Morphological traits determine detectability bias in North American grassland butterflies

KATHERINE C. KRAL-O’BRIEN,1† BROOKE M. KARASCH,2 TORRE J. HOVICK,2 RAYMOND A. MORANZ,3 AND JASON P. HARMON1

1Department of Entomology, North Dakota State University, Fargo, North Dakota 58102 USA
2Range Science, North Dakota State University, Fargo, North Dakota 58108 USA
3Department of Integrative Biology, Oklahoma State University, Stillwater, Oklahoma 74078 USA

Citation: Kral-O’Brien, K. C., B. M. Karasch, T. J. Hovick, R. A. Moranz, and J. P. Harmon. 2020. Morphological traits determine detectability bias in North American grassland butterflies. Ecosphere 11(12):e03304. 10.1002/ecs2.3304

Abstract. Ecological research often includes observing and counting individuals to obtain a population estimate or index that can inform conservation, management, and policy. However, the ability to accurately estimate wildlife populations has always been hindered by bias, and researchers aim to overcome bias by using correction factors that calculate the relationship between observer counts and true counts. Accurate estimates of butterfly populations, for example, are especially necessary to inform management and policy, with over 40 species listed or proposed for listing under the United States Endangered Species Act. Researchers can utilize methods like line-transect distance sampling (LTDS) to help account for detection bias and calculate more accurate estimates, but species-specific traits or behaviors may influence survey effectiveness. We used LTDS to detect nearly 35,000 individuals of 33 species across five studies to calculate butterfly species’ effective strip widths (ESW)—a type of correction factor—across grasslands in the Great Plains, USA. To better understand how species’ traits influence detectability, we modeled the influence of species’ morphological, life history, and behavioral traits on ESW. The average ESW was 5.42 m, but varied from 1.84 to 12.6 m. We found that morphology (size and color) impacted the ability of observers to detect butterflies, with larger and brightly colored butterflies detected farther away from observers compared to smaller and dull-colored butterflies. Additionally, observers were generally better at detecting individuals while they were flying and nectaring compared to resting. Surprisingly, species’ life history and ecological traits did not help explain detectability differences. As conservation efforts continue to increase for butterflies, improved estimates of their population size will be necessary to evaluate management strategies and aid conservation decision-making for future policy. Future surveys need to consider butterfly size and color, adjusting weather protocols when necessary, to minimize and account for bias associated with butterfly species, especially if accurate population estimates are a study goal.

Key words: butterfly; detection probability; effective strip width; Great Plains; imperfect detection; insect; pollinators.

Received 9 June 2020; revised 5 August 2020; accepted 17 August 2020; final version received 8 October 2020.

Corresponding Editor: Robert R. Parmenter.

Copyright: © 2020 The Authors. 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: kralx009@gmail.com

INTRODUCTION

A primary objective of ecological research often involves counting individuals to provide a population estimate (Krausman and Cain 2013). Researchers count individuals to determine population trends, assess management strategies or treatments, provide information for policy and potential listings (e.g., IUCN Red List), and to focus conservation efforts. Especially when creating policy, reliable methods are necessary to count wildlife and produce accurate population
estimates (Kéry and Plattner 2007, Henry et al. 2015, Kral et al. 2018a). When sampling wildlife, observers may assume raw counts are accurate and all individuals are detected. However, this is rarely, if ever, the case (Boulinier et al. 1998). Researchers often use visual surveys to count individuals, but these surveys are prone to observer biases. Additionally, counts are often complicated by additional species’ trait-based or behavioral factors that diminish detectability of individuals. Improvements in technology (trail cameras, GPS tracking, etc.) can reduce bias but as long as human observers continue to be a major tool for wildlife surveys, bias will never be eliminated entirely (Wellington et al. 2014). Instead, attempting to determine and account for bias, whether related to species’ biology or observer, is much more plausible. When not corrected, bias can cause results to differ substantially from true population sizes and has serious implications for conservation, management, or policy (Melero et al. 2016, Harabìs et al. 2019). Consequently, caution should be taken to understand how species’ traits and behaviors can influence wildlife surveys, and when known, surveys should incorporate ways to account for bias that impact species’ detection (Pellet et al. 2012).

Researchers may lack species-specific information on bias, but they can still account for bias when conducting wildlife surveys. A major source of bias comes from perceptibility. Perceptibility relates to the observer being able to detect an individual (Johnson 2008). Species’ perceptibility can be biased by observer experience (Boulinier et al. 1998, Kéry and Plattner 2007, van Swaay et al. 2008, Isaac et al. 2011), species’ mobility (Brown and Boyce 1998), species’ size and color (Dennis et al. 2006, Kéry and Plattner 2007), species’ behavior (Dennis et al. 2006, Kéry and Plattner 2007, Pellet et al. 2012), transect placement (Kéry and Plattner 2007, Harker and Shreeve 2008), weather conditions (van Swaay et al. 2008), vegetation structure (Brown and Boyce 1998, van Swaay et al. 2008), and vegetation type (Pellet et al. 2012). Some of these sources of biases can be reduced during survey design with timing, weather protocols, and stratifying site selection (Pellet 2008). However, others must be corrected for on a species-specific or project basis after data collection. Therefore, an understanding on the biases associated with multiple species would be meaningful in guiding survey design and garnering more accurate data for policy formation.

