Popul Ecol
DOI 10.1007/s10144-012-0316-7

ORIGINAL ARTICLE

How emergence and death assumptions affect count-based estimates of butterfly abundance and lifespan

Justin M. Calabrese

Received: 26 October 2011 / Accepted: 14 March 2012

Abstract Transect count data form the basis of many butterfly and other insect monitoring programs worldwide. A clear understanding of the limitations of such datasets, including the potential for biases in the statistical methods used to analyze them, is therefore crucial. The classical Zonneveld model (CZ) can extract estimates of a suite of demographic parameters from transect count datasets, and has also been used in theoretical analyses of protandry and reproductive asynchrony. The CZ relies on strong assumptions about the emergence and death processes underlying observed transect count datasets. Though reasonable as a starting place, a growing body of empirical evidence suggests these assumptions will, in many cases, not hold. Here, I explore how violations of these assumptions bias CZ-based estimates of two key population parameters: total population size and mean individual lifespan. To do this, I generalize the Zonneveld model by relaxing the symmetrical emergence distribution and constant death rate assumptions such that the generalized models contain the CZ as a special case. Using the generalized models as data generating processes, I then show that the CZ is able to closely mimic the shape of the abundance time course produced by either variant of the generalized model under a wide range of conditions, but produces highly biased estimates of population size and mean lifespan in doing so. My analysis therefore demonstrates both that the CZ is not robust to violations of its emergence and death assumptions, and that a good observed fit to transect count data does not mean these assumptions are satisfied.

Keywords Butterflies · Insect count analyzer · Monitoring · Phenology · Transect count data · Zonneveld model

J. M. Calabrese (✉)
Conservation Ecology Center, Smithsonian Conservation Biology Institute, National Zoological Park,
1500 Remount Rd., Front Royal, VA 22630, USA
e-mail: CalabreseJ@si.edu

J. M. Calabrese
Helmholtz Centre for Environmental Research-UFZ,
04318 Leipzig, Germany

Introduction

Butterflies are one of the most intensively studied taxa world-wide and are valuable indicators of biodiversity (Sisk et al. 1994), ecosystem health (Bouyer et al. 2007), and impacts of climate change (Parmesan and Yohe 2003; Parmesan 2007). As a result, efforts to systematically monitor butterfly populations have increased dramatically in recent decades (Thomas 2005; van Swaay et al. 2008). The bulk of these monitoring programs are based on transect counts of adult butterflies repeated at intervals throughout the flight season (Pollard 1977; Thomas 2005). The accumulating stockpile of transect count datasets represents a tremendous resource for understanding butterfly phenology (Roy and Sparks 2000; Parmesan 2007) and for detecting trends in population status over time(Roy et al. 2001; Warren et al. 2001; Crone et al. 2007). Most statistical methods for analyzing these datasets have focused on estimating an index of population abundance (Pollard 1977; Pollard and Yates 1993; Rothery and Roy 2001), but the sheer volume of such datasets warrants a deeper exploration of the potential for transect count data to reveal finer details about butterfly demography and phenology.

Building on earlier work by Manly (1974), Zonneveld (1991) modeled the adult butterfly abundance time series

Published online: 30 March 2012

produced by transect count surveys as a function of the emergence of new adults into the population and loss due to death. He coupled this basic emergence and death model with a simple Poisson sampling error assumption to yield a maximum likelihood approach to estimating the size, death rate, and emergence parameters of a focal population from observed transect count data. A number of studies have explored the statistical behavior of the Zonneveld model (Mattoni et al. 2001; Gross et al. 2007; Haddad et al. 2008), and the freely available INsect Count Analyzer (INCA, http://www.urbanwildlands.org/INCA/) makes the method accessible to a broad audience. Zonneveld's method is appealing because it potentially allows a suite of biologically meaningful demographic parameters, including an index of population size, to be estimated from simple transect count data. Elaborations of this basic model have also been used in theoretical studies of protandry and reproductive asynchrony (Zonneveld and Metz 1991; Zonneveld 1992, 1996a, b; Calabrese and Fagan 2004; Calabrese et al. 2008; Fagan et al. 2010).

Zonneveld's approach necessarily relies on strong assumptions about the functional forms of the emergence and death processes. His assumptions of logistically distributed (symmetrical) emergence events and exponentially distributed lifespans (constant death rate) strike a nice balance between biological realism and mathematical tractability, and are likely to be broadly applicable. However, it seems unlikely that these assumptions will always hold. For example, several butterfly phenology studies demonstrate emergence patterns that can be strongly right or left skewed (Brakefield 1982; Iwasa et al. 1983; Sims and Shapiro 1983; Xue et al. 1997). Other studies, often relying on effort intensive mark–recapture methods, have shown evidence of non-constant death rates over time (Schtickzelle et al. 2002; Auckland et al. 2004), an increasing death rate with number of matings (Kawagoe et al. 2001), or a death rate that increases with individual age (Brakefield 1982; Lederhouse 1983; Ban et al. 1990; Cushman et al. 1994; Zheng et al. 2007). The effects of these types of violations of the emergence and death assumptions of the classical Zonneveld model (CZ) on the bias in population size and mean lifespan estimates have not been studied.

