Of detectability and camouflage: evaluating Pollard Walk rules using a common, cryptic butterfly

Federico Riva [ID],[1],† Giorgio Gentile [ID],[2] Simona Bonelli [ID],[2] John H. Acorn [ID],[1]
Francisco V. Denes [ID],[1] Andrew D. Crosby [ID],[3] and Scott E. Nielsen [ID][1]

[1]Department of Renewable Resources, University of Alberta, General Services Building, Edmonton, Alberta T6G 2H1 Canada
[2]Dipartimento di Scienze della Vita e Biologia dei Sistemi, University of Turin, Turin Via Verdi 8-10124 Italy
[3]Department of Biological Sciences, University of Alberta, Biological Sciences Building, Edmonton, Alberta T6G 2H1 Canada

Abstract.   Estimating distribution and abundance of species depends on the probability at which individuals are detected. Butterflies are of conservation interest worldwide, but data collected with Pollard walks —the standard for national monitoring schemes—are often analyzed assuming that changes in detectability are negligible within recommended sampling criteria. The implications of this practice remain poorly understood. Here, we evaluated the effects of sampling conditions on butterfly counts from Pollard walks using the Arctic fritillary, a common but cryptic butterfly in boreal forests of Alberta, Canada. We used an open population binomial N-mixture model to disentangle the effects of habitat suitability and phenology on abundance of Arctic fritillaries, and its detectability by sampling different conditions of temperature, wind, cloud cover, and hour of the day. Detectability varied by one order of magnitude within the criteria recommended for Pollard walks (P varying between 0.04 and 0.45), and simulations show how sampling in suboptimal conditions increases substantially the risk of false-absence records (e.g., false-absences are twice as likely than true-presences when sampling 10 Arctic fritillaries at $P = 0.04$). Our results suggest that the risk of false-absences is highest for species that are poorly detectable, low in abundance, and with short flight periods. Analysis with open population binomial N-mixture models could improve estimates of abundance and distribution for rare species of conservation interest, while providing a powerful method for assessing butterfly phenology, abundance, and behavior using counts from Pollard walks, but require more intensive sampling than conventional monitoring schemes.

† **E-mail:** friva@ualberta.ca

Introduction

Species' distribution and abundance are the two most common state variables in ecology (Krebs 1972, Kéry and Schaub 2012). To estimate distribution and abundance, ecologists have historically drawn upon counts of organisms, treating them either as indices of abundance or censuses (Krebs 1972, Kéry and Schaub 2012). Yet, observers usually fail to detect all individuals present at a site, and when detection probability is not constant and adjusted for, inferences based on count data can be biased (Brown and Boyce 1998, MacKenzie et al. 2002, Pellet et al. 2012). Therefore, detection probability (thereafter "detectability") plays an important role in our

ability to understand the ecology, abundance, and distribution of organisms (Brown and Boyce 1998, MacKenzie et al. 2002, Ancona et al. 2017).

Because ignoring imperfect detection can result in biased estimates of diversity, occupancy, and abundance (Kéry and Plattner 2007, Jarzyna and Jetz 2016, Ancona et al. 2017), ecologists have developed techniques to account for detectability (MacKenzie et al. 2002, Nowicki et al. 2008). Often, the probability of detecting an individual depends on the conditions under which sampling is conducted (even for plants; see, e.g., Dennett and Nielsen, 2019) and thus can be modeled as a function of relevant covariates (MacKenzie et al. 2002, Kéry and Schaub 2012). Nevertheless, methods designed to incorporate the observation process are often not used, especially for insects (Nowicki et al. 2008, Kellner and Swihart 2014).

Butterflies (Lepidoptera: Papilionoidea) are one of the most studied and charismatic insect groups (Boggs et al. 2003, Thomas 2016), and losses in abundance and distribution of many butterfly species have prompted the launch of conservation initiatives worldwide (Melero et al. 2016, Thomas 2016, van Strien et al. 2019, Wepprich et al. 2019). Given widespread interest, the merits and limitations of different sampling protocols and analytical approaches have been widely scrutinized (Haddad et al. 2008, Nowicki et al. 2008, Isaac et al. 2011, Pellet et al. 2012, Schmucki et al. 2016). Today, Pollard walks (PW) emerged as the most widespread method for sampling butterflies (Nowicki et al. 2008).

Pioneered in the 1970s by Ernest Pollard (Pollard 1977), PW consist of repeated transect counts of adult butterflies conducted under specific sampling conditions (originally, samples between 10:45 h and 15:45 h, at temperatures >13°C with sunny weather and >17°C with cloudy weather). Since Pollard's seminal manuscript, only minor adjustments have been proposed to the original method, including controlling for the effects of wind and adapting time of sampling to different countries (Van Swaay et al. 2008). Today, decades of data from thousands of PW worldwide have been recorded (Dennis et al. 2013, Schmucki 2016, Wepprich et al. 2019), with this massive effort resulting in prominent contributions in ecology and conservation (Warren et al. 2001, Pateman et al. 2012, Macgregor et al. 2019).

Although originally conceived for estimating population trends (Pollard 1977), PW have also been used to assess other applied and theoretical questions (Nowicki et al. 2008, Dennis et al., 2013, 2016, Matechou et al. 2014, Schmucki et al. 2016). Counts from PW were historically analyzed with regression methods (e.g., linear models), using the sum of multiple site visits as a site abundance index. However, analyzing PW counts as abundance indices assumes that the proportion of butterflies counted is equal among sites. This proportionality assumption is violated when detectability varies, undermining comparisons between samples (Nowicki et al. 2008) and between species within the assemblage (Isaac et al. 2011, Pellet et al. 2012).

Importantly, butterfly detectability varies not only with species' characteristics (e.g., natural history, sex, behavior, and morphology), but also with sampling conditions and habitat assessed (Brown and Boyce 1998, Haddad et al. 2008, Isaac et al. 2011, Bried and Pellet 2012, Pellet et al. 2012). For instance, detection is easier when organisms move (Caro 2005), with ectotherm movement conditioned by the abiotic environment (Kevan and Shorthouse 1970). Therefore, variation in conditions during sampling affects PW counts, and the changes in these counts are not necessarily due to variation in the true abundance of butterflies within a transect, but rather to our ability to detect them (e.g., due to changes in activity patterns). Defining sampling criteria based on thresholds to environmental conditions at which PW should be conducted, PW implicitly assume that detectability does not vary substantially within recommended sampling criteria (e.g., temperature >13°C with cloud cover <50% and >17°C when cloud cover >50%, wind speed <5 Beaufort units, and hour of sampling approximately between ± 3.5 h from the time of sun peaking; Van Swaay et al. 2008). Yet, except for initial tests conducted from Pollard himself (Pollard 1977), this assumption has virtually remained untested (but see Harker and Shreeve 2008).

