# Monitoring small mammal abundance using NEON data: are calibrated indices useful?

ARIELLE W. PARSONS,<sup>1,4,\*,0</sup> JAMES S. CLARK,<sup>2</sup> AND ROLAND KAYS<sup>1,3,0</sup>

<sup>1</sup>Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina 27695, USA <sup>2</sup>Nicholas School of the Environment, Duke University, Durham, North Carolina 27710, USA <sup>3</sup>North Carolina Museum of Natural Sciences, Raleigh, North Carolina 27601, USA <sup>4</sup>Present address: Lincoln Park Zoo, Alexander Center for Applied Population Biology, Chicago, Illinois 60614, USA

\*To whom correspondence should be addressed: aparsons@lpzoo.org

Small mammals are important to the functioning of ecological communities with changes to their abundances used to track impacts of environmental change. While capture-recapture estimates of absolute abundance are preferred, indices of abundance continue to be used in cases of limited sampling, rare species with little data, or unmarked individuals. Improvement to indices can be achieved by calibrating them to absolute abundance but their reliability across years, sites, or species is unclear. To evaluate this, we used the US National Ecological Observatory Network capture-recapture data for 63 small mammal species over 46 sites from 2013 to 2019. We generated 17,155 absolute abundance estimates using capture-recapture analyses and compared these to two standard abundance indices, and three types of calibrated indices. We found that neither raw abundance indices nor index calibrations were reliable approximations of absolute abundance, with raw indices less correlated with absolute abundance than index calibrations (raw indices overall  $R^2 < 0.5$ , index calibration overall  $R^2 > 0.5$ 0.6). Performance of indices and index calibrations varied by species, with those having higher and less variable capture probabilities performing best. We conclude that indices and index calibration methods should be used with caution with a count of individuals being the best index to use, especially if it can be calibrated with capture probability. None of the indices we tested should be used for comparing different species due to high variation in capture probabilities. Hierarchical models that allow for sharing of capture probabilities over species or plots (i.e., joint-likelihood models) may offer a better solution to mitigate the cost and effort of large-scale small mammal sampling while still providing robust estimates of abundance.

Key words: abundance, abundance indices, Bayesian, capture-recapture, NEON, small mammal

Our planet is undergoing an unprecedented change in the global environment, creating novel conditions that could fundamentally alter the functioning of ecosystems of the world (Sage 2020). Monitoring species responses to global change is the first step to understanding and predicting ecosystem and community-level consequences, providing insight into the many ways in which global change can affect species persistence and community composition (e.g., Princé et al. 2013; Furnas 2020), migration and movements (e.g., Van Buskirk et al. 2009; Zhu et al. 2012), and disease transmission (e.g., Bosch et al. 2018; Ludwig et al. 2019). Maintaining long-term and large-scale monitoring projects and improving the use of generated data will improve the predictive ability of models to inform action to mitigate the effects of global change. Small mammals are important to the functioning of ecological communities and could be key indicators of ecosystem change over space and time in response to global environmental change. Small mammals directly affect plant community structure, energy and mineral flow, and arthropod communities through predation, soil movement, food storage, and nest building (Grant and French 1980; Schnurr et al. 2004; Dylewski et al. 2020). Due to their role as primary seed predators and dispersers in forest ecosystems, small mammals may be critical in determining which plant species will be successful in response to global stressors. The abundance and diversity of small mammals also directly affect the success of mesopredators for whom they are a major food source (e.g., Krebs et al. 2019). However, monitoring of small mammals can be challenging,

© The Author(s) 2022. Published by Oxford University Press on behalf of the American Society of Mammalogists, www.mammalogy.org.

due to the inherent cost and the heterogeneity of species detectability over space and time (Jones 2011). To reduce cost, studies of small mammals typically rely on measures of diversity, distribution, and relative abundance; however, estimation of absolute abundance (i.e., the actual number of individuals) is often required for addressing questions related to population dynamics, growth rate, viability, and species interactions.

Estimation of absolute abundance for small mammal populations typically requires individual identification and large sampling effort, capturing and marking as many individuals as possible, and applying capture-recapture models to estimate absolute abundance and capture probabilities (Otis et al. 1978). Typically, a minimum sample size is required (e.g., >25 individuals for traditional models; Otis et al. 1978) for these models to produce robust absolute abundance estimates through the estimation of capture probabilities. This minimum sample size can sometimes be challenging to meet, especially with rare species. Since rare species are also often the most vulnerable to environmental change, there is a need for alternative measures or models that can make meaningful inference for rare species while remaining robust to heterogeneity in capture probabilities. To meet this need, two main solutions have been put forth: (1) Bayesian methods of abundance estimation appropriate for small sample sizes (e.g., sometimes as few as three individuals captured; Reppucci et al. 2011); and (2) indices of relative abundance. Bayesian methods can be useful in stabilizing estimates for species with small sample sizes by pulling them in the direction of the average through data pooling and/or the use of informative priors. However, although Bayesian hierarchical capture-recapture models have allowed a reduction in the minimum necessary sample size relative to traditional maximum likelihood estimation (MLE) capture-recapture models, they require more expertise to implement and estimates for small sample sizes will be very sensitive to prior information and/or pooling decisions.

Abundance indices such as trap success rates or counts of individuals captured are simpler than capture–recapture to calculate and more commonly used because they allow monitoring of rare species with less rigorous sampling requirements and less computational expertise. However, because indices typically ignore capture probabilities that may vary over space, time, and species, they are prone to bias, leading to erroneous inference (Nichols and Pollock 1983; Sollmann et al. 2013). Nevertheless, several past studies have concluded that indices work well to monitor single-species population trends over time (Slade and Blair 2000; Hopkins and Kennedy 2004; Schwemm et al. 2018), but struggle to make comparisons across species when interspecific capture probabilities vary (Slade and Blair 2000; Hopkins and Kennedy 2004). Improvement of inference from indices may be achieved by calibration with absolute abundance, but this has only been done for single species at a single site (e.g., Dröge et al. 2020), and whether calibrations can then be applied across sites remains unclear (Lambin et al. 2000; Slade and Blair 2000). Indeed, there have been no studies that have looked at the use and generalizability of indices and index calibrations for small mammal monitoring over many (>11) species and (>3) geographically diverse sites. Furthermore, most validation studies of indices in small mammal trapping report only  $R^2$  values, with no reporting of bias (but see McKelvey and Pearson 2001) or variation in capture probabilities. Thus, there remains a need to explore and characterize the best alternatives to standard capture-recapture modeling that maximize the amount of data used for inference and minimize sampling effort while still providing unbiased estimates of abundance.