Here, I quantify how CZ-based parameter estimates of population size and mean lifespan degrade as the model's emergence and lifespan assumptions are violated to increasing degrees. To do this, I first generalize the Zonneveld model to incorporate more flexible assumptions about the emergence and death processes. I then use the emergence-generalized and lifespan-generalized Zonneveld models (hereafter EZ and LZ, respectively) as tools to explore how CZ-based parameter estimates respond to violations of these two core assumptions. It is important to note that the goal of this paper is not to perform a side-by-side comparison of the CZ, EZ, and LZ on the same data. The generalized models each contain an additional parameter that will be very difficult to estimate from transect count data, and thus they will be of little practical use for data analysis. I will address this issue in greater depth in the discussion. Instead, the EZ and LZ are considered data generating processes that allow the effects of asymmetric emergence distributions (EZ) and non-constant death rates (LZ) on the quality of CZ-based parameter estimates to be directly assessed.

To facilitate my analysis, I describe an alternative derivation of the CZ that leads to a novel technique for rapidly fitting the CZ directly to both of the generalized models. This technique allows a thorough exploration of the behavior of the CZ across a wide range of emergence and lifespan distribution shapes. Focusing on empirically based scenarios, I demonstrate that: (1) the CZ can very often closely approximate the generalized models even when emergence distributions are highly asymmetric (EZ) or when the death rate varies substantially over time (LZ), and (2) doing so results in consistently and often strongly biased CZ-based estimates of mean lifespan and total population size. These findings therefore suggest that the CZ is not robust to violations of its emergence and death rate assumptions, and that using it in situations where these assumptions do not hold can lead to severely biased parameter estimates. More insidiously, these results imply that a good observed fit of the CZ to a dataset cannot be taken as evidence that its underlying assumptions are met.

Methods

The generalized Zonneveld model

Assuming no net change in the population due to immigration or emigration, the within-season dynamics of adult population size can be written as

$$\frac{dx}{dt} = Nf_E(t) - M(t)x \tag{1}$$

where t is time, x is the adult population size at time t, N is the total number of adults that emerge during the flight period (hereafter, population size), $f_E(t)$ is a probability density function (PDF) specifying the emergence schedule, and $M(t)$ is the average mortality rate in the population at time t. Zonneveld chose the classical logistic distribution for $f_E(t)$ and a constant death rate, α, for $M(t)$. For consistency, I have designed my notation such that in the special case of the CZ it is identical to Zonneveld's (1991) notation. Here, I extend his model in two ways. First, I generalize the emergence component by specifying a three

parameter generalized logistic distribution (Wu et al. 2000) for the emergence distribution, with PDF

$$f_E(t;\mu,\beta,\delta) = \frac{\delta e^{(t-\mu)/\beta}}{\beta(1+e^{(t-\mu)/\beta})^{\delta+1}} \quad (2)$$

where μ is the location parameter, β is the scale parameter, and δ is a shape parameter affecting the skewness of the distribution. The classical logistic distribution is recovered when $\delta = 1$, the distribution is (slightly) left-skewed for $\delta > 1$, and is right-skewed for $\delta < 1$ (Fig. 1).

The second generalization is to allow an individual's mortality rate to depend on its age (senescence) such that lifespans have a Weibull distribution. The Weibull distribution is a standard tool in survival analysis and is widely used to model age-dependent individual death or failure rates (Johnson et al. 1994). As it contains the exponential as a special case, it provides a natural way to generalize the lifespan assumptions of the Zonneveld model. The Weibull lifespan distribution has PDF (Johnson et al. 1994)

$$f_L(l;\rho,\alpha) = \rho\alpha(\alpha l)^{\rho-1}e^{-(\alpha l)^{\rho}}$$

where ρ is a shape parameter and α is a scale parameter. When $\rho = 1$ the Weibull reduces to the exponential lifespan distribution implied by the constant death rate, α, in the CZ (Fig. 1). For $\rho > 1$, an individual's death rate increases with its age while $\rho < 1$ implies a death rate that decreases with age (Fig. 1).

