


Ignoring uncertainty in predictor variables leads to false confidence in results: a case study of duck habitat use

ADAM C. BEHNEY  †

Avian Research Section, Colorado Parks and Wildlife, 317 W Prospect Road, Fort Collins, Colorado 80526 USA

Citation: Behney, A. C. 2020. Ignoring uncertainty in predictor variables leads to false confidence in results: a case study of duck habitat use. *Ecosphere* 11(10):e03273. 10.1002/ecs2.3273

Abstract. An assumption of most regression analyses is that independent variables are measured without error. However, in ecological studies it is common to use independent variables that are derived from samples and therefore contain some uncertainty. For example, when assessing the assumption that energy availability on the landscape is the primary driver of duck distribution during nonbreeding seasons, investigators typically sample energy availability at sites and use the site-level means as a covariate to predict duck abundance. This strategy ignores uncertainty in the estimates of energy availability, which should be propagated into estimates of effects and predicted values of the response variable. I used Bayesian hierarchical models to include uncertainty in site-level covariates when modeling dabbling duck count data during the spring in northeastern Colorado, USA. I found that even after accounting for uncertainty in energy availability, it was an important predictor of dabbling duck use of sites. Counts were greater at sites with more energy available; however, credible intervals were substantially wider when uncertainty in predictor variables was included. Therefore, ignoring uncertainty leads to overly precise model outputs. Furthermore, I found that larger sites and those further east also supported more dabbling ducks. Using a sample as a covariate is common in ecological studies, and researchers can use the methods outlined here to account for this additional level of uncertainty. These case study results can be used by habitat managers and planners to guide how and where wetland restoration occurs with a more accurate idea of the uncertainty associated with various effects.

Key words: Bayesian; Colorado; ducks; habitat selection; hierarchical models; uncertainty.

Received 22 June 2020; **accepted** 29 June 2020; **final version received** 27 August 2020. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2020 The Author. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** adam.behney@state.co.us

INTRODUCTION

Uncertainty enters into our understanding of ecological processes through multiple sources. Increasingly, researchers are acknowledging various sources of uncertainty and attempting to account for it (Cressie et al. 2009). Process uncertainty represents incomplete knowledge or depiction of a phenomenon or the unpredictability of parameters, whereas observation uncertainty results from inability to measure perfectly or completely (i.e., sampling) and may include bias in measurements (Hilborn and Mangel 1997, Hobbs and Hooten 2015). Traditionally, process

uncertainty has been the main source of uncertainty included in statistical analyses, whereas observation uncertainty is not typically mentioned unless there was an obvious source of bias in measurements (e.g., Dorazio 2014). Specifically with regard to predictor variables, observation uncertainty can lead to bias in parameter estimates, loss of power to detect effects, and potentially false conclusions (Davies and Hutton 1975, Reeves et al. 1998, Carroll et al. 2006). A fundamental assumption of most regression models is that independent variables are measured without error (Sokal and Rohlf 1995). Therefore, methods to deal with observation

uncertainty in predictor variables have received much attention, and an entire class of models (measurement error or errors-in-variables models) have been developed (Fuller 1987). However, used in a likelihood-based framework, these models can be somewhat limiting in terms of how complex they can be. Achieving convergence and parameter estimation can be challenging due to their high dimensionality (Jitjareonchai et al. 2006, Bartlett and Keogh 2018).

Bayesian hierarchical models are being used more frequently by ecologists to model complex systems and represent one approach to address process and observation uncertainty. Hierarchical models allow uncertainty at one level (e.g., predictor variable) to be propagated into effect sizes and parameter estimates in another level (predictions of response variable; Gelman and Hill 2007, Bartlett and Keogh 2018). Furthermore, in a Bayesian framework with Markov Chain Monte Carlo methods, models can handle much greater complexity when compared with frequentist methods (Bartlett and Keogh 2018). Bayesian hierarchical models accounting for measurement error in covariates have been used in epidemiology (Richardson and Gilks 1993, Bartlett and Keogh 2018), occupancy modeling (Rota et al. 2016), and population dynamics (Calder et al. 2003, Clark and Bjornstad 2004) but have received limited use in other realms of ecology. The approach involves treating the true unobserved values of a predictor variable as a latent variable (Hoffmann et al. 2017). Measurement error can be modeled using either classical or Berkson errors. Classical measurement error models consist of the observed variable X^* conditional on the latent true variable X and an error term ($X^* = X + \epsilon$), whereas with Berkson errors, the latent true variable is conditional on the observed values plus an error term ($X = X^* + \epsilon$, Carroll et al. 2006). Classical errors arise when researchers have repeated measurements around a true value, whereas Berkson errors occur when a group's mean is assigned to each individual (Heid et al. 2004). I focus on classical measurement error in this paper.

A common scenario in ecological research involves taking samples of some covariate at the site (or individual, group, etc.) level that is suspected to explain variation in a response variable

among sites (or individuals, groups, etc.; Osborn et al. 2017, Nhu et al. 2019, Spaan et al. 2019). For example, in a study attempting to identify predictors of bird abundance among montane meadows, vegetation samples were taken within each site and the site-level means were used as covariates in regression models (Saveraid et al. 2001). Using site-level means as covariates seems to be the standard way information derived from samples is included as covariates in analyses (Kalies et al. 2012, Pilliod et al. 2013, Wimp et al. 2019, Lindstrom et al. 2020). Under this scenario, no bias is suspected because the covariate is based on multiple samples from each site. However, there is still uncertainty in the estimate of the site-level mean that should be accounted for but is rarely mentioned (Freckleton 2011). As with any sampling, this uncertainty is the result of data collected at a smaller scale (sample) than that at which inference is desired (population). As long as there is no systematic bias in sampling, parameter estimates should be unbiased but confidence intervals around predicted values and effects should widen. By taking multiple samples at each site, researchers have the ability to estimate the uncertainty associated with the site-level mean and include this information in their models.

Herein, I use a case study of duck habitat use among sites to show how observation uncertainty in predictor variables can be accounted for. Although I focus on predictors of habitat use among sites, these methods of accounting for uncertainty in predictor variables could be used in many other scenarios involving sampled variables that are used as predictors of variation among individuals, species, or any other grouping or classification. Methods have previously been developed to deal with observation uncertainty (Richardson and Gilks 1993, Fox and Glas 2003, Hobbs and Hooten 2015, Bartlett and Keogh 2018), and the effects of ignoring observation uncertainty have been documented (Linden and Knape 2009, Freckleton 2011), but they are rarely used in ecological research for the habitat use or selection scenario described in the case study. Therefore, my objective was to show how these methods can be applied to a common ecological research scenario and how they influence overall inference.