The conservation prioritization and concomitant research interest for butterflies has been increasing over recent decades (Kral et al. 2018a). Accordingly, more accurate population estimates are necessary to guide policy and conservation (Henry and Anderson 2016). Currently, 186 butterfly species are globally listed as Critically Endangered, Endangered, or Vulnerable (IUCN 2019), with over 40 species listed or proposed for listing under the Endangered Species Act in the United States (USFWS 2019). As with other wildlife, bias impacts the ability to visually detect butterflies and accurately estimate populations. In addition to these challenges, butterflies are fragile species, and, not only does detectability change between species (Boulinier et al. 1998, Dorazio et al. 2006), detectability changes for the same species across its range (Kéry and Plattner 2007). Consequently, research addressing mechanisms contributing to bias and detectability differences is necessary across multiple species in various ecosystems to accurately determine species’ abundance and distribution (Dennis et al. 2006).

Although butterflies are relatively easy to survey because they usually fly near the ground, are diurnal, and easy to identify, correction factors are still necessary for raw counts. Correction factors are multipliers used in combination with raw counts to account for bias (Table 1), but how they are calculated differs upon data availability and coverage. In Europe, observers frequently survey butterflies with a transect count method (often called a Pollard walk) where an observer walks along a transect and counts butterflies in the 5 × 5 m area in front of them (Pollard 1977, Kral et al. 2018b). The original method assumes equal detectability among species (Pellet et al. 2012), so researchers are encouraged to model correction factors after data collection using detection histories (Pellet 2008) or mark–recapture data (Gross et al. 2007, Harker and Shreeve 2008, Pellet et al. 2012). With a long history of use and wide coverage throughout the UK and Europe, researchers can determine species-specific correction factors with relative ease. In the USA, however, researchers are often
limited by poor coverage and the absence of long-term datasets. Therefore, some use line-transect distance sampling (LTDS) to survey butterflies and account for detectability differences during data collection (Buckland et al. 2001). Unlike Pollard walks, observers are not limited to an arbitrary distance. Instead, in addition to counting butterflies, observers estimate distances from individual butterflies to the transect in order to understand how detection patterns change between individuals at specific distances (Moranz et al. 2012, Henry et al. 2015). In general, species are less likely to be observed farther from the transect, so researchers can use distance data for each species to calculate an effective strip width or detection probability (correction factors; Buckland et al. 2001, 2004).

Correction factors are important sources for limiting bias in butterfly surveys. However, most information on butterfly detectability comes from Europe where correction factors—often detection probabilities (DP, Table 1)—vary based on seasonal timing of surveys (Pellet 2008), increase with observer experience, and tend to increase from univoltine to multivoltine species (Melero et al. 2016). In the USA, researchers utilizing LTDS report effective strip width (ESW, see Table 1) because DPs often differ based on the distance observers detect butterflies from a transect, which changes between species (Isaac et al. 2011, Henry and Anderson 2016; see Methods for more details). Effective strip width, like DPs, vary based on species but generally tend to increase with wingspan (Moranz 2010), color brightness (Isaac et al. 2011), and more open landscapes (Pocewicz et al. 2009). Overall, the butterfly research community has just begun answering questions about how environmental or biological factors influence detectability and inform bias, and consequently, a large number of butterflies lack correction factors (either DPs or ESWs). Since no consensus has been found on what consistently influences detectability, additional variables may be responsible and should be explored for butterflies using large datasets when possible.

As our understanding of species’ detectability increases, researchers will have more information on how to correct count data, inform survey design, and calculate accurate population estimates. Correcting for bias becomes increasingly important when comparing between treatments, studies, or landscape types (McNeil et al. 2019, Monroe et al. 2019). However, most species lack formal estimates of DP or ESW, and little is known how perceptibility biases impact detection. Accordingly, we utilized data from five different studies to create a large dataset (i.e., >34,000 detections) that will allow us to quantify the influence of a priori morphological, ecological, and behavioral traits on ESW. Specifically, our objectives were to (1) calculate and summarize ESWs for 33 grassland butterfly species in the northern Great Plains, (2) model the influence of species’ traits on ESW, and (3) assess the influence of butterfly behavior on ESW. This information will be vital in providing correction factors for a large number of species that can be used to better understand observer bias and species detectability. Moreover, this information will be critical for forming survey protocols for species of conservation concern to minimize bias and maximize effectiveness for successful management plans and policy formation.

Methods

Butterfly surveys

We combined survey data from five independent studies conducted across the northern Great Plains from 2015 to 2018 to address our three objectives. All studies were part of graduate student research projects from the same institution. In each study, researchers randomly placed transects in grasslands to survey butterfly abundance

Table 1. List of terms and abbreviations utilized in the manuscript related to detectability and distance sampling.

| Term or abbreviation | Definition |
|----------------------|------------|
| Correction factor    | A number or factor multiplied by raw counts to account for bias (Kissling and Garton 2006) |
| Detection probability (DP) | Probability of observing an object in a defined area (P; Thomas et al. 2010) |
| Effective strip width (ESW) | The distance from the line at which as many individuals are detected beyond as are missed within (Thomas et al. 2002:545) |
| Imperfect detection  | An individual is present but not detected (Kellner and Swihart 2014) |
| LTDS                 | Line-transect distance sampling |
and species richness, and all observers used the same distance sampling methods. Although study design varied slightly with regard to transect length, observers, and site visits per year (Appendix S1: Table S1), differences were minimal, and we were able to account for this in our modeling procedures (see Modeling in program distance). All studies used similar survey protocols to increase the likelihood of detecting butterflies. Surveys were conducted between 09:00 and 18:00 h (CDT) when weather conditions were as follows: temperature >15°C, wind speed <20 km/h, and cloud cover <50% (Royer et al. 1998, Kral et al. 2018b).

