. DePue, and C. S. Loftin. 2017. Pairing field methods to improve inference in wildlife surveys while accommodating detection covariance. Ecological Applications 27:2031-2047.

Conkin, J. A., and R. T. Alisauskas. 2013. Modeling probability of waterfowl encounters from satellite imagery of habitat in the central Canadian Arctic. The Journal of Wildlife Management 77:931-946.

Conroy, M. J., J. T. Peterson, O. L. Bass, C. J. Fonnesbeck, J. E. Howell, C. T. Moore, and J. P. Runge. 2008. Sources of variation in detection of wading birds from aerial surveys in the Florida Everglades. The Auk 125:731-743.

Cordts, S. D., G. G. Zenner, R. R. Koford. 2002. Comparison of helicopter and ground counts for waterfowl in Iowa. Wildlife Society Bulletin 30:317-326.

Cox, A. R., S. G. Gilliland, E. T. Reed, and C. Roy. 2022. Comparing waterfowl densities detected through helicopter and airplane sea duck surveys in Labrador, Canada. Avian Conservation and Ecology 17:24.

Davis, J. B., M. Guillemain, R. M. Kaminski, C. Arzel, J. M. Eadie, and E. C. Rees. 2014. Habitat and resource use by waterfowl in the northern hemisphere in autumn and winter. Wildfowl 4:17-69.

Davis, K. L, E. D. Silverman, A. L. Sussman, R. R. Wilson, and E. F. Zipkin. 2022. Errors in aerial survey count data: Identifying pitfalls and solutions. Ecology and Evolution 12:e8733.

Dennis, B., J. M. Ponciano, and M. L. Taper. 2010. Replicating sampling increases efficiency in monitoring biological populations. Ecology 91:610-620.

Dewitz, J., and U.S. Geological Survey. 2021. National land cover database (NLCD) 2019 Products (ver. 2.0 June 2021): US. Geological Survey data release, <https://doi.org/10.5066/P9KZCM54>. Accessed 1 May 2024.

Diefenbach, D. R., M. R. Marshall, J. A. Mattice, and D. W. Brauning. 2007. Incorporating availability for detection in estimates of bird abundance. The Auk 124:96-106.

Dittmer, E. M., R. J. Askren, H. M. Hagy, J. Hitchcock, and D. C. Osborne. 2023. Not all sanctuaries are created equal: variation in protected area selection by wintering mallards. The Journal of Wildlife Management 88:e22535.

Eggeman, D. R, and F. A. Johnson. 1989. Variation in effort and methodology for the midwinter waterfowl inventory in the Atlantic Flyway. Wildlife Society Bulletin 17:227-233.

Eggeman, D. R., F. A. Johnson, M. J. Conroy, and D. H. Brakhage. 1997. Evaluation of an aerial quadrat survey for monitoring wintering duck populations. The Journal of Wildlife Management 61:403-412.

Erwin, R. M. 1982. Observer variability in estimating numbers: An experiment. Journal of Field Ornithology 53:159-167.

ESRI 2020. ArcGIS Pro Desktop v2.5.0. Environmental Systems Research Institute. Redlands, CA, USA.

Fletcher, R. J., Jr., and R. L. Hutto. 2006. Estimating detection probabilities of river birds using double surveys. The Auk 123:695-707.

Fredrick, P. C., B. Hylton, J. A. Heath, and M. Ruane. 2003. Accuracy and variation in estimates of large numbers of birds by individual observers using aerial survey simulator. Journal of Field Ornithology 74:281-287.

Gilbert, A. D., C. N. Jacques, J. D. Lancaster, A. P. Yetter, and H. M. Hagy. 2020. Disturbance caused by aerial waterfowl surveys during the nonbreeding season. The Journal of Wildlife Management 84:1063-1071.

Gilbert, A. D., C. N. Jacques, J. D. Lancaster, A. P. Yetter, and H. M. Hagy. 2021. Visibility bias of waterbirds during aerial surveys in the nonbreeding season. Wildlife Society Bulletin 45:6-15.

Hagy, H., B. Van Druten, C. Bleke, N. Wirwa, M. Chouinard, K. Jensen, K. Schrover, E. Stinson, P. Stinson, B. Fortier, J. Talley, G. Wilderson, S. Yates, J. Rayfield, and J. Wortham. 2024. Midwinter aerial waterfowl surveys on national wildlife refuges in the Southeast, 2024. U.S. Fish and Wildlife Service Report.

Hagy, H. M., J. N. Straub, M. L. Schummer, and R. M. Kaminski. 2014. Annual variation in food densities and factors affecting wetland use by waterfowl in the Mississippi Alluvial Valley. Wildfowl 4:436-450.

Hagy, H. M., M. G. Brasher, J. Fleskes, J. N. Straub, A. Lafon, and E. Carrera. 2025. Important geographies for migrating and wintering waterfowl. *In* Ballard, B. M., M. G. Brasher, and J. P. Fleskes, editors. Migrating and Wintering Waterfowl. Texas A&M University Press.

Harrell Jr., F. 2024. _Hmisc: Harrell Miscellaneous_. R package version 5.1-3, <https://CRAN.R-project.org/package=Hmisc>.

Hartig, F. 2022. _DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models_. R package version 0.4.6, <https://CRAN.R-project.org/package=DHARMa>.

Heitmeyer, M. E., and P. A. Vohs, Jr. 1984. Distribution and habitat use of waterfowl wintering in Oklahoma. The Journal of Wildlife Management 48:51-62.

Henry, T. H. 2012. Results of the Tennessee River Valley shorebird initiative. Tennessee Valley Authority, Knoxville, TN, USA.

Johnson, F. A., K. H. Pollock, and F. Montalbano, III. 1989. Visibility bias in aerial surveys of mottled ducks. Wildlife Society Bulletin 17:222-227.

Johnson, D. H. 2008. In defense of indices: The case of bird surveys. The Journal of Wildlife Management 72:857-868.

Koneff, M. D., J. A. Royle, M. C. Otto, J. S. Wortham, and J. K. Bidwell. 2008. A double-observer method to estimate detection rate during aerial waterfowl surveys. The Journal of Wildlife Management 72:1641-1649.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, B. M. Bell. 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software 70:1-21.

Kumar, A. V., and M. B. Rice. 2021. Optimized survey design for monitoring protocols: A case study of waterfowl abundance. Journal of Fish and Wildlife Management 12:572-584.

Laake, J. L. 2013. RMark: An R Interface for analysis of capture-recapture data with MARK. AFSC Processed Rep. 2013-01, 25 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., Seattle, WA, USA.

Lougheed, L. W., A. Breault, and D. B. Lank. 1999. Estimating statistical power to evaluate ongoing waterfowl population monitoring. The Journal of Wildlife Management 63:1359-1369.

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.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. The Journal of Wildlife Management 53:1017-1024.

Martinson, R. K., and C. F. Kaczynski. 1967. Factors influencing waterfowl counts on aerial surveys, 1961-1966. U.S. Fish and Wildlife Service Special Science Report-Wildlife 105, Washington D.C., USA.

Masto, N. M. 2019. Evaluation of aerial surveys to monitor fall-winter waterbird populations in South Carolina. Thesis, Clemson University, Clemson, USA.

McClanahan, M. D., 2015. Habitat use and response to wetland management practices of non-breeding dabbling ducks in western Tennessee. Thesis, University of Tennessee-Knoxville, Knoxville, USA.

McKnight, S. K., and G. R. Hepp. 1998. Diet selectivity of gadwalls wintering in Alabama. The Journal of Wildlife Management 62:1533-1543.

Musgrove, A. J., and S. J. Holloway. 1997. Counting waterfowl on large estuaries at low tide. British Trust for Ornithology, National Centre for Ornithology, The Nunnery, Thetford, Norfolk, UK.

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. The Auk 117:393-408.

O'Donnel, K. M., F. R. Thompson III, and R. D. Semlitsch. 2015. Partitioning detectability components in populations subject to within-season temporary emigration using binomial mixture models. PLoS ONE 10: e0117216.

Osborn, J. M., H. M. Hagy, M. D. McClanahan, J. B. Davis, and M. J. Gray. 2017. Habitat Selection and activities of dabbling ducks during non-breeding periods. The Journal of Wildlife Management 81:1482-1493.

Osborne, D. C., R. J. Askren, S. D. Phelps, T. L. Osborne, J. R. Henson, and C. A. Nicolai. 2024. Variation in diurnal and nocturnal movements distances revealed in wintering dabbling ducks in the Mississippi Alluvial Valley. Wildlife Biology e01309.

Pagano, A. M., and T. W. Arnold. 2009*a*. Detection probabilities for ground-based breeding waterfowl surveys. The Journal of Wildlife Management 73:392-398.

Pagano, A. M., and T. W. Arnold. 2009*b*. Estimating detection probabilities of waterfowl broods from ground-based surveys. The Journal of Wildlife Management 73:686-694.

Paulus, S. L. 1984. Activity budgets of non-breeding gadwalls in Louisiana. The Journal of Wildlife Management 48:371-380.

Pearse, A. T. 2007. Design, evaluation, and applications of an aerial survey to estimate abundance of wintering waterfowl in Mississippi. Dissertation, Mississippi State University, Starkville, USA.

Pearse, A. T., S. J. Dinsmore, R. M. Kaminski, and K. J. Reinecke. 2008*a*. Evaluation of an aerial survey to estimate abundance of wintering ducks in Mississippi. The Journal of Wildlife Management 72:1413-1419.

Pearse A. T., P. D. Gerard, S. J. Dinsmore, R. M. Kaminski, K. J. Reinecke. 2008*b*. Estimation and correction of visibility bias in aerial surveys of wintering ducks. The Journal of Wildlife Management 72:808-813.

Pearse, A.T., R. M. Kaminski, K. J. Reinecke, and S. J. Dinsmore. 2012. Local and landscape associations between wintering dabbling ducks and wetland complexes in Mississippi. Wetlands 32:859-869.

Pollock, K. H., and W. L. Kendall. 1987. Visibility bias in aerial surveys: A review of estimation procedures. The Journal of Wildlife Management 51:502-510.

Powell, L. A., and G. A. Gale. 2015. Estimation of parameters for animal populations: A primer for the rest of us. Caught Napping Publications, Lincoln, Nebraska, USA.

Prenzlow, D. M., and J. R. Lovvorn. 1996. Evaluation of visibility correction factors for waterfowl surveys in Wyoming. The Journal of Wildlife Management 60:286-297.

Prenzlow, D. M., and J. R. Lovvorn. 1997. Design and results of a waterfowl breeding population survey for Wyoming. The Journal of Wildlife Management 61:758-767.

Reinecke, K. J., M. W. Brown, and J. R. Nassar. 1992. Evaluation of aerial transects for counting wintering mallards. The Journal of Wildlife Management 56:515-525.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ross, R. K. 1985. Helicopter versus ground surveys of waterfowl in the Boreal Forest. Wildlife Society Bulletin 13:153-157.

Roy, C., S. G. Gilliland, and E. T. Reid. 2022. A hierarchical dependent double-observer method for estimating waterfowl breeding pairs abundance from helicopters. Wildlife Biology 2022:e01003.

Schummer, M. L., A. D. Afton, S. S. Badzinski, S. A. Petrie, G. H. Olsen, M. A. Mitchell. 2018. Evaluating the waterfowl breeding and habitat survey for scaup. The Journal of Wildlife Management 82:1252-1262.

Sharp, D. E., K. L. Kruse, and P. P. Thorpe. 2002. The midwinter waterfowl survey in the Central Flyway. United States Fish and Wildlife Service, Division of Migratory Bird Management, Denver, CO, USA.

Shirkey, B. T., D. R. Luukkonen, and S. R. Winterstein. 2014. Application of distance sampling techniques for diving ducks on Lake St. Clair and western Lake Erie. Journal of Great Lakes Research 40:274-280.

Skalski, J. R., and D. S. Robson. 1992. Techniques for Wildlife Investigations: Design and Analysis of Capture Data. Academic Press, San Diego, California.

Smith, D. R., K. J. Reinecke, M. J. Conroy, M. W. Brown, and J. R. Nassar. 1995. Factors affecting visibility rate of waterfowl surveys in the Mississippi Alluvial Valley. The Journal of Wildlife Management 59:515-527.

Smith, G. W. 1995. A critical review of the aerial and ground surveys of breeding waterfowl in North America. United States Department of the Interior, National Biological Service, Washington D.C., USA.

Soulliere, G. J., B. W. Loges, E. M. Dunton, D. R. Luukkonen, M. W. Eichholz, and M. E. Koch. 2013. Monitoring waterfowl in the Midwest during the non-breeding period: Challenges, priorities, and recommendations. Journal of Fish and Wildlife Management 4:395-405.

Stoffel, M. A., S. Nakagawa, and H. Schielzeth. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods in Ecology and Evolution 8:1639-1644.

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8:246-257.

United States Fish and Wildlife Service [USFWS]. 2004. The history of North American waterfowl surveys 50 years and still counting. Division of Migratory Bird Management, Arlington, VA, USA.

United States Department of the Interior and Environment Canada. 1986. North American Waterfowl Management Plan. U.S. Department of the Interior. Washington D.C., USA.

United States Department of the Interior, Environment Canada, and Environment and Natural Resources Mexico. 1994. 1994 Update to the North American Waterfowl Management Plan Expanding the Commitment. U.S. Department of the Interior. Washington D.C., USA.

United States Department of the Interior, Environment Canada, and Environment and Natural Resources Mexico. 2018. 2018 North American Waterfowl Management Plan (NAWMP) Update, Connecting People, Waterfowl, and Wetlands. U.S. Department of the Interior. Washington D.C., USA.

United States Fish and Wildlife Service [USFWS]. 2007. Comprehensive conservation plan and environmental assessment for Wheeler National Wildlife Refuge Complex, Jackson, Lauderdale, Limestone, Madison, and Morgan counties, Alabama. U.S. Department of the Interior, Fish and Wildlife Service, Southeast Region, Atlanta, GA, USA.

United States Fish and Wildlife Service [USFWS]. 2022. Mid-winter waterfowl Survey. Division of Migratory Bird management, Washington D.C., USA.

United States Fish and Wildlife Service and Canadian Wildlife Service [USFWS and CWS]. 1987. Standard operating procedures for aerial waterfowl breeding ground population and habitat surveys in North America. U.S. Fish and Wildlife Service, Washington, D.C., USA.

Vrtiska, M. P., and L. A. Powell. 2011. Estimates of duck breeding populations in the Nebraska Sandhills using double observer methodology. Waterbirds 34:96-101.

Whitaker, J. M., K. M. Ringelman, J. R. Marty, W. Selman, and J. T. Linscombe. 2019. Changes in waterfowl abundance and species composition on Louisiana coastal wildlife management areas and refuges 2004-2016. Journal of the Southeastern Association of Fish and Wildlife Agencies 6:136-145.