Case study

Current paradigms of habitat management and planning for nonbreeding waterfowl are focused on providing enough energy on the landscape to support regional population goals (U.S. Department of the Interior and Environment Canada 1986, Central Valley Joint Venture 2006, Soulliere et al. 2007, Playa Lakes Joint Venture 2008). These habitat management and planning strategies assume that energy availability is a limiting factor during nonbreeding periods (Playa Lakes Joint Venture 2008), and therefore, migrating and wintering ducks select habitat based on energy availability. This assumption should be thoroughly tested, given the amount of management resources that are spent to increase food availability on the landscape (Colorado Division of Wildlife 2011, U.S. Fish and Wildlife Service 2013). Generally, previous research has shown that local-scale duck abundance is positively related to energy density of sites although the effects were variable (Brasher 2010, O'Shaughnessy 2014). In addition to energy availability, many other factors have been shown to influence duck use of sites during migration and winter including wetland type or vegetation patterns in wetlands, size, presence of adjacent wetlands, distance to rivers, and water depth (Brasher 2010, Webb et al. 2010, Hagy and Kaminski 2012, Beatty et al. 2014, O'Shaughnessy 2014, Osborn et al. 2017). If ducks do not respond to changes in energy availability as strongly as predicted due to selection for other factors, the actual number of ducks supported may be less than what planners anticipated.

Examining the effects of energy availability on duck abundance requires estimating energy availability. Like many ecological covariates, food sampling is a time and labor-intensive process (Williams et al. 2014). Therefore, researchers take a small number of food samples (usually about 5–10, Olmstead et al. 2013, O'Shaughnessy 2014, Osborn et al. 2017) at each site that they are interested in representing. These samples contain uncertainty (Behney et al. 2014) that is generally ignored in analyses that use site-level means of energy availability as predictor variables. Accounting for this uncertainty should broaden confidence intervals around the covariate's parameter estimate and overall model predictions, potentially to the point where the variable is no longer important.

In a Bayesian hierarchical modeling framework, I assessed the relationship between site-level characteristics and the intensity of duck use in northeastern Colorado during spring migration. My main objective was to examine the effect of accounting for observation uncertainty in energy availability while assessing its relationship with the intensity of duck use of sites. I expected other factors to influence site use as well so I also assessed how vegetation structure, size, and geographic location affected the intensity of duck use of sites. Like energy availability, vegetation density was also a sample and contained observation uncertainty for which I accounted. Vegetation structure or density can influence how ducks perceive predation risk (Behney et al. 2018). If perceived predation risk prevented ducks from using sites, I expected my measure of vegetation density to be a good predictor of use. There are energetic costs when traveling from roosting or loafing sites to foraging patches (Johnson et al. 2014). I assessed whether these energetic costs played an important role in shaping duck distribution by including variables quantifying the distance of each site to a large reservoir and the distance to a river, which are commonly used for roosting or loafing (Ringelman et al. 1989). Lastly, local-scale duck abundance is limited by duck abundance at larger spatial scales. If this large-scale flyway location effect is an important driver of local-scale duck abundance, then geographic location along an east–west gradient should be a good predictor of duck abundance.

METHODS

Study area and site selection

I conducted this study over a 3,331,501-ha area in northeastern Colorado in Sedgwick, Washington, Logan, Morgan, Weld, and Larimer counties. The area is generally classified as shortgrass prairie with intensive agriculture focusing on cattle production and row crop farming. Most water features in the region (other than playas) are associated with the South Platte River and were found within the river basin corridor. Therefore, to ensure a spatially balanced sample, I divided the South Platte River corridor (~10 km from river) into four quadrants (~70 river km per quadrant). Using wetland geographic information system

data, I compiled a list of all potential sites (properties, both public and private) within each quadrant for each type of water feature and randomly selected sites within each quadrant. I used 6 general types of water features: actively managed emergent wetland, passively managed emergent wetlands, sloughs, playas, small reservoirs (<200 ha), and large reservoirs (≥ 200 ha). These water feature types represented a substantial proportion of the overall wetland habitat base used in regional avian habitat planning models (Playa Lakes Joint Venture 2008). I randomly selected sites equally among quadrants and water feature type. I was unable to gain access to 3 randomly selected sites so I randomly selected new sites to replace them.

Playas are found farther from the South Platte River than other wetland types. Therefore, I selected study playas, by creating a grid (7.5×7.5 min topographic map quadrangles) across northeastern Colorado. I randomly selected two grid cells that contained at least 5 playas that were greater than 0.25 ha. I then randomly selected playas within each grid cell to sample. During this study, playas were infrequently inundated due to drought, which drastically reduced the sample size of playas.

Sampling methods

Duck counts.—In late winter, after most water bodies thawed and ducks started to arrive, I conducted weekly bird counts during morning (sunrise to 1000) between 28 February and 1 June 2016 and 2017 on all sites. I visited each site once per week in a random order and noted the number of each species of duck present. For most sites, I was able to conduct counts each week and collected 10 or 11 counts throughout each spring. However, for some sites, I was not able to visit them every week and I excluded sites from analyses with fewer than three samples in a year. I also excluded sites for which I did not collect energy availability or vegetation density data. I attempted to estimate detection probability during duck counts through double counting. First, I counted ducks from a vantage point or vehicle without disturbing them. Then, for water features less than ~ 2 ha, I walked the perimeter and out into the wetland to flush and count all ducks. For sloughs, I walked a 500 m stretch to flush and count all ducks. For reservoirs, I

randomly selected a quarter or an eighth of the water feature, and used aerial photographs in the field to identify the extent of, and counted all ducks within, the selected area. I also walked the bank or out into any vegetation to flush any ducks that were not initially visible. Farther from the bank, there was no vegetation that would have prevented detection of ducks. Observers used spotting scopes and binoculars to scan open water areas to search for ducks. Detection of ducks during flush counts was always \geq to that of vantage surveys so I only used flush counts in analyses and did not account for detection probability. By flushing all ducks in smaller water features and those in vegetation in reservoirs, I believe that any bias due to detection probability was minimal.