We used the same methodology, regardless of study, to conduct LTDS on 1,066 individual transects (Kral et al. 2018b). The observer would walk a transect marked with a measuring tape at a slow pace (10 m/min) at least twice a year between June and August. During a survey, the observer would scan 180 degrees in front of them between June and August. During a survey, the observer would record the species, estimate the perpendicular distance from the butterfly to the transect, and record the behavior (basking, flying, nectaring, mating, resting) of each individual butterfly. If observers could not identify a butterfly, they would either capture it or photograph it for later identification. Before each field season, all observers were trained to identify butterflies and conduct LTDS surveys. Specifically, observers were trained to (1) detect all butterflies on the line, (2) detect butterflies at their initial location, and (3) accurately record distances from butterflies to the line in order to meet assumptions of LTDS (Buckland et al. 2001).

**Modeling in program distance**

We estimated species ESWs and DPs using Program Distance (v. 7.2, Thomas et al. 2010) for all species detected ≥30 times (Stanbury and Gregory 2009). We pooled detections for species across all sites, studies, and years to increase the number of detections (Pocewicz et al. 2009). We created models using all possible combinations of (1) the half-normal or hazard rate key functions and (2) cosine, simple polynomial, or hermite polynomial series expansions (Thomas et al. 2010). We identified top models using a standard practice in distance modeling that combines Akaike’s information criterion (AIC) adjusted for small sample sizes (AICc), chi-square (χ²) goodness-of-fit tests, and visual evaluation of detection curves (Buckland et al. 2001, Burnham and Anderson 2002, Isaac et al. 2011). Top models had the lowest AICc (or within 2 AICc of the lowest AICc), passed χ² tests to ensure a better fit was not probable, and produced histogram curves with the expected shape of a large shoulder and decreasing tail (see Pocewicz et al. 2009 for examples). Of the six potential models, we selected the best models for the next modeling step. In addition, we truncated the data by initially removing 5% of the detections that occurred farthest from the transect as recommended by Thomas et al. (2010). The purpose of truncation is to reduce outliers for each species and create smooth detection curves that resemble classic detection curves with large shoulders and a gradually decreasing flatline. We adjusted the amount of truncation (≥5% ≤) based on χ² goodness-of-fit tests and visual evaluation of detection curves to ensure truncation levels were neither too stringent nor too lenient. Additionally, we created models that incorporated covariates that might influence detection including Julian date, time of day, wind speed, percent cloud cover, temperature, and observer (Appendix S1: Table S2).

For species with ≥30 detections in at least two of the five potential behavior categories (i.e., basking, flying, nectaring, mating, resting), we used the selected model from above and post-stratified ESW by behavior (Moranz 2010). Observers did not detect each species in every possible behavior. Therefore, we only post-stratified 14 different species by behavior, and we limited behaviors to those that occurred most often, including flying, nectaring, and resting. We combined basking with resting and removed mating due to low sample sizes (<30).

**Species traits**

We selected and categorized nine a priori butterfly traits that may change species’ perceptibility to influence ESW and DP (Table 2; Appendix S1: Table S2). Our primary morphological traits were wingspan and color. We established average wingspan (mm) for each species using a regional field guide (Royer 2003). Color can be difficult to standardize across butterfly species,
so we quantified color two different ways using a subjective and objective method. To quantify color subjectively, we took scanned images of museum specimens sourced from Royer (2003), resized them so all individuals were relatively the same size, and placed them on a universal camouflage background to survey 16 past and present butterfly surveyors (Appendix S1: Fig. S1A). Participants viewed 33 slides, each showing an individual butterfly, and rated how well they could differentiate the butterfly from the background (1, extremely dull; 2, moderately dull; 3, slightly dull; 4, neither dull nor bright; 5, slightly bright; 6, moderately bright; and 7, extremely bright), similar to the survey conducted by Dennis et al. (2006) to quantify visual apparency. If participants took the quiz more than once ($\bar{x} = 2$), we averaged their results to determine visual apparency of butterflies. To determine color objectively, we used one museum specimen photograph from Royer (2003), one photograph of an individual with wings closed for a ventral view, and one photograph of an individual with wings open for a dorsal view from an online source (Lotts and Naberhaus 2017). We then scaled these photographs to equivalent digital pixel measurements and used the colorDistance R package (Weller and Westneat 2019) to compare each of the species’ three photographs to a well-lit, neutral grassland background (Appendix S1: Fig. S1B). The final output value for objective color is a percentage of similarity to the grassland background.

In addition to morphological traits, we incorporated several behavioral and life history traits that could impact detection. We included flight style, mating habit, and oviposition preference as behavioral traits and month of adult emergence, voltinism, and duration of flight season as life history traits (Table 2). We determined most of these traits using field guides (Glassberg 2001, Royer 2003, Hébert-Allard 2013) and online resources (Lotts and Naberhaus 2017). However, we determined duration of flight season specific to our study region by using the average first and last day we observed a species across the four years of the five studies.

### Statistical analyses

First, we present the ESWs and DPs for the species with over 30 detections. This will be the largest compilation of species-specific ESW and DP for butterflies to date. By presenting individual species information, we will be able to depict the variation between species and begin modeling trait effects on ESW and DP. For the remainder of our analyses, we focused solely on ESW.

