

Robustness and uncertainty in estimates of butterfly abundance from transect counts

Kevin Gross · Eric J. Kalendra · Brian R. Hudgens · Nick M. Haddad

Received: 19 April 2006 / Accepted: 22 December 2006 / Published online: 22 February 2007
© The Society of Population Ecology and Springer 2007

Abstract Many butterfly populations are monitored by counting the number of butterflies observed while walking transects during the butterfly's flight season. Methods for estimating population abundance from these transect counts are appealing because they allow rare populations to be monitored without capture–recapture studies that could harm fragile individuals. An increasingly popular method for estimating abundance from transect counts relies on strong assumptions about the counting process and the processes that govern butterfly population dynamics. Here, we study the statistical performance of this method when underlying model assumptions are violated. We find that estimates of population size are robust to departures from underlying model assumptions, but that the uncertainty in these estimates (i.e., confidence intervals) is substantially underestimated. Alternative bootstrap and Bayesian methods provide better measures of the uncertainty in estimated population size, but are conditional upon knowledge of butterfly detectability. Because of these requirements, a mixed approach that combines data from small capture–recapture studies with transect counts strikes the best balance between accurate monitoring and minimal in-

jury to individuals. Our study is motivated by monitoring studies for St. Francis satyr (*Neonympha mitchelli francisci*), a rare and relatively immobile butterfly occurring only in the sandhills region of south-central North Carolina, USA.

Keywords Abundance · Bayesian statistics · Estimation · Parametric bootstrap · Population monitoring

Introduction

Butterfly populations are monitored extensively worldwide, both for their own sake and because they are considered indicator species that provide biological measures of ecosystem health (Thomas 2005). Most prominently, the British Butterfly Monitoring Scheme has monitored butterfly populations across the UK since 1976 (Pollard and Yates 1993). In many monitoring programs, one or more fixed transects are established that run through a butterfly's habitat, and observers walk these transects regularly during the butterfly's flight season while counting the number of individual butterflies observed. For butterflies with nonoverlapping generations, these transect counts produce a time series of count data over the course of a flight season.

A primary goal of transect counts is to measure changes in butterfly abundance across multiple generations. For fragile species such as butterflies, the ability to estimate population abundance from count data is desirable because transect counts do not risk harming individuals by capturing and/or marking them. Estimating abundance with transect counts is particularly

K. Gross · E. J. Kalendra
Department of Statistics, North Carolina State University,
Raleigh, NC, USA

B. R. Hudgens · N. M. Haddad
Department of Zoology, North Carolina State University,
Raleigh, NC, USA

K. Gross (✉)
Biomathematics Program, North Carolina State University,
CB 8203, Raleigh, NC 27695, USA
e-mail: gross@stat.ncsu.edu

appealing for rare or endangered butterfly species, where monitoring population trends is essential, and the cost of harming individuals is high (Singer and Wedlake 1981; Murphy 1988; Mattoni et al. 2001).

Several methods have been developed to estimate population abundance using transect counts. These methods do not estimate absolute population abundance, but instead estimate an index of abundance that is comparable across generations if monitoring protocols (transect placement, observer effort) and detectability of butterflies remain consistent. Historically, the most popular method of estimating abundance from transect counts has been to average counts during a week and to sum those weekly averages over the flight season, possibly standardizing for transect length, site size, and observer effort (Pollard et al. 1975; Pollard 1977; Thomas 1983; Pollard and Yates 1993). A second approach is to fit a smooth curve to count data and integrate the area under the curve (Rothery and Roy 2001). A third approach, distance sampling (Buckland et al. 2001), adjusts counts for differences in detectability (Brown and Boyce 1998; Boughton 2000); however, its implementation involves practical limitations in transect placement and distance measurement. A fourth approach, and the one that is the focus of this study, uses a population dynamics model to estimate an abundance index, the butterfly's mortality rate, the day of peak emergence, and the variance of the emergence times. This method was initially proposed by Manly (1974) and later modified by Zonneveld (1991) (henceforth MZ). The MZ approach is becoming increasingly popular because of its ability to estimate demographic rates, and because Zonneveld's version is implemented in the online computer program Insect Count Analyzer (<http://www.urbanwildlands.org/INCA>).

Because of its increasing use in monitoring programs (Mattoni et al. 2001; Zonneveld et al. 2003), the MZ model needs to be rigorously evaluated. In particular, the MZ model makes strong assumptions about both the processes governing butterfly population dynamics and the counting process. One goal of this paper is to study how the MZ method performs when its underlying assumptions are violated, and to determine which assumptions are most important for ensuring the reliability of population estimates. By doing so, we delineate some of the conditions under which count-based estimates of population size are trustworthy. In short, we find that estimates of population size are robust to departures from model assumptions, but that estimates of the uncertainty in these estimates (i.e., confidence intervals) are too small. A second goal is to describe modifications to the MZ model that provide more

accurate measures of the uncertainty in population estimates. The modifications that we describe, however, require a willingness to specify or assume the average detectability of butterflies. Because this information is best obtained from mark–recapture studies, our third goal is to show how data from multiple sources can be combined to strike a balance between accurate monitoring and reducing (though not eliminating) the need for mark–recapture studies.

Our study is motivated by our attempts to apply Zonneveld's (1991) version of the MZ model to modified Pollard–Yates transect counts for St. Francis satyr (SFS), an endangered butterfly occurring only in the sandhills region of south-central North Carolina, USA. We describe the SFS system in some detail first, and return to the SFS data near the conclusion of the paper to illustrate the methods we have developed.