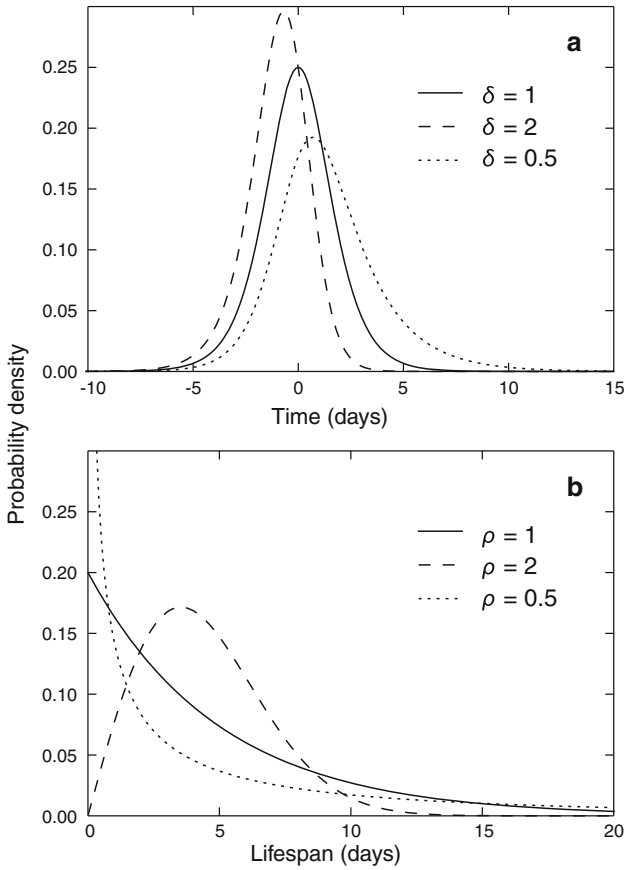

Fig. 1 The generalized logistic emergence distribution (**a**) and Weibull lifespan distribution (**b**). The generalized logistic reduces to the classical logistic when $\delta = 1$, is slightly left skewed for $\delta > 1$, and can have pronounced right skew for $\delta < 1$ (**a**). For all three curves in **a**, $\mu = 0$ and $\beta = 1$. The Weibull distribution simplifies to the exponential when $\rho = 1$, has an interior mode and shorter right tail when $\rho > 1$, and has a longer right tail when $\rho < 1$ (**b**). The $\rho > 1$ case corresponds to senescence, where the death rate increases with an individual's age. For all three scenarios in **b**, $\alpha = 0.2$. The Weibull distribution is also capable of producing a slightly left skewed shape for ρ large (not shown)

Equation (1) incorporates the loss of individuals from the active adult population as a (potentially) time varying death rate, and I now derive a time-dependent death rate function that produces Weibull-distributed lifespans. The hazard and survivorship functions of the Weibull distribution are required to do this. The hazard function, which specifies the relationship between death rate and individual age, a, is given by Evans et al. (2000)

$$H(a;\rho,\alpha) = \rho\alpha^{\rho}a^{\rho-1}.$$

The Weibull survivorship function gives the probability of an individual surviving to age a and is written (Evans et al. 2000)

$$S(a;\rho,\alpha) = e^{-(\alpha a)^{\rho}}.$$

The probability of an individual being a days old on day t of the season is $f_E(t-a;\mu,\beta,\delta)\,S(a;\rho,\alpha)$. Normalizing this quantity such that, for any time t, the integral over all values of a is one, the time-dependent age distribution is (Zonneveld 1992; Calabrese et al. 2008)

$$f_A(a,t;\Theta,\Phi) = \frac{f_E(t-a;\Theta)S(a;\Phi)}{\int_{-\infty}^{t} f_E(y;\Theta)S(t-y;\Phi)dy}$$

where $\Theta = \{\mu,\beta,\delta\}$, and $\Phi = \{\rho,\alpha\}$ are parameter vectors related to emergence and lifespan, respectively. The time-dependent age distribution, together with the Weibull hazard function, can now be used to obtain the average death rate in the population at any time t:

$$M(t;\Theta,\Phi) = \int_{-\infty}^{t} H(t-a;\Phi)f_A(t-a,t;\Theta,\Phi)da. \quad (3)$$

Notice that the average death rate depends on both the parameters of the lifespan distribution, Φ, as well as those of the emergence distribution, Θ. Substituting Eqs. (2) and (3) into Eq. (1), I obtain the generalized Zonneveld model, which can be solved numerically subject to initial condition

$$\lim_{t\to-\infty} x(t) = 0$$

to yield $x(t)$, the time course in abundance.

For tractability, I consider three special cases of the fully generalized model: (1) the CZ model ($\delta = 1$ and $\rho = 1$); (2) the emergence-generalized Zonneveld model (EZ, $\delta \neq 1$, $\rho = 1$); and (3) the lifespan-generalized Zonneveld model (LZ, $\delta = 1$, $\rho \neq 1$). The subscripts cz, ez, and lz are used to denote that a quantity comes from the classical Zonneveld model, the emergence-generalized model, or the lifespan-generalized model, respectively. For example, $x_{ez}(t)$ refers to the abundance time course from the emergence-generalized model. I use the subscript gz to denote a quantity that could come from either the EZ or the LZ.

Fitting the CZ to the data generating processes

The generalized models each relax one of the CZ's assumptions (emergence or death). Varying the additional shape parameter of each generalized model (δ or ρ) away from one thus represents an increasing degree to which the more restrictive corresponding assumption in the CZ is violated. Fitting the CZ to each generalized model under these conditions will then demonstrate how the assumption that has been relaxed affects: (1) the CZ's ability to fit an abundance time course generated by the generalized model, and (2) the quality of parameter estimates the CZ produces in accommodating a time course of that shape. I am particularly interested here in the how biased the estimates of population size and mean lifespan become when the CZ is fit to abundance time courses produced by the generalized models, and am not focusing on sampling error assumptions, small sample size performance, or issues of observability (for those aspects of the CZ, see Gross et al. 2007; Haddad et al. 2008).

To assess the performance of a statistical procedure, one would typically draw many random samples from a data generating process (here the EZ or LZ) with known parameters, fit the focal model (here the CZ) to each realization thus obtained, average the estimated parameters across realizations, and compare those averages to the "true" parameter values that generated the data. Fitting the CZ to transect count data via maximum likelihood involves a numerical search algorithm that repeatedly evaluates a computationally expensive likelihood function (Zonneveld 1991). It is therefore a slow process, and the number of separate fits that would be required to adequately assess the robustness of the CZ to violations of its emergence and death assumptions is prohibitively large. To work around this issue, I develop an alternative technique that allows the CZ to be fit directly to the data generating model (the EZ or LZ) without having to go through the intermediate steps of generating many random datasets, fitting the CZ to each, and then averaging across realizations. Though non-standard, this approach is extremely efficient and allows a very broad range of EZ and LZ scenarios to be considered.