White, E. R. 2019. Minimum time required to detect population trends: The need for long-term monitoring programs. BioScience 69:40-46.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Yetter, A. P., H. M. Hagy, M. M. Horath, J. D. Lancaster, C. S. Hine, R. V. Smith, and J. D. Stafford. 2018. Mallard survival, movements, and habitat use during autumn in Illinois. The Journal of Wildlife Management 82:182-191.

TABLES AND FIGURES

Table 1. Aerial Survey Dates. Four sets of aerial surveys were conducted each study year as weather and scheduling permitted. Each set contained an aerial survey conducted in the eastern and western portions of the Tennessee River Valley (TRV), and the Mobile Coast, Bay, and Delta (Mobile CBD) study areas in Alabama. Surveys that were missed due to weather conditions are noted by a dash. Where two dates are listed, the first indicates the TRV East survey and the second indicates the TRV West survey.

Aerial Survey Dates								
	Pre-Season		December		Midwinter		Late-Season	
Study Year	TRV	Mobile CBD	TRV	Mobile CBD	TRV	Mobile CBD	TRV	Mobile CBD
2021-22	Nov 17	Nov 17	Dec 13-14	Dec 15	Jan 5	Jan 6	Jan 25	Jan 26
2022-23	Dec 1	Nov 16	-	-	Jan 10	Jan 4	Jan 24	-
2023-24	Nov 16	Nov 27	Dec 12	Dec 14	Jan 2	Jan 4	Feb 2	-

Table 2. Observer Seat Position and Recording Device Rotation Used in Aerial Surveys. Each set of surveys is indicated by the season and study year it occurred in. Observers are indicated by the observer group they are categorized in for analyses. Missing flight sets are indicated by dashes.

Observer Seat Position and Recording Device Rotation Used in Aerial Surveys				
2021-2022				
	Front Seat		Rear Seat	
Survey Set	Observer Group	Recording Device Used	Observer Group	Recording Device Used
Pre-Season	A	Hodges H1	-	-
December	A	Hodges H1	B	Sony ICD-B300 IC
Midwinter	A	Hodges H1	B	Sony ICD-B300 IC
Late-Season	A	Hodges H1	B	Sony ICD-B300 IC
2022-2023				
	Front Seat		Rear Seat	
Survey Set	Observer Group	Recording Device Used	Observer Group	Recording Device Used
Pre-Season	A	Hodges H2	B	Hodges H1
December	-	-	-	-
Midwinter	A	Hodges H2	B	Hodges H1
Late-Season	B	Hodges H1	A	Hodges H2
2023-2024				
	Front Seat		Rear Seat	
Survey Set	Observer Group	Recording Device Used	Observer Group	Recording Device Used
Pre-Season	B	Hodges H1	C	Sony ICD-B300 IC
December	C	Sony ICD-B300 IC	B	WearHome V1.0
Midwinter	B	WearHome V1.0	C	Sony ICD-B300 IC
Late-Season	C	Sony ICD-B300 IC	B	WearHome V1.0

Table 3. Model-comparison results for the top, competing, and first non-competing models from AIC analyses of closed population mark-recapture Huggins models modeling detection probability estimates for double-observer crewed aerial surveys in Alabama.

Model	**#**	**AICc**	**DeltaAICc**	**Weight**	**Neg2LnL**
~Seat + Guild + Observer + Group Size	8	849.836	0.000	0.305	833.668
~Seat + Fine-Scale Group + Observer + Group Size	13	849.936	0.100	0.290	823.509
~Seat + Fine-Scale Group + Observer + Group Size + Season	16	850.649	0.813	0.203	818.008
~Seat + Guild + Observer + Group Size + Season	11	850.690	0.854	0.199	828.381
~Seat + Fine-Scale Group + Group Size + Season	14	862.125	12.289	0.001	833.631

Table 4. The beta (logit-scale) and real (probability-scale) coefficients from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. Intercept represents the estimate for ObsA in the front seat observing flocks in the dabbling duck guild.

Covariates	**β**	**SE**	**Lower CL**	**Upper CL**	**Real Estimate**
Intercept	-1.140	0.489	-2.096	-0.183	0.24
Back Seat	-0.361	0.135	-0.625	-0.098	0.41
Diving Duck	0.612	0.448	-0.266	1.49	0.65
Goose	2.02	0.499	1.043	2.997	0.88
Non-waterfowl	1.165	0.418	0.345	1.985	0.76
ObsB	-1.039	0.262	-1.552	-0.526	0.26
ObsC	-1.158	0.296	-1.738	-0.578	0.23
Group Size	0.318	0.071	0.179	0.457	0.58

Table 5. Detection probability estimates for all observers, seats, and groups from flock sizes 1, 100, and maximum observed group size from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. 95% Confidence limits are shown in parentheses.

ObsA						
	Front Seat Estimated Detectability for			**Back Seat Estimated Detectability for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1]**
Dabbling Duck	0.24 (0.11, 0.45)	0.58 (0.37, 0.76)	0.67 (0.46, 0.83)	0.18 (0.08, 0.38)	0.49 (0.28, 0.71)	0.59 (0.36, 0.79)
Diving Duck	0.37 (0.22, 0.56)	0.72 (0.56, 0.84)	0.81 (0.66, 0.90)	0.29 (0.15, 0.48)	0.64 (0.45, 0.79)	0.75 (0.55, 0.88)
Goose	0.71 (0.52, 0.85)	0.91 (0.82, 0.96)	0.92 (0.83, 0.97)	0.63 (0.41, 0.80)	0.88 (0.75, 0.95)	0.89 (0.76, 0.95)
Non-waterfowl	0.51 (0.36, 0.65)	0.82 (0.69, 0.90)	0.88 (0.76, 0.94)	0.42 (0.27, 0.58)	0.76 (0.59, 0.87)	0.83 (0.67, 0.93)
ObsB						
	Front Seat Estimated Detectability for			**Back Seat Estimated Detectability for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1]**
Dabbling Duck	0.10 (0.04, 0.22)	0.33 (0.19, 0.51)	0.42 (0.25, 0.62)	0.07 (0.03, 0.16)	0.25 (0.14, 0.42)	0.33 (0.19, 0.53)
Diving Duck	0.17 (0.09, 0.30)	0.48 (0.33, 0.62)	0.60 (0.43, 0.75)	0.13 (0.07, 0.22)	0.39 (0.27, 0.53)	0.51 (0.36, 0.68)
Goose	0.46 (0.29, 0.64)	0.79 (0.64, 0.89)	0.81 (0.66, 0.90)	0.37 (0.22, 0.55)	0.72 (0.55, 0.84)	0.75 (0.58, 0.86)
Non-waterfowl	0.27 (0.18, 0.37)	0.61 (0.48, 0.73)	0.72 (0.56, 0.84)	0.20 (0.14, 0.28)	0.52 (0.39, 0.65)	0.64 (0.47, 0.79)
ObsC						
	Front Seat Estimated Detectability for			**Back Seat Estimated Detectability for**		

Group	Individual Birds	Flock Size = 100	Maximum Observed Flock Size[1]	Individual Birds	Flock Size = 100	Maximum Observed Flock Size[1]
Dabbling Duck	0.09 (0.04. 0.20)	0.30 (0.17, 0.48)	0.39 (0.22, 0.59)	0.07 (0.03, 0.15)	0.23 (0.13, 0.39)	0.31 (0.17, 0.50)
Diving Duck	0.16 (0.08, 0.28)	0.45 (0.30, 0.60)	0.57 (0.39, 0.73)	0.11 (0.06, 0.21)	0.36 (0.24, 0.50)	0.48 (0.32, 0.65)
Goose	0.43 (0.26, 0.62)	0.77 (0.60, 0.88)	0.79 (0.63, 0.89)	0.35 (0.20, 0.53)	0.70 (0.52, 0.83)	0.72 (0.55, 0.85)
Non-waterfowl	0.24 (0.16, 0.35)	0.58 (0.45, 0.71)	0.70 (0.53, 0.83)	0.18 (0.12, 0.26)	0.49 (0.37, 0.62)	0.62 (0.45, 0.77)

[1]Maximum flock size observed for Dabbling Duck = 350, Goose = 150, all others reached or exceeded 500 birds

Table 6. Beta coefficients, standard errors (SE), and P-values from linear regression models evaluating the effect of sources of error and survey effort of waterfowl aerial surveys on values of coefficient of variation (CV), relative bias, mean error, and power.

	CV (%)		**Relative Bias (%)**		**Mean Error (Birds)**		**MSE (1,000,000 Birds^2)**		**Power**	
	Beta (SE)	P	Beta (SE)	P	Beta (SE)	P	Beta (SE)	P	Beta (SE)	P
Intercept	53.29 (0.78)	<0.001	-69.27 (1.43)	<0.001	29726.13 (587.78)	<0.001	834.85 (23.94)	<0.001	54.26 (0.97)	<0.001
High Detectability	-0.187 (0.55)	0.736	7.15 (1.01)	<0.001	-2251.59 (415.63)	<0.001	-94.84 (16.93)	<0.001	-0.028 (0.69)	0.967925
Low Count Error	-13.17 (0.55)	<0.001	-6.77 (1.01)	<0.001	174.84 (415.63)	0.675413	18.79 (16.93)	0.27121	-4.42 (0.69)	<0.001
High Availability	0.0033 (0.55)	0.995	34.08 (1.01)	<0.001	-10049.98 (415.63)	<0.001	-431.18 (16.93)	<0.001	-0.083 (0.69)	0.903986
2Observers	-10.48 (0.68)	<0.001	-0.35 (1.23)	0.776	-49.17 (509.04)	0.923346	-2.53 (20.73)	0.90338	-3.042 (0.84)	<0.001
2Flights	-12.60 (0.68)	<0.001	-0.78 (1.23)	0.531	-1249.89 (509.04)	0.016799	-47.27 (20.73)	0.02595	-2.04 (0.84)	0.018249
3Flights	-18.23 (0.68)	<0.001	-0.025 (1.23)	0.984	-1807.38 (509.04)	<0.001	-65.93 (20.73)	0.00227	-4.46 (0.84)	<0.001
Paired	10.43 (0.68)	<0.001	13.43 (1.23)	<0.001	-2866.48 (509.04)	<0.001	-123.44 (20.73)	<0.001	2.96 (0.84)	<0.001

Table 7. R^2 values from linear regression models evaluating the effect of sources of error and survey effort of waterfowl aerial surveys on values of coefficient of variation (CV), percent relative bias, mean error (ME), mean squared error (MSE), and power.

Terms	CV	% Rel Bias	ME	MSE	Power
Detectability	0.0	3.4	4.0	3.9	0.0
CountError	33.2	3.1	0.0	0.2	25.6
Availability	0.0	78.5	79.6	80.7	0.0
Permutation	64.1	10.6	8.8	8.3	35.9
Detectability:CountError	0.0	0.0	0.3	0.3	0.0
Detectability:Availability	0.0	0.4	0.1	0.0	0.1
Detectability:Permutation	0.0	0.9	0.6	0.6	0.6
CountError:Availabillity	0.0	0.2	0.9	0.9	0.0
CountError:Permutation	2.6	0.7	2.1	1.8	35.9
Availability:Permutation	0.0	1.1	1.5	1.8	0.5
Residuals	0.0	1.1	2.1	1.6	1.5

Table 8. Overall average rankings of simulated aerial waterfowl survey situations where row names represent levels of availability, detectability, and count error, respectively and column names represent the rank of each survey situation within those levels of sources of error. Cell values should be read as number of observers, number of surveys, paired (P) or unpaired (U) data, and (average rank). Cells are color coded such that average survey situation ranks ≤19 are dark green, 19< ranks ≥27 are light green, 27< ranks ≥35 have no fill, 35< ranks ≥43 are light red, and ranks >43 are dark red.

	1	2	3	4	5	6	7	8	9
High, High, Low	2,3,P (9.8)	2,3,U (10.5)	1,3,U (10.8)	2,2,P (10.8)	2,2,U (11)	2,1,P (19.5)	2,1,U (20.3)	1,2,U (21)	1,1,U (25.3)
High, Low, Low	2,3,P (11.5)	2,2,P (15)	1,3,U (18.5)	2,3,U (19.5)	2,2,U (20)	2,1,P (22.5)	1,2,U (23)	2,1,U (26)	1,1,U (32)
High, High, High	2,3,P (14.3)	2,3,U (15.5)	1,3,U (18.3)	2,2,U (19.3)	2,2,P (23.8)	1,2,U (24.8)	2,1,U (25.8)	2,1,P (29.5)	1,1,U (31.3)
High, Low, High	2,3,P (16.5)	2,3,U (20.3)	2,2,P (22.3)	2,2,U (22.8)	1,3,U (29.5)	2,1,P (30.3)	1,2,U (31.3)	2,1,U (33.5)	1,1,U (36)
Low, High, Low	2,3,P (27.5)	2,3,U (28)	2,2,P (30)	1,3,U (34)	2,2,U (34.3)	2,1,P (36.3)	1,2,U (38)	2,1,U (39.3)	1,1,U (43.5)
Low, High, High	2,3,P (31.8)	1,3,U (34)	2,3,U (34.5)	2,2,U (37.5)	2,1,P (38.5)	2,2,P (39.3)	2,1,U (42.3)	1,2,U (43)	1,1,U (49.3)
Low, Low, Low	2,3,P (32.8)	2,2,P (33.8)	2,3,U (35.5)	2,2,U (38)	2,1,P (40.8)	1,3,U (42)	1,2,U (44)	2,1,U (45.5)	1,1,U (52)
Low, Low, High	2,3,P (33)	2,3,U (37.3)	2,2,P (40.3)	2,2,U (42.3)	1,3,U (43.3)	2,1,P (44)	1,2,U (50)	2,1,U (51.3)	1,1,U (51.8)