Motivating example: St. Francis satyr (*Neonympha mitchelli francisci*)

St. Francis satyr (*Neonympha mitchelli francisci*) (Parshall and Kral 1989) is one of the rarest butterfly subspecies in the world; it is only known to occur on a small portion of Ft. Bragg Military Reservation in the sandhills region of North Carolina, USA. It is a wetland butterfly whose larval host plant is not known, but is likely a sedge in the genus *Carex*. SFS occurs as a metapopulation, a collection of distinct subpopulations between which movement by individuals is rare but not impossible. (In previous mark–recapture studies, only four marked individuals have been observed to move between subpopulations, out of several hundred individuals marked between 2002 and 2005.) SFS is bivoltine, with the first generation of adults flying from mid-May to mid-June, and the second generation flying from late July to late August.

SFS researchers have monitored all accessible subpopulations outside artillery impact areas for both flight periods since 2002 (Haddad et al. 2005). They conduct transect counts during the first flight period and capture–recapture studies during the second flight period. The unknown consequences of marking butterflies places a premium on count-based approaches, motivating this study. As we discuss later, capture–recapture studies provide valuable information about detectability and survivorship that can inform the techniques we develop.

SFS lives in mucky wetland meadows. Walking directly through the meadows damages vegetation, which could be detrimental to SFS. To minimize vegetation damage, transects were constructed by placing

5 × 15 cm wooden planks through strategically selected sites within each subpopulation that allowed views of all open butterfly habitat. Thus, for each subpopulation there was not a single contiguous transect, but instead there were a series of disconnected transect segments from which counts were pooled. Total transect length per subpopulation was standardized to approximately 300 m/ha. Because the transects were not placed randomly, and because it was difficult to measure accurate distances without trampling vegetation, distance sampling is not a viable sampling technique for this species.

Transect counts were conducted on most days of the flight period, except during rain. Counts were conducted during times of the day that were optimal for SFS activity (11 am–6 pm; Haddad et al. 2005). The duration of each transect count was standardized to habitat area, at the rate of 200 observer minutes per ha, with a minimum time of 40 observer minutes per subpopulation. This time period is long relative to typical transect counts, and was necessary because SFS are rare and immobile. This effort was typically divided among two observers to keep better track of multiple butterflies simultaneously.

Each flight season, counts were initiated a few days before cumulative degree day records predicted that the earliest adults would begin emerging. The data reported here were collected for four subpopulations during the first flight period in 2004; these subpopulations are designated A1, D1, D3, and E2 (Fig. 1). During each transect count, investigators counted the number of distinct adults observed, using morphological (e.g., wing wear or coloration) and behavioral differences among butterflies to determine gender and to avoid counting the same individual multiple times during a single walk. Transect counts were conducted throughout the flight period until no butterflies were observed in three consecutive counts.

Robustness studies of the MZ model

To lay the groundwork for what follows, we briefly restate Manly's (1974) and Zonneveld's (1991) models. Our construction differs slightly from the constructions originally presented, but the resulting models are the same.

Both Manly's and Zonneveld's models assume that adult butterfly emergence times are independent and identically distributed (iid) random variables drawn from a normal (Manly) or logistic (Zonneveld) probability distribution with unknown mean and variance. Both models assume that each butterfly lives

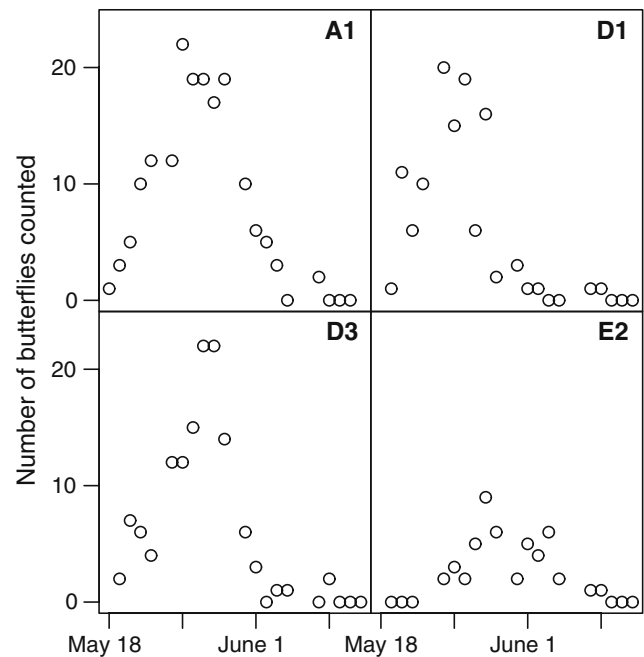


Fig. 1 Time series of transect counts for four subpopulations of St. Francis' satyr

for an exponentially distributed length of time, with mortality rate α . Combining these two assumptions yields an ordinary differential equation (ODE) for $x(t)$, the expected number of butterflies counted at time t

$$\frac{dx(t)}{dt} = N^* \times f(t; \mu, \sigma^2) - \alpha x(t) \quad (1)$$

where $f(t; \mu, \sigma^2)$ is the probability density function of the emergence distribution and N^* is the index of abundance. N^* equals the absolute abundance N (the total number of butterflies in the population) multiplied by the average number of times that a single butterfly is counted during a single transect count, c . Solving Eq. 1 gives an expression for $x(t)$ that is a function of four unknown parameters: N^* , μ , σ^2 , and α . Manly estimated these parameters by minimizing the sum of squared deviations between $x(t)$ and the data, while Zonneveld assumed that each transect count had a Poisson distribution with mean equal to $x(t)$ and maximized the resulting likelihood function. Zonneveld also found confidence intervals for each parameter by inverting the estimated information matrix, which is also based on the likelihood. Throughout the remainder of the paper, we work with Zonneveld's version of the MZ model.