To develop a method to rapidly fit the CZ to the generalized models, I exploit an alternative view of these models. The area under the curve of the CZ is N/α (Zonneveld 1991), where I have assumed perfect observability (see Gross et al. 2007, for an exploration of the issue of observability in the CZ). More generally, when the death rate is not constant, the area under the curve of a Zonneveld-type model is $N\langle l\rangle$, where $\langle l\rangle$ is the mean lifespan. In the CZ and EZ, $\langle l\rangle = 1/\alpha$. For the LZ presented here, the area under the curve is

$$\mathrm{AUC}_{lz} = \frac{N\Gamma(1 + 1/\rho)}{\alpha} \tag{4}$$

where $\Gamma(\bullet)$ is the Gamma function. Equation (4) reduces to N/α when $\rho = 1$. By multiplying $x(t)$ by the reciprocal of the area under the curve, any Zonneveld-type model can be made to integrate to one across its domain, and thus behave like a PDF. In Appendix A, I show that the abundance time course of the normalized form of the CZ (EZ) can be derived as the distribution of the sum of exponential and logistic (generalized logistic) random variables.

In Appendix B, I use this alternative view to obtain analytical expressions for the first three cumulants ($\kappa^{(1)}$ = mean, $\kappa^{(2)}$ = variance, and $\kappa^{(3)}$ = third central moment) of the abundance time courses of the normalized CZ and the normalized EZ as functions of their parameters. The cumulants of the CZ can then be used to obtain the following estimators of its parameters via the method of moments (Clark 2007) (Appendix C):

$$\begin{aligned}
\hat{\mu}_{cz} &= \kappa_{gz}^{(1)} - \left(\frac{\kappa_{gz}^{(3)}}{2}\right)^{1/3} \\
\hat{\beta}_{cz} &= \frac{\sqrt{3}}{\pi}\sqrt{\kappa_{gz}^{(2)} - \left(\frac{\kappa_{gz}^{(3)}}{2}\right)^{2/3}} \\
\hat{\alpha}_{cz} &= \left(\frac{2}{\kappa_{gz}^{(3)}}\right)^{1/3}.
\end{aligned} \tag{5}$$

As the population size, N, does not affect the shape of the abundance time course, it does not enter into the moment estimators. Given $\hat{\alpha}_{cz}$, the corresponding population size estimate can be calculated by forcing the area under the curve of the fitted CZ to equal that of the variant of the generalized model to which it is fit:

$$\hat{N}_{cz} = N_{gz}\langle l_{gz}\rangle\hat{\alpha}_{cz}.$$

As analytical expressions for the cumulants of the EZ are also tractable (Appendix C), it is possible to solve directly for the parameters that allow the CZ to most closely mimic the EZ by substituting the expressions for the cumulants of the EZ into the right hand side of Eqs. (5) yielding Eqs. (14). The cumulants of the LZ can easily be obtained by

numerical integration, and these values can then be substituted into Eqs. (5) to get the best-fitting CZ parameters (Appendix C).

The moment estimators provide a rapid way to obtain, for any choice of parameters of a more complicated generalized model, the corresponding set of CZ parameters that provide the best fit to it. Due to its efficiency, this approach facilitates a thorough exploration of the behavior of the CZ's mean lifespan and population size estimates over a very broad range of conditions in the data generating processes. Notice that the moment estimators are used here merely as a computational device to facilitate my analysis. They are not an alternative to the maximum likelihood approach described by Zonneveld (1991) and implemented in INCA for estimating CZ parameters from real transect count data.

Empirically based scenarios

Each scenario I consider features a base set of parameter values (N, α, μ, and β) for the data generating process (EZ or LZ). The additional shape parameter each generalized model introduces is set to one initially, in which case both generalized models reduce to the CZ. Fitting the CZ to either the EZ or LZ in this situation simply returns the base parameter set as the models are identical. I then systematically varied the value of either δ (EZ) or ρ (LZ) away from one in both directions, fit the CZ at every step using the moment estimators, and calculated the percent error in the CZ estimates of mean lifespan and population size relative to the true values of these quantities in the data generating process. This allowed me to quantify how the percent error in the CZ estimates of N and $\langle l \rangle$ changed as a function of the additional shape parameter of each data generating process, and thus as a function of increasing violations of the CZ's emergence or lifespan assumptions. To base my analysis on biologically reasonable parameter values, I used two published sets of empirically estimated values of the parameters N, α, μ, and β for the butterflies *Coenonympha pamphilus* and *Aricia agestis* (Zonneveld 1991), as base parameter sets for the data generating processes. The two scenarios represent broadly different estimated population sizes ($\hat{N}$) and mean lifespans ($= 1/\hat{\alpha}$).