Here, our objective was to assess how sampling conditions influence PW counts for an abundant, but cryptic (i.e., overall body color resembling the general color of the organism

habitat; Caro 2005) species—the Arctic fritillary (*Boloria chariclea*; Fig. 1). We hypothesized that changes in the abiotic environment would affect detectability of this species primarily through changes in its activity patterns due to constraints and relationships to body temperature (i.e., thermoregulation behavior). Specifically, we predicted that detectability would (1) increase with air temperature, because arctic fritillaries are active at body temperatures that are generally high (~26°–31°C) relative to the study area; (2 and 3) decrease in windy and cloudy conditions that hinder the activity of butterflies; and (4) follow a unimodal trend with time of the day, because butterflies are generally more commonly observed a few hours before and after noon (Kevan and Shorthouse 1970, Van Swaay et al. 2008). To test recommended sampling conditions for PW, we used a hierarchical, open population binomial N-mixture model, treating the true abundance of butterflies as a latent (unobserved) state process influenced by habitat suitability and phenology, and the observation (detection) process associated with sampling conditions. Finally, we use simulations to quantify the implications of the original PW assumptions on estimated abundance and occupancy of Arctic fritillaries.

METHODS

Study area

This study was conducted within boreal treed peatlands dominated by black spruce (*Picea mariana*), in the Wood Buffalo region of Alberta, Canada (56°37′22″ N, 111°58′71″ W, Fig. 1; see Riva et al. 2018*a* for additional information on the study area). The Wood Buffalo region has been subject to widespread anthropogenic disturbance associated with in situ oil sand developments, particularly the clearing of narrow (<10 m wide) linear corridors involved in seismic assessments of the underground bitumen reserve (seismic lines). A series of studies conducted in the area revealed effects of seismic lines on abundance, behavior, and responses to wildfire in
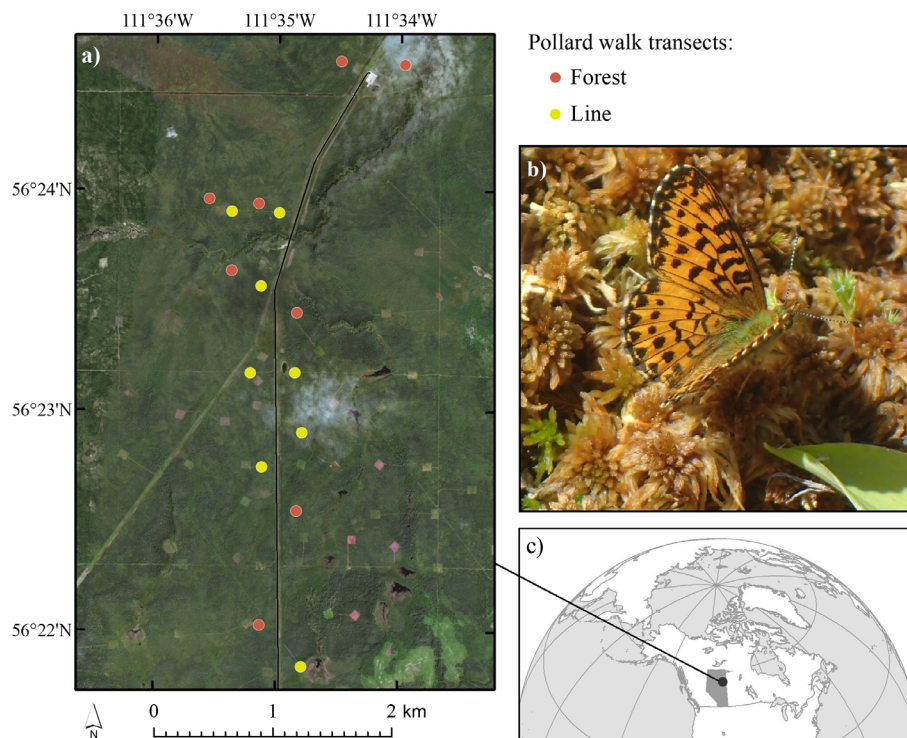

Fig. 1. (a) Map of the study area; (b) Arctic fritillary basking on boreal peatland mosses. In these forests, Arctic fritillaries are hard to detect unless active; (c) location of study area and Alberta in North America.

butterflies (Riva et al., 2018*a*, *b*, *c*, 2019), as well as responses in other organisms (Dabros et al. 2018, Fisher and Burton 2018).

Model species

We studied the Arctic fritillary (*Boloria chariclea*), a Holarctic species previously used to assess climate change, thermoregulation, and behavior (Kevan and Shorthouse 1970, Bowden et al. 2015, Riva et al. 2018*c*). In the boreal forests of Alberta, Canada, Arctic fritillary is univoltine and is the most common peatland butterfly, being more abundant on seismic lines than in the surrounding peatland forests (Riva et al. 2018*b*, 2019). Seismic lines in boreal Alberta are rich in many of the larval host plants for this generalist species, especially *Salix* spp. (Riva et al. 2019). Given the similarity in color between boreal peatland mosses and the dorsum of Arctic fritillaries (Fig. 1), and its relatively small size (wingspan ~38 mm; Burke et al. 2011), we considered it cryptic in these forests and thus appropriate for an investigation of the effects of detectability on PW counts. Specifically, we expected for Arctic fritillaries' larger differences between minimum and maximum detectability in comparison with other species that are easier to detect (e.g., larger or more colorful), particularly while not flying. Arctic fritillaries often bask to thermoregulate (Kevan and Shorthouse 1970), here on *Sphagnum* spp. or the understory vegetation, where they are poorly detected (Fig. 1). Conversely, when observed during their gentle flight, Arctic fritillaries become more detectable. Therefore, changes in the abiotic environment affect this species' activity patterns, thereby affecting its detectability. Information on longevity and dispersal of *Boloria chariclea* is scarce, but three *Boloria* spp. are estimated to live between 3 and 12 d (Bubová et al. 2016), whereas *Boloria napaea* has a mean dispersal distance of ~100–150 m (Ehl et al. 2019).