Food and vegetation sampling.—I used core sampling (Williams et al. 2014) to estimate food density during fall and spring. Details of food sampling and estimates of food and energy density and depletion for these study sites are presented in Behney (2020). Briefly, I sampled in late winter (23 February through 16 March 2016 and 2017), as soon as wetlands began thawing to estimate food density at the beginning of spring migration. I randomly distributed seven core samples (Behney et al. 2014) throughout portions of water features with water depth less than 50 cm. I used this cutoff because dabbling ducks (Genus *Anas*, *Spatula*, *Mareca*, *Aix*) generally do not feed in water deeper than 50 cm (Behney 2014). Core samples were processed in the laboratory where I visually searched through the material and picked out any seed, tuber, or invertebrate and identified items to lowest taxonomic level possible, generally, to genus for plant matter and to class or order for invertebrates. I dried all material at 60°C to a constant mass (about 48 h) and weighed to the nearest 0.00001 g to estimate food density (kg/ha) and facilitate conversion to energy density (kcal/ha) based on published true metabolizable energy values for duck food items at the lowest taxonomic level possible (Livolsi et al. 2015). I converted energy density to total available energy at a site by multiplying energy density by the area of each site for which water depth was shallow enough to facilitate feeding by dabbling ducks (<50 cm) presented in Behney (2020). At each core location, I estimated vegetation density/structure using a modified

Robel pole technique for which I noted the number of 10 cm bands on a 1.5 m tall pole (2.54 cm diameter) that were completely visible from 5 m, repeated for each of the four cardinal directions.

Statistical analyses

To assess factors that influence dabbling duck use of sites during spring, I modeled weekly dabbling duck counts (pooled across species) in a Bayesian hierarchical analysis. I used a negative binomial distribution to represent duck counts because initial data exploration suggested a Poisson distribution could not adequately represent the variation present in the duck count data. I limited analyses and inference to dabbling ducks because they are the primary focus of management activity in the region (Playa Lakes Joint Venture Waterfowl Team 2005). I included week as a count-level (i.e., individual-level) predictor and total available energy (all energy found at water depths less than 50 cm), an index of vegetation density (Robel pole measurements), site size (ha), distance to the nearest large reservoir (km), distance to the South Platte River (km), and UTM easting coordinate (km) as site-level (i.e., group-level) predictors. To calculate the distance from each site to large reservoirs and the river, I calculated the distances between the centroid of each site and the centroid of the nearest large reservoir or the centerline of the river in ArcMap (Esri, Redlands, California, USA).

Energy density and vegetation density metrics included some uncertainty because they, themselves, were samples. Therefore, in addition to the count model, I included energy and vegetation density models in the overall hierarchical model and used the parameters estimated from these models as covariates in the duck count model. I modeled energy using a lognormal distribution:

$$\log(\text{energy}) \sim \text{normal}\left(\mu_{jk}^{\text{energy}}, \sigma_{\text{energy}}^2\right)$$

where energy_{ijq} is the energy estimated from food sample i at site j in year q .

For vegetation density, I used a binomial distribution:

$$\begin{aligned} \text{veg}_{dijq} &\sim \text{binomial}\left(\text{inv.logit}\left(\alpha 2_{ijq}^{\text{veg}}\right), 15\right) \\ \alpha 2_{ijq}^{\text{veg}} &\sim \text{normal}\left(\alpha 1_{jq}^{\text{veg}}, \sigma_{\text{veg}}^2\right) \end{aligned}$$

where veg_{dijq} is the number of sections of the pole that were visible out of 15 total sections, read from direction d , at sample location i , within site j , in year q . I included direction d as an additional level for the vegetation model but not in the energy availability model because at each sample location, a vegetation density reading was taken from 4 directions, whereas only 1 food sample was taken at each sample location. Sample location was nested within site so I used the vegetation density mean of the site as the prior mean for the sample location.

I modeled weekly duck counts (k) at site j in year q using a negative binomial distribution where the intercept was allowed to vary by site and year. I included site-level covariates to explain variability in the intercept. The global count model took the form:

$$\begin{aligned} \text{count}_{kjq} &\sim \text{negative binomial}\left(p_{kjq}, r\right) \\ p_{kjq} &= \frac{r}{r + \lambda_{kjq}} \\ \log(\lambda_{kjq}) &= \alpha_{jq} + \beta_1 \text{ week}_k + \beta_2 \text{ week}_k^2 \\ \alpha_{jq} &\sim \text{normal}\left(\mu_{jq}^\alpha, \sigma_\alpha^2\right) \\ \mu_{jq}^\alpha &= \tau_0 + \tau_1 \mu_{jk}^{\text{energy}} + \tau_2 \alpha 1_{jq}^{\text{veg}} \\ &\quad + \tau_3 \text{ res dist}_j + \tau_4 \text{ riv dist}_j \\ &\quad + \tau_5 \text{ easting}_j + \tau_6 \text{ size}_j \end{aligned}$$

where μ_{jk}^{energy} and $\alpha 1_{jq}^{\text{veg}}$ were estimated in the energy and vegetation density submodels, respectively. Full model posterior and joint distributions are shown in Appendix S1.

By modeling energy and vegetation density in the overall hierarchical model, uncertainty in their estimates is propagated into the duck count model. All priors were vague; coefficients (β , τ) were assigned normal(0, 0.0001[precision]) except in the vegetation density model where coefficients ($\alpha 1_{jq}^{\text{veg}}$) were assigned normal(0, 0.5[precision]) priors because it used a logit link function, and I wanted priors to be vague on the probability scale (Northrup and Gerber 2018) rather than the scale of the link function. Standard deviations were assigned uniform (0, 5) priors. I centered all predictor variables before inclusion in the model.