As the method of moments approach I used to fit the CZ to the data generating processes is non-standard, I used INCA to spot check the percent error results with computationally intensive (slow) large-sample maximum likelihood fits of the CZ to each variant of the generalized model for particular values of δ and ρ (Appendix D). To determine if the quality of the fit of the CZ could be used to judge if its assumptions were met, I also examined the ability of the CZ to mimic the shape of the generalized models across the ranges of δ and ρ considered in the percent error analysis. Finally, I searched over a much broader range of parameter values to identify abundance time course shapes produced by the EZ and LZ that the CZ was not able to mimic.

Results

For both empirical scenarios the percent error in CZ estimates of N and $\langle l \rangle$ increased rapidly as either δ (EZ) or ρ (LZ) were varied away from unity (Fig. 2). The error could reach several hundred percent above (positive) or below (negative) the true value of the parameter being estimated. The results for the bias in N and $\langle l \rangle$ incurred by fitting the CZ to the LZ were very similar for both parameter sets (Fig. 2b, d), whereas they differed sharply between the two scenarios when EZ was the data generating process (Fig. 2a, c).

The difference between the two EZ-based scenarios is due to the left skewness the generalized logistic emergence distribution introduces into the abundance time course. For the *A. agestis* scenario in Fig. 2c, the death rate is high ($\alpha = 0.362$), lifespans are short, and the abundance time course has only a small amount of right skewness (see Fig. 1c in Zonneveld 1991). When δ is increased above one in the EZ, the emergence distribution becomes left skewed and when $\delta = 1.579$, this left skewness exactly cancels out the right skewness due to the lifespan distribution, and the abundance time course is symmetrical. For δ values exceeding this threshold, any attempt to fit the CZ to the EZ, either by the method of moments or by maximum likelihood, will fail because the CZ cannot produce a left-skewed abundance time course. Though the Weibull lifespan distribution is also technically capable of introducing a small amount of left skewness into the abundance time course (for ρ large), it was not an issue across the range of parameter values explored here.

Figure 2 demonstrates that fitting the CZ either of the generalized models when its underlying assumptions are violated produces strongly biased estimates of population size and mean lifespan. However, the inherent flexibility of the CZ still allows it to closely mimic most of the abundance time course shapes the EZ and LZ are capable of producing (Fig. 3). All four fits shown in Fig. 3 occur in situations that yield large percent errors in both N and $\langle l \rangle$ estimates (Fig. 2), and are typical of the ability of the CZ to closely approximate either of the generalized models. This indicates that a good fit to an observed abundance time course does not mean that underlying model assumptions are met, and may yield heavily biased parameter estimates.

The CZ either completely fails to fit the generalized models or provides a very poor fit in two distinct situations