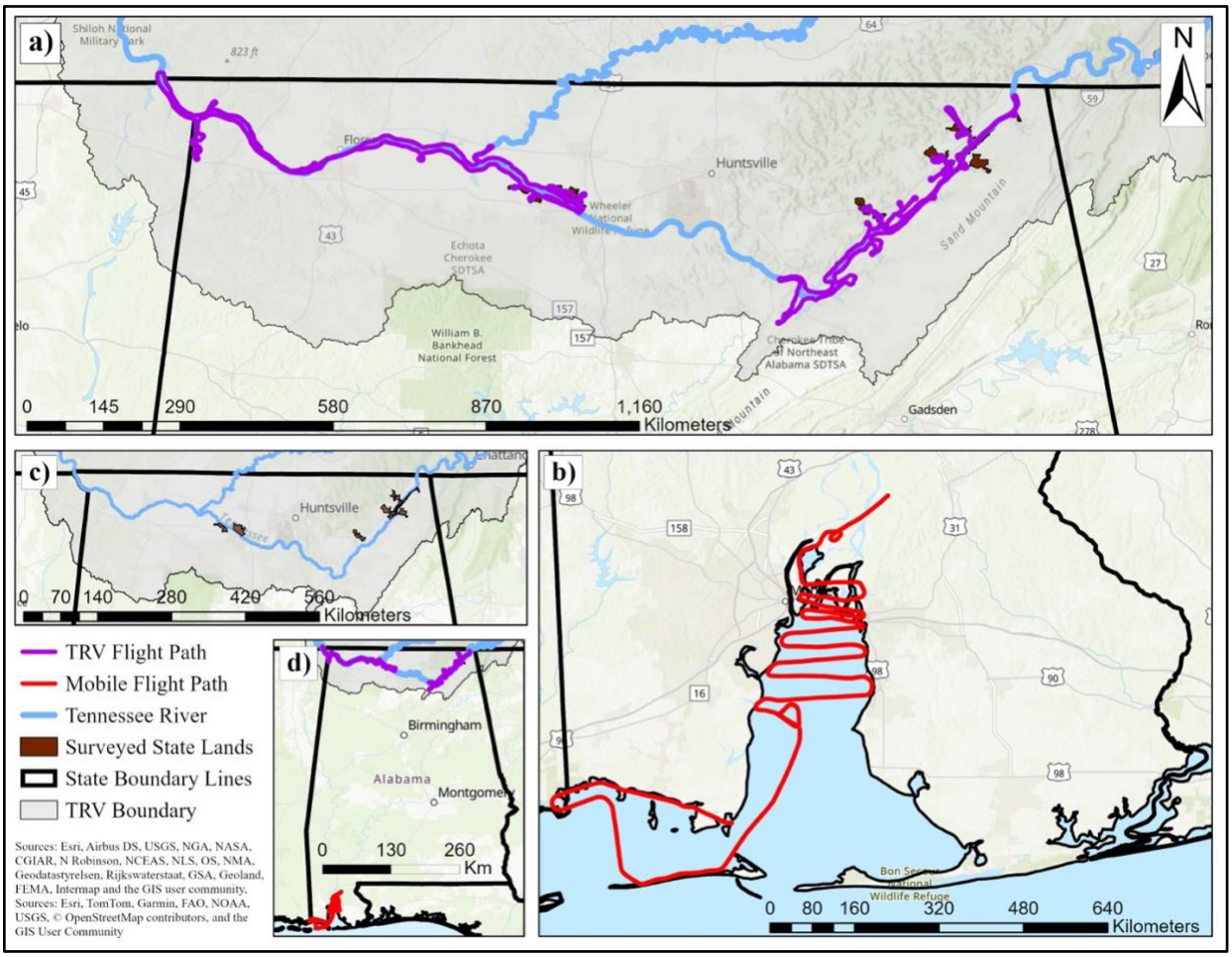

Figure 1. Map of aerial survey locations of wintering waterfowl population flights conducted by the Alabama Wildlife and Freshwater Fisheries Division from November–February 2021–2024. Pane a) Tennessee River Valley (TRV) East and West aerial survey flight paths. Pane b) Mobile Coast, Bay and Delta aerial survey flight path. Pane c) Tennessee River, TRV, and additional surveyed state lands of interest. Pane d) Alabama state boundary with TRV and flight path layers superimposed.

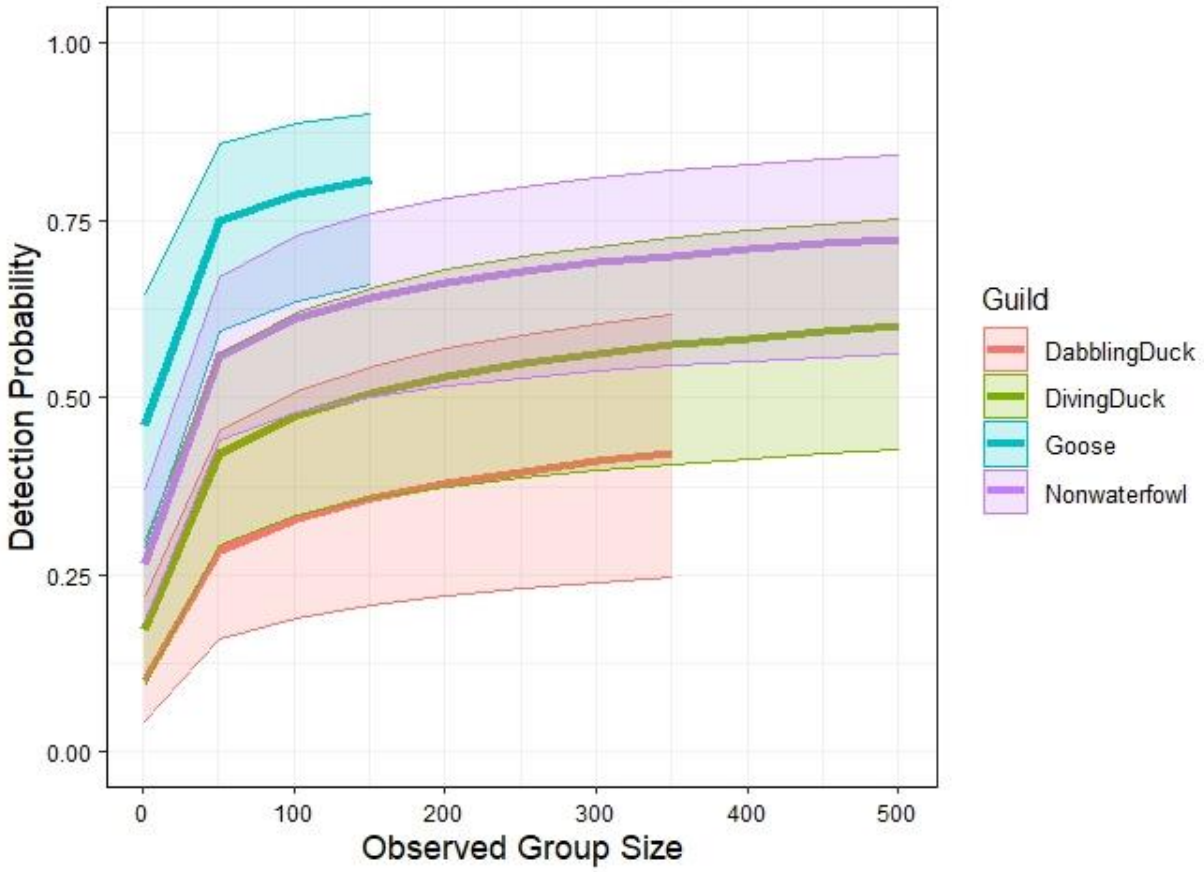

Figure 2. Detection probability estimates of four observed guilds over increasing group sizes from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front, observer constant at group A, and for modeled group sizes closest to that of the largest group size observed.

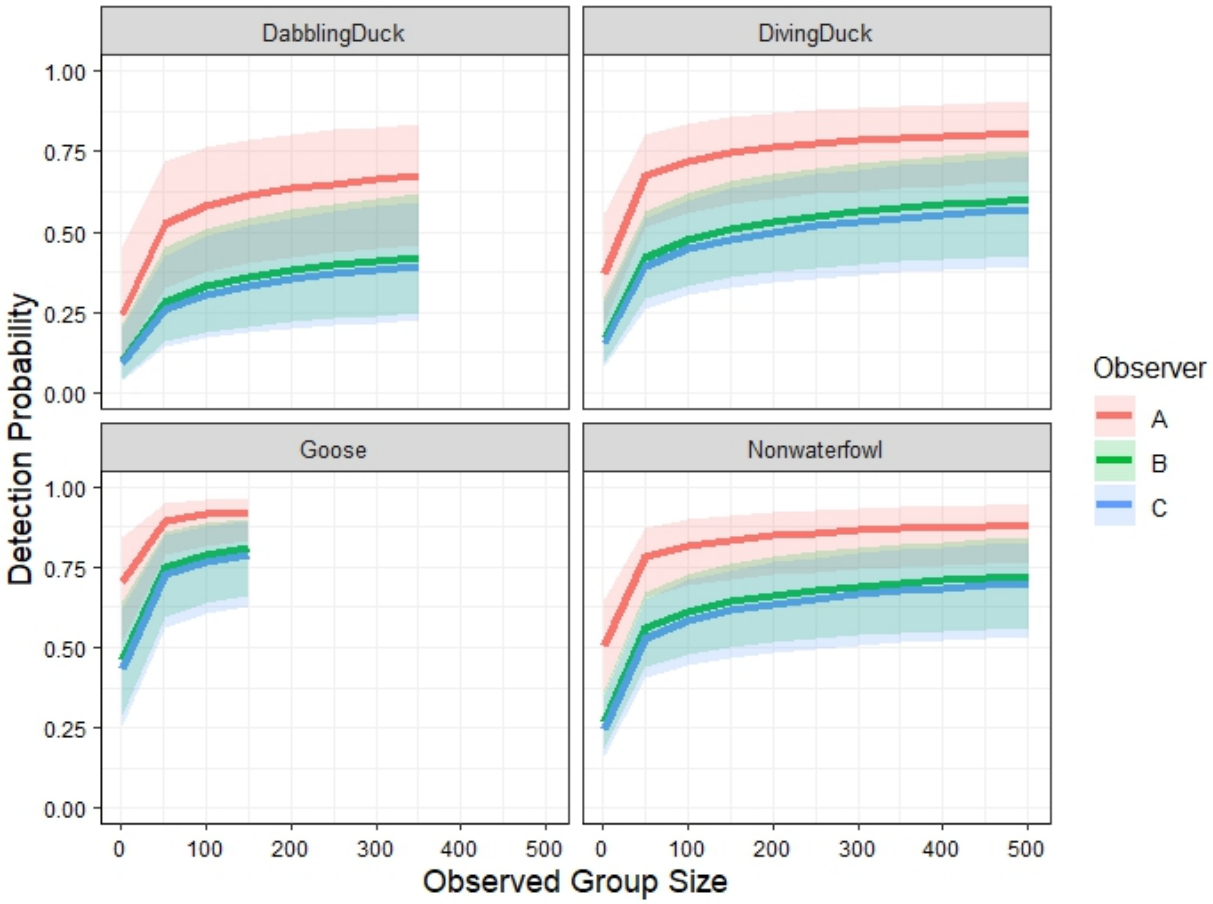

Figure 3. Detection probability estimates of three observer groups over increasing group sizes by guild from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front and for modeled group sizes closest to that of the largest group size observed.

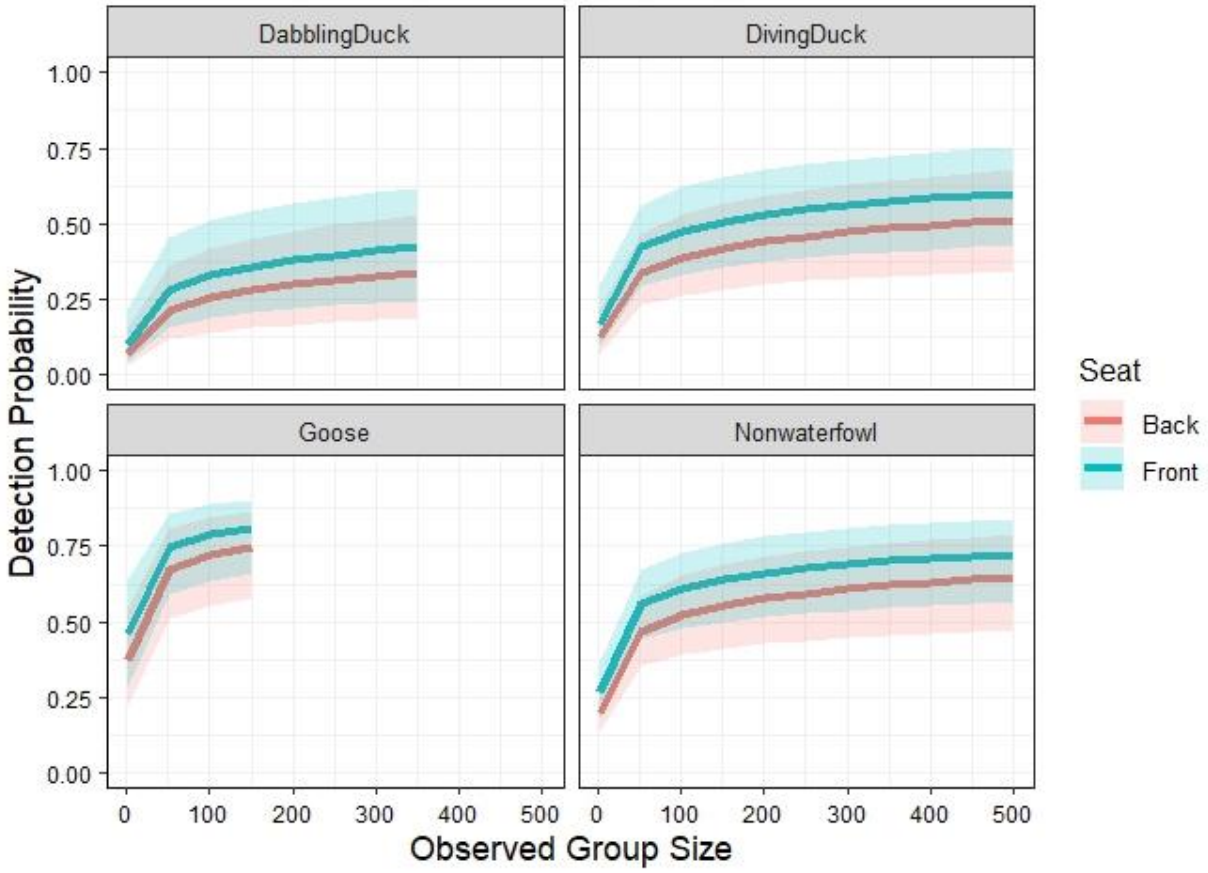

Figure 4. Detection probability estimates of two seat positions over increasing group sizes by guild from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding observer constant at group A, guild held constant at Goose, and for modeled group sizes closest to that of the largest group size observed.