**Table 2. Summary statistics and descriptions for all a priori morphological, life history, and ecological covariates used for modeling differences in effective strip width among butterflies surveyed in the northern Great Plains, USA, from 2015 to 2018.**

| Covariate          | Type       | Mean (SE) | Range       | Description                                                                 |
|--------------------|------------|-----------|-------------|-----------------------------------------------------------------------------|
| Objective color    | Continuous | 0.37 (0.02) | 0.17–0.70   | Percent similarity to neutral grassland background (Weller and Westneat 2019) |
| Subjective color   | Continuous | 2.67 (0.16) | 1.06–4.39   | Visual apparency against camouflage background rated from extremely dull to extremely bright (Dennis et al. 2006) |
| Duration of flight season | Continuous | 54.4 (3.11) | 26.7–95.0 | Average number of days adults of a species were detected in North Dakota |
| Flight style       | Categorical |           |             | Movement pattern of a species flying (erratic, strong, bouncy, or gliding; Brock and Kaufman 2003, Royer 2003) |
| Mating habit       | Categorical |           |             | Male courtship display (Patrol, perch, or both; Lotts and Naberhaus 2017) |
| Month of emergence | Categorical |           |             | Month when the first adults emerge (May, June, or July; Royer 2003) |
| Oviposition behavior | Categorical |           |             | Female choice of egg-laying location (not on host, under host, or on host) |
| Voltinism†         | Categorical |           |             | Number of broods each year (Univoltine, bivoltine, or multivoltine; Royer 2003) |
| Wingspan           | Continuous | 45.2 (3.40) | 19.5–97.5   | Average wingspan of individual species (mm; Royer 2003)  

† Covariate not included in model creation due to high correlations ($r \geq 0.60$) with additional covariates.
dependent upon users’ choices regarding truncation (Henry and Anderson 2016), leading us to assume DP may be unreliable for modeling the effects of species’ traits. 

Next, we used generalized linear modeling (GLM) to model the impact of species’ traits on ESW in the R statistical environment (v. 3.5.3; Pocewicz et al. 2009, R Development Core Team 2015). Prior to creating models, we tested for correlations between traits and eliminated redundant covariates that were highly correlated ($r \geq 0.60$; Kral et al. 2018b). Duration of flight season was correlated with voltinism ($r = 0.66$), so we chose to retain duration of flight season since this covariate was more specific to our survey region. Then, we created univariate models for each covariate and compared them to the null model. We created additive models from univariate models only if AIC$_c$ was less than the null and within 2 AIC$_c$ of the best model to reduce the number of model combinations (Burnham and Anderson 2002, Loss and Blair 2011). Finally, we selected the final model based on model weights and AIC$_c$ scores (Burnham and Anderson 2002). We validated our models using a best subset model selection procedure with the MuMIn package in R (Bartoń 2019). This procedure creates models with all possible covariate combinations, in our case one to nine covariates. The best subset model selection procedure does not eliminate covariates in each subset until all models have been created. All models within 2 AIC$_c$ of the best model were competitive and moved to the next step where they were compared to the top models in other subsets. The biggest difference with this approach is that all covariates can be utilized in subsequent subsets. With our original model creation, we eliminated univariate covariates if they were not within 2 AIC$_c$ of the top model. The best-ranked models from both approaches were the same, so we will present the results from the a priori modeling procedure.

Finally, we compared differences in ESW based on behavior using the 14 species with $\geq 30$ detections in at least two different behaviors. We used an ANOVA to make comparisons after confirming all assumptions of ANOVA had been met. If we found a significant difference ($\alpha \leq 0.05$), we used Tukey’s HSD for pairwise comparisons in R.

RESULTS

Program distance summary

We detected 34,531 butterflies representing 60 species when combining the five studies. Thirty-three of these species met our minimum requirement of $\geq 30$ detections, and species used in modeling had 30–11,248 detections (Appendix S1: Table S2). Additionally, the maximum width of transects (i.e., the farthest distance observers detected an individual) varied across species from 12 to 81 m prior to truncation. Within Program Distance, we commonly used the hazard rate key function in 26 species’ models, with the remaining using half-normal. Hermite polynomial and cosine were the most common series expansions, with the remainder using simple polynomial (Appendix S1: Table S2). We truncated 26 of the 33 species at $\leq 5\%$, but several species required a larger truncation (i.e., 5 with 6–10% truncation and 2 with >10% truncation; Appendix S1: Table S2). The final widths of the transects after truncation varied from 6 to 36 m (Appendix S1: Table S2).

We included additional detection covariates for 11 of the 33 species, but no models included more than one detection covariate (Appendix S1: Table S2). Temperature, cloud cover, and observer were each included once, Julian date was included twice, and wind was included six times to improve species’ models (Appendix S1: Table S2). Most improvements were minimal. For example, reducing cloud cover from 25% to 5% made the detection probability for Polites mystic (W.H. Edwards; Hesperiidae) at 4 m increase from 0.38 to 0.44. However, some weather variables altered detection probabilities more drastically. Increasing temperatures from 22°C to 31°C improved detectability of Lycaena hylus (Cramer; Lycaenidae) at 2 m from 0.50 to 0.95. Since wind was included for more than one species’ model, the overall results are less consistent. Minimal increases in wind speeds (e.g., 6–10 kph) improved detection probabilities for five species including Limenitis archippus (Cramer; Nymphalidae), Papilio polyxenes (Fabricius; Papilionidae), Phyciodes batesii (Reakirt; Nymphalidae), Polites themmistocles (Latreille; Hesperiidae), and Pyrgus communis (Grote; Hesperiidae), Polites peckius (W. Kirby; Hesperiidae) was the only species where detection probabilities decreased with increasing...
wind speeds within our current weather protocol. Detection probabilities at 2 m increased from 0.42 to 0.60 when winds decreased from 9 to 6 kph.