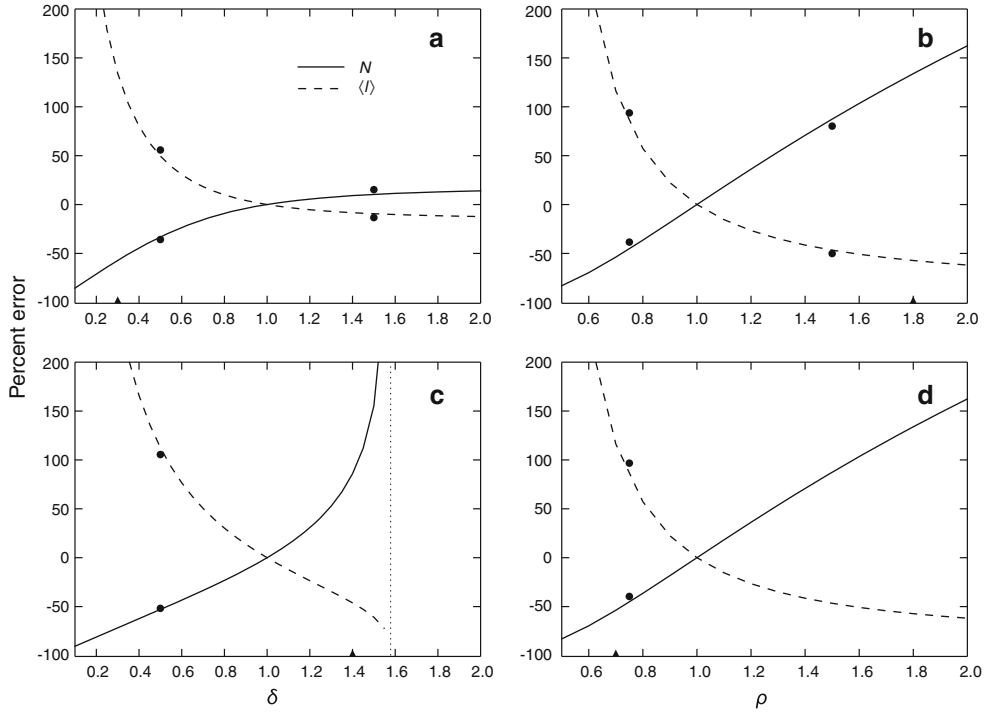

Fig. 2 The percent error between estimated mean lifespan and true mean lifespan (*dashed curves*) and estimated population size and true population (*solid curves*) when the CZ is fitted to the EZ (**a**, **c**) or the LZ (**b**, **d**). The true values of the target parameters are those that are used in the generalized model serving as the data generating process. **a**, **b** The *C. pamphilus* scenario with base parameter set: $N = 769$, $\mu = 15.9$, $\beta = 5.52$, and $\alpha = 0.126$. **c**, **d** The *A. agestis* base parameter set with $N = 162$, $\mu = 10.2$, $\beta = 2.93$, and $\alpha = 0.362$. *The vertical dotted line* in **c** occurs at $\delta \approx 1.579$ and is the point at which the EZ time course for that scenario becomes symmetrical. For δ values exceeding this threshold, the EZ time course is left skewed and the moment estimators fail because the CZ cannot produce a left-skewed shape. The *triangles* on the *x*-axis of each panel indicate the points for which the fits of the CZ to the corresponding generalized model are shown in Fig. 3. The *points along the percent error curves* are the large sample (asymptotic) maximum likelihood results for that parameter combination (Appendix D), demonstrating that the moment-estimator based results are reasonable

(Fig. 4). The first was mentioned above and occurs when there is left skewness in the abundance time course (Fig. 4a). In this case the CZ fit is not possible because it is not capable of producing a left-skewed abundance time course and there is no information from which the death rate, α, can be estimated. The second situation occurs in the LZ, when the individual-level mortality rate is very low until an individual is quite old and then accelerates sharply thereafter. This results in an abundance time course with broad "shoulders" and a large plateau that the CZ simply cannot mimic (Fig. 4b). Of the two, the case of left skewness (or at least too little right skewness) is the more likely to occur in real data. In all other situations, the CZ was able to closely approximate the shape of either generalized model, indicating that relying on goodness of fit as an indication of model appropriateness may be dangerous with the CZ.

Discussion

Here, I have defined a nested family of models by generalizing the CZ to accommodate an asymmetrical emergence distribution and an age-dependent death rate that leads to Weibull distributed lifespans. I have then used the emergence-generalized (EZ) and lifespan-generalized (LZ) models as data generating processes to explore the CZ's ability to approximate abundance time courses that deviate from its assumptions, and to quantify the bias in population size and mean lifespan estimates incurred by doing so. These analyses demonstrate that the CZ is remarkably flexible and can mimic a wide range of abundance time course shapes. The cost of this flexibility is that the CZ is sensitive to violations of both its emergence and death rate assumptions and is capable of producing highly biased estimates of key population parameters when its assumptions are not met. The fact that the model can still fit well even when it yields massively biased parameter estimates means that a good fit to an observed transect count dataset cannot be taken as evidence that the model's assumptions hold.

A more subtle message that emerges is that the CZ is pushing the limits of the information content of transect count data. One could envision fitting the CZ, EZ, LZ and the fully generalized model ($\delta \neq 1$, $\rho \neq 1$) to the same transect count data via maximum likelihood and using

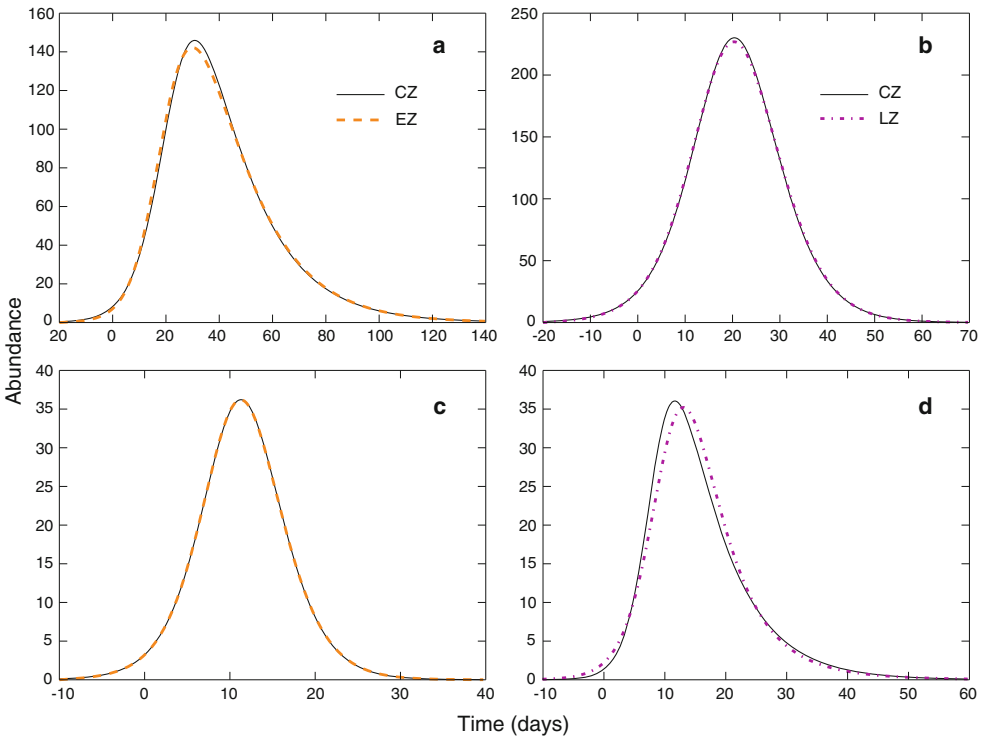

Fig. 3 A demonstration of the ability of the CZ to provide close approximations to both the EZ and LZ when the emergence and death rate assumptions of the CZ do not hold. The structure of the figure is the same as for Fig. 2, with **a** and **b** being the *C. pamphilus* scenario and **c** and **d** representing the *A. agestis* scenario. Similarly, **a** and **c** are CZ fits to the EZ, while **b** and **d** are fits to the LZ. Each fit shown corresponds to that occurring at the triangles on the *x*-axis of the corresponding panel of Fig. 2. Note that the scales of both axes differ among panels