I used a multi-phase modeling strategy. Firstly, I assessed the best form of the week variable

without including any site-level predictors. I included week as a cubic, quadratic, and linear predictor of duck counts and included an intercept-only null model in the model set. I compared models using deviance information criterion (DIC) and by assessing whether 95% credible intervals of the coefficients contained zero. Many options exist for Bayesian model selection but I chose DIC because it is simple to calculate and is similar to AIC, which is widely used and understood by ecologists (Hooten and Hobbs 2015). Secondly, using the best form of the week variable, I assessed whether energy density or vegetation density should be included in the final model of duck counts. I included these variables in this separate stage of modeling because they were samples themselves and required a mixture of models, each with their own parameters. I fit a model including vegetation and energy density in an additive relationship and determined that a variable should be in the final model if the 95% credible interval for its coefficient excluded zero (e.g., Hobbs et al. 2012). I did not use DIC to compare models in this stage because DIC measures the fit of a model penalized by the effective number of parameters (Gelman and Hill 2007). The goal of incorporating uncertainty in covariates is not necessarily to improve model fit, but to more realistically represent uncertainty that is present in the data, which requires estimating additional parameters. Therefore, a model incorporating predictor uncertainty will likely have a worse DIC than a model ignoring this uncertainty. Lastly, using the best model from stage two, I compared models including all additive combinations of distance to large reservoir, distance to river, and easting using DIC and assessing 95% credible intervals. In this final stage, I also included two models incorporating interactions between energy density, distance to reservoir, and distance to river (energy \times reservoir and energy \times river) to test for differences in how ducks select foraging sites based on how far they are from roost locations (Kennedy and Gray 1997). Because the response variable was raw counts (not density), to account for differences in area among sites, I included a size (ha) variable in all models during this last stage of modeling. None of the predictors were highly correlated (all Pearson correlation coefficients < 0.5).

I ran three chains for 30,000 Markov Chain Monte Carlo (MCMC) iterations following 5,000 iterations for burn-in and 5,000 iterations for adaptation for optimal sampling and fit models using Jags via the RJags package (Plummer 2018) in program R (R Core Team 2019). To assess convergence, I visually examined trace plots and calculated Gelman and Rubin's convergence diagnostic in the CODA package (Plummer et al. 2006) to ensure it was close to 1 (Gelman and Rubin 1992, Brooks and Gelman 1998, Gelman and Hill 2007).

To check whether models adequately fit the data, I used posterior predictive checks with Bayesian p values based on means, standard deviations, and discrepancy (sums of squares) of the data and from simulated data from the fitted models (Gelman and Hill 2007, Hobbs and Hooten 2015). Bayesian P values were calculated as the percentage of iterations for which the metric (i.e., mean, SD, discrepancy) from the observed data exceeded that of the simulated data where extreme values (close to zero or one) indicate poor fit.

RESULTS

I was able to include 498 counts over 28 sites in analyses. The three most common species observed were mallard (*Anas platyrhynchos*), northern shoveler (*Spatula clypeata*), and gadwall (*Mareca strepera*). For all models, trace plots indicated convergence for all parameters and Gelman and Rubin's convergence diagnostic was ≤ 1.01 for all parameters. Little evidence of lack of fit was revealed for models from posterior predictive checks; all Bayesian p values were between 0.5 and 0.7.

For the first stage of modeling, including week as a quadratic or a cubic term resulted in very similar DIC scores (within 0.1 DIC; Appendix S2: Table S1). Both models were substantially better than models including week in linear form (Δ DIC = 32.6) or a null, intercept-only model (Δ DIC = 148.7). The cubic coefficient's 95% credible interval overlapped zero (-0.001 – 0.005 ; Appendix S2: Table S1), and the predicted values were very similar between cubic and quadratic models. Therefore, I used the quadratic form of week as a count-level predictor in subsequent modeling.

For the second stage of modeling, the 95% credible interval for the energy availability coefficient excluded zero (0.49, 95% CI 0.26–0.66), whereas the credible interval for the vegetation density coefficient included zero (0.17, 95% CI –0.84–1.03). Therefore, I only included energy availability as a site-level predictor in subsequent modeling.

I included 10 models in the third stage of modeling. Based on DIC, the best model included additive effects of energy availability, distance to reservoir, and easting (Table 1). There were four competing models (Table 1) so I identified the model on which to base inference by looking at whether credible intervals overlapped zero. In the top model, credible intervals for energy availability (0.49, 0.30–0.67) and site size (0.004, 0.002–0.006) excluded zero, easting was borderline (0.01, 0.0005–0.01), and distance to reservoir included zero (–0.02, –0.06–0.01). In the second best model, the credible interval for the interaction term between distance to reservoir and energy availability included zero (0.009, –0.004–0.03). Therefore, I generated predictions and based inference on the model including energy availability, site size, and easting in an additive relationship. Energy availability, easting, and site size were positively related to duck counts (Fig. 1). Accounting for uncertainty in energy availability resulted in substantially wider credible intervals for both predicted duck counts and the group-level effect of energy availability on duck counts (Fig. 2).

Table 1. Model selection results for predicting duck counts based on energy density (food), site size, distance to nearest large reservoir, distance to river, and UTM coordinate easting (east).

| Model | Δ DIC | Penalty |
|---------------------------------|--------------|---------|
| Food + reservoir + east | 0 | 103.4 |
| Food \times reservoir | 0.6 | 105.2 |
| Food | 0.7 | 103.8 |
| Food + east | 0.7 | 103.4 |
| Food \times river | 1.4 | 106.0 |
| Food + reservoir | 1.9 | 104.5 |
| Food + reservoir + river | 2.3 | 104.9 |
| Food + river | 2.7 | 105.1 |
| Food + river + east | 5.2 | 106.1 |
| Food + reservoir + river + east | 6.8 | 106.6 |

Note: Size was included in all models.

DISCUSSION

Increasingly, investigators are recognizing the value of accounting for various sources of uncertainty in ecological studies. Observation uncertainty in predictor variables is one source that is rarely accounted for in ecology but can have substantial influence on overall inference (Carroll et al. 2006). Using a case study of duck habitat use among sites, I found that accounting for observation uncertainty in site-level covariates resulted in substantially less confidence in model predictions and effect sizes than if this uncertainty was ignored. The case study presented in this paper represents a common situation in ecological research where site-level predictors contain observation uncertainty, and my analyses confirm that researchers should account for this uncertainty in predictors or risk severely overestimating precision.

With the increasing use of Bayesian analyses, many researchers who are familiar with coding Bayesian models should find it straightforward to add a term to account for predictor uncertainty in their analyses. A simple rule that researchers can use to indicate whether accounting for predictor uncertainty is required is if they were planning on calculating site or some other group-level means of a predictor variable before inclusion in a model, then there is uncertainty in that variable and it should be accounted for. Studies involving vegetation sampling are one of the most common occurrences of predictor uncertainty because researchers typically collect a large amount of vegetation data to attempt to accurately represent a site and then use site-level means as covariates (e.g., Saveraid et al. 2001, Behney et al. 2018). In this situation, researchers can simply include a submodel in the overall hierarchical model that predicts the site-level vegetation means and use the predictions as a covariate in the primary model.