Sampling design

We sampled 16 PW transects within the study area following Pollock's robust design (Pollock 1982), stratified into eight undisturbed forests vs. eight 9 m wide seismic lines. PW were 50 m long and 5 m wide, with centroids separated by at least 200 m to reduce spatial autocorrelation among samples (Fig. 1). Arctic fritillaries were

common in the area (i.e., hundreds; Riva et al., 2018*b*, *c*, 2019), and 50-m transect lengths were sufficient to obtain adequate sample sizes. Each of the 16 PW was visited once per day during 12 sampling days between 20 July and 9 August 2018, for a total of 192 PW transects within 21 d. Samples were conducted between ~07:00 h and 19:00 h by the same observer (FR), randomizing sampling order within day, at any condition of air temperature and wind, but avoiding rain. Temperature (°C), wind speed (Beaufort units, BU, measured at a height of ~1.5 m), and hour of the day (h) were recorded for each PW using a Kestrel 5500 Pocket Weather Meter. We also visually estimated whether cloud cover was >50% (binary covariate). To assess the recommended sampling conditions for PW, we considered time between 09:00 and 15:00, that is, 3 h before and after peak sun. Note that this choice is conservative in relation to the recommended guidelines from Butterfly Conservation Europe (i.e., 3.5 h before and after peak sun; Van Swaay et al. 2008).

Statistical model

We used a hierarchical open population binomial N-mixture model (hereafter N-mixture model; Kéry and Schaub 2012) to assess the effects of habitat suitability and phenology on abundance of Arctic fritillaries and to determine how environmental conditions affect detectability. The model estimates butterfly abundance N_{ik} at site i during each primary sampling period k, where $N_{ik} \sim$ Poisson (λ_{ik}). We modeled expected abundance (λ_{ik}) as a function of treatment type (forest or seismic line) and the quadratic of Julian day, so that

$$\log(\lambda_{ijk}) = \alpha_0 + \varepsilon_i + \alpha_1 \text{Treatment} + \alpha_2 \text{Day} + \alpha_3 \text{Day}^2$$

where ε_i is the random effect of site i on abundance and $\varepsilon_i \sim$ Normal (0, σ^2).

In our model, N_{ik} is a latent, unobserved variable estimated by modeling detection probability at each survey j done within primary sampling period k at site i, so that each count $y_{ijk} \sim$ Binomial (N_{ik}, p_{ijk}), where the number of individuals N_{ik} is the number of trials and p_{ijk} is the probability of detecting a given individual during the survey. The model assumes sites are closed to changes in abundance within each

primary sampling period k (Kéry and Schaub 2012). We modeled p_{ijk} as a logit function of survey level environmental variables, so that

$$\begin{aligned} \mathrm{logit}(p_{ijk}) = \ & \beta_0 + \beta_1 \mathrm{Temp} + \beta_2 \mathrm{Wind} + \beta_3 \mathrm{Time} \\ & + \beta_4 \mathrm{Time}^2 + \beta_5 \mathrm{Cloud} + \beta_6 \mathrm{Cloud} \\ & \times \mathrm{Temp}, \end{aligned}$$

where Temp is temperature, Wind is wind speed on the Beaufort scale, Time is the time of day expressed in hours from midnight (0–24 h), and Cloud is a binary covariate representing cloud cover >50%. We used temperature, wind speed, hour of the day, and cloud cover because controlling for these factors has been shown to affect butterfly detectability and is recommended for monitoring schemes based on PW (Brown and Boyce 1998, Van Swaay et al. 2008, Pellet et al. 2012). The interaction between temperature and cloud cover was included to test the hypothesis that temperature is especially important in cloudy weather—indeed, it is recommended that PW should only be performed in cloudy weather if temperature is >17°C (Van Swaay et al. 2008).

Analysis

We fit the model in a Bayesian framework using programs JAGS version 4.3.0 (Plummer 2003) and R version 3.6.1 (R Core Team 2019) with the package R2jags (Su and Yajima 2015). We standardized continuous covariates to facilitate model convergence and allow comparisons between the effects of different covariates (Kéry and Schaub 2012). MCMC simulations were run for 600,000 iterations, retaining 400,000 iterations after discarding the first 200,000. We evaluated convergence of the MCMC chains using the Gelman and Rubin R-hat statistic (Brooks and Gelman 1998). Model fit was evaluated by simulating new data from the posterior distribution of model parameters and comparing these to the estimates from the actual data (Kéry and Schaub 2012). We used diffuse prior distributions for all estimated parameters. See Data S1 for model code and prior distributions.

Simulations

To assess the interplay between abundance and detectability in determining estimates of species' counts (and thus presence) in PW, we supplemented our analysis with simulations.

Specifically, we generated butterfly detection data according to binomial distributions with N trials, representing site-level abundance, and probability P, representing detectability. Note that this stochastic process is identical to that assumed in the N-mixture model, where the number of butterflies counted is treated as the realization of a random variable from a binomial distribution. Simulations were set to represent PW conducted on transects with $N = 10$, 20, and 40 Arctic fritillaries and probabilities of detections of $P = 0.04$, 0.20, and 0.40. For each combination of N and P, we simulated 100,000 realizations of the stochastic process. We chose values of N and P based on the results of the hierarchical model to represent sites where Arctic fritillaries are rare, moderate, or high in abundance, and samples that were poor, average, or good in terms of sampling conditions.

Results

We counted 669 Arctic fritillaries during 192 PW, 253 in forests and 416 in seismic lines. On average, we observed 2.6 Arctic fritillaries per transect in forests (min = 0, max = 8) and 4.5
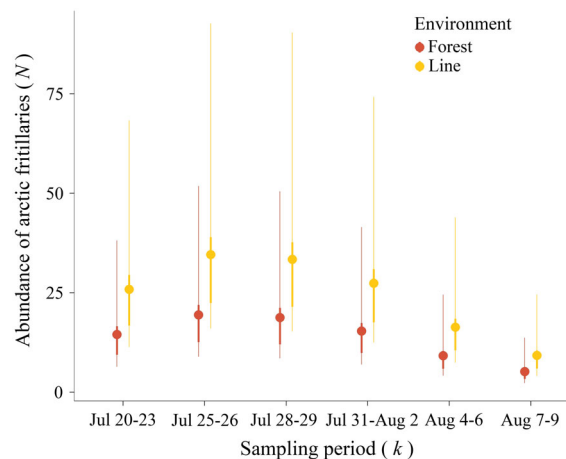

Fig. 2. Predicted abundance N of Arctic fritillary butterflies in forest and seismic line environments, for 6 closure periods k in the summer of 2018. A unimodal effect of sampling date was fitted to represent the trend in abundance typical of the univoltine brood strategy of Arctic fritillaries in the study area. Circles represent posterior mean predicted abundances, with lines representing credible intervals (CRI; thick lines = 50% CRI; narrow lines = 95% CRI).

Arctic fritillaries per transect in seismic lines (min = 0, max = 11). PW were conducted at temperatures between 12.9° and 33.4°C (mean and standard deviation: 24.1° ± 4.1°C), wind speeds were between 1 and 5 units on the Beaufort scale (0.9 ± 1.2 BU), and hours between 07:02 and 18:27 (11:55 ± 2:45). Cloud cover was >50% in 82 of 192 PW (~43% of the samples). Data are provided in Data S1.