Using Zonneveld's model to analyze our SFS data piqued our curiosity about two of its assumptions. First,

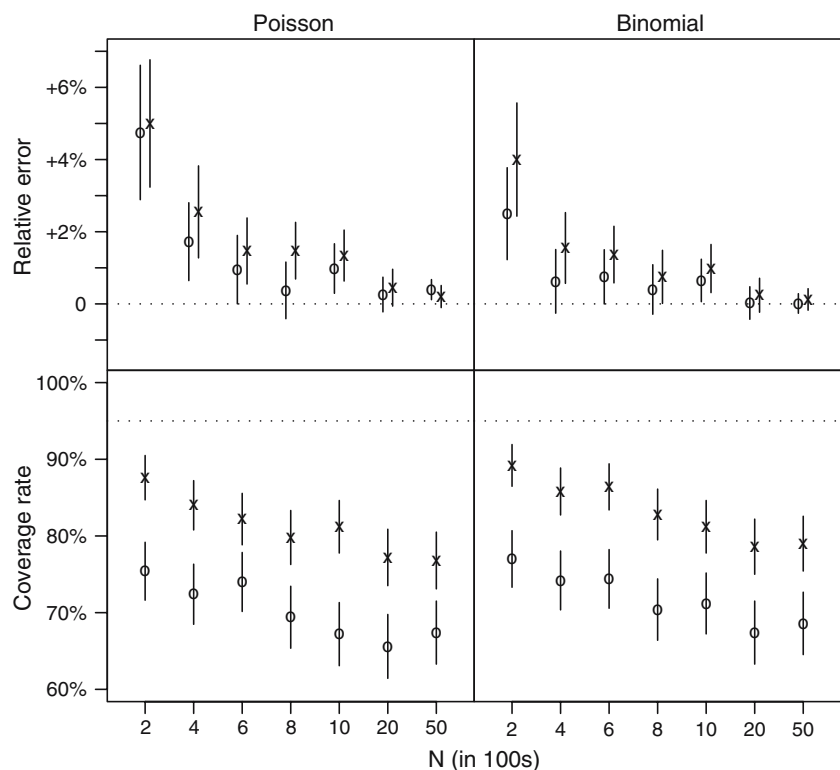
although the model is built on a probabilistic description of butterfly population dynamics (emergence times and lifetimes are assumed to be iid random variables), the model effectively converts a stochastic model of population dynamics into a deterministic one by working on the expected number of butterflies counted. Second, Zonneveld’s model assumes that the number of butterflies counted during a single transect walk has a Poisson probability distribution, which is reasonable when populations are large enough that observers are likely to count the same individual multiple times during a single transect walk. However, with small populations such as SFS, observers are reasonably confident that they can avoid counting the same individual multiple times during the same transect count. Under these conditions, a binomial distribution is a more reasonable model for the count data than a Poisson distribution.

Thus, our first two questions are: how reliable are the estimated indices of population abundance from Zonneveld’s model when population dynamics are stochastic, or when count data are binomially instead of Poisson distributed? We investigated these questions with a simulation study. The simulation varied three factors: the absolute population size, N (ranging from 200 to 5,000), the probability distribution for the transect counts (Poisson and binomial), and the average number of times that a single butterfly is counted

during a single transect walk, c ($c = 0.5$ and 1). The combination of a binomial probability distribution for transect counts and $c = 1$ is the limiting case where the transect counts are exactly equal to the number of butterflies alive on that particular day. For all simulations, we assumed that emergence times were independent and logistically distributed with mean $\mu = 10$ and $\sigma^2 = 7.4$ (this corresponds to a shape parameter of $\beta = 1.5$ in Zonneveld’s notation), and we assumed lifetimes were exponentially distributed with a mortality rate of $\alpha = 0.25 \text{ day}^{-1}$. For each parameter combination, we generated 500 data sets and fit Zonneveld’s model to each, using the inverse of the estimated information matrix to estimate nominal 95% confidence intervals (CIs) for N^* . As Fig. 2 shows, neither binomial counting nor stochastic population dynamics caused much of a bias in N^* , the estimated index of population abundance. On the other hand, coverage rates of 95% CIs were appreciably less than the nominal 95% coverage rate in all cases, ranging from 66 to 89%. Estimates of other model parameters behaved similarly (data not shown): biases were small ($\hat{\alpha}$ had a 3–6% positive bias when $N = 200$), but actual coverage rates of 95% CIs were too low, ranging from 56 to 89%.

This simulation suggests that Zonneveld’s method is surprisingly unaffected by whether count data have a Poisson or binomial distribution. This result can be

Fig. 2 Simulation study of \hat{N}^* from data sets with stochastic population dynamics and Poisson (left panels) or binomial (right panels) probability distributions for transect counts. Upper panels relative error, $(\hat{N}^* - N^*)/N^*$. Lower panels coverage rates of nominally 95% confidence intervals. Symbols (open circle $c = 1$; cross $c = 0.5$) and line segments indicate point estimates and 95% confidence intervals, respectively. $n = 500$ data sets for each parameter combination



explained by quasi-likelihood theory (Wedderburn 1974; McCullach and Nelder 1989). In a nutshell, if the number of times that any single butterfly is counted during a single transect walk is an iid random variable for all butterflies and all transect counts, then the variance of the transect counts will be linearly proportional to the mean. Quasi-likelihood shows that when the variance of the data is linearly proportional to the mean, point estimates of model parameters can be found by maximizing an objective function that is identical to a Poisson likelihood. Thus, we have the pleasing result that Zonneveld’s method has theoretical and empirical support for any probability distribution of counts per butterfly per transect walk, as long as the number of counts per butterfly per walk is iid for all butterflies and all walks. If counts per butterfly are not independent (for example, butterflies occurring in clusters or aggregations), the quasi-likelihood argument does not apply.