We used data from small mammal surveys across the United States generated by the National Ecological Observatory Network (NEON) to evaluate the use of indices for small mammal population and community monitoring. These NEON data sets include capture-recapture data for small mammal communities using multiple trapping grids per site per year over the continental United States, Alaska, and Puerto Rico. With the goal of maximizing the amount of data used for inference while reducing sampling effort, we tested how well two abundance indices (1) count of individuals captured per three trapnights (hereafter: "iCap," aka: "M<sub>t+1</sub>") and (2) trap success (total captures per three trap-nights, ignoring individual ID) reflected absolute abundance estimates from capture-recapture analysis (Table 1). We assessed both indices in their raw form and calibrated them using two methods: (1) regression-based calibration where a linear relationship is established between absolute abundance (i.e., capture-recapture) and the index; and

**Table 1.**—Index and index calibration measures used to test the performance of indices against capture–recapture (CR) for monitoring small mammals. Each metric is named and defined, giving an example reference for their use and recommendations based on our results for when they might be used (with caution) when CR methods are not possible. \* indicates methods requiring information from CR or other methods of estimation of true abundance and/or capture probabilities (*p*).

Metric	Definition	References	When to use
iCap	The count of individuals captured (aka:	Slade and Blair (2000);	Short-term snapshot of single
	$M_{t+1}$ ) per 3 nights of sampling	McKelvey and Pearson (2001)	species at single site
Trap success	The count of total captures, ignoring	Slade and Blair (2000);	Do not use
	individual ID per 3 nights of sampling	McKelvey and Pearson (2001)	
Regression-calibrated	A linear relationship is established between	Brown et al. (1996); Ramesh	Use in place of <i>p</i> -corrected,
abundance*	absolute abundance (i.e., CR) and the index	et al. (2013)	but performance poor
p-corrected iCap	iCap corrected using CR P estimates from	Steinhorst and Samuel (1989);	Long-term monitoring, single
(conservative)*	same species at same site	Pollock et al. (2002)	species, single/multiple sites
<i>p</i> -corrected iCap	iCap corrected using CR P estimates from	Steinhorst and Samuel (1989);	Long-term monitoring, single
(flexible)*	same or different species at different site	Pollock et al. (2002)	species, single/multiple sites

(2) capture probability correction where an index of individual counts (iCap) is corrected using capture probabilities estimated from capture–recapture (Table 1). We tested two types of iCap, one where capture probabilities used were specific to each species at each geographic site, and one that allowed capture probabilities for a species to be used from the same species at a different site or a similar species (Table 1). We tested correlation, predictive ability, consistency, and generalizability of each index and calibration method relative to capture–recapture abundance estimates.

# MATERIALS AND METHODS

NEON data description and processing.—The NEON data set we used contains capture–recapture data for 63 species over 46 sites from 2013 to 2019, representing 39% of North American small mammal species. Each site contains 3–8 (mean = 6) replicate trapping arrays of 100 traps set in grids with 10-m spacing (Supplementary Data SD1, Fig. 1; NEON 2020a). Field scientists used baited Sherman live traps for animal capture, checked daily, and set within 10 days before or after the new moon (Thibault 2019). Traps were typically run at monthly intervals for 6 months during the growing season at a subset of sites (core sites; NEON 2020a), and 3–4 months for the rest of the sites. At each site, half of the trap arrays were run for multiple nights (mean = 3) and the other half were run for a single night. Trapped individuals were tagged with either an individually identifiable ear tag or PIT tag (Thibault 2019). NEON trapping targets small rodents including cricetids (New World rats and mice, lemmings, voles), dipodids (jumping mice), heteromyids (kangaroo rats, pocket mice), small sciurids (squirrels, chipmunks), and introduced murids (Old World rats and mice, gerbils; Thibault 2019). Sampling did not target lagomorphs (rabbits, hares, pikas), mustelids (weasels), large squirrels, or soricids (shrews). Although some of these taxa were incidentally captured, they were removed from this analysis. Some captured individuals (n = 453) were incorrectly identified at the species level in the field, and we used NEON DNA barcoding data (NEON 2020b) to correct those identifications before analysis.

Capture-recapture analysis.--Under the NEON sampling protocol, about half of the trapping grids at each site were sampled multiple nights per trapping session, while other grids were sampled only a single night per month. Traditional capture-recapture abundance estimates require that multiple nights of sampling are conducted, omitting data from grids that are sampled only a single night. To accommodate all the data, we used a Bayesian hierarchical approach wherein trapping arrays sampled for multiple nights share a likelihood with arrays that ran only one night, thus providing information on capture probability that exploits the information available both in repeat- and single-night counts. The approach of Royle et al. (2012) was adapted to estimate abundance at replicate trapping arrays within each geographic site during each month and year of sampling (Royle et al. 2012; Royle and Converse 2014; Sollmann et al. 2015). The model assumes that the abundance



**Fig. 1.**—National Ecological Observatory Network (NEON) small mammal trapping design. Trapping took place at 46 geographic sites (circles) throughout the United States including Alaska and the territory of Puerto Rico. Each site contained 3-8 (mean = 6) unique trapping arrays (triangles) at which 100 traps were run over 7 years (2013–2019), with samples being taken 1–8 months each year (mean = 4.43).