The open population binomial-mixture model fitted the data well (posterior predictive check of model adequacy = 1.08), and MCMC mixing was adequate (all R-hat values = 1). Abundance of Arctic fritillaries was higher in seismic lines than forest and peaked during the second sampling period (July 25–26; Fig. 2). Detectability peaked around 12:00, decreased with cloud cover >50%, and increased with temperature at cloud cover >50% (Fig. 3; Tables 1, 2). The Bayesian

credible intervals for two covariates, the effects of temperature when cloud cover was <50% and wind, included zero, suggesting weak effects of these environmental conditions on detectability (Table 1).

The estimated abundance of Arctic fritillaries among sites and primary sampling periods varied between 5 (CRI 2.2–12.7) and 18.9 (CRI 8.9–47.9) in forests and between 9 (CRI 4–22.9) and 33.7 (CRI 16.1–85.7) in seismic lines (Table 2, Fig. 2). Probability of detection P was on average 0.22 (CRI 0.07–0.38), with averages between 0.13 (CRI 0.04–0.23) and 0.31 (CRI 0.09–0.52) in different closure periods (Table 2).

Evaluating the recommended sampling conditions for PW (i.e., samples conducted between 9:00 and 15:00, at temperatures >13°C or 17°C if cloud cover >50%, and with wind speed <5 BU), we found that P varies between a minimum of
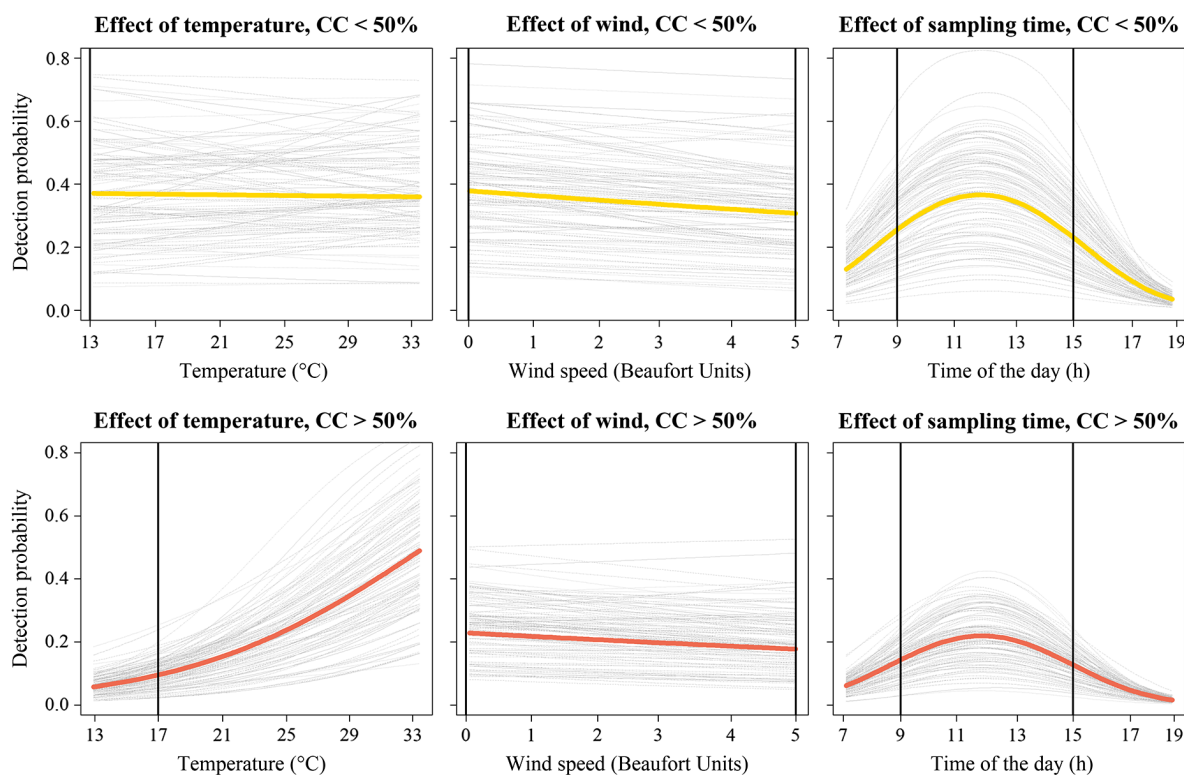

Fig. 3. Trends in detectability with temperature, hour of samples, and wind speed at cloud cover (CC) >50% (top row, in gold) and <50% (bottom row, in red). Predictions are calculated holding the other covariates at their average value (i.e., temperature ~ 24°C, wind speed ~ 1 BU, time of the day ~ noon). Black vertical lines represent the sampling criteria recommended for PW for monitoring schemes (i.e., temperature >13°C with cloud cover <50% and >17°C when cloud cover >50%; wind speed <5 BU; sampling hour between 9:00 and 15:00).

Table 1. Posterior mean, standard deviation (SD), credible interval (CRI, i.e., interval between 2.5% and 97.5% percentiles), potential scale reduction factors (R-hat), and effective sample sizes (n.eff) of parameters representing the effects of model covariates on detection (β coefficients) and abundance (α coefficients) of Arctic fritillaries.

Model	Parameter	Variable	Mean	SD	CRI	R-hat	n.eff
Abundance	ε_i, μ	Site (random effect; μ)	0.00	0.02	(−0.05; 0.05)	1.00	26,000
	ε_i, σ	Site (random effect; σ)	0.08	0.06	(0; 0.23)	1.00	3400
	α_0	Intercept	2.63	0.44	(1.98; 3.69)	1.00	1300
	α_1	Seismic line	0.58	0.10	(0.37; 0.79)	1.00	60,000
	α_2	Ordinal day	−0.50	0.06	(−0.62; −0.37)	1.00	60,000
	α_3	Ordinal day^2	−0.32	0.06	(−0.46; −0.19)	1.00	11,000
Detection	β_0	Intercept	−0.52	0.65	(−1.94; 0.61)	1.00	1500
	β_1	Temperature	−0.01	0.10	(−0.21; 0.19)	1.00	60,000
	β_2	Wind speed	−0.07	0.06	(−0.2; 0.04)	1.00	60,000
	β_3	Time	−0.11	0.06	(−0.25; 0)	1.00	60,000
	β_4	Time^2	−0.53	0.10	(−0.74; −0.35)	1.00	3500
	β_5	Cloud Cover	−0.72	0.17	(−1.09; −0.41)	1.00	7100
	β_6	Cloud Cover × Temperature	0.57	0.15	(0.27; 0.9)	1.00	3800

Table 2. Counts (mean and standard deviation), and estimated abundance and detectability (mean and credible interval) of Arctic fritillaries in Pollard Walks conducted in forest or seismic line environments, for six closure (i.e., of assumed constant abundance) periods k_i.