model selection techniques to identify the most parsimonious variant. Such an approach could, in principle, be used to infer the functional forms of the emergence distribution and death rate function. Unfortunately, the ability of the CZ to closely approximate the more complicated models, coupled with its smaller parameter count, strongly suggests that it would consistently win such model comparisons even when it is the "wrong" model. I have made preliminary attempts to fit the EZ and LZ to transect count data via maximum likelihood, but was not able to get these models to reliably converge. This again suggests that the information content of transect count data is limited and that the new parameters that define the EZ and LZ are not consistently identifiable from such datasets.

In addition to providing a useful computational device, the moment estimators developed here shed light on the way the CZ uses the information in transect count time series. The location of the time series along the time axis provides information about both the mean emergence date, μ and the death rate, α. The width of the abundance time course provides information about the scale parameter of the emergence distribution, β, and α. The skewness of the time course (and not just the right tail) provides direct information about α. Finally, the height of the abundance time course informs the population size index, N.

The only source of skewness in the CZ is the exponential lifespan distribution. Both the EZ and LZ introduce additional features that can also affect the skewness of the abundance time courses they produce. The effect these additional features have on the skewness is the source of the bias in the CZ estimates. Any feature that accentuates the right skewness of the abundance time course will lead to overestimates of $\langle l \rangle$ and underestimates of N. Conversely, any feature that decreases the right skewness of the time course would cause the CZ to underestimate $\langle l \rangle$ and overestimate N. The biases in these two parameters go in opposite directions because once the mean lifespan is determined by the shape of the time course, the population size must compensate such that the area under the curve is conserved. This inverse relationship between the two parameters can be clearly seen in Fig. 2 and has been noted before (Mattoni et al. 2001; Haddad et al. 2008).

The estimation of any features added to the CZ, such as skewed emergence distributions or non-constant death rates, would rely on finer and finer details of the shape of the time course. For example, in the moment framework described here, estimation of δ (EZ) or ρ (LZ) would involve the kurtosis of the time course. Such fine details would be extremely difficult to resolve from noisy transect count time series. It is important to note that the maximum likelihood approach typically used to fit the CZ does not use the moments of the abundance time course, but instead uses the shape of the entire time course. Still, the idea that finer and finer details of that shape would be required to identify additional model features is general and points back to the idea that the CZ is at the limit of the information transect counts can provide.

As many butterfly monitoring programs mainly collect transect count data, the potential for bias in CZ-based

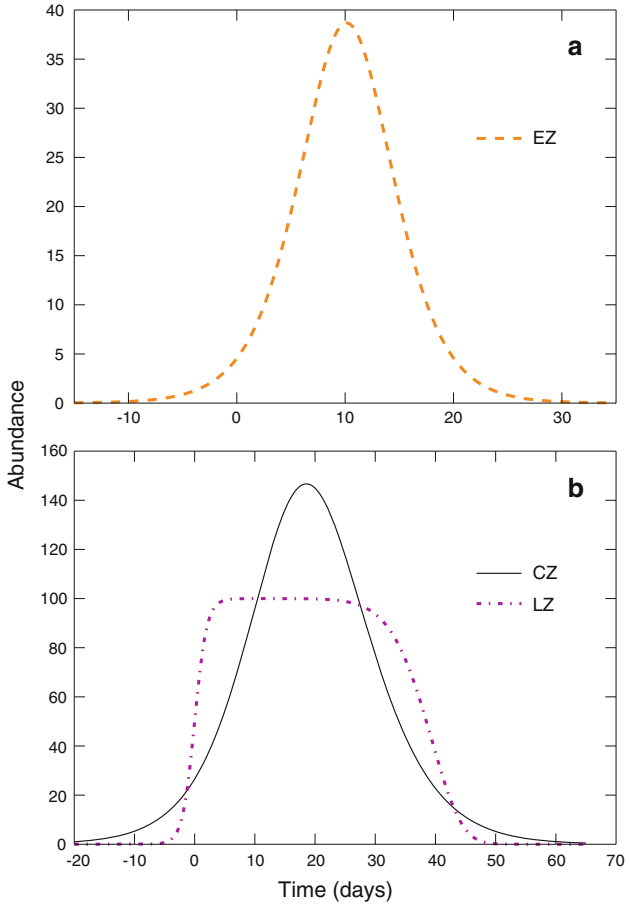

Fig. 4 Situations for which the CZ either fails to fit (**a**) or provides a very poor approximation of the shape of the generalized model (**b**). **a** The situation occurring to the right of the vertical line in Fig. 2c. Here, the abundance time course of the EZ is slightly left-skewed and the CZ fails to fit it entirely. **b** Deviates from the two empirically based scenarios and depicts a situation in the LZ where individuals all live long lives and achieve lifespans that are close to the mean lifespan. This results in an abundance time course with very broad shoulders and a plateau. Though the CZ can be fit to such a time course (the best fit is shown), it cannot provide a close approximation to this shape. Note that the scales of both axes differ between panels