at a trapping array g at a given geographic site s is a Poisson random variable with an array-specific mean  $(\lambda_{o})$ :

$$N_g \sim Poisson(\lambda_g)$$

with

$$\log(\lambda_g) = \beta_0$$

where  $\beta_0$  has prior distribution:

$$\beta_0 \sim Normal(0, 0.01)$$

We condition on the total population size  $N_{Total}$  over all G plots such that the  $N_g$  variables have a multinomial distribution:

$$N = (N_1 \dots N_G) | N_{Total} \sim Multinom(\pi | N_{Total})$$

with  $N_{total} = \sum_{g} N_{g}$  and multinomial probabilities:

$$\pi_g = \frac{\lambda_g}{\sum_g \lambda_g}$$

We use data augmentation to fix the dimensionality of the model (i.e., make an unknown sample size "known") by choosing an arbitrarily large super-population size M of 700 individuals for each species at each trapping array during each closed session, representing a density of roughly 70,000 individuals/km<sup>2</sup>. We embed the multinomial for  $N_g$  into a multinomial of the same dimension but with larger, fixed sample size by introducing a latent super-population variable  $U_g$ , the sum of which is equal to M, giving U a multinomial distribution:

$$U|M \sim Multinom(M; \pi)$$

where

$$\pi_g = \frac{\lambda_g}{\sum_g \lambda_g}$$

are the same probabilities for target multinomial **N**. We specify Bernoulli variables that differentiate between "real" and "pseudo-" individuals (i) by:

$$z_i \sim Bernoulli(\psi)$$
 for  $i = 1, 2, \dots, M$ 

where

$$\psi = \frac{1}{M} \sum_{g} \lambda$$

The capture histories of all individuals of a given species captured at a geographic site are pooled (i.e., capture histories for each individual over all arrays at the site) and we then add an individual-level covariate describing the array membership,  $g_i$ :

$$g_i \sim Categorical(\pi)$$

Both  $z_i$  and  $g_i$  are latent for augmented individuals. Estimates of abundance at each replicate array within the geographic site can be obtained by summing over the individuals with  $z_i = 1$ associated with each array.

The likelihood for the site-level capture history data is therefore given by:

$$y_{ik} \sim Bernoulli(z_i p)$$

where *p* has prior distribution:

$$p \sim Unif(0, 1)$$

where k is the number of capture periods (3) and p is the sitelevel capture probability for a given species, assumed constant over the 3 days of sampling and between individuals.

Because sampling at each trapping array took place over a restricted time period (1-3 days), the assumption of constant detection probability *p* over time is reasonable. While accounting for individual heterogeneity in capture probabilities has been shown to be important in small mammal studies (e.g., Davis et al. 2003), because we are making inference about individuals for which we have no individual detection data (i.e., those captured in arrays running single nights only) we could not adequately model individual heterogeneity in capture probabilities.

We ran our model separately for each species within a site/ month/year closed session in JAGS (Plummer 2003) via package runjags (Denwood 2016) in Program R (Version 3.5.3; R Core Team 2017). We skipped species during closed sessions with <10 individuals captured at a geographic site and those site/month/years with <3 trapping arrays running, for a total of 9,226 models. We assessed model fit for each species closed session with posterior predictive checks (Kéry and Schaub 2011; Gelman et al. 2014) by calculating the sum of squared Pearson residuals. We calculated a Bayesian P-value from posterior simulations and assumed adequate fit if 0.1 < P < 0.9. Most models converged (i.e., Gelman-Rubin statistic < 1.1, confirmed by examining traceplots) after a burn-in of 5,000 iterations and sampling of 300,000 iterations, thinning every 100 samples. Those that did not converge after that number of iterations were discarded (n = 114). Resulting abundance estimates are archived and publicly available on Dryad (https://doi. org/10.5061/dryad.v41ns1rw2; Parsons et al. 2022)

*Index calculation.*—We calculated two commonly used indices of abundance: the count of individuals captured (iCap) and trap success, both standardized to count/3 days, to see how well they aligned with estimates of absolute abundance estimated from capture–recapture. iCap is a count of the unique individuals captured and relies on the ability to identify individuals. Trap success is the count of animals divided by the sampling effort and does not rely on individual marks. We calculated each index at the trapping array level.

*Regression–calibration.*—We used a linear modeling approach to examine the relationship between each index of abundance and absolute abundance (i.e., abundance estimated from capture–recapture; hereafter regression–calibration). We weighted this regression for the precision of the absolute

abundance estimates such that more precise estimates received a higher weight and contributed more to the relationship (Jhala et al. 2011). We tested three models, one with iCap as the sole predictor of absolute abundance, one with trap success as the sole predictor, and one that considered additive effects of iCap and trap success. We ranked our models using Akaike's Information Criterion (AIC) with the lowest AIC representing the best-fit model (Burnham and Anderson 2002) which we then used to generate predictions of absolute abundance.

*p-corrected iCap.*—We calibrated one index, iCap, by using the average capture probability of a species at a site estimated from capture–recapture analysis to correct iCap counts (hereafter *p*-corrected iCap) such that:

$$\widehat{N_g} = C_g / \overline{p_s}$$

where  $\hat{N}_g$  is the *p*-corrected abundance estimate for a given species at trap array *g*,  $C_g$  is the iCap count, and  $\overline{p_s}$  is average capture probability per capture period estimated by capture– recapture for a given species at geographic site *s* (Pollock et al. 2002). We note that iCap is simply a special case of the Huggins capture–recapture estimator (Huggins 1989). We tested two types of *p*-correction for iCap: (1) estimates of capture probability for a given species at a given site could be taken from that species, or a species of the same genus, at different sites if estimates from the same site were lacking and (2) estimates of capture probability had to come from the same species at the same site/month/year (hereafter "restricted").

Index calibration performance, consistency, and generalizability.—We assessed the overall performance of our raw indices and index calibrations by summarizing the linear correlation of regression-calibrated and *p*-corrected abundance estimates with absolute abundance estimates from capture–recapture. We summarized these correlations through  $R^2$ , relative bias, and root mean squared error (RMSE; calibrated for different species by dividing by the input data for easy comparison). We assessed overall performance across all species and trapping arrays and then separately by species to determine how generalizable the relationships were.

We assessed the consistency of the indices and index calibrations for six widespread species with data from many sites. We chose them from two groups to represent the bestcase and worst-case scenarios: three with high variation in capture probabilities over time and space (Myodes gapperi, *Peromyscus keeni*, and *P. leucopus*;  $SD_p = 0.17, 0.18$ , and 0.17, respectively) and three with low variation in capture probabilities (Perognathus parvus, Neotamias townsendii, and Neotoma *albigula*;  $SD_{p} = 0.1, 0.1, and 0.09$ , respectively) and for which we had abundance estimates at a minimum of 30 arrays/month/ years. For each species, we randomly chose 10 closed array/ month/year sessions from the data to train our index or index calibration model, then evaluated how well the resulting model predicted abundance at the remaining array/month/years, given index data. We replicated this leave-out validation 10 times for each species, choosing a new random 10 array/month/years for training each time and plotted the resulting RMSE and  $R^2$ values.