	Ecological state process (abundance)								Observation process (detection)	
	Forest				Seismic line				Forest and seismic line	
	Counts (y)		Abundance (λ)		Counts (y)		Abundance (λ)		Detectability (P)	
Period	Mean	SD	Mean	CRI	Mean	SD	Mean	CRI	Mean	CRI
k_1, July 20–23	1.4	1.2	14.1	6.4; 35.7	2.9	2.4	25.2	11.4; 63.9	0.13	0.04; 0.23
k_2, July 25–26	3.4	2.1	18.9	8.9; 47.8	6.7	2.8	33.6	16.1; 85.7	0.24	0.07; 0.41
k_3, July 28–29	4.2	1.8	18.2	8.5; 46.7	7.6	2.2	32.4	15.4; 83.5	0.31	0.09; 0.52
k_4, July 31–August 2	3.8	2	14.9	6.9; 38.5	4.9	2	26.6	12.5; 68.5	0.22	0.07; 0.39
k_5, August 4–6	2.2	2.2	8.9	4.1; 22.7	2.4	1.5	15.9	7.5; 40.6	0.21	0.06; 0.36
k_6, August 7–9	0.8	0.9	5	2.3; 12.7	1.4	1.3	8.9	4; 22.9	0.22	0.07; 0.38

Notes: Note that detectability did not differ between forests and seismic lines. For parameter notation (i.e., y, λ, and P), refer to *Methods: Statistical model*.

0.04 (CRI 0.01–0.09) when samples are conducted at 15:00 h, with 17°C, wind at 5 BU, and cloud cover >50% vs. a maximum of 0.45 (CRI 0.15–0.72) when samples are conducted at noon, with 31°C, wind at 0 BU, and cloud cover >50%. The interactive effect between cloud cover and temperature predicts that at temperatures lower than 30°C, cloud cover >50% decreases the probability of detecting Arctic fritillaries, while for the few samples that occurred at temperatures higher than 30°C ($n = 12$), the probability of detecting Arctic fritillaries was higher at cloud cover of >50%.

Simulations demonstrate that for PW with $N = 10$ butterflies and $P = 0.04$, 0.20, and 0.40, the probability of false-absence (F) is, respectively, 0.66, 0.10, and ~0; with $N = 20$ and $P = 0.04$, 0.20, and 0.40, F is 0.46, 0.01, and ~0; and with $N = 40$ and at $P = 0.04$, 0.20, and 0.40, F is 0.20, ~0, and ~0 (Fig. 4).

DISCUSSION

This study demonstrates that, even within recommended sampling criteria, environmental conditions have substantial effects on
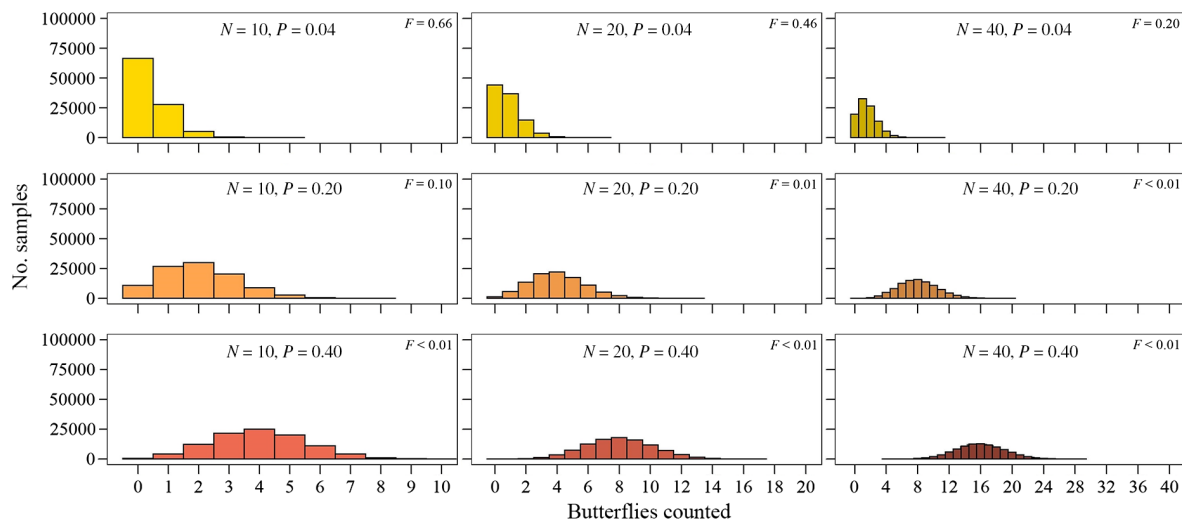
Fig. 4. Histograms summarizing the results of simulations conducted using binomial distributions $B(N,P)$ for each combination of abundance (N) = 10, 20, and 40 Arctic fritillaries at detection probabilities (P) = 0.04, 0.20, and 0.40. Probability of false-absence F, that is, the proportion of simulations where no butterflies were counted, is reported for each scenario (top-right corner). Note that the average number of butterflies sampled is equal to $N \times P$.

detectability and thus the number of butterflies counted in PW. Specifically, we showed that an open population binomial N-mixture model can disentangle the effects of habitat suitability and phenology on abundance of butterflies from the effects of sampling conditions on their detectability. As expected for a cryptic species, detectability was generally low (average of ~0.22) and varied by an order of magnitude (~0.04 vs. 0.45) between poor and optimal sampling conditions within the recommended criteria for PW (i.e., samples conducted between 9:00 and 15:00, at temperatures >13°C or 17°C if cloud cover >50%, and with wind speed <5 BU; Fig. 3; Van Swaay et al. 2008). To our knowledge, this is the first comprehensive assessment of how changes in environmental conditions within the recommended PW criteria affect butterfly counts, despite several studies dealing with detectability in butterflies (Brown and Boyce 1998, Kéry and Plattner 2007, Bried and Pellet 2012, Pellet et al. 2012, Matechou et al. 2014, Melero et al. 2016). Results are especially important in relation to initiatives such as the European Butterfly Monitoring scheme (https://butterfly-monitoring.net/ebms), a continental effort that includes more than 6200 PW transects that have been sampled for

decades, recording more than 400 species and used to inform the protection and management of habitats and species in Europe (e.g., using butterflies as indicators; van Strien et al., 2019).