Unfortunately, however, stochastic population dynamics make the CIs for model parameters misleadingly narrow, regardless of the probability distribution for the data. We can confidently attribute the reduced CI coverage to stochastic population dynamics because other simulations (not shown here) with butterfly dynamics generated directly by Eq. 1 produced CIs with correct coverage rates. These results imply that ignoring stochasticity in butterfly dynamics makes it impossible to quantify the precision of \hat{N}^* , which is not a trivial concern. Below, we propose two modifications to Zonneveld’s method that yield accurate CIs.

Modification 1: parametric bootstrap for large populations

A parametric bootstrap that correctly adjusts CIs for \hat{N}^* is easy to implement, although it requires both a relatively large population and a willingness to specify butterfly detectability. Given c , the average number of times that a butterfly is counted during a single transect walk, an estimate of absolute abundance can be found by $\hat{N} = \hat{N}^*/c$. To generate a bootstrap data set, parameter estimates are used to simulate birth and

death times for each of \hat{N} butterflies. With this collection of birth and death times, a bootstrap data set can be produced by simulating transect counts with the same timing as the actual data. Fitting Zonneveld’s model to the bootstrap data produces bootstrap estimates of all model parameters. Repeating this procedure many times yields a distribution of bootstrap estimates for N^* , from which a CI can easily be calculated (Efron and Tibshirani 1993).

We conducted a simulation study to investigate the performance of this bootstrap. The simulation used five different combinations of N and probability distributions for transect counts. As in the first simulation study, $\alpha = 0.25$, $\mu = 10$, and $\sigma^2 = 7.4$ throughout. For each pair of N and probability distribution, we generated 100 data sets, and for each used 1,000 bootstrap data sets to calculate percentile-based bootstrap CIs for N^* . Table 1 compares coverage rates and widths of bootstrap CIs to CIs that assume deterministic dynamics (the standard errors shown are standard errors of estimation for the median CI width). Coverage rates of bootstrap CIs are much closer to the nominal 95% coverage rate than the CIs that assume deterministic dynamics. The bootstrap CIs are appreciably wider, reflecting the uncertainty in \hat{N}^* caused by stochastic dynamics. A bootstrap also successfully adjusts the CIs for α , μ , and σ^2 to properly quantify the uncertainty in each estimate (data not shown).

In theory, the bootstrap can be applied to all data sets. In practice, however, when the number of butterflies counted during transect walks is small (maximum count <30), there is an appreciable chance of producing a data set to which Zonneveld’s model cannot be fit. Zonneveld’s model fails with these data sets because the right tail of the time series is not sufficient to estimate the mortality rate α . The possibility of model failure dooms the bootstrap for small population sizes, because even if the actual data set produces reasonable parameter estimates, it is almost inevitable that an appreciable fraction of the bootstrap data sets will not yield parameter estimates. (Discarding the problematic data sets corrupts the bootstrap sample.) Thus, we propose a second, Bayesian method for analyzing data sets with small population sizes. The

Table 1 Observed coverage rates and median widths of 95% confidence intervals for N^* with stochastic population dynamics

N	Counting distribution	Coverage rates (%)		Median width (\pm SE)	
		Naive	Bootstrap	Naive	Bootstrap
200	Binomial ($c = 0.5$)	82	94	46.7 \pm 1.5	79.0 \pm 2.8
200	Binomial ($c = 0.8$)	77	89	56.0 \pm 1.5	109.9 \pm 4.3
200	Poisson ($c = 1$)	66	93	65.4 \pm 2.1	154.3 \pm 5.1
200	Poisson ($c = 2$)	58	96	88.9 \pm 2.2	291.3 \pm 9.9
400	Poisson ($c = 1$)	72	91	85.9 \pm 2.1	197.3 \pm 4.5

Bayesian model will work for any population size, but it requires prior information about both the butterfly's detectability and mortality rate (alternatively, the butterfly's average life span). We sketch the main ideas of the model below, and present a more detailed description in the [Appendix](#).

Modification 2: Bayesian model for small populations

Bayesian inference treats unknown parameters as random variables whose probability distributions reflect the uncertainty in those parameters. Prior and posterior distributions reflect uncertainty in parameters before and after data are observed, respectively. Priors are updated to posteriors via a likelihood function for the data, which in this case has three components: a probability model for emergence times, a probability model for lifetimes, and a probability model for transect counts. As in Zonneveld's model, we assume that emergence times are iid with mean μ and variance σ^2 . In keeping with the historical roots of the MZ model, we introduce a binary shape parameter ζ that controls whether the emergence distribution is normal or logistic. For technical reasons, we construct the Bayesian model in discrete time, which requires reparameterizing from a mortality rate, α , to a daily survival probability, φ . We assume that lifetimes have a geometric distribution, which is the discrete-time analog of an exponential distribution. As in the bootstrap, we assume that butterfly detectability is either known or assumed.