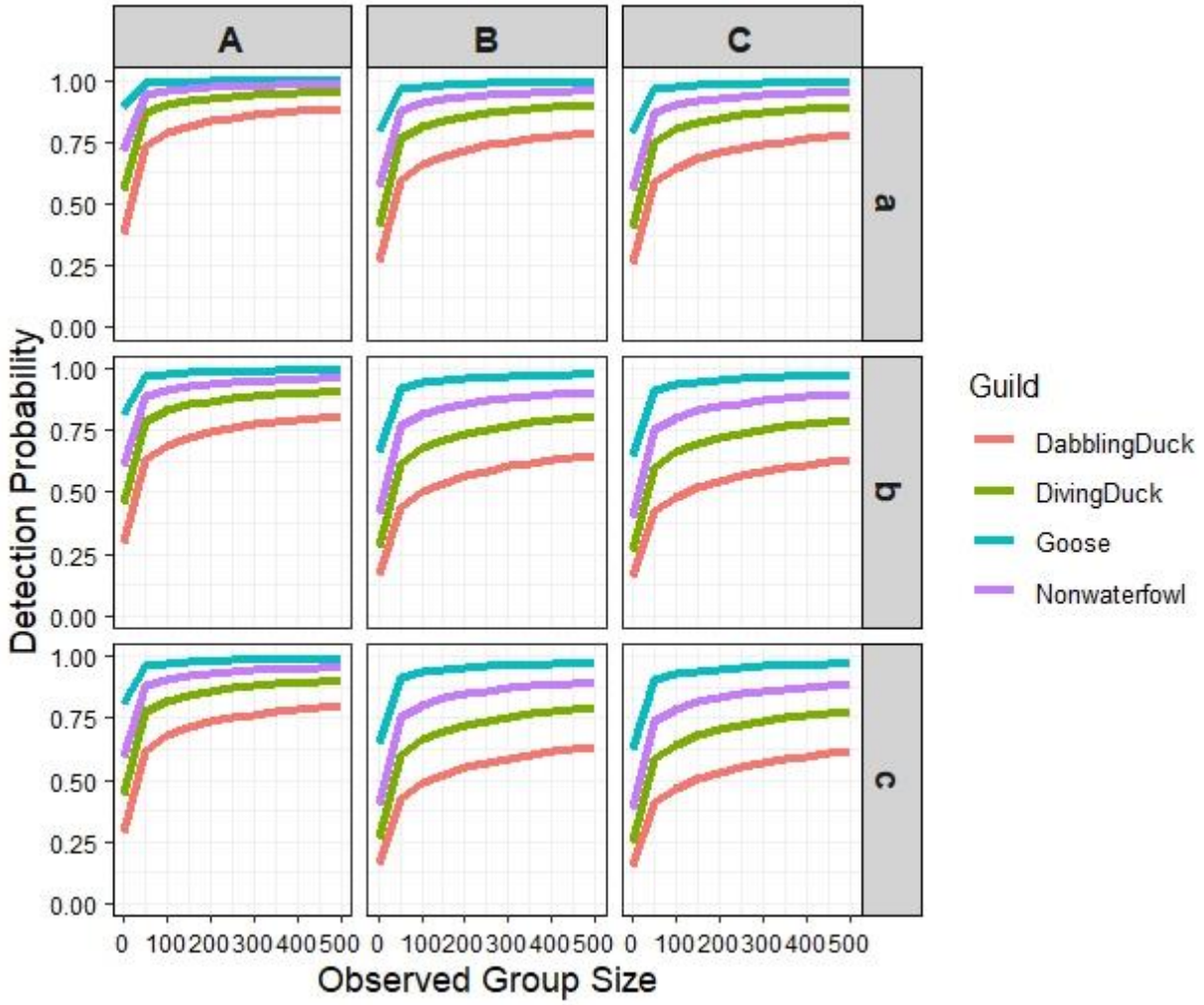

Figure 5. Overall detection probability estimates of four guilds over increasing group sizes from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Letters denote observer group pairs (A – high experience, B – medium experience, and C – low experience) for the front (capital letters) and back (lower case) seat positions in aerial surveys. The middle diagonal from top left to bottom right represents hypothetical overall detection probability if the same observer or one of equal experience surveyed in both the front and back seat.

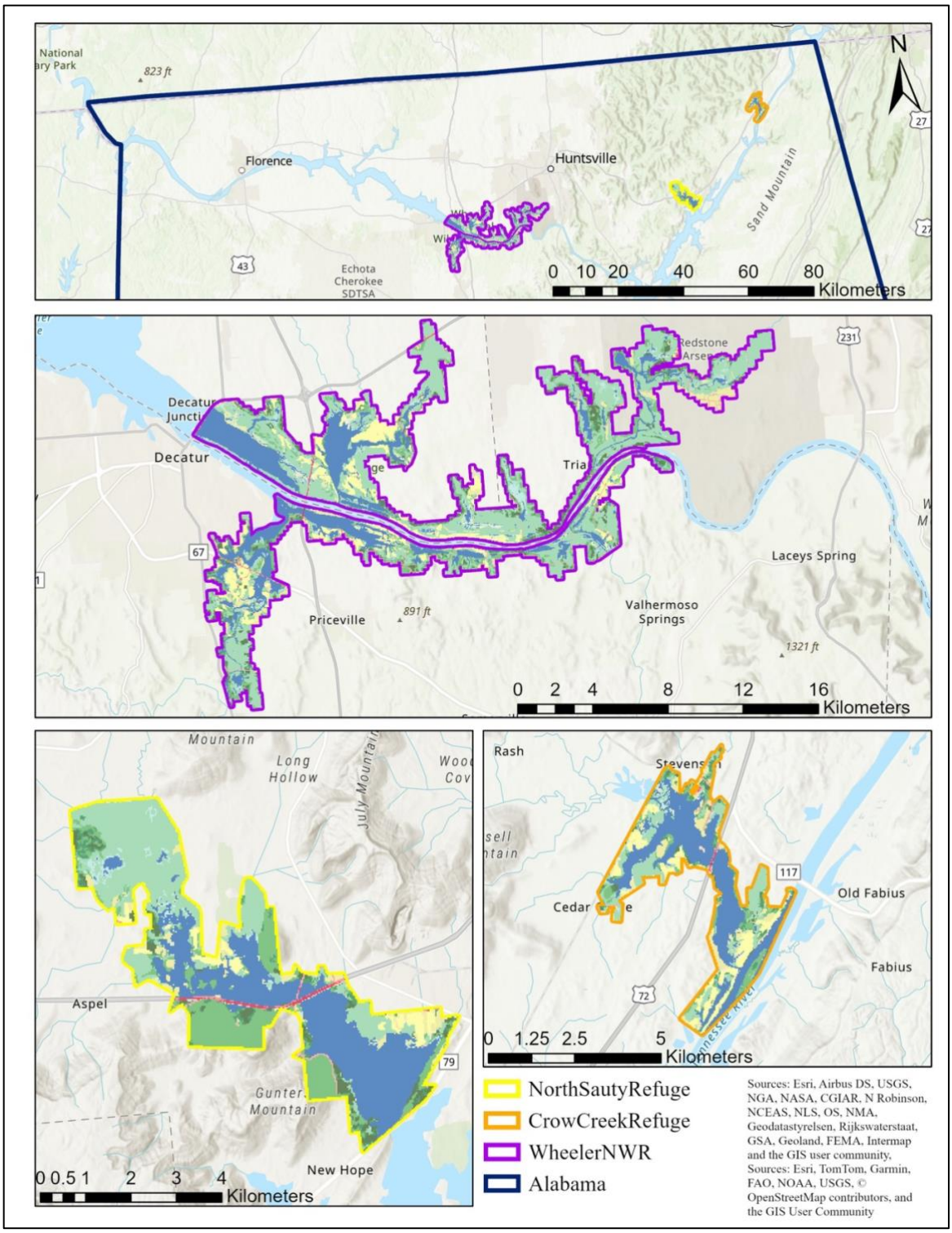

Figure 6. Gadwall trapping and tracking sites: Wheeler National Wildlife Refuge (NWR), North Sauty Refuge, and Crow Creek Refuge (top pane) November–February 2017–2019 to assess gadwall selection of areas within aerial waterfowl survey areas. Middle Pane: Wheeler NWR

with USDA National Land Cover designations. Bottom left pane: North Sauty Refuge with USDA National Land Cover designations. Bottom right pane: Crow Creek Refuge with USDA National Land Cover designations.

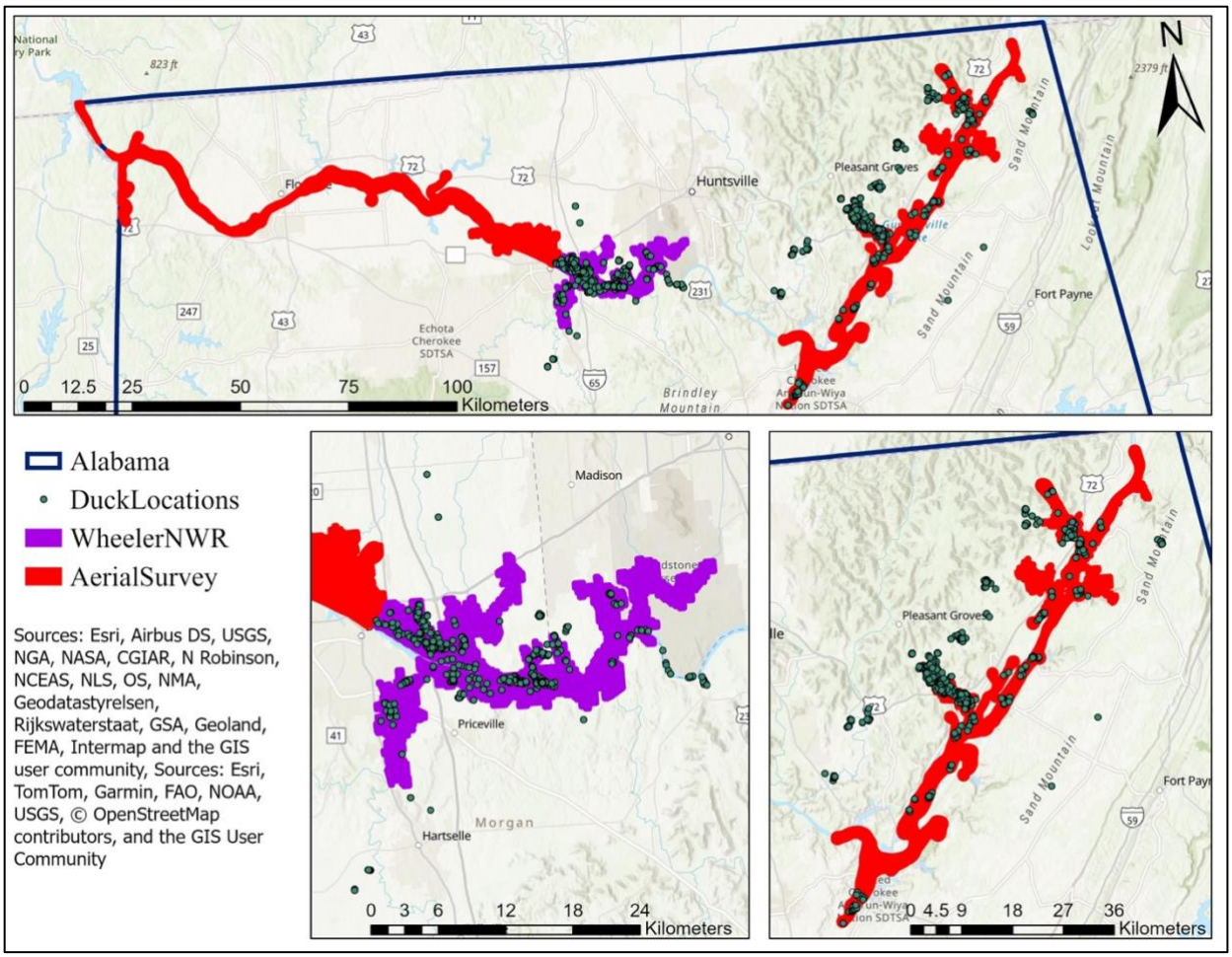

Figure 7. Diel locations of tracked gadwall in north Alabama November–February 2017–2019 to assess their selection of areas within aerial waterfowl survey areas. Locations are mapped over the Wheeler National Wildlife Refuge (NWR) boundary and aerial waterfowl survey areas. Top pane: Tennessee River Valley (TRV) aerial survey area with Wheeler NWR and diel gadwall locations. Bottom left pane: Wheeler NWR and diel gadwall locations. Bottom right pane: TRV East aerial survey area with gadwall locations.

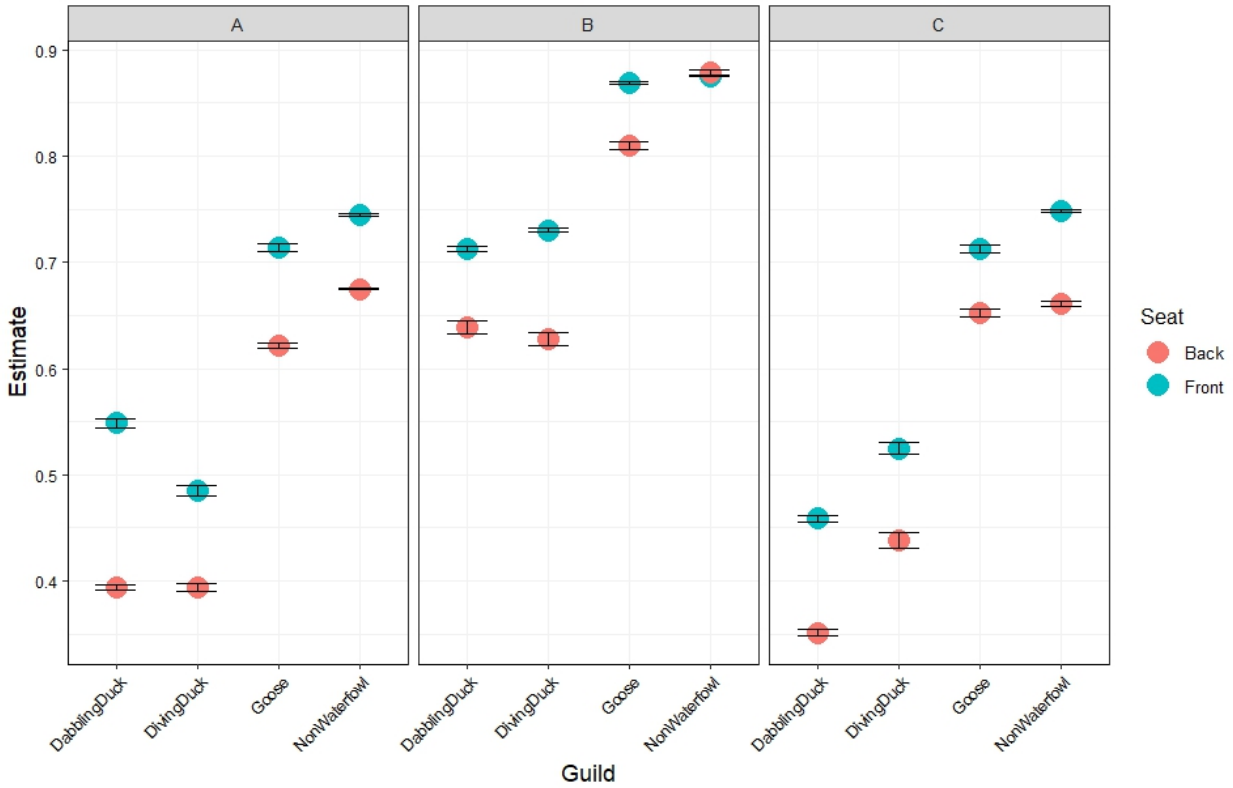

Figure 8. Survey-level detection probability estimates by observer (A, B, C), seat (front, back), and guild (dabbling duck, diving duck, goose, and non-waterfowl). Estimates were calculated from overall weighted average group-level detection estimates of all groups present in an aerial survey. Group-level detection was estimated using mark-recapture Huggins models fit to crewed aerial survey data in Alabama during November–February 2021–2024. Confidence intervals are shown at the 95% level.