**Effective strip width and detection probability summary**

From each species’ final model, we extracted the ESW from Program Distance (Fig. 1). The average ESW was 5.42 m (±0.39 SE), ranging from 1.84 to 12.6 m (Fig. 1). Over 80% of ESW were between 3.00 and 7.00 m (Fig. 1). In addition to ESW, we determined DP for each species (Appendix S1: Fig. S2). The average DP was 0.34 (±0.02 SE), ranging from 0.20 to 0.51 (Appendix S1: Fig. S2).

**Influence of species’ traits on effective strip width**

The best-ranked model for ESW included the morphological traits of wingspan and subjective color (Fig. 2; Appendix S1: Table S4). This was the top model for both modeling procedures. For each mm increase in wingspan, ESW increased by 0.04 m. When comparing species with the smallest to largest wingspan, the ESW would increase by ~3 m. Conversely, as butterflies became duller by one rating (e.g., 6, moderately bright to 5, slightly bright), ESW decreased 0.88 m. When comparing species with the dullest to brightest color ratings, the ESW would also increase by ~3 m. However, we determined using ANOVA that these variables did not interact ($F_{3,29} = 0.54, P = 0.60$). The remaining behavioral (mating habit, oviposition behavior, and flight style) and life history traits (month of adult emergence, voltinism, and duration of flight season) were not included in our best-ranked models, although mating habit was included in an unselected competitive model (Appendix S1: Table S4).

**Influence of behavior on effective strip width**

We found that ESW differed between behaviors ($F_{2,38} = 7.63, P = 0.002$), increasing from resting to flying to nectaring (Fig. 3). We determined the average flying ESW was 5.58 m (±0.50 SE). Since most butterflies were observed flying (Appendix S1: Table S5), the flying ESW was the closest to the average total ESW (5.42 m). The ESW for eight of the fourteen species significantly increased when they were observed nectaring, and the average nectaring ESW was 7.25 m (±1.28 SE). Conversely, when butterflies were detected resting, their ESWs decreased and averaged 2.21 m (±0.34 SE; Fig. 3; Appendix S1: Table S5).

**DISCUSSION**

In North America, few previous studies have included correction factors for butterfly abundance estimates in an attempt to determine what alters detectability. Therefore, our main objectives were to calculate correction factors and determine how species’ traits and behaviors influence them. In doing so, we expect our findings to be applied to future butterfly surveys in order to showcase the importance of including correction factors and help improve abundance estimates for grassland butterflies, despite the fact that no corrections would every work perfectly. Without accurate estimates of abundance, new policy and guidelines for conservation may be misinformed and decrease the effectiveness of limited money devoted to conservation. Using one of the largest datasets ever assembled in North America that accounts for detectability, we found (1) species’ ESWs to be heterogeneous, (2) the variation in ESW for species was primarily due to wingspan and color, and (3) behavior influenced ESW with active individuals being detected farther away than resting individuals. Based on our findings, researchers need to account for detectability changes due to imperfect detection, along with species’ morphology and behavior. In the field, observers need to be particularly aware of small and dull individuals which may go undetected, especially if these species are specifically being pursued.

The application of our results may commonly focus on informing survey design for individual species monitoring projects. For individual species, our ESWs can determine how far observers may be expected to effectively survey species. As an example, surveys for *Danaus plexippus* (Linnaeus; Nymphalidae) can be designed to search areas beyond traditional 2.5 m or even 5 m, since they are larger, brighter, and the calculated ESW for open grasslands in this study was over 8 m. Additionally, researchers particularly interested in threatened and endangered species can use our results to estimate survey effort necessary to
Fig. 1. Effective strip width (ESW) calculated in Program Distance for 33 species detected ≥30 times in the northern Great Plains from 2015 to 2018. We combined detections for each species across behaviors. The traditional Pollard walk truncation (2.5 m) is shown with a solid red line, and the overall average ESW for the 33 species is shown with a dashed black line. The totals at the bottom of the figure depict the number of species which fell within each range. Butterfly six letter codes: Polites peckius (POLPEC), Polites themistocles (POLTHE), Euptoicta claudia (EUPCLA), Pyrgus communis (PYRCOM), Phyciodes cocyta (PHYCOC), Polites mystic (POLMYS), Coenonympha tullia (COETUL), Glaucopsyche lygdamus (GLALYG), Plebejus melissa (PLEMEL), Phyciodes batesii (PHYBAT), Phyciodes tharsus (PHYTHA), Vanessa atalanta (VANATA), Chlosyne gorgone (CHLGOR), Speyeria aphrodite (SPEAPH), Cupido comyntas (CUPCOM), Satyrides eurydice (SATEUR), Cercoyonis pegala (CERPEG), Speyeria cybele (SPECYB), Celastrina neglecta (CELNEG), Papilio polyxenes (PAPPOL), Speyeria idiaria (SPEIDA), Lycaena hyllus (LYCHYL), Boloria bellona (BOLBEL), Lycaena helleides (LYCHEL), Boloria selene (BOLSEL), Vanessa cardui (VANCAR), Colias philodice (COLPHI), Pontia protodice (PONPRO), Pieris rapae (PIERAP), Vanessa virginiensis (VANVIR), Danaus plexippus (DANPLE), Colias eurytheme (COEUR), Limenitis archippus (LIMARC).
achieve the number of required detections or train observers based on species-specific morphological differences to increase detection rates. Since most butterfly species lack detectability estimates (Kéry and Plattner 2007), our dataset is important for future research and conservation for individual species, but the results can also be used for multi-species projects. Multi-species projects can utilize our data to evaluate how their count data may be misrepresenting accurate population estimates—that is, gauge how large or small corrections may be for certain species—and provide an example of effectively utilizing LTDS for multi-species monitoring and incorporating correction factors into survey methods. Corrections may not be necessary for all species (Isaac et al. 2011), and some methods may assume equal detectability for all species. However, our DPs were never over 0.51, and ESWs varied greatly between species (Fig. 2; Appendix S1: Fig. S2), signifying a need to use correction factors for grassland butterflies. To illustrate this point, we can use an example from our dataset for two species, *D. plexippus* and *Coeonympha tullia* (Müller; Nymphalidae). Both species had similar raw counts—715 detections