We assessed the generalizability of index calibrations by examining variation in capture probabilities over space, time, and species and their relationship with index calibration performance. We also examined the ability of indices and index calibrations to detect the same population changes over time, space, and between species as was detected by capture–recapture, evaluating the correlation between trend lines by calculating Pearson correlation coefficients.

### RESULTS

Out of 63 target species captured over a possible 4,900 spatially and temporally closed trapping arrays (i.e., 321 arrays sampled in each of 1–41 months), we generated 9,112 (81% success) capture–recapture abundance estimates for 47 species over 4,275 array replicates. Model fit for capture–recapture models was good, with all species closed sessions having Bayesian *P*-values between 0.1 and 0.9, with most around 0.5 indicating excellent fit (Supplementary Data SD2; Gelman et al. 2014). Data augmentation was sufficient as assessed by mean inclusion probabilities (z; Supplementary Data SD3). Abundance estimates were not attempted for species with low total captures and recaptures due to a tendency for unstable estimates and standard errors (i.e., very large or very small; 19%, <10 individuals per geographic site; Supplementary Data SD1).

Both raw indices were positively but weakly correlated with absolute abundance, with the iCap index performing better ( $R^2 = 0.4$ ) than trap success ( $R^2 = 0.3$ ). Index calibrations all performed better than raw indices in terms of fit with *p*-corrected iCap performing better in terms of  $R^2$ , RMSE, and relative bias (overall  $R^2 > 0.6$ ) than regression-calibrated indices (overall  $R^2 = 0.47$ ; Fig. 2; Supplementary Data SD4). When comparing two approaches for calibrating iCap with capture probabilities, we found that using values from the same species at the same site during the same month/year improved the performance of *p*-corrected iCap and trap success through linear regression, a model including terms for both iCap and trap success performed best (Supplementary Data SD5).

Heterogeneity in capture probabilities was important to how well and how consistently index calibrations predicted capturerecapture abundance estimates. Capture probability was highly variable over sites and over time for some species, though low for others (mean.SE = 0.05, range 0.01–0.15; Supplementary Data SD4) with spatial variation being generally higher than temporal variation (Fig. 3). When we based our indices and calibrations on 10 trapping array/month/year closed samples (i.e., training) and used the resulting models to predict capturerecapture abundance estimates in the remaining array/month/ year closed samples (range 28-158; i.e., testing), we found lower fit and higher levels of variation in fit for species with high variation in capture probability, with higher fit and lower levels of variation in fit for species with low variation in capture probabilities (Supplementary Data SD6). We found that index calibrations were more consistent in their estimates than raw indices, especially when indices were corrected for variation



Fig. 2.-Correlation of absolute abundance estimates from capturerecapture analysis with five indices of abundance for small mammals. (A) iCap: the count of individuals captured and (B) total count: trap success shows raw abundance indices, both scaled to count/3 days. The other estimates are calibrated index abundance estimates including capture probability (p) – corrected iCap (C, E) and regression-calibrated indices (N ~ iCap + trap success; D). Capture probabilities were taken from the same species at the same site only (restricted; C) or allowed to come from the same species at a different site or a species of the same genus (E). The solid black line shows a 1:1 relationship, the dotted gray line shows best fit line, and the shading shows the density of points. Plots show 17,155 capture-recapture estimates for 47 species at 321 trapping arrays located at 46 unique geographic sampling sites (mean = 6 arrays per site) over 7 years (2013–2019). Both types of p-corrected iCap had similar relative bias, only the restricted p-correction is shown in panel F. Both types of p-correction had lower relative bias than regression-calibrated indices (F). The index with the highest  $R^2$ , lowest root mean squared error (RMSE), and low relative bias was p-corrected iCap using a restricted capture probability estimate.

in capture probability (*p*-corrected iCap; Supplementary Data SD6). Overall, species with higher average capture probabilities tended to have better fit of calibrated indices to capture–recapture abundance estimates (Supplementary Data SD7).

The raw and calibrated indices were capable of mirroring trends in capture–recapture abundance over time and space, achieving Pearson correlation of >0.6 of at least one raw index/ calibration with capture–recapture in 50% of example cases



**Fig. 3.**—Boxplots showing how capture probability varies over sites within a single month (September 2018; top panel) and over time at a single site (BART; bottom panel). Capture probability was estimated by capture–recapture analysis. Species shown are those detected at a minimum of three sites over at least 5 months.

(Fig. 4). Over all indices/calibrations, raw iCap had the highest level of correlation to capture–recapture estimates 62.5% of the time, followed by *p*-corrected iCap 37.5% of the time (Fig. 4). However, raw and calibrated indices were prone to errors when assessing relative changes in the populations of different species with relative community changes over sites, sometimes differing widely from those estimated by capture–recapture (Supplementary Data SD8).

# DISCUSSION

The NEON small mammal data set provides the opportunity to explore variation in capture probabilities and the appropriateness of indices over an unprecedentedly large number of species, sites, and time periods. Despite the challenge of having some trapping arrays sample only a single night, using a Bayesian hierarchical approach allowed us to generate capture–recapture abundance estimates for most (81%) of species/ array/month/year replicates with associated capture probability estimates aligning well with previously published estimates (Hammond and Anthony 2006). While our joint-likelihood



**Fig. 4.**—Ability of indices and index calibrations to reflect population changes over time (A) and space (B). For (A), we selected four example species at one site over 6 years (2014–2019) for which we had at least 6 months per year of capture–recapture estimates. For (B), we selected four example species during 1 month of 1 year for which we had at least four sites at which capture–recapture estimates could be generated. Dashed lines show raw indices (iCap and trap success). Based on Pearson's correlation coefficient (top left of each subplot), indices were well correlated with capture–recapture abundance estimates through time and space for some species (e.g., *P. parvus, P. maniculatus*) but poorly correlated for others (e.g., *P. leucopus, T. striatus*). Overall, iCap was the best index in terms of Pearson correlation to capture–recapture estimates 62.5% of the time, followed by *p*-corrected iCap (37.5%).