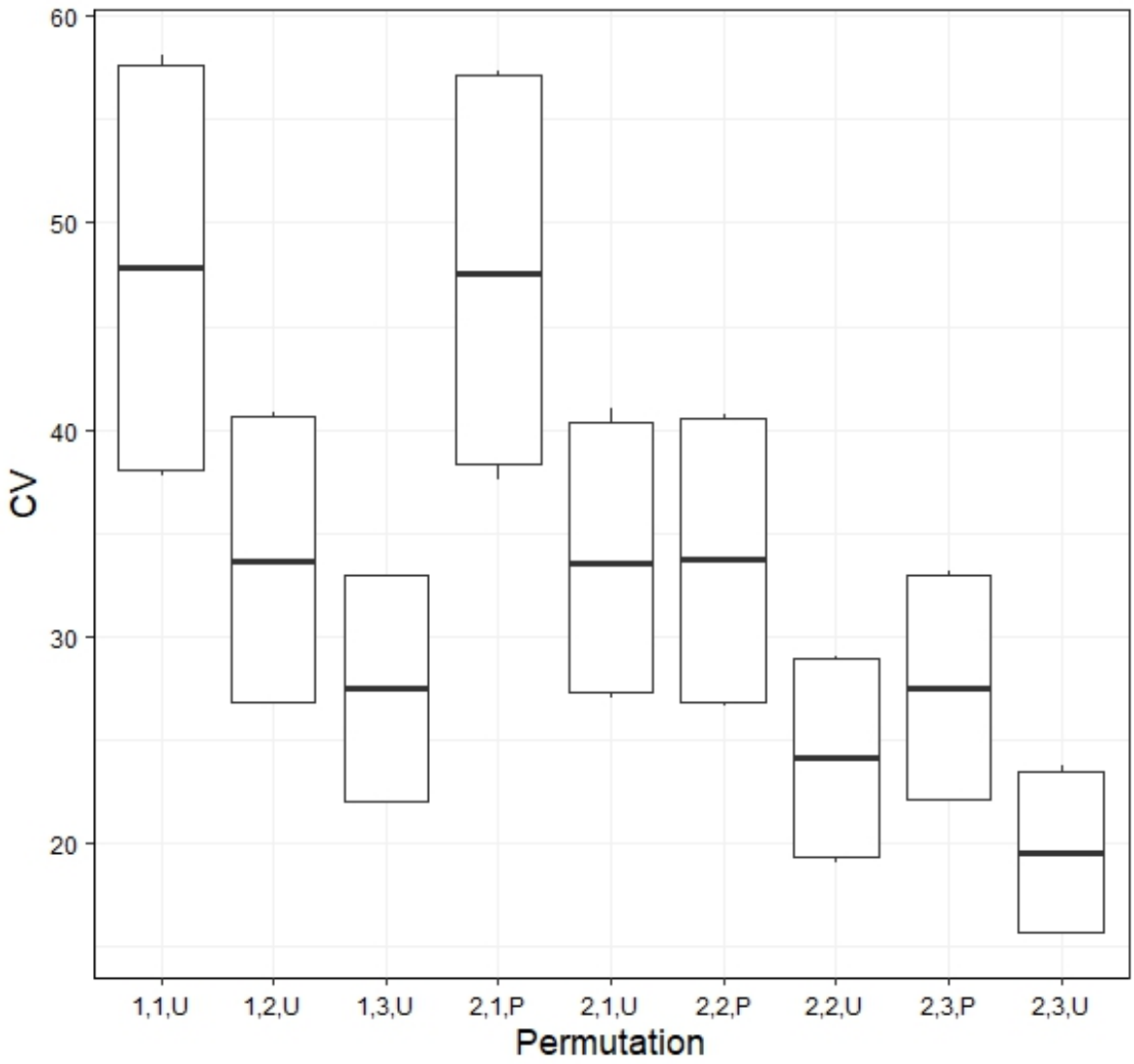

Figure 9. Coefficients of variation (CV) among survey designs ("permutation") from a linear regression modeling the effect of each survey design aspect from simulated aerial waterfowl surveys on CV. Survey designs are read as the number of observers (1-2), number of surveys (1-3), and if data was paired (P; double-observer mark-recapture estimation assumed) or unpaired (U; no mark-recapture estimation).

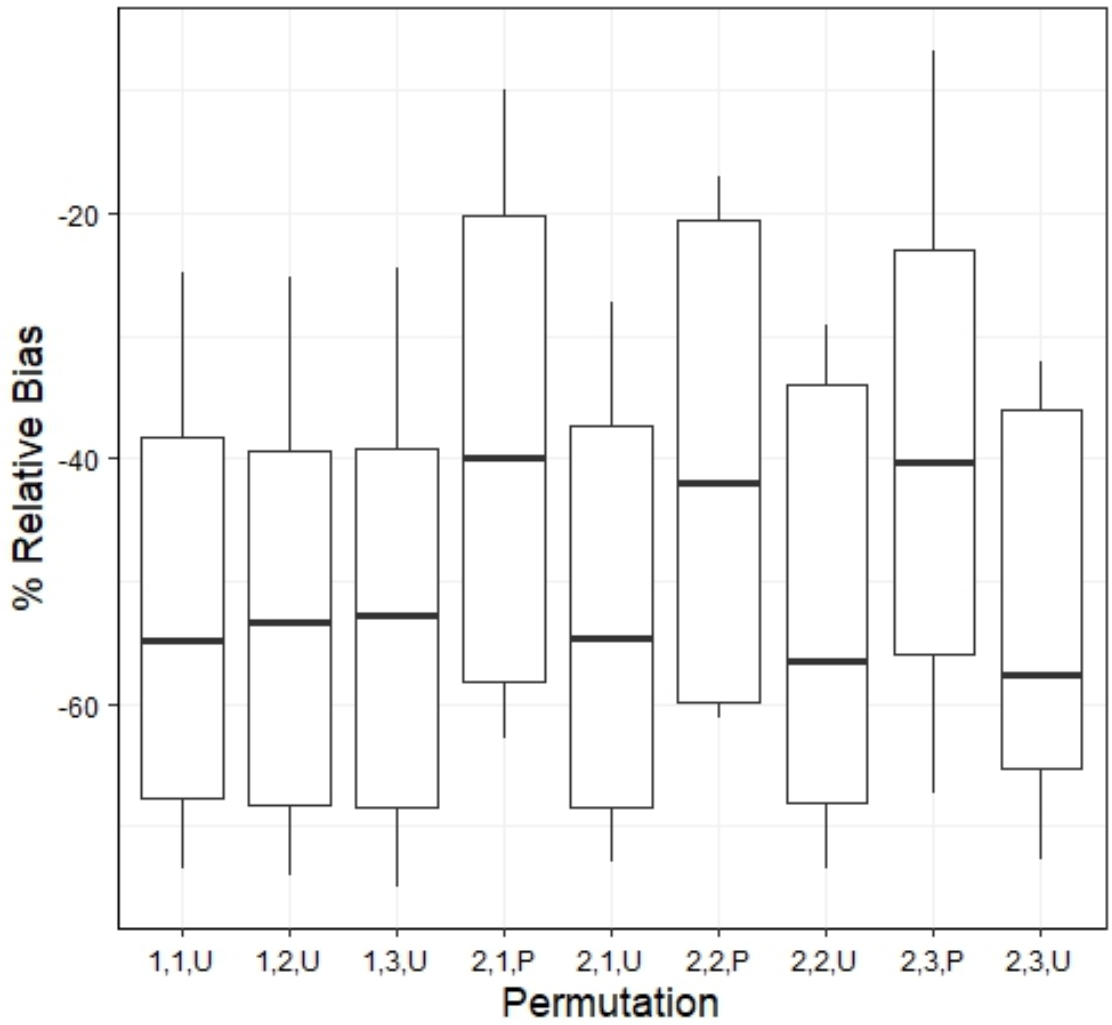

Figure 10. Percent relative bias by survey design ("permutation") from a linear regression modeling the effect of each survey design aspect from simulated aerial waterfowl surveys on percent relative bias. Survey designs are read as the number of observers (1-2), number of surveys (1-3), and if data was paired (P; double-observer mark-recapture estimation assumed) or unpaired (U; no mark-recapture estimation).

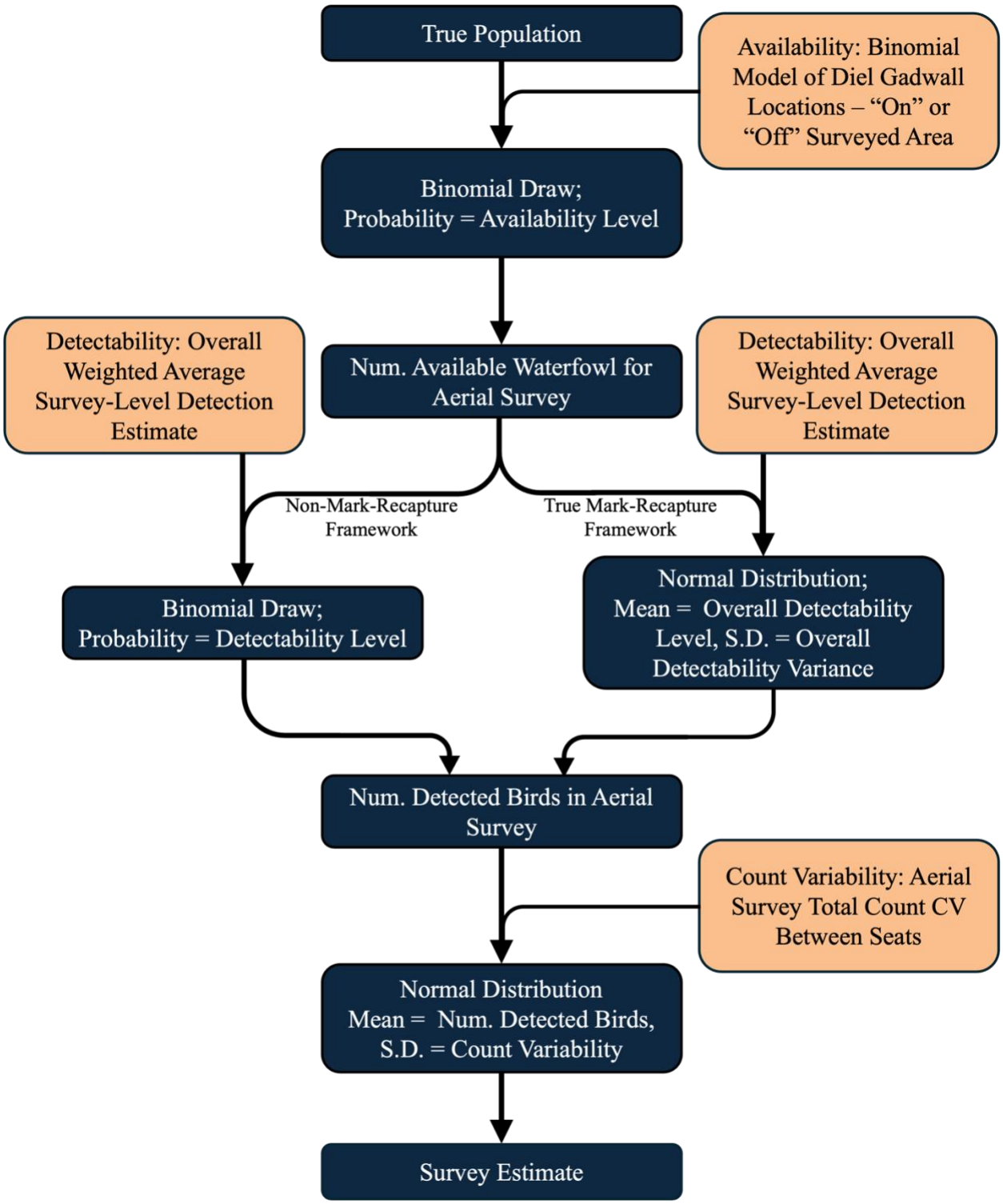

Figure 11. Graphic representation of the steps (dark-blue boxes) by which waterfowl aerial survey sources of error (availability, detectability, and count variability; peach-colored boxes) were used as inputs to simulate waterfowl aerial survey estimates to assess the relative influence

of the sources of error on waterfowl aerial survey estimates in Alabama 2017-2024. Arrows represent direction of flow. For each box with “binomial draw” or “normal distribution” titles, the boxes directly above represent the inputs for that “binomial draw” or “normal distribution” and the box directly below represents the output from that “binomial draw” or “normal distribution”.

Appendix A. Condensed Results From Full Data Set Analyses.

From AIC analyses conducted on the closed population mark recapture Huggins full (non-subset) model set, there is clear support that a fine-scale-level grouping variable, seat position, a grouped observer variable, and a log-transformed observed group size variable have significant effects on group-level detectability (Table 1A). The following results are from the full analysis and will discuss the top model and its beta coefficients (Table 2A).

We estimated detection probability estimates for nine fine-scale groups within our aerial surveys: *Anas*, Other Dabblers, *Aythya*, *Bucephala*, *Branta* + *Grus*, *Fulica*, *Haliaeetus*, *Pelecanus*, and *Phalacrocorax*. When observer and seat were held constant and at flock size = 100 birds, *Pelecanus* were estimated to have the highest detection probability of all taxonomic groups at 0.79. The *Branta* + *Grus* group had the next highest estimated detectability at 0.75 followed by *Bucephala* at 0.64, *Phalacrocorax* at 0.54, *Fulica* at 0.49, *Anas* at 0.40, Other Dabblers at 0.36 and lastly, *Aythya* at 0.34 (Figure 1A). Individual *Haliaeetus* were estimated to have a 0.30 detectability. *Pelecanus* were 1.26 times as likely to be detected compared to flocks of the *Branta* + *Grus* group and 7.35 times as likely as the *Aythya* group. All combinations of comparisons of detection probabilities between fine-scale groupings can be found in Table 3A.

Observer group A (ObsA) had the highest detectability in both seats, and across all fine-scale groups and flock sizes (Figure 2A). With other covariates held constant, groups were 0.88 and 0.55 times as likely to be observed by ObsB or ObsC than by ObsA respectively, and detection was 0.45 times as likely for ObsC than ObsB. Detection probability estimate comparisons between observers, seats and flock sizes for all groups can be found in Table 4A. Group detection was 0.73 times as likely in the back seat compared to the front seat when observer and guild were held constant; however, this changed among groups and observers.