Fig. 2. Influence of wingspan and subjective color brightness on effective strip width (ESW) for 33 grassland butterfly species in the northern Great Plains from 2015 to 2018. Wingspan was the average wingspan of individual species (mm; Royer 2003), and subjective color was the visual appearance of species against a camouflage background rated from extremely dull to extremely bright (Dennis et al. 2006). Black markers represent individual species. Species in the upper right-hand corner of the graph are more likely to be detected farther away by observers, while species in lower left-hand corner are less likely to be detected farther away.

Fig. 3. Average effective strip width (ESW) for all species combined (all species) and for the 13 species with enough detections in each behavior (flying, nectaring, or resting) to stratify ESW by behavior. ESWs for individuals flying and nectaring were larger compared to resting individuals ($F_{2,38} = 7.63$, $P = 0.002$). Different letters denote a significant difference between behaviors ($P \leq 0.05$).
for *D. plexippus* and 1,005 for *C. tullia*. This may lead to the assumption that they occur in relative equal proportions in grassland communities. However, the ESW for *D. plexippus* was over double the ESW for *C. tullia* (8.37 and 3.49 m, respectively). This means we could effectively survey *D. plexippus* over twice the distance of *C. tullia*, and observers missed more *C. tullia* than *D. plexippus* when surveying. We can use these correction factors to determine the density of individuals per hectare, and over a 200-ha area, the corrected counts become 240 *D. plexippus* individuals and 806 *C. tullia* individuals. Therefore, correction factors may be especially important in research concerning butterfly communities to ensure general relationships are accurately depicted. Although we find it more useful for researchers to collect data that can inherently produce correction factors, our ESWs and DP's can be constructive for approximating estimates or gauging the number of species undetected but present in grassland landscapes. They are particularly useful when long-term datasets are lacking or when making comparisons among treatments or habitat types are necessary (McNeil et al. 2019). Correcting for imperfect detection and bias are imperative moving forward to accurately depict vulnerable butterfly populations.

We found wingspan and color explained the most variability in species’ ESW. These results were similar to other studies that found ESW increased with wingspan and color brightness (Moranz 2010, Isaac et al. 2011). Wingspan and color relate to species’ detectability, and additional covariates (e.g., mating habit, duration of flight period) that may also cause observer bias were not included in best-ranked models. Categorizing butterfly behavioral traits such as their inclination to patrol (Pellet et al. 2012) or activity height (Dennis et al. 2006) have previously been helpful in partitioning observer bias. Therefore, we expected to include other variables such as flight style, although our categorization may not have been sufficient to detect inter-species differences. However, our results support the previous research identifying color and size as major drivers of observer bias for butterflies.

Behavior influenced species’ ESWs, with observers detecting species farther away when they were flying and nectaring compared to resting. Behavior is expected to influence detectability (Dennis et al. 2006), and we anticipated nectaring to universally improve ESW because butterflies would be more conspicuous when nectaring on flowerheads (Moranz 2010). Although species were not always easier to detect nectaring and flying, this information can additionally inform survey protocols. Observers cannot control the behavior of butterflies, but they can try to reduce the chances of encountering certain behaviors (e.g., resting or basking individuals). For example, surveying during specific portions of the day or specific parts of the flight season when butterflies are more active can reduce the probability of encountering resting individuals, improving ESW. Flight season patterns will change from species to species, but protocols for surveying during the time of day when butterflies are more active can likely be addressed with weather protocols.