approach allowed inference on more plots than would be possible otherwise, we did not attempt to estimate abundance for sites with very few captures or very few replicate trapping arrays. Although we were quite conservative with our hierarchical framework, this model could be extended to allow additional pooling across species, sites, months, and years to increase the power to generate estimates for rare species. We caution, however, that estimates for rare species could be driven by more common species which may become problematic over months and years given the sometimes drastic cycles in small mammal populations. Furthermore, for the rarest of species detected at single sites, within single months and/or years, pooling of data will be of little benefit. In these cases, practitioners are left with a choice: ignore those sites and species where abundance cannot be estimated, potentially missing important information about the most at-risk species, or employ other methods less dependent on a minimum sample size but potentially sensitive to variation in capture probabilities over time, space, and species (i.e., an index). The use of indices to monitor small mammals has been much debated. Our study supports past research suggesting that if indices are to be used for monitoring, accounting for variation in detection probability (i.e., iCap) is preferred over indices that ignore detection probability (i.e., trap success; Slade and Blair 2000; McKelvey and Pearson 2001). In general, iCap captured a larger amount of variation and was better correlated with absolute abundance estimates from capture-recapture; however, individual species varied widely in the strength of that relationship. For individual species, iCap worked best when N was small, also consistent with past research which has found that the fluctuations of large populations over a wide range are more challenging for indices to track, possibly a result of trap saturation (Schwemm et al. 2018). Nevertheless, indices, especially iCap, worked well for capturing fluctuations in abundance of single species over time and space in 50% of our example cases, with correlations being lower over sites. Indices and calibrations tracked capture-recapture abundance estimates less well for fluctuations of several species simultaneously, supporting the concern that indices may not be appropriate to monitor communities (Slade and Blair 2000; McKelvey and Pearson 2001).

The inability of indices to capture similar patterns to capture-recapture, especially across sites and between species, was likely due to variation in capture probabilities (Nichols 1986; Watkins et al. 2010). Some species showed substantial variation in capture probability over time and space, which resulted in poor correlation and correlation consistency (i.e., the ability to predict out-of-sample abundance) between index and absolute abundance, although high overall capture probability tended to improve correlation. An inability to account for variation in capture probability is the main reason why use of raw indices is discouraged (i.e., Nichols and Pollock 1983) despite examples of good correlation for some species. Index calibrations that make specific use of capture probability estimates can help account for heterogeneity in capture probabilities and improve the use of indices for monitoring of single and multiple species, but the choice of which capture probabilities to use where they are lacking for rare species is critical. The use of capture probabilities to correct iCap not only provided

better fit to capture-recapture data overall but performed as well or better at capturing variation in abundance over time, space, and species with lower relative bias than raw indices or regression-calibrated indices. This makes sense since capturerecapture abundance estimates are at their core a correction of iCap for capture probabilities. The strength of using a p-corrected index is that capture-recapture modeling could be done on a smaller subset of plots within a site to provide the information for estimation of capture probabilities to be used with other plots less rigorously sampled but for which iCap could still be obtained (much in the way of NEON sampling; aka "double sampling"; Pollock et al. 2002). Capture probabilities to correct iCap could be used from different sites, or potentially different but similar species; however, performance of such a p-correction was inferior to a more restricted approach, while still outperforming raw indices across all species, sites, and month/years. This suggests that databases containing capture probability estimates for rare species from different sites over time throughout their range may be useful to facilitate index calibration and improve our understanding of population status and trajectories but cannot substitute site/time/species-specific capture probabilities where absolute abundance is needed. We found that regression-based calibrations were inferior to p-corrected indices in terms of fit, relative bias, and reflection of capture-recapture trends in abundance over space and time. This is consistent with past studies that have found  $R^2$ -based index calibrations to be unreliable (e.g., Gopalaswamy et al. 2015).

Our work supports others that have cautioned against the use of indices where capture probabilities can be expected to vary over time, space, and species. Most researchers that have written on the subject have concluded that capture-recapture estimates remain preferable where possible because of their ability to account for heterogeneity in capture probabilities (Slade and Blair 2000; Wiewel et al. 2009). However, a major issue remains of how to monitor populations of rare species since captures will be too low for traditional models. Some authors have suggested combining capture-recapture models and iCap, with the latter being used for situations where capture-recapture is not possible (e.g., Hanley and Barnard 1999; Slade and Blair 2000). However, as pointed out by McKelvey and Pearson (2001), this solution is risky because of the differential bias where capture-recapture tends to be biased high and iCap low, which may result in incomparable estimates. Fortunately, capture-recapture models have come a long way in the last several decades, with Bayesian methods able to leverage prior information and accommodate pooling (i.e., share information about capture probability over sites and/or species) to improve estimates of abundance for species with sparse data. Nevertheless, we did not attempt to generate abundance estimates for sites at which fewer than 10 individuals were captured, below that recommended for MLE capture-recapture (Otis et al. 1978) but above some of the lowest sample sizes for Bayesian capturerecapture analysis published (e.g., Reppucci et al. 2011). It is possible that with additional pooling we could have generated abundance estimates for species at sites with extremely low captures, but we feel that a more conservative approach allowed us to more rigorously explore the specific question of indices and index calibration to make specific recommendations about their use. Nevertheless, we advise that capture–recapture modeling be used whenever possible and caution that shortcuts via the use of indices and index calibrations can provide misleading results.

Based on our findings, we can make several recommendations in terms of design and analysis for different types of small mammal studies (Table 1). Specifically, we consider: (1) species with few captures; (2) comparing multiple species; (3) long-term monitoring at a single site; (4) long-term monitoring at multiple geographically separate sites simultaneously; and (5) short-term snapshots at many geographically separate sites.