We performed a simulation study to investigate the performance of the Bayesian model. In the simulations, populations contained either $N = 100$ or 200 butterflies. Butterfly emergence times were normally distributed with mean $\mu = 10$ and variance $\sigma^2 = 7.4$. Daily butterfly survival was set to one of three values ($\varphi = 0.577, 0.669, \text{ or } 0.761$) corresponding to the 20th, 50th, and 80th percentiles of the prior distribution for φ . To add realism, we assumed that count data were beta-binomially distributed. A beta-binomial is an "overdispersed" version of a binomial distribution that is appropriate when detectability varies among transect counts. The degree of overdispersion is controlled by a parameter η ; if the number of butterflies counted during a single transect count is beta-binomially distributed with parameters n (the number of butterflies flying on that day), p (the average detectability) and η , then the expected number of butterflies counted is np and the variance in the number of butterflies counted is $np(1-p)[1+(n-1)\eta]$ (McCullagh and Nelder 1989). In this simulation, we set the average detectability

equals to either 30 or 35%, and η equal to 0, 0.05, or 0.1. All told, we simulated and fit 50 data sets for each of $2 \times 2 \times 3 \times 3 = 36$ different parameter combinations.

To fit the Bayesian model, we used the same non-informative prior distributions for N , μ , and σ^2 and the same informative prior for φ that we do for the SFS data (see below). The model fit assumed (usually incorrectly) that the count data were beta-binomially distributed with an average detectability of 35% and $\eta = 0.05$.

For each data set, we estimated posterior means, posterior medians, and 95% posterior intervals (the Bayesian equivalent of CIs) for N , μ , σ^2 , and φ . Table 2 reports the average posterior median for N , the standard error of the median (i.e., the standard deviation of the median for the 50 data sets), the average width of a 95% posterior interval for N , the standard error of the width of the posterior interval, and the coverage rate of the posterior interval. We used a posterior median instead of a posterior mean as a point estimate for N because posterior distributions were right-skewed, and thus the posterior median was slightly closer to the true value of N . Table 2 shows that posterior medians of N were biased high by about 5% when model assumptions matched actual values. Posterior medians tended to underestimate N when actual detectability was lower than assumed detectability, and under(over)estimated N when actual overdispersion was lower (higher) than assumed. Coverage of confidence intervals for N was near the nominal 95% coverage rate in all cases, suggesting that the Bayesian intervals accurately measure the uncertainty in estimated population size even under mild departures from model assumptions. Posterior inference for μ and σ^2 was adequate for all parameter combinations, while posterior inference for φ was influenced by the prior, as we would expect when the prior is informative (data not shown).

Application of the Bayesian model to SFS data

Because SFS population sizes were small, the Bayesian model was our best option for analyzing SFS data. We used noninformative priors for μ , σ^2 , and N to reflect ignorance of these parameter values (technical details in the [Appendix](#)). For the daily survival probability φ we used an informative prior based on capture–recapture studies for these same subpopulations from 2002 to 2003. We used a beta-binomial distribution to model the probability distribution of the transect counts, with average detectability for each subpopula-

Table 2 Simulation study results for the Bayesian modification of the MZ model

	$\varphi = 0.577$	$\varphi = 0.669$	$\varphi = 0.761$
(a) $N = 100$, average detectability = 30%			
$\eta = 0$			
\hat{N} (avg \pm SE)	75.2 \pm 10.5	89.7 \pm 14.7	97.0 \pm 21.8
CI width	81.2 \pm 8.7	98.2 \pm 12.6	105.6 \pm 26.0
CI coverage	49/50	49/50	50/50
$\eta = 0.05$			
\hat{N}	74.9 \pm 14.1	93.4 \pm 16.1	102.4 \pm 22.6
CI width	81.2 \pm 10.9	98.6 \pm 14.8	112.1 \pm 25.3
CI coverage	44/50	50/50	50/50
$\eta = 0.10$			
\hat{N}	78.0 \pm 19.9	98.8 \pm 21.6	106.8 \pm 26.0
CI width	81.4 \pm 16.2	100.4 \pm 18.0	114.9 \pm 27.4
CI coverage	44/50	49/50	48/50
(b) $N = 100$, average detectability = 35%			
$\eta = 0$			
\hat{N}	86.7 \pm 14.7	103.0 \pm 18.6	115.4 \pm 25.5
CI width	89.0 \pm 10.6	107.6 \pm 15.7	123.6 \pm 27.6
CI coverage	50/50	50/50	48/50
$\eta = 0.05$			
\hat{N}	91.3 \pm 17.8	105.8 \pm 23.1	123.4 \pm 24.1
CI width	92.0 \pm 13.4	107.2 \pm 17.8	124.7 \pm 24.8
CI coverage	49/50	50/50	47/50
$\eta = 0.10$			
\hat{N}	95.2 \pm 22.9	106.3 \pm 21.4	126.8 \pm 27.3
CI width	94.7 \pm 16.8	108.8 \pm 18.7	132.9 \pm 29.9
CI coverage	49/50	49/50	43/50
(c) $N = 200$, average detectability = 30%			
$\eta = 0$			
\hat{N}	152.0 \pm 19.8	184.8 \pm 25.0	183.4 \pm 33.1
CI width	143.0 \pm 15.2	170.5 \pm 21.1	169.4 \pm 39.3
CI coverage	46/50	49/50	49/50
$\eta = 0.05$			
\hat{N}	166.5 \pm 28.9	186.2 \pm 34.0	201.8 \pm 34.7
CI width	151.3 \pm 19.3	176.3 \pm 25.6	177.1 \pm 38.7
CI coverage	48/50	50/50	48/50
$\eta = 0.10$			
\hat{N}	169.1 \pm 38.4	188.4 \pm 44.3	209.8 \pm 58.9
CI width	146.2 \pm 27.0	174.4 \pm 29.9	188.2 \pm 52.5
CI coverage	45/50	47/50	43/50
(d) $N = 200$, average detectability = 35%			
$\eta = 0$			
\hat{N}	179.3 \pm 23.4	212.3 \pm 33.3	221.3 \pm 33.9
CI width	163.3 \pm 19.4	193.1 \pm 27.6	196.9 \pm 41.7
CI coverage	49/50	49/50	50/50
$\eta = 0.05$			
\hat{N}	185.5 \pm 38.7	209.4 \pm 38.3	231.0 \pm 55.9
CI width	165.0 \pm 26.3	185.8 \pm 33.7	207.6 \pm 55.5
CI coverage	48/50	49/50	40/50
$\eta = 0.10$			
\hat{N}	205.4 \pm 41.4	216.8 \pm 41.4	234.3 \pm 50.8
CI width	178.0 \pm 29.7	194.7 \pm 29.0	202.6 \pm 56.6
CI coverage	47/50	47/50	41/50