Weather covariates influenced observer bias in our study. We expected to retain covariates such as observer and date in best-ranked models, since previous studies have found them to impact detectability (Harker and Shreeve 2008, Melero et al. 2016), but they were rarely used. Additionally, we did not expect to include weather covariates in our best-ranked models, since we developed weather protocols to eliminate this bias during survey design (see Methods). However, weather covariates (wind, temperature, and cloud cover) were included for 8 of 11 species with additional detection covariates in Program Distance. Although we used weather protocols to minimize bias, we suggest researchers consider using stricter weather protocols if targeted species are particularly difficult to detect in windier (e.g., *P. peckius*) or cloudier (e.g., *P. mystic*) conditions, as long as survey protocols do not become too restrictive (i.e., reducing the number of available survey days). Since some threatened and endangered butterfly species currently listed in North American grasslands—*Hesperia dacotae* (Linnaeus; Hesperiidae) and *Oarisma powershiek* (Parker; Hesperiidae)—are both small and dully colored, slight changes could be valuable for improving detection probability. Either way, reducing bias at multiple points through protocol changes and modeling procedures allows researchers to increasingly improve species estimates (McNeil et al. 2019).
As research on butterflies increasingly uses correction factors, trends begin to emerge for butterfly families. We averaged ESWs within butterfly families to informally compare them with average ESWs for butterfly families from previous studies. Pierid ESWs from other studies averaged 7.12 m (Moranz 2010, Isaac et al. 2011), and within our dataset, they averaged 7.83 m. Nymphalid ESWs from other studies averaged 5.65 m (Moranz 2010, Isaac et al. 2011) and 5.60 m from our data. The largest difference was for lycaenid ESWs, which averaged 1.87 m from other studies (Brown and Boyce 1998, Moranz 2010, Isaac et al. 2011, Henry and Anderson 2016) and 5.02 m in our study. This difference is likely due to some of the studies occurring in woody areas (Brown and Boyce 1998, Henry and Anderson 2016) instead of open grasslands found in our study region. Moving from open to closed canopy habitats generally decreases species’ ESW (Pocewicz et al. 2009). More research will continue to enumerate the relationship between raw counts and corrections to accurately depict populations (Bart et al. 2004).

Although we tested a wide range of species’ traits, we acknowledge some limitations with our large dataset. First, we were unable to test some covariates used in other studies such as flight height (Dennis et al. 2006) because we did not measure this data nor is it readily available for North American butterflies. Determining flight height would be very valuable to add in case flight height and typical observer scanning techniques did not match up. However, given that butterflies generally fly low in grasslands, this may be less important in our study system. Second, our metric for abundance may not have been detailed enough to aid our models. Long-term datasets with butterfly count data are generally absent in the northern Great Plains or are not detailed enough to allow us to utilize them. Third, we did not include phylogenetic histories in our modeling to help explain variation between species, but phylogenetics are not always useful for detecting patterns among butterflies (Pavoine et al. 2014). Even though some covariates were excluded, our data generally aligned with similar research while offering new insights for butterfly surveying based on a large dataset. Fourth, we were not able to use distance sampling for low abundance species since Program Distance calculates more accurate species’ estimates with approximately 60 detections (Buckland et al. 2001). This issue has previously been acknowledged by other researchers using distance sampling (Henry and Anderson 2016), but distance sampling does improve the accuracy of estimates (Bart et al. 2004), which is especially important for species of conservation concern or where long-term datasets may not be available (Isaac et al. 2011, Kral et al. 2018a). More work will be necessary to improve our ability to accurately capture population estimates, particularly for species with fewer detections.

Ecologists and conservationists have long recognized that visually observing wildlife is biased by observer perceptibility. Corrections factors, such as ESW, can be used to calculate more accurate estimates from count data, but they are not consistently used nor is it widely known how these change across individuals in multiple regions and ecosystems due to species’ morphology, behavior, and ecology. We found large variability between ESWs for grassland butterflies in the northern Great Plains, and several species were difficult to detect even when within 2 m of an observer. We attributed ESW variability to detectability biases associated with butterfly morphology (size and color) and behavior. Others can apply our results to their own research by carefully considering how the wingspan, color, and behavior of target species impacts their ability to detect individuals. Using our results, researchers can specially train observers on how wingspan and color will influence the number of detections, adjust weather protocols, and understand how including their own correction factors will increase the accuracy of estimates. Increasing the use of ESW, or some other type of correction factor for detectability, in count data will improve our ability to determine population estimates, measure treatment or management impacts accurately, and effectively direct future conservation efforts and policy for butterfly species.

ACKNOWLEDGMENTS

We would like to thank everyone involved with the five projects, especially graduate students (Adrienne Antonsen, Cayla Bendel, Cameron Duquette, Chyna Pei, and Jasmine Cutter) and technicians that provided
and collected the data. Funding for this project was provided by the USDA National Institute of Food and Agriculture (Hatch project number ND02394). KCK, BMK, TJH, RAM, and JPP developed the concept of the manuscript. KCK and BMK led the writing and analysis. All authors contributed to editing and gave approval for the final manuscript. The authors declare no conflicts of interest.

LITERATURE CITED

Barto, J. 2019. Package MuMIn. https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf

Boulinier, T., J. D. Nichols, J. R. Sauer, J. E. Hines, and R. L. H., T. G. Shreeve, N. J. B. Isaac, D. B. Roy, Burnham, K. P., and D. R. Anderson. 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Dennis, R. L. H., T. G. Shreeve, N. J. B. Isaac, D. B. Roy, P. B. Hardy, R. Fox, and J. Asher. 2006. The effects of visual apparency on bias in butterfly recording and monitoring. Biological Conservation 128:486–492.

Dorazio, R. M., J. A. Royle, B. Söderström, and A. Glimskär. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87:842–854.

Glassberg, J. 2001. Butterflies through binoculars: The West. Oxford University Press, New York, New York, USA.

Gross, K., E. J. Kalendra, B. R. Hudgens, and N. M. Haddad. 2007. Robustness and uncertainty estimates of butterfly abundance from transect counts. Population Ecology 49:191–200.

Harabiš, F., P. Jakubec, and J. Hronková. 2019. Catch them if you can! Do traits of individual European dragonfly species affect their detectability? Insect Conservation and Diversity 13:303–312.

Harker, R. J., and T. G. Shreeve. 2008. How accurate are single site transect data for monitoring butterfly trends? Spatial and temporal issues identified in monitoring Lasiommata megera. Journal of Insect Conservation 12:125–133.