In my case study, even after accounting for observation uncertainty, I detected a positive effect of energy availability on duck use. My results are consistent with other studies of duck habitat use during the nonbreeding season in that food or energy density was positively related to duck use but other factors contribute to how ducks distribute themselves among sites (Brasher 2010, O'Shaughnessy 2014, Osborn et al.

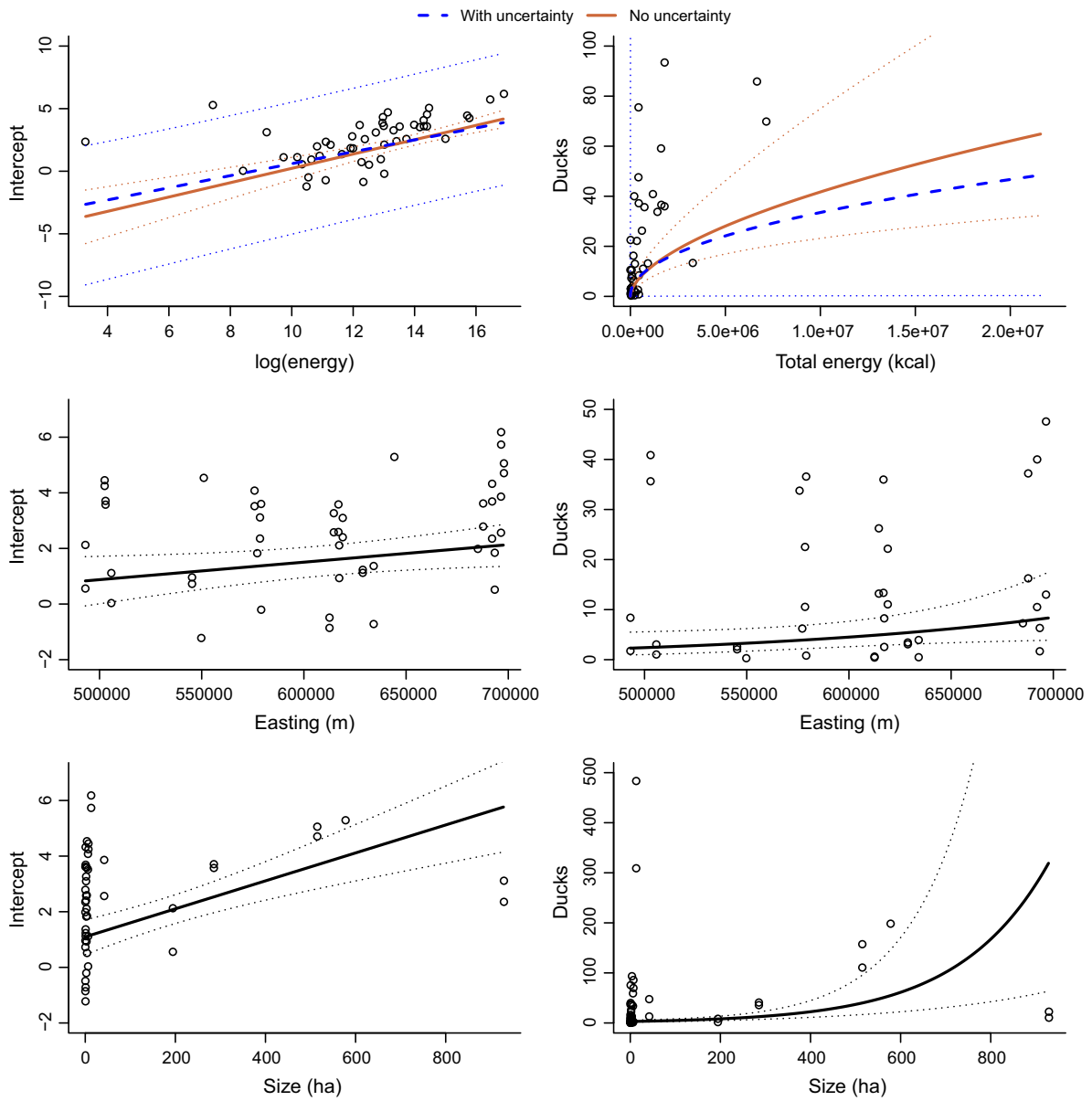


Fig. 1. The intercept in the count model was allowed to vary by site and site-level predictors were used to explain variation in the intercept (left column). The intercept was on the log scale, because I used a log link function in the count model. Back-transformed values of predicted number of ducks site⁻¹ d⁻¹ at the mean date during spring are shown in the right column. Predicted values are shown with heavy solid or dashed lines and 95% credible intervals are shown with dotted lines. Each circle represents the intercept for that site and its measured energy availability, size, or easting. The top row shows the effects of total energy at a site that is found at water depths less than 50 cm. The second row shows the effect of UTM easting coordinate, and the third row shows the effects of site size. All predictions were made from the model including energy + easting + size. When showing the effects of individual variables, other variables were held constant at their mean.

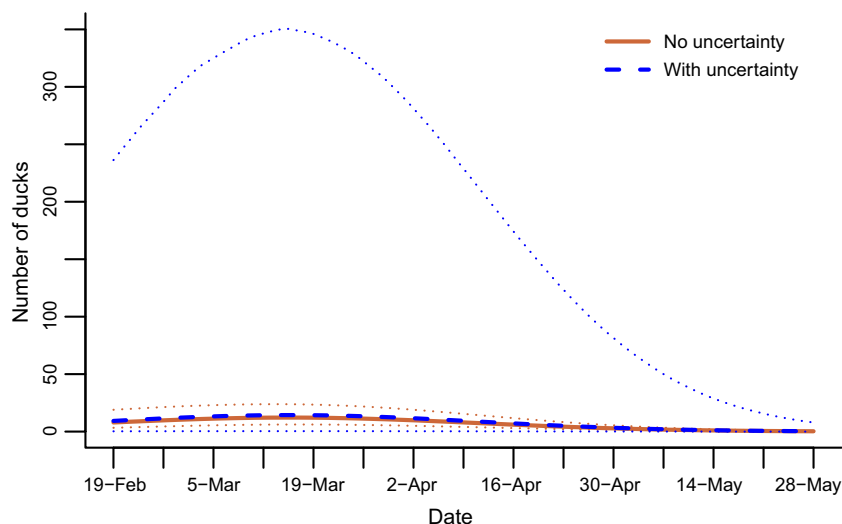


Fig. 2. Predicted number of ducks $\text{site}^{-1} \text{d}^{-1}$ during the spring from a model incorporating uncertainty in estimates of energy availability and a model ignoring uncertainty in energy availability estimates. Relationship between energy availability and the intercept of the count model incorporating and ignoring uncertainty in sampled energy availability (right column). Predicted values are shown with heavy solid or dashed lines and 95% credible intervals are shown with dotted lines.