Detection differences between the front and back seat were smallest when flocks were identified as being in the *Aythya* group and were greatest when flocks were identified as being in the *Pelecanus* group when holding observer constant. At constant observer and guild, front and back seat detection probability estimates were 0.46 and 0.38 for individual birds in each seat respectively. These estimates increased to 0.79 and 0.73 for flocks 100 birds in size, and up to 0.86 and 0.82 for flocks 500 birds in size (Figure 3A).

TABLES AND FIGURES

Table 1A. Model-comparison results for the top, competing, and first non-competing models from AIC analyses of closed population mark-recapture Huggins models modeling detection probability estimates for double-observer crewed aerial surveys in Alabama from a full dataset.

Model	#Parameters	AICc	DeltaAICc	Weight	Neg2LnL
~ Seat + Fine-Scale + Observer + Group Size	13	2828.32	0.00	0.69	2802.19
~ Seat + Fine-Scale + Observer + Group Size + Season	16	2829.88	1.55	0.32	2797.68
~ Seat + Fine-Scale + Group Size + Season	14	2864.20	35.87	1.11E-08	2836.05

Table 2A. The beta (log-scale) and real (probability-scale) coefficients from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. Intercept represents the estimate for ObsA in the front seat observing flocks in the *Anas* fine-scale taxonomic group.

Covariates	**β**	**SE**	**Lower CL**	**Upper Cl**	**Real Estimate**
Intercept	-1.78	0.42	-2.61	-0.95	0.14
Back Seat	-0.32	0.08	-0.48	-0.17	0.42
Aythya	-0.26	0.51	-1.25	0.73	0.44
Branta + *Grus*	1.50	0.43	0.67	2.34	0.82
Bucephala	0.99	0.48	0.04	1.93	0.73
Fulica	0.20	0.42	-0.63	1.03	0.55
Haliaeetus	1.06	0.54	0.00	2.11	0.74
Other Dabblers	-0.18	0.46	-1.08	0.73	0.46
Pelecanus	1.74	0.43	0.90	2.57	0.85
Phalacrocorax	0.55	0.42	-0.27	1.37	0.63
ObsB	-0.13	0.12	-0.37	0.11	0.47
ObsC	-0.59	0.13	-0.84	-0.33	0.36
Group Size	0.32	0.04	0.24	0.41	0.58

Table 3A. Fine-Scale Group Detectability Comparisons. Group names in the vertical column are cell value times as likely to be detected compared to group names in the corresponding horizontal column when observer is held constant to group A and seat is held constant to front.

	Pelecanus	***Branta* + *Grus***	***Bucephala***	***Phalacrocorax***	***Fulica***	***Anas***	**Other Dabblers**	***Aythya***	***Haliaeetus***
Pelecanus	-	1.26	2.12	3.28	4.65	5.67	6.77	7.35	1.97
Branta + *Grus*	0.79	-	1.70	2.60	3.69	4.50	5.37	5.83	1.56
Bucephala	0.47	0.60	-	1.55	2.20	2.68	3.20	3.47	0.93
Phalacrocorax	0.30	0.38	0.65	-	1.42	1.73	2.06	2.24	0.60
Fulica	0.22	0.27	0.46	0.71	-	1.22	1.46	1.58	0.42
Anas	0.18	0.22	0.37	0.58	0.82	-	1.19	1.30	0.35
Other Dabblers	0.15	0.19	0.31	0.48	0.69	0.84	-	1.09	0.29
Aythya	0.14	0.17	0.29	0.45	0.63	0.77	0.92	-	0.27
Haliaeetus	0.51	0.64	1.07	1.66	2.36	2.88	3.43	3.73	-

Table 4A. Detection probability estimates for all observers, seats, and groups from flock sizes 1, 100, and maximum observed group size from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. Dashes indicate calculated estimates with no biological meaning.

ObsA						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**
Pelecanus	0.49	0.81	0.88	0.41	0.76	0.84
Branta + *Grus*	0.43	0.77	0.84	0.36	0.71	0.79
Bucephala	0.31	-	0.62	0.25	-	0.54
[3]*Phalacro*	0.23	0.57	0.69	0.17	0.49	0.61
Fulica	0.17	0.48	0.61	0.13	0.40	0.53
Anas	0.14	0.43	0.53	0.11	0.35	0.45
Other Dabblers	0.12	0.39	0.52	0.09	0.31	0.44
Aythya	0.12	0.37	0.50	0.09	0.30	0.42
Haliaeetus	0.33	-	0.33	0.26	-	0.26
ObsB						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**
Pelecanus	0.46	0.79	0.86	0.38	0.73	0.82
Branta + *Grus*	0.40	0.75	0.82	0.33	0.68	0.77
Bucephala	0.29	-	0.59	0.22	-	0.51
[3]*Phalacro*	0.21	0.54	0.66	0.16	0.45	0.58
Fulica	0.15	0.45	0.58	0.12	0.37	0.50
Anas	0.13	0.40	0.50	0.10	0.33	0.42

Other Dabblers	0.11	0.36	0.48	0.08	0.29	0.40
Aythya	0.10	0.34	0.46	0.08	0.27	0.38
Haliaeetus	0.30	-	0.30	0.24	-	0.24

ObsC

Group	Front Seat Estimated Detection for Individual Birds	Front Seat Estimated Detection for Flock Size = 100	Front Seat Estimated Detection for Maximum Observed Flock Size[1,2]	Back Seat Estimated Detection for Individual Birds	Back Seat Estimated Detection for Flock Size = 100	Back Seat Estimated Detection for Maximum Observed Flock Size[1,2]
Pelecanus	0.35	0.70	0.80	0.28	0.63	0.74
Branta + *Grus*	0.30	0.65	0.75	0.23	0.58	0.68
Bucephala	0.20	-	0.47	0.15	-	0.39
[3]*Phalacro*	0.14	0.42	0.55	0.11	0.34	0.47
Fulica	0.10	0.34	0.46	0.07	0.27	0.38
Anas	0.09	0.30	0.39	0.06	0.23	0.31
Other Dabblers	0.07	0.26	0.37	0.05	0.20	0.30
Aythya	0.07	0.24	0.35	0.05	0.19	0.28
Haliaeetus	0.21	-	0.21	0.16	-	0.16

[1]Maximum flock size observed for *Anas* = 350, *Branta* + *Grus* = 400, *Bucephala* = 55, *Haliaeetus* = 12, all others reached or exceeded 500 birds.

[2]Estimates shown from modeled group sizes closest to that of the largest group size observed

[3]*Phalacro* short for *Phalacrocorax*

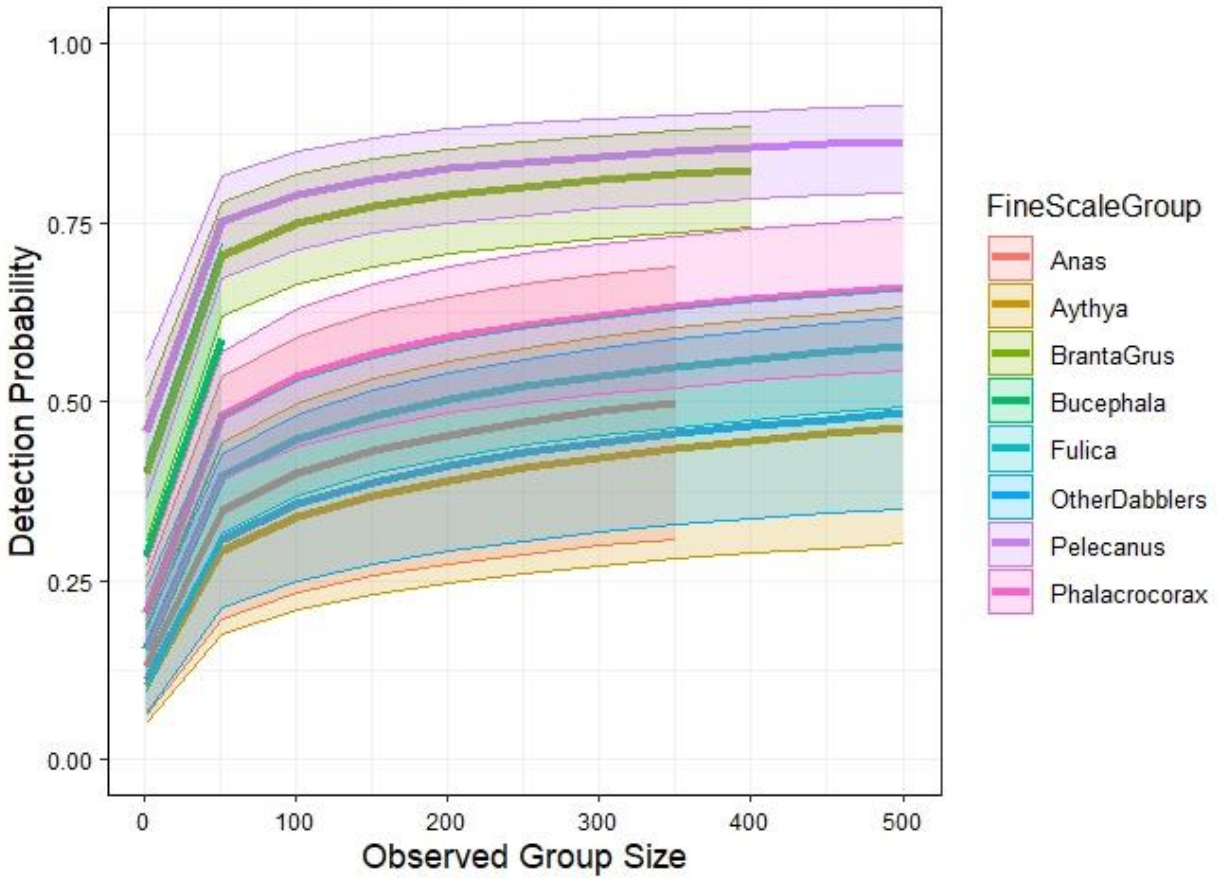

Figure 1A. Detection probability estimates of eight observed fine-scale groups over increasing group sizes from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front, observer constant at group A, and for modeled group sizes closest to that of the largest group size observed. *Haliaeetus* was removed from this figure as it is not biologically relevant to estimate detection probabilities for large group sizes of bald eagles.

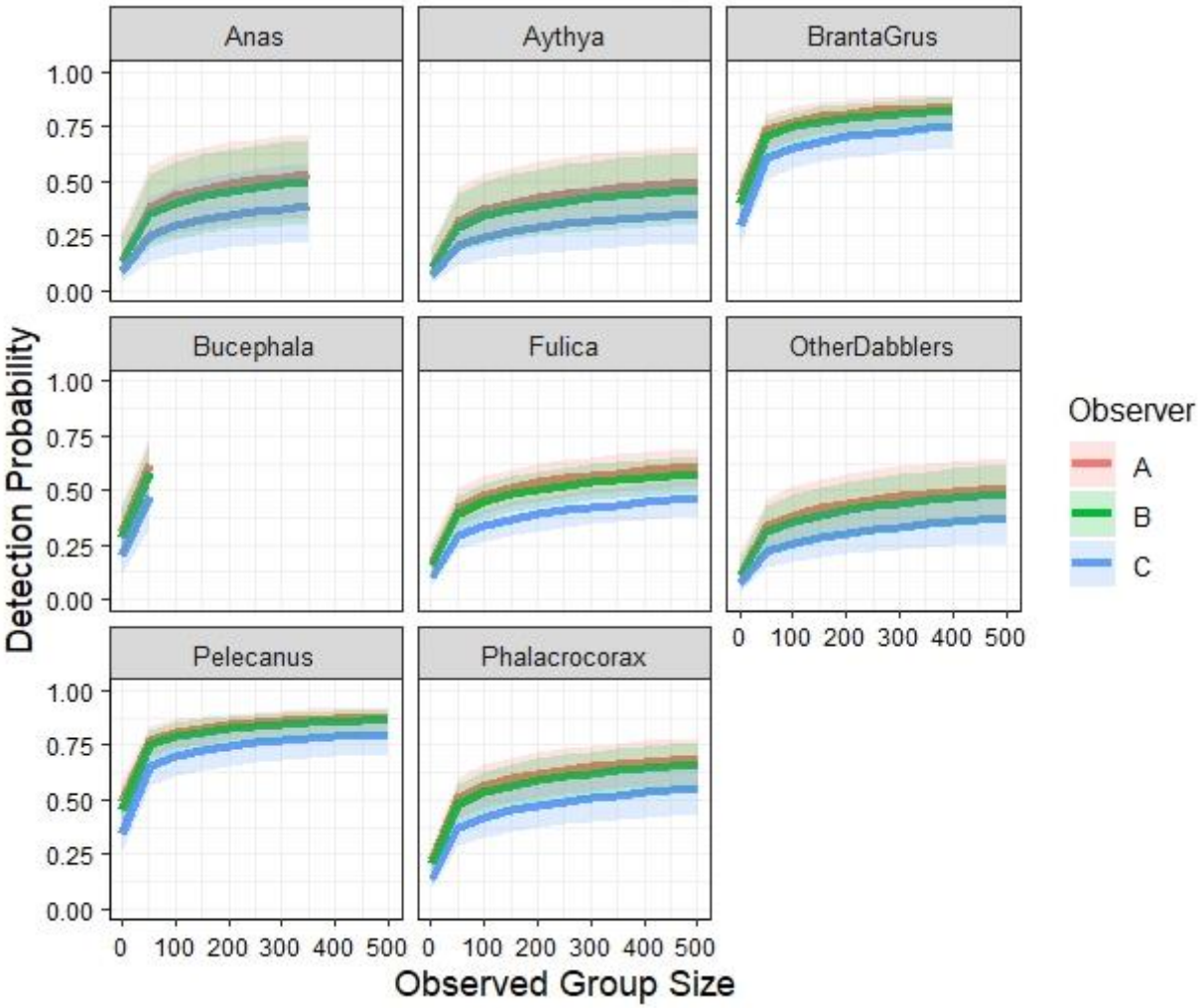

Figure 2A. Detection probability estimates of three observer groups over increasing group sizes by guild from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front and for modeled group sizes closest to that of the largest group size observed. *Haliaeetus* was removed from this figure as it is not biologically relevant to estimate detection probabilities for large group sizes of bald eagles.