tion determined by each subpopulation’s average estimated detectability from 2002 to 2003 capture–recapture studies: 32, 49, 22, and 27% for subpopulations A1, D1, D3, and E2, respectively. The overdispersion parameter η was set to 0.05 based on observers’ knowledge of daily variability in detectability.

Posterior medians and intervals for all subpopulations and all parameters are listed in Table 3. Although the time series of counts are roughly similar for subpopulations A1, D1, and D3 (Fig. 1), the estimated size of subpopulation D1 is smaller because detectability in D1 is assumed to be higher. The estimated size of subpopulation E2 is smallest because fewer butterflies were observed during transect counts despite comparable detectability with subpopulations A1 and D3. Posterior distributions for the daily survival probability are similar to the prior, suggesting that the data contain little additional information about butterfly survival (this is the same lack of information that causes Zonneveld’s model to fail with small population sizes). The date of peak emergence was well determined for all four subpopulations.

Discussion

Count-based estimates of population size (or of indices of population size) are appealing because they permit population monitoring without mark–release–recapture studies that may have negative impacts on the health and longevity of fragile individuals. This is particularly true for rare or endangered species, where population monitoring is essential and the cost of harming individuals is high (Mattoni et al. 2001). While methods exist for estimating population abundance from count data, these methods necessarily rely heavily on assumptions, and it is important to know how trustworthy these count-based estimates are in the face of departures from the assumptions on which they are based. The contribution of this paper is to delineate some of the conditions under which count-based estimates of population size are reliable, and to propose two alternative methods that measure the statistical uncertainty in these estimates of population size more accurately.

Our results suggest that Zonneveld’s model accurately measures the uncertainty in the estimated abundance index only when butterfly dynamics are deterministic. Although one could imagine mechanisms that would produce deterministic dynamics (e.g., systematic and coordinated variation among butterflies in both emergence times and lifetimes), the assumption

Table 3 Posterior medians and 95% posterior intervals of parameters from a Bayesian model for four SFS subpopulations

	Subpopulation			
	A1	D1	D3	E2
N	216 (127, 358)	111 (70, 169)	255 (148, 425)	73 (34, 137)
Survival (φ)	0.645 (0.454, 0.781)	0.619 (0.431, 0.756)	0.637 (0.439, 0.769)	0.672 (0.462, 0.828)
μ^a	6.9 (5.2, 8.4)	4.5 (2.7, 5.8)	5.7 (3.6, 7.3)	9.5 (7.2, 11.5)
σ	3.4 (2.3, 4.9)	3.0 (1.9, 4.8)	3.4 (2.2, 5.3)	2.8 (1.3, 4.7)
Shape ^b (ξ)	0.52	0.49	0.46	0.54

^a Day 1 is May 18

^b Probability of a logistic emergence distribution

of stochastic dynamics through independent emergence and death times seems more biologically reasonable. Moreover, the effect of stochastic dynamics does not disappear for larger populations; indeed, the underestimate of the uncertainty in estimated population size is actually worse for larger populations (Fig. 2). Thus, in most cases accounting for stochastic dynamics is essential for accurately measuring the precision of population estimates.

The two modifications we have described that correct for stochastic dynamics—a parametric bootstrap and a Bayesian model—both require a willingness to specify or assume the detectability of individuals. The Bayesian model also requires some prior information about butterfly survival. Without a willingness to specify or assume detectability (and survival for small populations), there is no good way to measure the statistical uncertainty in the estimated population size. In a sense, this is the “cost” of this type of count-based analysis. Count-based analyses offer the benefit of being able to estimate population abundance from count data alone, but the cost is that measures of the statistical uncertainty in these estimates depend on the validity of the specified detectability. Whether or not these costs will outweigh the benefits will depend on factors specific to the situation at hand. Under some circumstances, a preferable approach may be to use covariates to model differences in detectability (K. Pollock, personal communication).