Species with low captures.-For species with too few captures for capture-recapture models, individual marks remain important since iCap, which relies on individual marks, performed far better than simple trap success. iCap did not perform as well as index calibration methods over all species/closed sessions; however, it did well in comparison to p-corrected iCap in monitoring trends over time and space for a subset of species. Depending on the goal of the monitoring, iCap may be adequate, especially for monitoring trends over time. If absolute abundance is the goal, captures could be increased such that capture-recapture becomes possible simply by increasing the number of sampling days per closed session. If done at a subset of sites where the species is present, then a joint-likelihood approach could become feasible for minimal extra effort or traditional capture-recapture models could be used to estimate capture probability and/or establish a regression-calibration for index values, thus improving the overall correlation of the index with absolute abundance (i.e., double sampling; Pollock et al. 2002). Some recently developed capture-recapture models allow the use of passive detectors (i.e., camera traps) and low (or no) marked animals (e.g., Chandler and Royle 2013; Carter et al. 2019) which could further reduce effort associated with estimating abundance for species with low captures.

*Comparing multiple species.*—Our results suggest caution when using indices to compare relative abundance of multiple species because capture probabilities vary widely between species. Methods that account for this variation will provide the best inference to compare population trends of different species over time and space. Using a joint-likelihood approach or *p*-corrected iCap with some subset of sites/years being sampled for capture–recapture for each species would account for variation in capture probabilities; however, both would rely in part on adequate capture data. For relatively common species, sampling effort could be reduced by using a simple Lincoln–Petersen two-sample estimator to estimate capture probabilities instead of a multi-sample estimator (Pollock et al. 2002).

Long-term monitoring of a single species at a single site.— We found that indices and their calibrations better correlated trends in capture–recapture abundance estimates over time than over space. If monitoring a site over a long period of time during which capture probabilities would not be expected to change, an index calibration method (*p*-corrected iCap) could be adequate, given initial calibration via capture–recapture estimates of capture probability. We note, however, that this assumption of constant capture probabilities becomes less likely with longer time periods where species-specific behaviors may change over time due to changes in the small mammal community, changes to the predator community or habitat, and climate change. Thus, for longer-term studies using indices we strongly recommend a periodic recalibration of the index.

Long-term monitoring a single species at multiple geographically separate sites simultaneously.—Our results suggest that variation in capture probabilities over sites is greater than variation over time, at least for the time scale considered by our study. Thus, an assumption of constant capture probabilities over sites is not as reasonable as constant capture probabilities over time. In these cases, practitioners using indices risk compounding variation in capture probability over space with variation over time. Thus, we suggest the best option is to formally account for those sources of variation through capturerecapture modeling. In situations where that is not possible, indices and their calibrations should be used with caution, and independently at each site to account for capture probability variation across sites (i.e., p-corrected iCap). Just as for a single site being monitored over time, the index should be recalibrated regularly over time so ensure changing conditions impacting capture probabilities are accounted for as much as possible.

Short-term snapshots of a single species at many geographically separate sites.-In this scenario we are concerned most with variation in capture probabilities over space, but not time. Since site-specific factors can be complex in the different ways in which they affect capture probabilities, the best course is a formal capture-recapture model with site-specific habitat covariates. However, this may not be possible where the snapshot occurs over a single night without repeat visits (i.e., a Bioblitz). In this case, if there is no option for capture-recapture, iCap is the best option but should be used with caution because correlation with absolute abundance will vary depending on the site. As an alternative, we suggest a double-sampling approach where capture-recapture sampling be undertaken at a small number of trapping grids at each site for several nights spanning the wider snapshot to allow estimation of capture probabilities and abundance modeling using a joint-likelihood approach or p-corrected iCap.

*Conclusions.*—Making inference for populations of rare species is challenging, and indices are an attractive solution. We show that accuracy of indices is species-specific and can vary over space and time. Calibrated indices are more closely correlated with absolute abundance and more reliable but require some information from capture–recapture analysis which may not always be feasible to obtain. Efficient designs for capture–recapture (i.e., double sampling) and leveraging of Bayesian methods and joint likelihoods may improve our ability to generate absolute abundance estimates while reducing effort, but the rarest of species will remain problematic. Indices for those species, if used, should be used with caution and calibrated and/ or validated against absolute abundance as much as possible to ensure accurate, meaningful inference.

#### ACKNOWLEDGMENTS

The National Ecological Observatory Network (NEON) is a program sponsored by the National Science Foundation and

operated under cooperative agreement by Battelle Memorial Institute.

## FUNDING

This material is based in part upon work supported by the National Science Foundation through the NEON Program and NSF award #582434.

Submitted 14 April 2021. Accepted 24 July 2022.

# SUPPLEMENTARY DATA

Supplementary data are available at *Journal of Mammalogy* online.

**Supplementary Data SD1.**—Summary of National Ecological Observatory Network (NEON) small mammal trapping data of 63 species over 46 sites containing  $3-8\ 10\ m \times 10\ m$  (mean = 6) replicate trapping arrays of 100 traps each, from 2013 to 2019. Species are listed in descending order of most commonly trapped, showing the number of individuals captured over all sites, months and years, and the number of total recaptures of those individuals. Incidental captures of nontarget species are not included. Species names marked with \* (n = 18) did not have sufficient captures during any site/month/year for capture–recapture abundance estimation.

**Supplementary Data SD2.**—Range of abundance estimates generated from capture–recapture analysis over all geographic sites in each month and year of sampling. The range of credible interval limits is given in parentheses. The range of goodness-of-fit Bayesian *P*-values, calculated from the sum of Pearson residuals, is also given for each species. Suitable fit is indicated by values between 0.1 and 0.9.

**Supplementary Data SD3.**—The distribution of mean inclusion probabilities (z) associated with data augmentation for each species at each geographic site. Data come from 46 National Ecological Observatory Network (NEON) sampling sites across the United States containing 3–8 trapping arrays each. Separate abundance estimates were generated each array, month and year of sampling, totaling 4,026 possible estimates for each species. The actual number of estimates possible for each species varied based on species range and capture frequency across sites and trapping arrays (cutoff: <10 individuals/ site).