Detection and abundance of Arctic fritillaries

The most important covariates in determining detectability were cloud cover and hour of the day, while wind speed was less important, and temperature had an effect only when interacting with cloud cover (Table 1, Fig. 3). While rather surprising, these results are consistent with the ecology of this system. As ectotherms and behavioral thermoregulators, temperature plays an important role in determining the activity, and thus detection, of butterflies (Kevan and Shorthouse 1970, Bried and Pellet 2012). Indeed, Arctic fritillaries are known to perform dorsal basking, reaching body temperatures between 26° and 32°C (Kevan and Shorthouse 1970). Assuming that the primary determinant of changes in detectability here was butterfly activity, our model suggests that, as long as sunlight is available, temperatures in the range of those observed when cloud cover was <50% (i.e., 17.1°–33.4°C) allow for equal Arctic fritillary activity.

Presumably, Arctic fritillaries are efficient at solar basking, thus reaching the body temperature necessary to fly regardless of the air temperature when sun light is available (Kevan and Shorthouse 1970). This is not surprising given that this species inhabits much colder, temperature limited environments, for example, Zackenberg Research Station in Greenland, with an average temperature in July of ~6°C vs. ~16°C in our study area (Bowden et al. 2015). Conversely, when cloud cover was >50% (and thus basking was presumably more difficult), higher air temperatures were more important for Arctic fritillaries to reach the body temperature necessary to fly (Fig. 3). Interestingly, with all else being equal, Arctic fritillaries had a slight preference for flying in the morning. Finally, the smaller role of wind on detectability might reflect the fact that all assessed PW were embedded in forests. Wind speed was measured at a height of ~1.5 m, and while we did record a few windy PW, the presence of shrubs and trees buffered wind speed at lower heights (<0.5 m) where Arctic fritillaries usually fly (Kevan and Shorthouse 1970).

With respect to abundance estimates, our results confirm that PW counts that are not corrected for detection probability substantially underestimate the abundance of cryptic species (Brown and Boyce 1998, Isaac et al. 2011, Pellet et al. 2012). Here, we counted approximately one fifth of the butterflies estimated by the model, but we note that some PW were conducted on purpose out of the recommended sampling criteria, for example, in the early morning or evenings. Interestingly, because detectability did not change between forests and seismic lines, we confirmed the results of previous studies reporting increases in abundance of Arctic fritillaries on seismic lines (Riva et al. 2018b, 2019). Therefore, even canonical regression techniques applied to PW counts can be effective to address management questions, but only when detectability does not vary with factors of interest (e.g., space, time or habitat; Nowicki et al. 2008, Pellet et al. 2012). It has been argued that modeling detectability is not always necessary (Welsh et al. 2013, Hutto 2016, but see Guillera-Arroita et al. 2014, Marques et al. 2017). Our results suggest that the context, constraints, and objectives of a study should inform whether hierarchical models are the most appropriate approach.

Contextualization and implications of our study

How sampling conditions influence PW counts through variation in butterfly detectability remains poorly understood, especially at a fine temporal scale. Previous studies generally assess the effects of single covariates on the detection process, for example, temperature (Bried and Pellet 2012, Matechou et al. 2014), or the effects of site, time, or habitat type (Brown and Boyce 1998, Isaac et al. 2011, Melero et al. 2016), thereby ignoring the more complex effects documented here (Fig. 3). Other analyses using binomial-mixture models (e.g., Kéry and Schaub 2012, Melero et al. 2016) had different objectives, with coarser temporal scales. Our study fills therefore a knowledge gap and was possible thanks to an exhaustive sampling effort over a short period, with 192 PW samples in 21 d.

Complementing our study with simulations, we show how species at low densities and poorly detectable are likely to be overlooked under suboptimal sampling conditions. For instance, when sampling a transect containing 10 Arctic fritillaries at the lowest detectability here observed within PW conditions ($P = 0.04$), false-absences are twice as likely than true-presences (Fig. 4). Since species assemblages are usually uneven, composed by a few common species and many uncommon species (Preston 1948), the implications of imperfect detection often interest the majority of species, especially in areas of high biodiversity. Furthermore, we observed large variations in abundance within a period as short as one week (e.g., ~threefold decreases between 2 and 9 August; Fig. 2, Table 2). Therefore, when weekly visits (sampling frequency required by most monitoring schemes) do not capture the peak of a species flight curve, the risk of false negatives for species that are poorly detectable, low in abundance, and with short flight periods is exacerbated. We acknowledge that detectability estimates varied here broadly (e.g., 95% credible interval between 0.15 and 0.72 for the best sampling conditions), but rather than focusing on the accuracy of abundance estimates, we stress that low detection probabilities often result in false-absences when estimating the occupancy of rare species based on weekly PW counts. There are indeed success stories of rare butterflies recovering from the verge of extinction (Haddad 2018), such that it is undesirable to

overlook even small populations of species of conservation concern.

Notably, the implications of this study should be evaluated in the light of the assumptions underlying the open binomial-mixture model: (1) no false positives recorded (including species misidentification); (2) equal detectability among Arctic fritillary individuals; (3) appropriateness of a Poisson distribution for the ecological state model; and (4) the ecological state did not change within each closure period (Kéry and Schaub 2012). The first three assumptions are reasonably met as (1) Arctic fritillaries were distinctive in the field (only species of the genus *Boloria* flying at the end of July in the study area), and attention was given to avoid counting the same individual multiple times; (2) no reason to expect substantial variation in detectability between individuals (e.g., no pronounced sexual dimorphism; however, sexes may behave differently, and whether the sex ratio differs between seismic lines and forests was here unknown. For instance, because male butterflies tend to cover larger distances than females when dispersing (Ehl et al. 2019), changes in sex ratio between sampling environments can affect detectability estimates due to more severe violations of the closure assumption for PW with higher proportions of males); and (3) a Poisson distribution is appropriate because we usually counted few butterflies (no sign of overdispersion) and model fit was good (posterior predictive check of model adequacy = 1.08; Melero et al. (2016) considered adequate all models in the interval 0.8–1.2).

Conversely, the closure assumption (5) is likely violated for both the spatial (no emigration or immigration) and demographic (no recruitments or deaths) domains. Arctic fritillaries are common throughout these forests with both seismic lines and forests providing seemingly suitable habitat rich in willow (*Salix* spp.), the larval host plant (Riva et al. 2019). Therefore, Arctic fritillaries most likely moved freely within the study area, leaving or colonizing PW. Concurrently, butterfly life span is shorter than their flight period, such that only a subset of the population is present on any given day. This temporal fragmentation, typical of butterflies (Nowicki et al. 2008), implies that Arctic fritillaries emerged or died within closure periods (longevity for three European *Boloria* spp. was estimated between 3 and 12 d; Bubová et al. 2016). However, we imposed short closure periods (i.e., the six k periods that lasted on average 2.8 $\pm$ 0.7 d, min = 2, max = 4; Fig. 2). Therefore, we consider it reasonable for this analysis to assume that the number of Arctic fritillaries on a transect was relatively constant within 2–3 d. Kéry and Schaub (2012) report that mild violations of the closure assumption lead to inflated estimates of abundance and that in such cases the estimated abundance can be interpreted as the number of individuals that used a PW during the surveys. In our case, it is possible that abundance was overestimated (thus underestimating detectability), but the effects of sampling conditions observed here are substantial and we consider our inference robust.