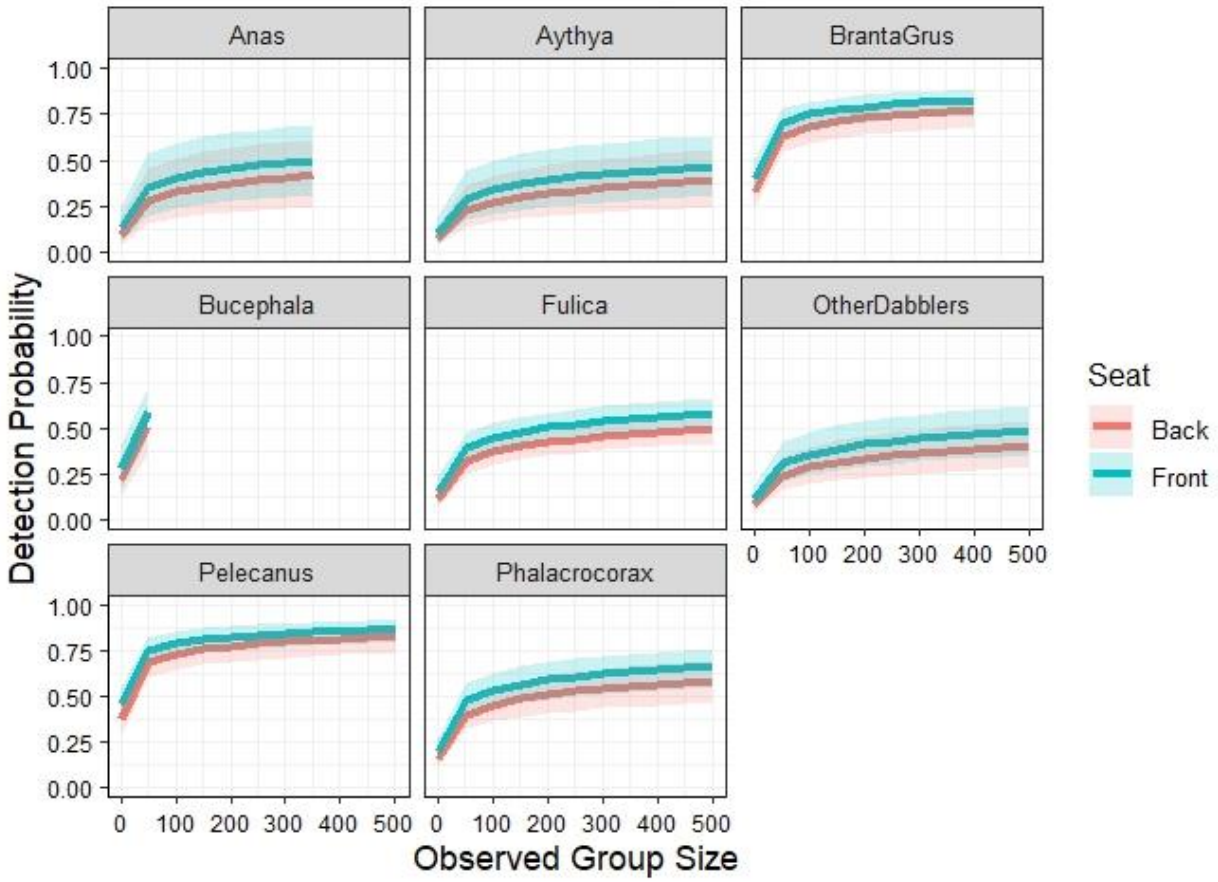

Figure 3A. Detection probability estimates of two seat positions over increasing group sizes by guild from closed population mark-recapture Huggins models modeling crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding observer constant at group A and for modeled group sizes closest to that of the largest group size observed. *Haliaeetus* was removed from this figure as it is not biologically relevant to estimate detection probabilities for large group sizes of bald eagles.

Appendix B. Logistic Regression Results.

We used paired observations for both the subset analysis in the main text and the full analysis in Appendix A. As such, we ran logistic regressions to investigate whether factors such as observer or group size influenced whether or not an observation was paired. We ran seven binomial models, each with a random effect of survey. Single covariate analyses modeled observation pair as a function of 1) a log-transformed continuous group size variable, 2) a fine-scale binned group size variable, and 3) an acting observer pair variable. Double covariate analyses modeled observation pair as a function of each of the group size variables added to the observer pair variable, and the last two analyses modeled observation pair as a function of each of the group size variables interacting with the observer pair variable. The model with the interaction between binned group size and observer pair did not converge therefore it was removed from consideration. We ran an AIC analysis on the six remaining models. The top model was the interaction between the log-transformed group size variable and observer pair. There were no competing models.

Within the top model there was meaningful variance explained by the survey random effect: 4.49 (SD = 2.12). Holding observer pair constant, expected probability of pairing an observation is 0.49 times as high (SE = 0.034, $P < 2e-16$) for each 1 unit increase in log(group size) indicating that larger groups were less likely to be paired than smaller groups. However, for each 1 unit increase in log(group size), expected probability of pairing an observation from two observer combinations from year 3 is 0.25 times as high (SE = 0.046, P = 3.63e-8) and 0.22 times as high (SE = 0.053, P = 2.39e-5) as the original observer combination in all of year 1 of the study. Therefore, within year 3, certain observer combinations were more likely to have larger groups paired than the observer combination in year 1. All other modeled covariates

(interactions and observer pair combinations) were insignificant. We ran a Dharma goodness of fit test on the top model and found insignificant KS, dispersion, and outlier (P = 0.99, 0.89, 0.40, respectively).

Appendix C. Supplementary Analyses of Subset Data Results.

As there was unclear support of a fine-scale group and a season covariate within the competing models from subset analyses reported on in the main text of this paper, we included results from the third overall (second competing, DeltaAICc = 0.813) model here. That model included seat, observer, fine-scale group, season, and group size as covariates, whereas the top model did not include season and included guild rather than fine-scale group. This appendix will only include results from the season and fine-scale covariates from the third overall model.

We estimated detection probability estimates for nine fine-scale groups within our aerial surveys: *Anas*, Other Dabblers, *Aythya*, *Bucephala*, *Branta* + *Grus*, *Fulica*, *Haliaeetus*, *Pelecanus*, and *Phalacrocorax*. When observer, seat, and season were held constant and at flock size = 100 birds, *Branta* + *Grus* were estimated to have the highest detection probability of all taxonomic groups at 0.81. The *Pelecanus* group had the next highest estimated detectability at 0.80 followed by *Phalacrocorax* at 0.74, *Bucephala* at 0.73, *Anas* at 0.47, *Aythya* at 0.42, *Fulica* at 0.41, and Other Dabblers at 0.33 (Figure 1C). Individual Haliaeetus were estimated to have a 0.27 detectability. *Branta* + *Grus* were 1.09 times as likely to be detected compared to flocks of the *Pelecanus* group and 8.65 times as likely as the Other Dabblers group. All combinations of comparisons of detection probabilities between fine-scale groupings can be found in Table 1C. Additionally, detection probability estimate comparisons between observers, seats and flock sizes for all groups can be found in Table 2C.

Detectability estimates were highest in the Late Season surveys and lowest in the Preseason surveys. The December survey detection estimates were slightly lower than the Late Season surveys and the Midwinter detection estimates were between the December and Preseason estimates (Figure 2C). Detection was 1.22 times as likely in the Late Season compared

to the December season, 1.53 times as likely in the December season compared to the Midwinter Season, 1.61 times as likely in the Midwinter season compared to the Preseason, and 3.01 times as likely in the Late Season survey as the Preseason survey. Generally, with seat and fine-scale group held constant, detection probabilities tightened from the first survey conducted within a season to the last. Detection probability of flocks 100 birds in size were 0.62 in the Preseason, 0.80 in December, 0.72 in the Midwinter and 0.83 in the Late Season holding guild and observer constant (Figure 3C). All combinations of comparisons of detection probabilities between seasons can be found in Table 3C.

TABLES AND FIGURES

Table 1C. Fine-scale group detection probability estimates for all observers, seats, and groups from flock sizes 1, 100, and maximum observed from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. Dashes indicate estimates beyond maximum observed group size. Season held constant to December.

ObsA						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**
Pelecanus	0.60	0.91	0.96	0.50	0.88	0.94
Branta + *Grus*	0.62	0.92	0.96	0.52	0.89	0.94
Bucephala	0.50	-	0.85	0.41	-	0.79
Phalacro[3]	0.51	0.88	0.94	0.42	0.84	0.91
Fulica	0.21	0.65	0.79	0.15	0.56	0.72
Anas	0.25	0.71	0.80	0.19	0.62	0.74
Other Dabblers	0.16	0.57	0.64	0.11	0.48	0.55
Aythya	0.21	0.66	0.79	0.16	0.57	0.72
Haliaeetus	0.50	-	0.50	0.40	-	0.40
ObsB						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Group	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**	**Individual Birds**	**Flock Size = 100**	**Maximum Observed Flock Size[1,2]**
Pelecanus	0.36	0.80	0.89	0.27	0.73	0.84
Branta + *Grus*	0.38	0.81	0.89	0.29	0.75	0.84
Bucephala	0.28	-	0.67	0.21	-	0.58

[3]*Phalacro*	0.28	0.74	0.85	0.21	0.66	0.79
Fulica	0.09	0.41	0.58	0.06	0.32	0.49
Anas	0.11	0.47	0.60	0.08	0.38	0.51
Other Dabblers	0.07	0.33	0.40	0.05	0.25	0.31
Aythya	0.09	0.42	0.59	0.06	0.33	0.49
Haliaeetus	0.27	-	0.27	0.20	-	0.20

ObsC

Group	Front Seat Estimated Detection for Individual Birds	Front Seat Estimated Detection for Flock Size = 100	Front Seat Estimated Detection for Maximum Observed Flock Size[1,2]	Back Seat Estimated Detection for Individual Birds	Back Seat Estimated Detection for Flock Size = 100	Back Seat Estimated Detection for Maximum Observed Flock Size[1,2]
Pelecanus	0.34	0.79	0.88	0.26	0.71	0.83
Branta + *Grus*	0.36	0.80	0.88	0.28	0.73	0.83
Bucephala	0.26	-	0.65	0.19	-	0.56
[3]*Phalacro*	0.27	0.72	0.84	0.20	0.64	0.78
Fulica	0.08	0.39	0.56	0.06	0.31	0.47
Anas	0.10	0.45	0.59	0.07	0.36	0.49
Other Dabblers	0.06	0.32	0.38	0.04	0.24	0.30
Aythya	0.08	0.40	0.57	0.06	0.31	0.47
Haliaeetus	0.25	-	0.25	0.19	-	0.19

[1]Maximum flock size observed for *Anas* = 350, *Branta* + *Grus* = 400, *Bucephala* = 55, *Haliaeetus* = 12, Other Dabblers = 200, all others reached or exceeded 500 birds.

[2]Estimates shown from modeled group sizes closest to that of the largest group size observed

[3]*Phalacro* short for *Phalacrocorax*

Table 2C. Fine-Scale Group Detectability Comparisons. Group names in the vertical column are cell value times as likely to be detected compared to group names in the corresponding horizontal column when observer is held constant at Observer group A seat is held constant to front, and season held constant to December.

	Pelecanus	***Branta* + *Grus***	***Bucephala***	***Phalacrocorax***	***Fulica***	***Anas***	**Other Dabblers**	***Aythya***	***Haliaeetus***
Pelecanus	-	0.91	1.46	1.40	5.67	4.44	7.92	5.55	1.50
Branta + *Grus*	1.09	-	1.59	1.53	6.20	4.85	8.65	6.07	1.64
Bucephala	0.69	0.63	-	0.96	3.89	3.05	5.44	3.18	1.03
Phalacrocorax	0.71	0.65	10.4	-	4.04	3.16	5.64	3.96	1.07
Fulica	0.18	0.16	0.26	0.25	-	0.78	1.40	0.98	0.26
Anas	0.23	0.21	0.33	0.32	1.28	-	1.78	1.25	0.34
Other Dabblers	0.13	0.12	0.18	0.18	0.72	0.56	-	0.70	0.19
Aythya	0.18	0.16	0.26	0.25	1.02	0.80	1.43	-	0.27
Haliaeetus	0.67	0.61	0.97	0.94	3.79	2.97	5.29	3.71	-

Table 3C. Season-level detection probability estimates for all observers, seats, and seasons from flock sizes 1, 100, and 500 from the top closed population mark-recapture Huggins model modeling detection probability estimates for double-observer crewed aerial surveys in Alabama. Fine-scale group held constant to *Pelecanus*.

ObsA						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Season	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**
Preseason	0.38	0.81	0.90	0.29	0.75	0.85
December	0.60	0.91	0.96	0.50	0.88	0.94
Midwinter	0.49	0.87	0.93	0.40	0.83	0.90
Late Season	0.64	0.93	0.96	0.55	0.90	0.95
ObsB						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Season	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**
Preseason	0.18	0.62	0.76	0.13	0.52	0.68
December	0.36	0.80	0.89	0.27	0.73	0.84
Midwinter	0.26	0.72	0.84	0.20	0.64	0.78
Late Season	0.40	0.83	0.91	0.31	0.77	0.87
ObsC						
	Front Seat Estimated Detection for			**Back Seat Estimated Detection for**		
Season	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**	**Individual Birds**	**Flock Size = 100**	**Flock Size = 500**
Preseason	0.17	0.60	0.75	0.12	0.50	0.67
December	0.34	0.79	0.88	0.26	0.71	0.83
Midwinter	0.25	0.71	0.83	0.19	0.62	0.76
Late Season	0.38	0.82	0.90	0.30	0.75	0.86

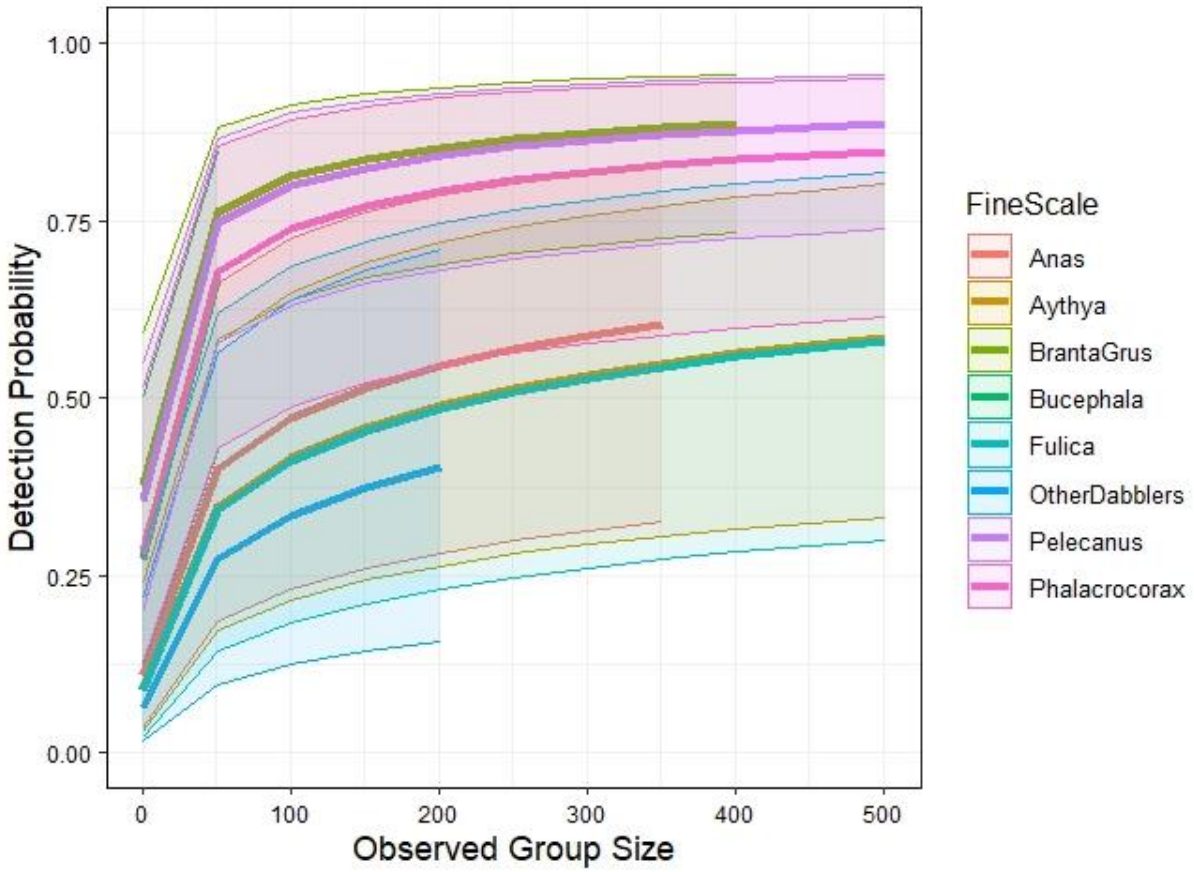

Figure 1C. Detection probability estimates of eight observed fine-scale groups over increasing group sizes from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front and observer constant at group A and for modeled group sizes closest to that of the largest group size observed. *Haliaeetus* was removed from this figure as it is not biologically relevant to estimate detection probabilities for large group sizes of bald eagles.