**Supplementary Data SD4.**—Relationships between two calibrated indices and absolute abundance for 32 species of small mammal. The species shown had at least five abundance estimates that could be used to estimate the  $R^2$  and root mean squared error (RMSE). We used two different calibrated indices: (1) we modeled the correlation between an index (the number of individuals captures [iCap] at an array combined with the raw count of animals [including double counts] at that array, both scaled to count/3 days) and true abundance estimated from capture–recapture, and (2) we corrected iCap with estimated capture probabilities from capture–recapture (*p*-correction). We assessed the ability of the model to predict abundance by assessing the  $R^2$  value of the predictions versus estimates, and the RMSE, standardized for the input data, both are shown as

"Regression–calibrationl*p*-correction." We also list the average capture probability (*p*) and standard error for each species estimated from capture–recapture. Data come from 46 National Ecological Observatory Network (NEON) sampling sites across the United States containing 3–8 trapping arrays each. Separate abundance estimates were generated each array, month and year of sampling, totaling 4,026 possible estimates for each species. The actual number of estimates possible for each species varied based on species range and capture frequency across sites and trapping arrays (cutoff: <10 individuals/site). Species are sorted by regression-calibrated  $R^2$ .

Supplementary Data SD5.—Correlation of absolute abundance estimates from capture-recapture analysis and model-predicted abundance from regression-calibrated indices: the number of unique individuals captured (iCap) and trap success (counts without individual identification), both standardized to count/3 days. Three models were tested, one with iCap as sole predictor, one with trap success as sole predictor, and one with both indices as predictors. The red line shows a 1:1 relationship and the blue line shows best-fit line. The model with both indices was the best-fitting model based on AIC ( $\sim$ iCap = 50,347, ~Trap success = 50,103, ~iCap + Trap success = 49,937) with highest  $R^2$  and lowest root mean squared error (RMSE). All models performed better in terms of AIC than an intercept-only model (AIC = 54,386). Models were based on 47 species over 46 sampling sites across the United States containing 3-8 trapping arrays each. Sites were sampled over 7 years (2013–2019), with samples being taken 4-12 months each year (mean = 9).

Supplementary Data SD6.—Evaluation of how well indices and index calibrations from a subset of trap arrays can be used to estimate abundance at other trap arrays (over different geographic sites, months and years), as measured by error (root mean squared error [RMSE]) and fit  $(R^2)$ . To represent the bestand worst-case scenarios, we chose six widespread species captured at minimum 30 arrays, three of which had low (SD  $\leq$ 0.1) and three of which had high ( $SD \ge 0.18$ ) variation in capture probability (p) over space and time. For each species, we randomly chose 10 arrays from the data to train indices (Trap success (A) and iCap (B)) and index calibration models (regression-calibrated (C), unrestricted p-correction (D), restricted p-correction (E)), then evaluated how well indices and calibrations predicted abundance estimated from capture-recapture at the remaining arrays. We replicated this leave-out validation 10 times for each species, choosing a new random 10 arrays for training each time and plotting the distribution of resulting RMSE and  $R^2$  across the 10 replicates. Median fit tended to be higher for species with low variation in capture probability (i.e., high  $R^2$ , low RMSE), and variation in fit (i.e., smaller IQR) was lowest for index calibrations (especially *p*-corrected calibrations) compared to raw indices, indicating greater consistency of calibrated abundance estimates.

**Supplementary Data SD7.**—Relationship between regression-calibrated index fit to capture–recapture data and average capture probability. (A) shows examples of correlations between capture–recapture estimated abundance and calibrated index estimated abundance for three species. The three example

species vary in the fit of calibrated indices to capture–recapture estimates and in their average capture probability. (B) shows how the fit of calibrated index models as measured by  $R^2$  varies with average capture probability (*p*) over 33 species for which capture probability could be estimated over at least two trapping arrays. Species with higher average capture probabilities tended to have better fit of index calibrations to capture–recapture abundance estimates. All species-level results are presented in Supplementary Data SD4.

**Supplementary Data SD8.**—Ability of indices and index calibrations to detect relative population changes between species over time (A) and space (B). For (A), we selected four example species at one site (HARV) over 4 years (2014–2018) for which we had at least 4 months per year of capture–recapture estimates for each species. For (B), we selected four example species during 1 month of 1 year (8/2019) for which we had at least four sites at which capture–recapture estimates could be generated. Indices and index calibrations did better reflecting relative changes between species over time than over space, but some example species peaks in abundance did not align with peaks estimated from capture–recapture (i.e., *P. maniculatus*). Relative changes over sites sometimes differed widely from those estimated by capture–recapture (i.e., *Myodes rutilus*).

## LITERATURE CITED

- Bosch J., Fernández-Beaskoetxea S., Garner T.W., Carrascal L.M. 2018. Long-term monitoring of an amphibian community after a climate change-and infectious disease-driven species extirpation. Global Change Biology 24:2622–2632.
- Brown K.P., Moller H., Innes J., Alterio N. 1996. Calibration of tunnel tracking rates to estimate relative abundance of ship rats (*Rattus rattus*) and mice (*Mus musculus*) in a New Zealand forest. New Zealand Journal of Ecology 20:271–275.
- Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, USA.
- Carter A., Potts J.M., Roshier D.A. 2019. Toward reliable population density estimates of partially marked populations using spatially explicit mark–resight methods. Ecology and Evolution 9:2131–2141.
- Chandler R.B., Royle J.A. 2013. Spatially explicit models for inference about density in unmarked or partially marked populations. The Annals of Applied Statistics 7:936–954.
- Davis S.A., Akison L.K., Farroway L.N., Singleton G.R., Leslie K.E. 2003. Abundance estimators and truth: accounting for individual heterogeneity in wild house mice. The Journal of Wildlife Management 67:634–645.
- Denwood M.J. 2016. runjags: an R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. Journal of Statistical Software 71:1–25.
- Dröge E., Creel S., Becker M.S., Loveridge A.J., Sousa L.L., Macdonald D.W. 2020. Assessing the performance of index calibration survey methods to monitor populations of wide-ranging low-density carnivores. Ecology and Evolution 10:3276–3292.
- Dylewski L., Ortega Y.K., Bogdziewicz M., Pearson D.E. 2020. Seed size predicts global effects of small mammal seed predation on plant recruitment. Ecology Letters 23:1024–1033.