Notably, it is generally assumed that changes in detectability do not determine changes in butterfly counts when analyzing data from monitoring schemes (but see Matechou et al. 2014). Analyses of monitoring schemes usually assess trends over time for species' abundance, commonly evaluating yearly indices calculated using the approach of area under the flight period curve (Rothery and Roy 2001, Dennis et al. 2013, Schmucki et al. 2016, Wepprich et al. 2019). These analyses assume that the effects of detectability on PW counts are negligible in comparison with changes in time due to population trends. Indeed, national-scale trends are robust, with declines in species also corresponding to loss in their distribution (Thomas 2016). However, site-level estimates are likely less accurate due to changes in detectability (Isaac et al. 2011), limiting our ability to manage populations of rare and/or poorly detectable species. The European Butterfly Monitoring Scheme is expanding in the Mediterranean region, one of the biodiversity hotspots in Europe, and when an assemblage contains several rare species, the risks described in this paper are exacerbated. Therefore, approaches currently applied in other monitoring schemes might be unsuitable in the Mediterranean region. For instance, Italy has one of the richest diversity of butterflies in Europe (i.e., ~300 native species, more than 4 times the British butterfly fauna; Bonelli et al. 2018), and distinguishing extinctions/colonization dynamics from false-absences might be problematic for uncommon species if using data recorded following

conventional monitoring approaches. A hybrid approach where transects of interest are sampled twice within closure periods (ideally one day) might be a better use of survey effort for informing the species status in each transect site.

Finally, even in cases of absolute spatial and demographic closure, it is not possible to estimate true population sizes using only PW counts analyzed with a binomial-mixture model, because information on butterfly longevity cannot be inferred from PW counts (Nowicki et al. 2008). However, in cases where marking individuals to obtain longevity estimates is not feasible (e.g., dealing with rare or delicate butterflies; Haddad et al. 2008), hierarchical models represent an effective approach to account for detectability. Substituting PW counts with true abundances estimated with an open population binomial-mixture model within the framework of area under the flight period curve (Dennis et al. 2013, Schmucki et al. 2016) would allow one to correct for the effect of different environmental conditions on PW counts, and thus compare more accurate indices of abundance for one species in space and time. This would require double visits of each PW within reasonable closure periods (i.e., ideally 1–2 d), but might have important implications when assessing rare or poorly detectable species. However, we note that because this model does not account for changes in longevity, it is thus inaccurate when mortality changes substantially within or between species, in time or with environmental characteristics.

CONCLUSIONS

Discriminating how much variation observed within samples is due to true differences in populations—and not in detectability—is a critical challenge for biodiversity management and conservation (MacKenzie et al. 2002, Kéry and Schaub 2012, Marques et al. 2017). Our study addresses this matter for butterflies, where imperfect detectability has been identified as a conservation priority, especially for species that are rare, poorly detectable, and with limited distribution (Brown and Boyce 1998, Haddad et al. 2008, Nowicki et al. 2008, Isaac et al. 2011). We focused on Pollard Walks (PW), the most widely employed sampling approach for butterflies, demonstrating that ignoring species detectability

can have important consequences in estimating occurrence and abundance of species even when sampling protocols follow recommended sampling conditions. Results would presumably differ when evaluating other systems and/or species (e.g., wind might be more important when sampling open habitats, or temperature when assessing thermophilous butterflies), but our work highlights the necessity to critically evaluate the assumptions underlying many studies that use PW. Here, we focused on a cryptic species, but we stress that detectability is an issue even for species easily detectable because variation in detectability with sampling conditions biases estimates of abundance and distribution regardless of the average detectability of a species.

To be clear, we are not suggesting that PW be abandoned. Ecologists routinely capitalize on all sorts of data, including opportunistic data that are arguably far more challenging that PW data. Instead, PW have been invaluable in providing some of the most important long-term, standardized field data. However, analyzing PW counts while disregarding detectability will yield unbiased results only when variation in detectability among samples is negligible. Other methods that incorporate detectability (e.g., distance sampling or mark–recapture techniques) are more appropriate when detectability varies substantially, and a reconciliation between such methods and PW has long been sought. For instance, Nowicki et al. (2008) stated that "there is apparently little space for reducing the labor requirements (for mark-recapture)... it is, rather, refinements of transect surveys that should be sought. A vital advance would be a field method making possible the estimation of detection probability of individuals counted on transects." Hierarchical models akin to that here presented seem the natural candidate to bridge the chiasms between PW and the analysis of detectability but require multiple samples within meaningful closure periods of assumed constant butterfly abundance.

The implications of low detectability are especially misleading when dealing with poorly detectable (e.g., cryptic) species that are low in abundance. When repeating PW counts in short periods is not feasible, we suggest that PWs should be conducted at more conservative conditions or conditions better representing idiosyncrasies in species ecologies, especially in areas of

high diversity or habitats potentially suitable for rare species. Further studies evaluating more species across a gradient of environments and ecological preferences will be necessary to understand the relation between environmental conditions and PW counts, and thus provide more robust recommendations in terms of sampling conditions.

LITERATURE CITED

Ancona, S., F. V. Dénes, O. Krüger, T. Székely, and S. R. Beissinger. 2017. Estimating adult sex ratios in nature. Philosophical Transactions of the Royal Society B: Biological Sciences 372:20160.

Boggs, C. L., W. B. Watt, and P. R. Ehrlich. 2003. Butterflies: Ecology and evolution taking flight. University of Chicago, Chicago, Illinois, USA.

Bonelli, S., L. P. Casacci, F. Barbero, C. Cerrato, L. Dapporto, V. Sbordoni, S. Scalercio, A. Zilli, A. Battistoni, C. Teofili, C. Rondinini, and E. Balletto. 2018. The first red list of Italian butterflies. Insect Conservation and Diversity 11:506–521.

Bowden, J. J., A. Eskildsen, R. R. Hansen, K. Olsen, C. M. Kurle, and T. T. Høye. 2015. High-arctic butterflies become smaller with rising temperatures. Biology Letters 11:201505.

Bried, J. T., and J. Pellet. 2012. Optimal design of butterfly occupancy surveys and testing if occupancy converts to abundance for sparse populations. Journal of Insect Conservation 16:489–499.

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Brown, J. A., and M. S. Boyce. 1998. Line transect sampling of Karner blue butterflies (*Lycaeides melissa samuelis*). Environmental and Ecological Statistics 5:81–91.