For our SFS studies, we have used estimates of detectability and survival from mark–capture–recapture studies conducted in previous flight seasons. We speculate that combining data from multiple sources may ultimately strike the best balance between minimizing harm to fragile individuals and accurate population monitoring (Pollock et al. 2002). Although mark–capture–recapture studies are not avoided altogether in this mixed approach, they can be smaller and less frequent than they would be if mark–capture–recapture studies were used exclusively, thus reducing (if

not eliminating) the fraction of individuals that are captured and/or marked. In any case, Bayesian statistics seems particularly well suited to this task of combining data from multiple sources.

Acknowledgments We thank Justin Calabrese, Chris Damiani, Sujit Ghosh, Travis Longcore, Bryan Manly, Ken Pollock, Leslie Ries, Cor Zonneveld, and two anonymous reviewers for helpful comments and discussion. This work was funded by Fort Bragg’s Endangered Species Branch and Natural Resources Division and the National Science Foundation, with administrative support provided by the North Carolina Cooperative Fish and Wildlife Research Unit.

Appendix: technical description of Bayesian model

Let $d = 1, \dots, D$ be a daily time index, where days 1 and D are the days of the first and last transect counts, respectively. Let $X(d)$ be the number of butterflies that emerge on day d , let $Y(d)$ be the number of butterflies alive on day d , and let $Z(d)$ (the data) be the number of butterflies counted on day d if a transect count was conducted. To make the notation more compact, we use the vector notation $\mathbf{X} = [X(1), \dots, X(D)]$, $\mathbf{Y} = [Y(1), \dots, Y(D)]$, and $\mathbf{Z} = [Z(t_1), \dots, Z(t_n)]$, where t_1, \dots, t_n are the days of the n transect counts. In keeping with standard Bayesian practice, we parameterize the variance of the emergence distribution by its inverse $\tau = \sigma^{-2}$. Also, let ξ be a shape parameter that indicates whether the emergence distribution is normal ($\xi = 0$) or logistic ($\xi = 1$). Let $\theta = (N, \mu, \tau, \varphi, \xi)$ be a vector that contains the unknown parameters.

Let $\pi(\cdot)$ denote a probability distribution. Our goal is to estimate $\pi(\theta|\mathbf{Z})$, the posterior distribution of the unknown parameters given the data. This marginal posterior distribution can be found by integrating \mathbf{X} and \mathbf{Y} out of the full posterior $\pi(\theta, \mathbf{X}, \mathbf{Y}|\mathbf{Z})$. The full posterior is proportional to the product of likelihood of the data and the prior distribution

$$\pi(\theta, \mathbf{X}, \mathbf{Y}|\mathbf{Z}) \propto \pi(\mathbf{Z}|\theta, \mathbf{X}, \mathbf{Y})\pi(\theta, \mathbf{X}, \mathbf{Y}). \quad (2)$$

The terms on the right-hand side of Eq. 2 can be simplified by observing that the distribution of the data \mathbf{Z} depends only on \mathbf{Y} , and by factoring the prior:

$$\pi(\boldsymbol{\theta}, \mathbf{X}, \mathbf{Y}|\mathbf{Z}) \propto \pi(\mathbf{Z}|\mathbf{Y})\pi(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{3}$$

The first term on the right of Eq. 3 is the probability distribution of the data given the actual (unobserved) dynamics, and can be written as

$$\pi(\mathbf{Z}|\mathbf{Y}) = \prod_{i=1}^n \pi[Z(t_i)|Y(t_i)] \tag{4}$$

where $\pi[Z(t_i)|Y(t_i)]$ is the appropriate probability model for the transect counts. For SFS, we use a beta-binomial distribution with a fixed overdispersion parameter η .

The second term on the right of Eq. 3 is the probability distribution of the actual butterfly dynamics given unobserved parameters, and comes directly from a modified version of Zonneveld’s model. First, the vector of emergence times \mathbf{X} has a multinomial distribution with parameters N and p_1, \dots, p_D , where p_d is the probability of emergence on day d . The emergence probabilities p_1, \dots, p_D are determined by discretizing the underlying continuous distribution for emergence times,

$$p_d = \frac{\int_{d-1}^d f(t; \mu, \tau) dt}{\int_0^D f(t; \mu, \tau) dt}, \quad d = 1, \dots, D. \tag{5}$$

Second, the number of butterflies flying on day d , $Y(d)$, is the sum of the butterflies that emerged on day d , $X(d)$, plus those butterflies that were flying on day $d-1$ and survived. In equations,

$$\begin{aligned} Y(1) &= X(1), \\ Y(d) &= X(d) + \text{binomial}[X(d-1), \varphi] \quad d = 2, \dots, D. \end{aligned} \tag{6}$$

For each SFS time series, we used the following proper noninformative priors for μ , τ , and N

$$\begin{aligned} \mu &\sim \text{Unif}(1, D), \\ \tau &\sim \text{Exp}(1), \\ N &\sim \text{Unif}(1, N_0) \end{aligned}$$

where N_0 is a relatively large value (5,000). ξ had equal prior probability of indicating either a logistic or normal distribution. The prior for the daily survival probability φ was based on capture–recapture studies for the same subpopulations in 2002 and 2003

$$\pi(\varphi) \propto N(\mu = 0.669, \sigma = 0.109) \times I_{(0,1)}.$$

We approximated the posterior distribution numerically using the Markov chain Monte Carlo (MCMC) algorithm described below. We used separate Metropolis–Hastings updating steps for the parameters μ , τ , and ξ , as well as the parameter blocks $(\varphi, N, \mathbf{X}, \mathbf{Y})$, $(N, \mathbf{X}, \mathbf{Y})$, (\mathbf{X}, \mathbf{Y}) , and \mathbf{Y} . N , \mathbf{X} , and \mathbf{Y} were updated as blocks because of the conditional dependence structure of the model, while φ was updated in a block because of its strong posterior correlation with N . Updates of the single parameters used standard proposal distributions (folded normals for μ and τ ; because ξ is a binary variable, its proposal was whatever the current value was not), but block updates used non-standard proposals, and so we describe these here. In what follows, we use primes ($'$) to denote proposed parameter values. In all cases, substantial cancellation follows from the model factorization (Eq. 3), which eases the calculation of the acceptance ratio considerably.