2017). I also found that site size and easting both contributed to explain duck counts in northeastern Colorado. Within the range of site-level predictor values observed in this study, energy availability and site size appeared to have the greatest effect on the number of ducks observed at a site. In other studies, factors that have been shown to influence duck use of sites during migration and winter include wetland type or vegetation patterns in wetlands (Hagy and Kaminski 2012, Beatty et al. 2014, O'Shaughnessy 2014, Osborn et al. 2017) and wetland structure/morphology (Brasher 2010, Osborn et al. 2017) and landscape characteristics (Brasher 2010, Beatty et al. 2014). These studies did not account for observation uncertainty in predictor values which may be why they were able to detect an effect of vegetation density and I was not. Energetic carrying capacity models used for allocating resources for habitat acquisition or restoration assume that energy provided through these projects will be available and consumed by ducks (Central Valley Joint Venture 2006, Soulliere et al. 2007, Playa Lakes Joint Venture 2008). However, this and other research has shown that a variety of factors (e.g., geographic location, site size, vegetation patterns) can influence duck use

of sites other than energy availability. If these factors are ignored, sites may not be exploited by ducks as predicted, resulting in a suboptimal use of resources for creation/enhancement of habitat projects.

In this study, site size and easting were both positively related to intensity of duck use of sites. Larger sites have previously been shown to support more ducks (Brasher 2010), and based on my results, the finding is not due to larger sites containing more total energy. My estimates of total energy availability only included energy found at depths shallow enough to be accessible to dabbling ducks. Larger sites contained proportionally less shallow water habitat than smaller, shallower sites (Behney 2020), and thus, energy availability was not positively correlated with size ($r = -0.47$). The relationship between site size and intensity of duck use was positive so it appears that ducks were selecting larger sites for reasons other than energy availability of the site. Ducks generally select larger, open water sites for roosting (Hopper 1968, Ringelman et al. 1989), possibly due to the lower perceived predation risk in open water (Behney et al. 2018), which may explain why we observed more ducks at larger sites. Furthermore, larger sites

generally support greater species richness (MacArthur and Wilson 1967, Rosenzweig 1995, Connor et al. 2000), which may result in more total individuals at each site.

My finding of more ducks using sites further east suggests there is a gradient of regional duck abundance ranging from more ducks in far eastern Colorado and fewer ducks as one moves west. Colorado lies at the far western edge of the Central Flyway, and if fewer ducks use the fringe of the flyway, I would expect a positive relationship with easting within Colorado. Alternatively, there tend to be more water features at the western edge (i.e., Colorado Front Range) of my study area, which may also result in fewer ducks on each individual water feature. However, the effect of easting was more a function of the extreme east sites supporting more ducks rather than the extreme western sites supporting fewer (Fig. 1).

Interestingly, I did not detect effects of distance to reservoir or river on duck counts. Ducks are known to select large reservoirs for roosting habitat (Hopper 1968, Ringelman et al. 1989) and would forage nearby if their goal was to minimize energy expenditure. In a meta-analysis, mean foraging flight distance for dabbling ducks was 5.1 km (Johnson et al. 2014). Many of my sites ($n = 28$) were farther than this reported mean foraging flight distance (range = 0–50 km), which may explain why I did not observe a relationship with distance to reservoir. Ducks are known to use the South Platte River, particularly during cold temperatures when other types of water features (e.g., reservoirs, wetlands) freeze (Hopper 1968, Ringelman et al. 1989). My study took place during spring, after water features thawed, so there likely was little selective pressure for ducks to use the river when so many other water features were available.

I attempted to estimate detection probability during duck counts through double counting. However, I always detected at least as many ducks during flush counts than vantage surveys so I only used flush counts in analysis and did not estimate detection probability. I acknowledge that failing to account for detection probability may have resulted in some bias in the count data but believe this to be minimal for two reasons. Firstly, the number of ducks I typically counted was low (75% of observations were ≤ 20 ducks,

median = 5), which makes it easier to accurately count all individuals. Secondly, the main factor that might limit an observer's ability to see and count all ducks present at a site was vegetation that limited visibility. By flushing ducks out of any vegetation, observers could accurately count ducks as they flushed into the air. Flush counts are a relatively common way of counting ducks in areas where detection probability may otherwise be < 1 (Linz et al. 1998, St. James et al. 2013, Lindstrom et al. 2020).

Inferences from this study are limited to spring migration and should not be used to infer patterns of habitat use during fall migration. Disturbance from waterfowl hunting plays a large role in habitat selection during the fall (Lancaster et al. 2015, Palumbo et al. 2019) but is absent during the spring. Therefore, managers wishing use this information to guide where habitat restoration takes place should keep in mind that habitat selection during fall migration may be different than what I present here. There is some evidence that ducks migrate faster and have shorter stop-over durations during spring than fall (Nilsson et al. 2013), presumably to reach the breeding grounds earlier to increase reproductive success (Dzus and Clark 1998). Therefore, they may be using different migration corridors during spring to reach the breeding grounds early. I recommend future research incorporating both fall and spring habitat selection in the same region to discern any differences between the two periods.

These results can be incorporated into habitat planning strategies for nonbreeding ducks in a number of ways. Firstly, they confirm that providing abundant energy for ducks to consume should be a primary goal of habitat conservation or restoration projects. Secondly, factors other than energy can also be used to guide where and how habitat projects are carried out. For example, all other factors being equal, projects that are closer to reservoirs and farther east in Colorado will tend to provide the greatest value to ducks during spring. Furthermore, I show that Bayesian hierarchical models represent a good way to account for uncertainties that arise at multiple levels (i.e., hierarchical). This approach could be incorporated into many ecological studies to more accurately estimate uncertainty. Not only does appropriately accounting for such uncertainty present a more accurate depiction of

reality but also results in better predictions (Cresie et al. 2009).