Hébert-Allard, S. 2013. Manitoba butterflies: a field guide. Turnstone Press, Winnipeg, Manitoba, Canada.

Henry, E. H., and C. T. Anderson. 2016. Abundance estimates to inform butterfly management: double-observer versus distance sampling. Journal of Insect Conservation 20:505–514.

Henry, E. H., N. M. Haddad, J. Wilson, P. Hughes, and B. Gardner. 2015. Point-count methods to monitor butterfly populations when traditional methods fail: a case study with Miami blue butterfly. Journal of Insect Conservation 19:519–529.

Isaac, N. J., K. L. Cruickshanks, A. M. Weddle, J. M. Rowcliffe, T. M. Brereton, R. L. Dennis, D. M. Shuker, and C. D. Thomas. 2011. Distance sampling and the challenge of monitoring butterfly populations. Methods in Ecology and Evolution 2:585–594.

IUCN. 2019. The IUCN Red List of Threatened Species. Version 2019–1. http://www.iucnredlist.org

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

Kellner, K. F., and R. K. Swihart. 2014. Accounting for imperfect detection in ecology: a quantitative review. PLoS ONE 9:e11436.

Kéry, M., and M. Plaffner. 2007. Species richness estimation and determinants of species detectability in butterfly monitoring programmes. Ecological Entomology 32:53–61.

Kissling, M. L., and E. O. Garton. 2006. Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. Auk 123:735–752.

Kral, K. C., J. P. Harmon, R. F. Limb, and T. J. Hovick. 2018a. Improving our science: evolution of butterfly sampling and surveying methods over time. Journal of Insect Conservation 22:1–14.

Kral, K. C., T. J. Hovick, R. F. Limb, and J. P. Harmon. 2018b. Multi-scale considerations for grassland butterfly conservation in agroecosystems. Biological Conservation 226:196–204.

Krausman, P. R., and J. W. Cain. 2013. Wildlife management and conservation: contemporary principles and practices. JHU Press, Baltimore, Maryland, USA.
Loss, S. R., and R. B. Blair. 2011. Reduced density and nest survival of ground-nesting songbirds relative to earthworm invasions in northern hardwood forests. Conservation Biology 25:983–992.

Lotts, K., and T. Naberhaus, coordinators. 2017. Butterflies and moths of North America. http://www.butterfliesandmoths.org/ (Version 07 June 2019).

McNeil, D. J., C. R. Otto, E. L. Moser, K. R. Urban-Mead, D. E. King, A. D. Rodewald, and J. L. Larkin. 2019. Distance models as a tool for modelling detection probability and density of native bumblebees. Journal of Applied Entomology 143:225–235.

Melero, Y., C. Stefanescu, and J. Pino. 2016. General declines in Mediterranean butterflies over the last two decades are modulated by species traits. Biological Conservation 201:336–342.

Monroe, A. P., G. T. Wann, C. L. Aldridge, and P. S. Coates. 2019. The importance of simulation assumptions when evaluating detectability in population models. Ecosphere 10:e02791.

Moranz, R. A. 2010. Effects of ecological management on tallgrass prairie butterflies and their nectar sources. Dissertation. Oklahoma State University, Stillwater, Oklahoma, USA.

Moranz, R. A., D. M. Debinski, D. A. McGranahan, D. M. Engle, and J. R. Miller. 2012. Untangling the effects of fire, grazing, and land-use legacies on grassland butterfly communities. Biodiversity and Conservation 21:2719–2746.

Pavoine, S., M. Baguette, V. M. Stevens, M. A. Leibold, C. Turlure, and M. B. Bonsall. 2014. Life history traits, but not phylogeny, drive compositional patterns in a butterfly metacommunity. Ecology 95:3304–3313.

Pellet, J. 2008. Seasonal variation in detectability of butterflies surveyed with Pollard walks. Journal of Insect Conservation 12:155–162.

Pellet, J., J. T. Bried, D. Parietti, A. Gander, P. O. Heer, D. Cherix, and R. Arlettaz. 2012. Monitoring butterfly abundance: beyond Pollard walks. PLoS ONE 7:e41396.

Pocewicz, A., P. Morgan, and S. D. Eigenbrode. 2009. Local and landscape effects on butterfly density in northern Idaho grasslands and forests. Journal of Insect Conservation 13:593–601.

Pollard, E. 1977. A method for assessing changes in the abundance of butterflies. Biological Conservation 12:115–134.

R Development Core Team. 2015. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria.

Royer, R. A. 2003. Butterflies of North Dakota: an atlas and guide. Science Monograph Number Two, Minot State University, Minot, North Dakota, USA.

Royer, R. A., J. E. Austin, and W. E. Newton. 1998. Checklist and “Pollard walk” butterfly survey methods on public lands. American Midland Naturalist 140:358–371.

Stanbury, A., and R. Gregory, 2009. Exploring the effects of truncated, pooled and sexed data in distance sampling estimation of breeding bird abundance. Bird Study 56:298–309.

Thomas, L., S. T. Buckland, K. P. Burnham, D. R. Anderson, J. L. Laake, D. L. Borchers, and S. Strøberg. 2002. Distance sampling. Pages 544–552 in A. H. El-Shaarawi and W. W. Piegrsch, editors. Encyclopedia of environmetrics. Volume 1. Wiley and Sons, Chichester, UK.

Wellington, K., C. Bottom, C. Merrill, and J. A. Litvaitis. 2014. Identifying performance differences among trail cameras used to monitor forest mammals. Wildlife Society Bulletin 38:634–638.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3304/full