- Furnas B.J. 2020. Rapid and varied responses of songbirds to climate change in California coniferous forests. Biological Conservation 241:108347.
- Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. 2014. Bayesian data analysis. 3rd ed. CRC Press, Boca Raton, Florida, USA.
- Gopalaswamy A.M., Delampady M., Karanth K.U., Kumar N.S., Macdonald D.W. 2015. An examination of index-calibration experiments: counting tigers at macroecological scales. Methods in Ecology and Evolution 6:1055–1066.
- Grant W.E., French N.R. 1980. Evaluation of the role of small mammals in grassland ecosystems: a modelling approach. Ecological Modelling 8:15–37.
- Hammond E.L., Anthony R.G. 2006. Mark–recapture estimates of population parameters for selected species of small mammals. Journal of Mammalogy 87:618–627.
- Hanley T.A., Barnard J.C. 1999. Spatial variation in population dynamics of Sitka mice in floodplain forests. Journal of Mammalogy 80:866–879.
- Hopkins H.L., Kennedy M.L. 2004. An assessment of indices of relative and absolute abundance for monitoring populations of small mammals. Wildlife Society Bulletin 32:1289–1296.
- Huggins R.M. 1989. On the statistical analysis of capture experiments. Biometrika 76:133–140.
- Jhala Y., Qureshi Q., Gopal R. 2011. Can the abundance of tigers be assessed from their signs? Journal of Applied Ecology 48:14–24.
- Jones J.P. 2011. Monitoring species abundance and distribution at the landscape scale. Journal of Applied Ecology 48:9–13.
- Kéry M., Schaub M. 2011. Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press, Cambridge, Massachusetts, USA.
- Krebs C.J., Boonstra R., Gilbert B.S., Kenney A.J., Boutin S. 2019. Impact of climate change on the small mammal community of the Yukon boreal forest. Integrative Zoology 14:528–541.
- Lambin X., Petty S.J., Mackinnon J.L. 2000. Cyclic dynamics in field vole populations and generalist predation. Journal of Animal Ecology 69:106–119.
- Ludwig A., Zheng H., Vrbova L., Drebot M.A., Iranpour M., Lindsay L.R. 2019. Climate change and infectious diseases: the challenges: increased risk of endemic mosquito-borne diseases in Canada due to climate change. Canada Communicable Disease Report 45:91–97.
- McKelvey K.S., Pearson D.E. 2001. Population estimation with sparse data: the role of estimators versus indices revisited. Canadian Journal of Zoology 79:1754–1765.
- NEON. 2020a. Small mammal box trapping RELEASE-2021 (DP1.10072.001). https://doi.org/10.48443/j1g9-2j27. Accessed 26 March 2021.
- NEON. 2020b. Small mammal sequences DNA barcode (DP1.10076.00), RELEASE-2021. https://doi.org/10.48443/ zhr5-ea60. Accessed 26 March 2021.
- Nichols J.D. 1986. On the use of enumeration estimators for interspecific comparisons, with comments on a 'trappability' estimator. Journal of Mammalogy 67:590–593.
- Nichols J.D., Pollock K.H. 1983. Estimation methodology in contemporary small mammal capture–recapture studies. Journal of Mammalogy 64:253–260.
- Otis D.L., Burnham K.P., White G.C., Anderson D.R. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:3–135.
- Parsons A.W., Clark J., Kays R. 2022. Monitoring small mammal abundance using NEON data: are calibrated indices useful? Dryad. doi:10.5061/dryad.v41ns1rw2.

- Plummer M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing 124:1–10.
- Pollock K.H., Nichols J.D., Simons T.R., Farnsworth G.L., Bailey L.L., Sauer J.R. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13:105–119.
- Princé K., Lorrilliere R., Barbet-Massin M., Jiguet F. 2013. Predicting the fate of French bird communities under agriculture and climate change scenarios. Environmental Science & Policy 33:120–132.
- R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramesh D., Home C., Jhala Y.V., Qureshi Q. 2013. Calibration of a burrow count index for the Indian desert jird, *Meriones hurrianae*. Population Ecology 55:241–245.
- Reppucci J., Gardner B., Lucherini M. 2011. Estimating detection and density of the Andean cat in the high Andes. Journal of Mammalogy 92:140–147.
- Royle J.A., Converse S.J. 2014. Hierarchical spatial capture–recapture models: modelling population density in stratified populations. Methods in Ecology and Evolution 5:37–43.
- Royle J.A., Converse S.J., Link W.A. 2012. Data augmentation for hierarchical capture-recapture models. arXiv, arXiv:1211.5706, preprint.
- Sage R.F. 2020. Global change biology: a primer. Global Change Biology 26:3–30.
- Schnurr J.L., Canham C.D., Ostfeld R.S., Inouye R.S. 2004. Neighborhood analyses of small-mammal dynamics: impacts on seed predation and seedling establishment. Ecology 85:741–755.
- Schwemm C.A., Drost C.A., Orrock J.L., Coonan T.J., Stanley T.R. 2018. Comparison of estimators for monitoring long-term population trends in deer mice, *Peromyscus maniculatus*, on the California Channel Islands. Western North American Naturalist 78:496–509.
- Slade N.A., Blair S.M. 2000. An empirical test of using counts of individuals captured as indices of population size. Journal of Mammalogy 81:1035–1045.
- Sollmann R., Mohamed A., Samejima H., Wilting A. 2013. Risky business or simple solution—relative abundance indices from camera-trapping. Biological Conservation 159:405–412.
- Sollmann R., White A.M., Gardner B., Manley P.N. 2015. Investigating the effects of forest structure on the small mammal community in frequent-fire coniferous forests using capture-recapture models for stratified populations. Mammalian Biology 80:247–254.
- Steinhorst R.K., Samuel M.D. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. Biometrics 45:415–425.
- Thibault K. 2019. TOS protocol and procedure: small mammal sampling. National Ecological Observatory Network.
- Van Buskirk J., Mulvihill R.S., Leberman R.C. 2009. Variable shifts in spring and autumn migration phenology in North American songbirds associated with climate change. Global Change Biology 15:760–771.
- Watkins A.F., McWhirter J.L., King C.M. 2010. Variable detectability in long-term population surveys of small mammals. European Journal of Wildlife Research 56:261–274.
- Wiewel A.S., Adams A.A.Y., Rodda G.H. 2009. Evaluating abundance estimate precision and the assumptions of a count-based index for small mammals. The Journal of Wildlife Management 73:761–771.
- Zhu K., Woodall C.W., Clark J.S. 2012. Failure to migrate: lack of tree range expansion in response to climate change. Global Change Biology 18:1042–1052.