Bubová, T., M. Kulma, V. Vrabec, and P. Nowicki. 2016. Adult longevity and its relationship with conservation status in European butterflies. Journal of Insect Conservation 20:1022–1032.

Burke, R. J., J. M. Fitzsimmons, and J. T. Kerr. 2011. A mobility index for Canadian butterfly species based on naturalists' knowledge. Biodiversity and Conservation 20:2273–2295.

Caro, T. 2005. The adaptive significance of coloration in mammals. BioScience 55:125–125.

Dabros, A., M. Pyper, and G. Castilla. 2018. Seismic lines in the boreal and arctic ecosystems of North America: environmental impacts, challenges, and opportunities. Environmental Reviews 16:1–16.

Dennett, J. M., and S. E. Nielsen. 2019. Detectability of species of Carex varies with abundance, morphology, and site complexity. Journal of Vegetation Science 30:352–361.

Dennis, E. B., S. N. Freeman, T. Brereton, and D. B. Roy. 2013. Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. Methods in Ecology and Evolution 4:637–645.

Dennis, E. B., B. J. T. Morgan, S. N. Freeman, T. M. Brereton, and D. B. Roy. 2016. A generalized abundance index for seasonal invertebrates. Biometrics 72:1305–1314.

Ehl, S., S. I. J. Holzhauer, N. Ryrholm, and T. Schmitt. 2019. Phenology, mobility and behaviour of the arcto-alpine species Boloria napaea in its arctic habitat. Scientific Reports 9:3912.

Fisher, J. T., and A. C. Burton. 2018. Wildlife winners and losers in an oil sands landscape. Frontiers in Ecology and the Environment 16:323–328.

Guillera-Arroita, G., J. J. Lahoz-Monfort, D. I. MacKenzie, B. A. Wintle, and M. A. McCarthy. 2014. Ignoring imperfect detection in biological surveys is dangerous: A response to "fitting and interpreting occupancy models". PLOS ONE 9: e99571.

Haddad, N. M. 2018. Resurrection and resilience of the rarest butterflies. PLoS Biology 16:e2003488.

Haddad, N. M., B. Hudgens, C. Damiani, K. Gross, D. Kuefler, and K. Pollock. 2008. Determining optimal population monitoring for rare butterflies. Conservation Biology 22:929–940.

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.

Hutto, R. L. 2016. Should scientists be required to use a model-based solution to adjust for possible distance-based detectability bias? Ecological Applications 26:1287–1294.

Isaac, N. J. B., K. L. Cruickshanks, A. M. Weddle, J. Marcus Rowcliffe, T. M. Brereton, R. L. H. 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.

Jarzyna, M. A., and W. Jetz. 2016. Detecting the Multiple Facets of Biodiversity. Trends in Ecology and Evolution 31:527–538.

Kellner, K. F., and R. K. Swihart. 2014. Accounting for imperfect detection in ecology: A quantitative review. PLOS ONE 9:e111436.

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

Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS. Elsevier, Amsterdam, The Netherlands.

Kevan, P. G., and J. D. Shorthouse. 1970. Behavioural Thermoregulation by High Arctic Butterflies. Arctic 23:268–279.

Krebs, C. 1972. The experimental analysis of distribution and abundance. Harper and Row, New York, New York, USA.

Macgregor, C. J., et al. 2019. Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year. Nature. Climate Change 10:4455.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, A. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

Marques, T. A., et al. 2017. Model-based approaches to deal with detectability: A comment on Hutto (2016a). Ecological Applications 27:1694–1698.

Matechou, E., E. B. Dennis, S. N. Freeman, and T. Brereton. 2014. Monitoring abundance and phenology in (multivoltine) butterfly species: A novel mixture model. Journal of Applied Ecology 51:766–775.

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.

Nowicki, P., J. Settele, P.-Y. Henry, and M. Woyciechowski. 2008. Butterfly Monitoring Methods: The ideal and the Real World. Israel Journal of Ecology and Evolution 54:69–88.

Pateman, R. M., J. K. Hill, D. B. Roy, R. Fox, and C. D. Thomas. 2012. Temperature-dependent alterations in host use drive rapid range expansion in a butterfly. Science 336:1028–1030.

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.

Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian. *In* K. Hornik, F. Leisch, and A. Zeileis, editors. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria . https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf

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

Pollock, K. H. 1982. A capture-recapture design robust to unequal probability of capture. The Journal of Wildlife Management 46:752–757.

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology 29:254–283.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Riva, F., J. H. Acorn, and S. Nielsen. 2018*a*. Distribution of Cranberry Blue Butterflies (*Agriades optilete*) and Their Responses to Forest Disturbance from In Situ Oil Sands and Wildfires. Diversity 10:112.

Riva, F., J. H. Acorn, and S. E. Nielsen. 2018*b*. Localized disturbances from oil sands developments increase butterfly diversity and abundance in Alberta's boreal forests. Biological Conservation 217:173–180.

Riva, F., J. H. Acorn, and S. E. Nielsen. 2018*c*. Narrow anthropogenic corridors direct the movement of a generalist boreal butterfly. Biology Letters 14:20170770.

Riva, F., J. Pinzon, J. H. Acorn, and S. E. Nielsen 2019. Composite effects of cutlines and wildfire result in fire refuges for plants and butterflies in boreal treed peatlands. Ecosystems:13.

Rothery, P., and D. B. Roy. 2001. Application of generalized additive models to butterfly transect count data. Journal of Applied Statistics 28:897–909.

Schmucki, R., et al. 2016. A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes. Journal of Applied Ecology 53:501–510.

Su, Y.-S., and M. Yajima. 2015. Package "R2jags" (Version:0.5-7). http://cran.r-project.org/package=R2jags

Thomas, J. A. 2016. Butterfly communities under threat. Science 353:216–218.

van Strien, A. J., C. A. M. van Swaay, W. T. F. H. van Strien-van Liempt, M. J. M. Poot, and M. F. WallisDeVries. 2019. Over a century of data reveal more than 80% decline in butterflies in the Netherlands. Biological Conservation 234:116–122.

Van Swaay, C. A. M., et al. 2019. The EU Butterfly Indicator for Grassland species: 1990-2017: Technical Report. Butterfly Conservation Europe & ABLE/eBMS. www.butterfly-monitoring.net

Warren, M. S., et al. 2001. Rapid responses of British butterflies to opposing forces of climate and habitat change. Nature 414:65–69.

Welsh, A. H., D. B. Lindenmayer, and C. F. Donnelly. 2013. Fitting and Interpreting Occupancy Models. PLOS ONE 8:e52015.

Wepprich, T., J. R. Adrion, L. Ries, J. Wiedmann, and N. M. Haddad. 2019. Butterfly abundance declines over 20 years of systematic monitoring in Ohio, USA. PLOS ONE 14:e0216270.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3101/full

Data S1