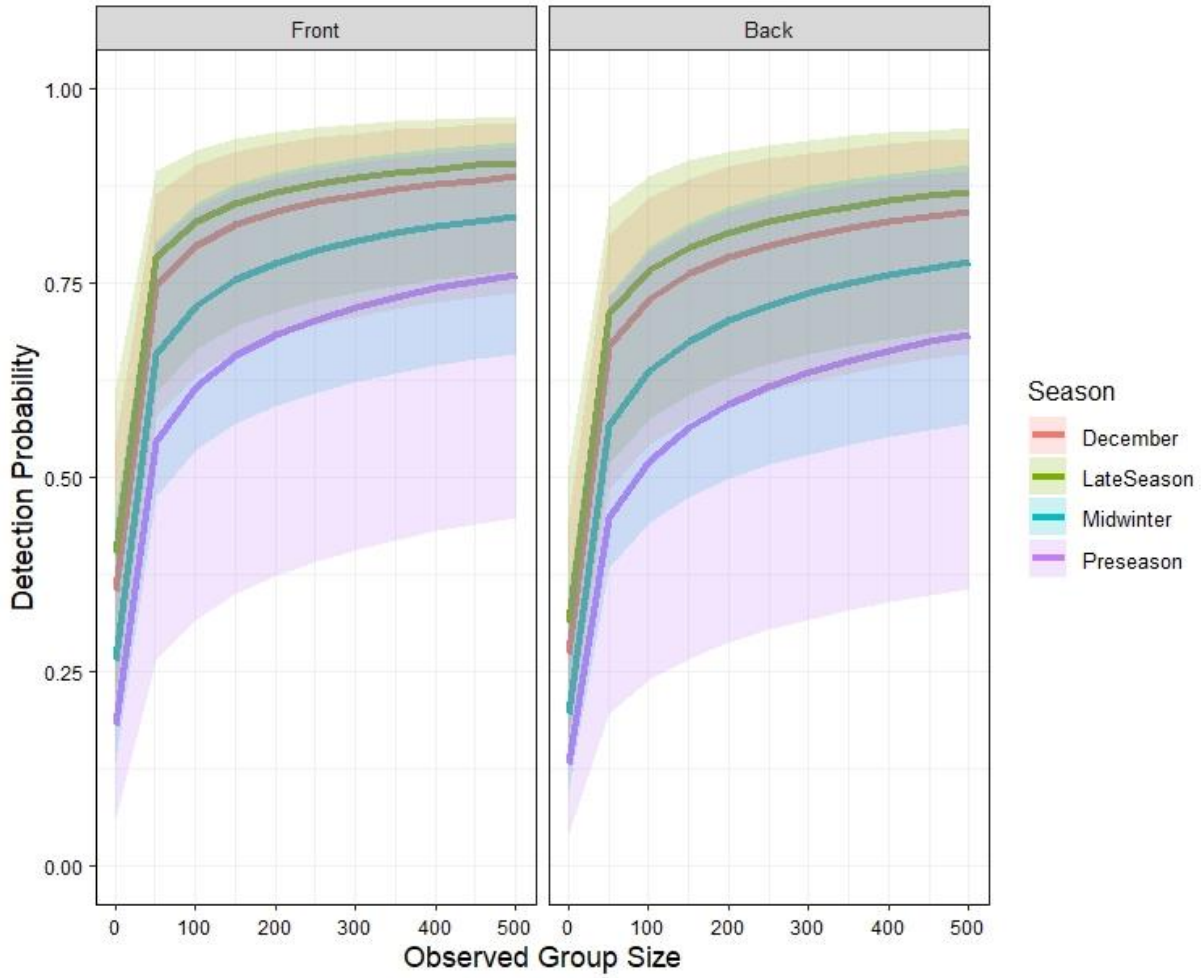

Figure 2C. Detection probability estimates of 4 season groups over increasing group sizes by seat position from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding observer constant at group A and Fine-Scale Group at *Pelecanus*.

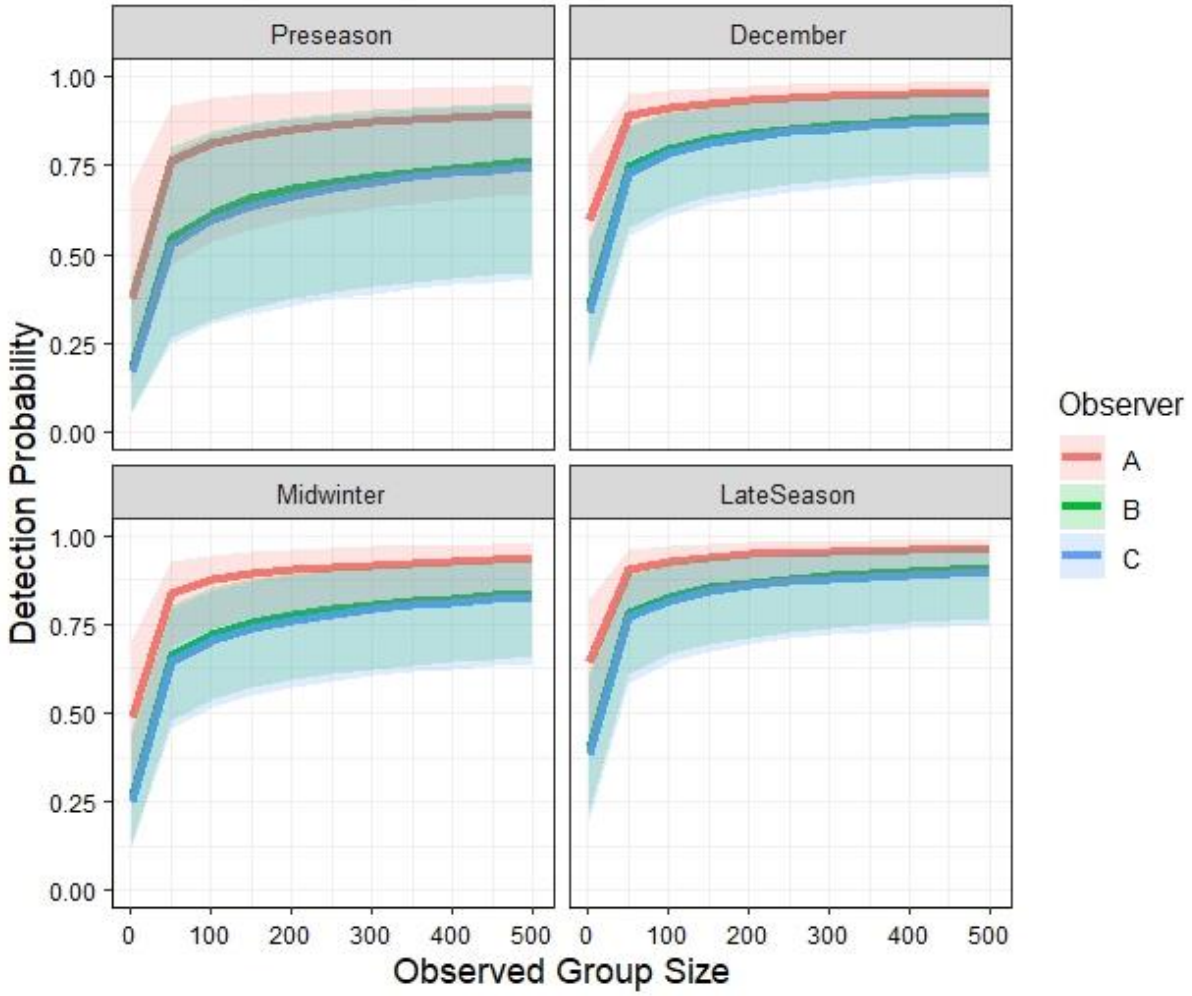

Figure 3C. Detection probability estimates of 3 observer groups over increasing group sizes by season from closed population mark-recapture Huggins models fit to crewed aerial survey data from Alabama. Confidence limits are shown at the 95% level. Estimates are shown from holding seat position constant at Front and Fine-Scale Group at *Pelecanus*.

Appendix D. Supplemental Error Metric Results.

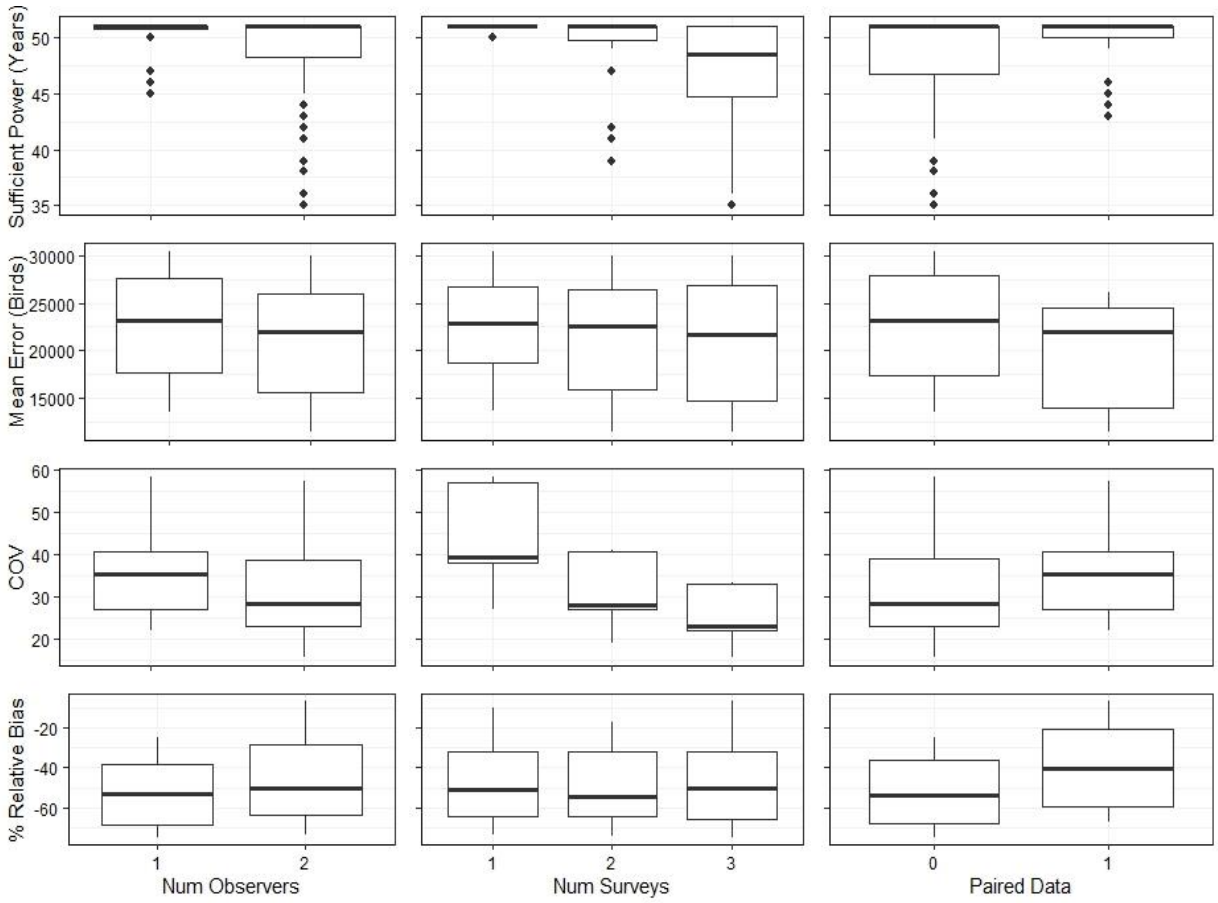

Figure 1D. Power, mean error, coefficient of variation (CV) and percent relative bias values from linear regressions modeling the impact of multiple simulated aerial waterfowl survey situations on listed error metrics. Columns are number of observers, number of surveys conducted, and paired (1) vs. unpaired (0) data, and rows are evaluated error metrics.

Appendix E. Supplementary Availability Results.

Birds trapped in NSR were 0.44 (SE = +/- 0.74) times as likely to be on surveyed areas as birds trapped on CCR (P = 0.55), and birds trapped on Wheeler NWR were 2.19 (SE = +/- 0.78) times as likely to be on surveyed areas as birds trapped on CCR (P = 0.0049). Birds trapped in NSR were -1.75 (SE = +/- 0.54) times as likely to be on surveyed areas as birds trapped on Wheeler NWR (P = 0.001). Additionally, for each day after Nov 15th, log odds of use of aerial survey areas was -33.56 (SE = +/- 4.04) times as likely per day (P < 0.001). For the polynomial date scaled variable, log odds of use of aerial survey areas was 10.89 (SE = +/- 2.79) times as likely per squared day after Nov 15th (P < 0.001). The variance intercept for the random effect of individual was 2.16 (SD = 1.47).

Birds trapped on Wheeler NWR were the most likely to be within survey areas, followed by birds trapped on NSR then CCR. This may be due to the respective sizes of each unit and the land-cover types within the units. Specifically, Wheeler was 14,164 ha in area, compared to 2,027 ha for NSR and 1,354 for CCR. Wheeler NWR also had the greatest proportion of forested wetland and wetland impoundment cover types, which were strong predictors of waterfowl resource selection in Tennessee (Osborn et al. 2017).