ACKNOWLEDGMENTS

Funding for this research was provided by Colorado Parks and Wildlife's Wetland Wildlife Conservation Program, Avian Research Program, and a Playa Lakes Joint Venture ConocoPhillips grant. I would like to thank all the technicians who worked on this project, the landowners who graciously allowed access to their property, and Colorado Parks and Wildlife field staff for logistical support. J. Gammonley, J. Lemly, N. Marymor, B. Noonan, M. Reddy, G. Stoebner, B. Sullivan, and T. Verquer provided valuable input on the design or logistics of this project. I thank the National Socio-Environmental Synthesis Center (SESYNC) for statistical training. K. Aagaard and J. Runge provided valuable input on statistical analyses. This manuscript benefitted greatly from reviews by K. Aagaard and B. Walker.

LITERATURE CITED

- Bartlett, J. W., and R. H. Keogh. 2018. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. *Statistical Methods in Medical Research* 27:1695–1708.
- Beatty, W. S., E. B. Webb, D. C. Kesler, A. H. Raedeke, L. W. Naylor, and D. D. Humburg. 2014. Landscape effects on mallard habitat selection at multiple spatial scales during the non-breeding period. *Landscape Ecology* 29:989–1000.
- Behney, A. C. 2014. Foraging behavior of ducks during spring migration in the Wabash River Region, Illinois. Dissertation. Southern Illinois University, Carbondale, Illinois, USA.
- Behney, A. C. 2020. The influence of water depth on energy availability for ducks. *Journal of Wildlife Management* 84:436–447.
- Behney, A. C., R. O'Shaughnessy, M. W. Eichholz, and J. D. Stafford. 2014. Influence of item distribution pattern and abundance on efficiency of benthic core sampling. *Wetlands* 34:1109–1121.
- Behney, A. C., R. O'Shaughnessy, M. W. Eichholz, and J. D. Stafford. 2018. Indirect risk effects reduce feeding efficiency of ducks during spring. *Ecology and Evolution* 8:961–972.
- Brasher, M. G. 2010. Duck use and energetic carrying capacity of actively and passively managed wetlands in Ohio during autumn and spring migration. Dissertation. The Ohio State University, Columbus, Ohio, USA.
- Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434–455.
- Calder, C., M. Lavine, P. Muller, and J. S. Clark. 2003. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology* 84:1395–1402.
- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. 2006. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, New York, New York, USA.
- Clark, J. S., and O. N. Bjornstad. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology* 85:3140–3150.
- Colorado Division of Wildlife. 2011. *Statewide strategies for wetland and riparian conservation: strategic plan for the wetland wildlife conservation program*. Colorado Division of Wildlife, Denver, Colorado, USA. <https://cpw.state.co.us/Documents/LandWater/WetlandsProgram/CDOWWetlandsProgramStrategicPlan110804.pdf>
- Connor, E. F., A. C. Courtney, and J. M. Yoder. 2000. Individuals-area relationships: the relationship between animal population density and area. *Ecology* 81:734–748.
- Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* 19:553–570.
- Davies, R. B., and B. Hutton. 1975. The effect of errors in the independent variables in linear regression. *Biometrika* 62:383–391.
- Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography* 23:1472–1484.
- Dzus, E. H., and R. G. Clark. 1998. Brood survival and recruitment of mallards in relation to wetland density and hatching date. *The Auk* 115:311–318.
- Fox, J., and C. A. W. Glas. 2003. Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika* 68:169–191.
- Freckleton, R. P. 2011. Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology* 65:91–101.
- Fuller, W. 1987. *Measurement error models*. John Wiley and Sons, New York, New York, USA.
- Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge, UK.
- Gelman, A., and D. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–511.

- Hagy, H. M., and R. M. Kaminski. 2012. Winter water-bird and food dynamics in autumn-managed moist-soil wetlands in the Mississippi Alluvial Valley. *Wildlife Society Bulletin* 36:512–523.
- Heid, I. M., H. Kuchenhoff, J. Miles, L. Kreienbrock, and H. E. Wichmann. 2004. Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology* 14:365–377.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective*. Princeton University Press, Princeton, New Jersey, USA.
- Hobbs, N. T., H. Andren, J. Persson, M. Aronsson, and G. Chapron. 2012. Native predators reduce harvest of reindeer by Sami pastoralists. *Ecological Applications* 22:1640–1654.
- Hobbs, N. T., and M. B. Hooten. 2015. *Bayesian models: a statistical primer for ecologists*. Princeton University Press, Princeton, New Jersey, USA.
- Hoffmann, S., E. Rage, D. Laurier, P. Laroche, C. Guihenneuc, and S. Ancelet. 2017. Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners. *Radiation Research* 187:196–209.
- Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. *Ecological Monographs* 85:3–28.
- Hopper, R. M. 1968. *Wetlands of Colorado*. Technical Publication No. 22. Colorado Department of Game, Fish, and Parks, Denver, Colorado, USA.
- Jitjareonchai, J. J., P. M. Reilly, T. A. Duever, and D. B. Chambers. 2006. Parameter estimation in the error-in-variables models using the Gibbs sampler. *Canadian Journal of Chemical Engineering* 84:125–138.
- Johnson, W. P., P. M. Schmidt, and D. P. Taylor. 2014. Foraging flight distances of wintering ducks and geese: a review. *Avian Conservation and Ecology* 9:2.
- Kalies, E. L., B. G. Dickson, C. L. Chambers, and W. W. Covington. 2012. Community occupancy responses of small mammals to restoration treatments in ponderosa pine forests, northern Arizona, USA. *Ecological Applications* 22:204–217.
- Kennedy, M., and R. D. Gray. 1997. Habitat choice, habitat matching and the effect of travel distance. *Behaviour* 134:905–920.
- Lancaster, J. D., J. B. Davis, R. M. Kaminski, A. D. Afton, and E. J. Penny. 2015. Mallard use of a managed public hunting area in Mississippi. *Journal of the Southeastern Association of Fish and Wildlife Agencies* 2:281–287.
- Linden, A., and J. Knappe. 2009. Estimating environmental effects on population dynamics: consequences of observation error. *Oikos* 118:675–680.
- Lindstrom, J. M., M. W. Eichholz, and A. C. Behney. 2020. Effect of habitat management on duck behavior and distribution during spring migration in Indiana. *Journal of Fish and Wildlife Management* 11:80–88.
- Linz, G. M., C. P. McMurl, H. J. Homan, and D. L. Bergman. 1998. Comparison of point-count and wade-flush methods for counting ducks. *Prairie Naturalist* 30:243–249.
- Livolsi, M. C., K. M. Ringelman, J. M. Coluccy, M. T. Dibona, and C. K. Williams. 2015. Implications of uncertainty in true metabolizable energy estimates for estimating wintering waterfowl carrying capacities. *Wildlife Society Bulletin* 39:827–833.
- MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey, USA.
- Nhu, Y. D. T., N. T. Hoang, P. K. Lieu, H. Harada, N. Brion, D. V. Hieu, N. V. Hop, and H. O. Venterink. 2019. Effects of nutrient supply and nutrient ratio on diversity-productivity relationships of phytoplankton in the Cau Hai Lagoon, Vietnam. *Ecology and Evolution* 9:5950–5962.
- Nilsson, C., R. H. G. Klaassen, and T. Alerstam. 2013. Differences in speed and duration of bird migration between spring and autumn. *American Naturalist* 181:837–845.
- Northrup, J. M., and B. D. Gerber. 2018. A comment on priors for Bayesian occupancy models. *PLOS ONE* 13:e0192819.
- O’Shaughnessy, R. 2014. *Factors determining habitat selection by spring migrating waterfowl along the Wabash River, Illinois*. Dissertation. Southern Illinois University, Carbondale, Illinois, USA.
- Olmstead, V. G., E. B. Webb, and R. W. Johnson. 2013. Moist-soil seed biomass and species richness on wetland reserve program easements in the Mississippi Alluvial Valley. *Wetlands* 33:197–206.
- 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. *Journal of Wildlife Management* 81:1482–1493.
- Palumbo, M. D., S. A. Petrie, M. Schummer, B. D. Rubin, and S. Bonner. 2019. Mallard resource selection trade-offs in a heterogeneous environment during autumn and winter. *Ecology and Evolution* 9:1798–1808.
- Pilliod, D. S., C. S. Goldberg, R. S. Arkle, and L. P. Waits. 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA

- from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences* 71:1123–1130.
- Playa Lakes Joint Venture Waterfowl Team. 2005. Waterfowl Conservation Strategy for the Playa Lakes Joint Venture. Playa Lakes Joint Venture, Lafayette, Colorado, USA.
- Plummer, M. 2018. rjags: Bayesian graphical models using MCMC. R package version 4-8. <https://cran.r-project.org/web/packages/rjags/index.html>
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6:7–11.
- R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reeves, G. K., D. R. Cox, S. C. Darby, and E. Whitley. 1998. Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine* 17:2157–2177.
- Richardson, S., and W. R. Gilks. 1993. A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology* 138:430–442.
- Ringelman, J. K., W. R. Eddleman, and H. W. Miller. 1989. High plains reservoirs and sloughs. Pages 311–340 in L. M. Smith, R. L. Pederson and R. M. Kaminski, editors. *Habitat management for migrating and wintering waterfowl in North America*. Texas Tech University Press, Lubbock, Texas, USA.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
- Rota, C. T., M. A. R. Ferreira, R. W. Kays, T. D. Forrester, E. L. Kalies, W. J. McShea, A. W. Parsons, and J. J. Millspaugh. 2016. A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution* 7:1164–1173.
- Saveraid, E. H., D. M. Debinski, K. Kindscher, and M. Jakubauskas. 2001. A comparison of satellite data and landscape variables in predicting bird species occurrences in the Greater Yellowstone Ecosystem, USA. *Landscape Ecology* 16:71–83.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry: the principles and practice of statistics in biological research*. Third edition. W. H. Freeman and Co., New York, New York, USA.
- Soulliere, G. J., B. A. Potter, J. M. Coluccy, R. C. Gatti, C. L. Roy, D. R. Luukkonen, P. W. Brown, and M. W. Eichholz. 2007. Upper Mississippi River and Great Lakes Region Joint Venture Waterfowl Habitat Conservation Strategy. U.S. Fish and Wildlife Service, Fort Snelling, Minnesota, USA.
- Spaan, R. S., C. W. Epps, V. O. Ezenwa, and A. E. Jolles. 2019. Why did the buffalo cross the park? Resource shortages, but not infections, drive dispersal in female African buffalo (*Syncerus caffer*). *Ecology and Evolution* 9:5651–5663.
- St James, E. A., M. L. Schummer, R. M. Kaminski, E. J. Penny, and L. W. Burger. 2013. Effect of weekly hunting frequency on duck abundances in Mississippi wildlife management areas. *Journal of Fish and Wildlife Management* 4:144–150.
- U.S. Department of the Interior and Environment Canada. 1986. North American waterfowl management plan: a strategy for cooperation. U.S. Fish and Wildlife Service, Washington, D.C., USA.
- U.S. Fish and Wildlife Service. 2013. Colorado partners for fish and wildlife summary information January 2013. U.S. Fish and Wildlife Service Partners for Fish and Wildlife, Lakewood, Colorado, USA.
- Central Valley Joint Venture. 2006. Central Valley Joint Venture Implementation Plan – Conserving Bird Habitat. U.S. Fish and Wildlife Service, Sacramento, California, USA.
- Playa Lakes Joint Venture. 2008. Area Implementation Plan for the Shortgrass Prairie Bird Conservation Region (18) of Colorado. Playa Lakes Joint Venture, Lafayette, Colorado, USA.
- Webb, E. B., L. M. Smith, M. P. Vrtiska, and T. G. Lagrange. 2010. Effects of local and landscape variables on wetland bird habitat use during migration through the Rainwater Basin. *Journal of Wildlife Management* 74:109–119.
- Williams, C. K., et al. 2014. Estimating habitat carrying capacity for migrating and wintering waterfowl: considerations, pitfalls and improvements. *Wildfowl Special Issue* 4:407–435.
- Wimp, G. M., L. Ries, D. Lewis, and S. M. Murphy. 2019. Habitat edge responses of generalist predators are predicted by prey and structural resources. *Ecology* 100:e02662.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3273/full>