To update the block $(N, \mathbf{X}, \mathbf{Y})$, we first flip a fair coin to decide whether to add ($N' = N + 1$) or subtract ($N' = N - 1$) an individual, and then roll a fair D -sided die to determine the day on which an individual is added or subtracted. We add or subtract the individual from both \mathbf{X} and \mathbf{Y} . To keep the proposal density symmetric, moves to regions of zero prior probability [e.g., $X'(d) = -1$] are allowed, but are automatically rejected because the acceptance ratio is zero. To update the block $(\varphi, N, \mathbf{X}, \mathbf{Y})$, we propose a value φ' separately, but evaluate the entire proposed move to $\varphi', N', \mathbf{X}'$, and \mathbf{Y}' at once.

We update the block (\mathbf{X}, \mathbf{Y}) by moving a randomly selected individual’s emergence time forward or backward by a day. To do so, we first choose the day from which an individual is moved, d^* , with probability $P(d^* = d) = X(d)/N$. If $2 \leq d^* \leq N - 1$, we flip a fair coin to determine whether to move that individual’s emergence time forward or backward by a day. If $d^* = 1$ or $d^* = N$, we move the individual’s emergence time ahead or back by a day, respectively. For example, if we choose to move an individual’s emergence time one day later, then $X'(d^*) = X(d^*) - 1$, $X'(d^* + 1) = X(d^* + 1) + 1$, $Y'(d^*) = Y(d^*) - 1$, and $Y'(d^* + 1) = Y(d^* + 1) + 1$. Not all moves of this type result in a legal \mathbf{X} and \mathbf{Y} ; such moves are considered but are automatically rejected for lack of prior support. The proposal distribution is not symmetric, and must be factored into the acceptance ratio.

To update \mathbf{Y} , d^* is selected at random from $2, \dots, N$ (each with equal probability), and a fair coin is flipped

to determine whether to increase or decrease $Y(d^*)$ by 1. Again, not all possible moves result in legal combinations of \mathbf{Y} and \mathbf{X} , and illegal moves are considered but are automatically rejected for lack of prior support. The proposal distribution is symmetric.

Ten separate chains were run from different starting values, and convergence was monitored by Gelman and Rubin's (1992) potential scale reduction factor. For the fits to SFS data, Monte Carlo standard errors of posterior medians were less than 2.5% of posterior standard errors in all cases.

References

- Boughton DA (2000) The dispersal system of a butterfly: a test of source–sink theory suggests the intermediate-scale hypothesis. *Am Nat* 156:131–144
- Brown JA, Boyce MS (1998) Line transect sampling of Karner blue butterflies (*Lycaeides melissa samuelis*). *Env Ecol Stat* 5:81–91
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling—estimating abundance of biological populations. Oxford University Press, Oxford
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, London
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–511
- Haddad NM, Kuefler D, Hudgens BR (2005) Monitoring protocol for St. Francis' satyr: final report for the 2004 field season. Report to the Department of Defense. NC State University, Raleigh, NC
- Manly BFJ (1974) Estimation of stage-specific survival rates and other parameters for insect populations developing through several stages. *Oecologia* 15:277–285
- Mattoni R, Longcore T, Zonneveld C, Novotny V (2001) Analysis of transect counts to monitor population size in endangered insects. *J Insect Conserv* 5:197–206
- McCullach P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall, London
- Murphy DD (1988) Are we studying our endangered butterflies to death? *J Res Lep* 26:236–239
- Parshall DK, Kral TW (1989) A new subspecies of *Neonympha mitchelli* (French) (Satyridae) from North Carolina. *J Lep Soc* 43:114–119
- Pollard E (1977) A method for assessing change in the abundance of butterflies. *Biol Conserv* 12:115–132
- Pollard E, Yates TJ (1993) Monitoring butterflies for ecology and conservation. Chapman & Hall, London
- Pollard E, Elias DO, Skelton MJ, Thomas JA (1975) A method of assessing the abundance of butterflies in Monks wood national nature reserve in 1973. *Entomol Gaz* 26:79–88
- Pollock KH, Nichols JD, Simons TR, Farnsworth GL, Bailey LL, Sauer JR (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119
- Rothery P, Roy DB (2001) Application of generalized additive models to butterfly transect count data. *J Appl Stat* 28:897–909
- Singer MC, Wedlake P (1981) Capture does affect probability of recapture in a butterfly species. *Ecol Entomol* 6:215–216
- Thomas JA (1983) A quick method for estimating butterfly numbers during surveys. *Biol Conserv* 27:195–211
- Thomas JA (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philos Trans R Soc B* 360:357
- Wedderburn RWM (1974) Quasi-likelihood functions, generalized linear models and the Gauss–Newton method. *Biometrika* 61:439–447
- Zonneveld C (1991) Estimating death rates from transect counts. *Ecol Entomol* 16:115–121
- Zonneveld C, Longcore T, Mulder C (2003) Optimal schemes to detect the presence of insect species. *Conserv Biol* 17:476–487