parameter estimates is especially troubling because it is not currently possible to use such data to assess the validity of model assumptions. While these results are somewhat discouraging, it is important to know the limitations of available data sources and analytical techniques. The constant death rate/exponential lifespan assumption is widely employed in insect population biology, and is likely to be a reasonable assumption in many cases (Zonneveld 1991). This is particularly so for populations that experience very strong, constant external mortality. Similarly, assuming a symmetrical emergence distribution is a reasonable staring place. These assumptions are, however, rarely tested explicitly and the consequences of serious violations of them for inferential procedures are rarely explored. My results indicate that, at least for CZ-based parameter estimation, more caution is warranted. When these assumptions are met, the CZ has been shown to produce reasonable parameter estimates under some conditions (Mattoni et al. 2001; Gross et al. 2007; Haddad et al. 2008). The problem is that it will be difficult to know from transect counts alone when the CZ's emergence and death assumptions are satisfied and when they are not.

The analysis of transect count data has focused primarily on developing an index of abundance that can be used to detect population trends over time (Pollard 1977; Pollard and Yates 1993; Rothery and Roy 2001). In contrast, mark recapture studies are currently considered the "gold standard" for estimating butterfly demographic parameters, but are widely acknowledged as being too effort intensive for large-scale monitoring programs (Gross et al. 2007; Haddad et al. 2008). Zonnevelds (1991) approach represents an attempt to extract some of the information provided by a full mark–recapture study from simple transect count data. The limitations of the CZ demonstrated in this and other studies suggest that, while sometimes appropriate, it will not consistently be able to estimate demographic parameters from transect count data alone (Mattoni et al. 2001; Gross et al. 2007; Haddad et al. 2008). Both Gross et al. (2007) and Haddad et al. (2008) have suggested a tiered approach, where count datasets are, at some sites and in some years, combined with limited mark–recapture studies. Another possibility might be to augment transect counts with more detailed observational information, perhaps wing-wear-based estimates of population age structure throughout the flight period, to allow better estimates of population parameters or more detailed models to be fit.

The generalized models presented here should also find use beyond their roles in probing the limits of count-based inference. Theoretical analyses of reproductive phenology focusing on protandry/protogyny and reproductive asynchrony, many of which are based on variations of the CZ (Zonneveld and Metz 1991; Zonneveld 1992, 1996a, b; Calabrese and Fagan 2004; Calabrese et al. 2008; Fagan et al. 2010), could directly leverage the machinery developed here. The results of such studies would depend primarily on the shapes of the abundance time courses, and not on the specific assumptions generating them. While I have shown that the CZ can generate a wide range of shapes, some occur only at biologically implausible parameter values (hence the bias in its estimates when assumptions are not met). Theoretical studies focusing on biologically reasonable scenarios could therefore employ generalizations of the CZ to explore, and justify, how a wider range of time course shapes affects individual reproductive success, population-level growth rates, and the optimality of alternative phenological strategies.

Acknowledgments I thank B. Fagan, H. Lynch, and L. Ries for helpful comments on early versions of the manuscript.

Appendix A: Alternative derivation of the EZ and CZ

The normalized abundance time course for the CZ, denoted $\dot{x}_{cz}(t)$, follows from Eq. (2) and the third (unlabeled) equation in Zonneveld (1991) and is written

$$\dot{x}_{cz}(t) = \alpha e^{-\alpha(t-\mu)} \int_0^{e^{(t-\mu)/\beta}} \frac{r^{\alpha\beta}}{(1+r)^2} dr. \quad (6)$$

Equation (6) integrates to unity on $t \in (-\infty, \infty)$.

Now I show that the normalized abundance time course of the EZ can be derived as the sum of a generalized logistic random variable and an exponential random variable. As the classical logistic is a special case of the generalized logistic, this implies that the CZ can be derived as the sum of logistic and exponential random variables. For convenience I use slightly different notation here than in the main text, and so begin by defining the PDFs of these distributions. Let U have an exponential distribution with PDF

$$f_U(u) = \begin{cases} \alpha e^{-\alpha u} & u \geq 0 \\ 0 & u < 0 \end{cases}$$

where α is the rate parameter. Let V have a generalized logistic distribution with PDF

$$f_V(v) = \frac{\delta e^{(v-\mu)/\beta}}{\beta(1+e^{(v-\mu)/\beta})^{\delta+1}}$$

where μ is the location parameter, β is the scale parameter, and δ is the shape parameter.

The PDF of the sum $T = U + V$ can be written as the convolution of $f_U(u)$ and $f_V(v)$ (Casella and Berger 2002)

$$f_T(t) = \int_0^{\infty} f_U(w) f_V(t-w) dw \quad (7)$$

where the lower limit of integration follows from the fact that $f_U(u) = 0$ for $u < 0$. Substituting the exponential and generalized logistic densities into the convolution integral, I have

$$f_T(t) = \int_0^{\infty} \alpha e^{-\alpha w} \frac{\delta e^{(t-w-\mu)/\beta}}{\beta(1+e^{(t-w-\mu)/\beta})^{\delta+1}} dw. \quad (8)$$

Next, let

$$r = e^{(t-w-\mu)/\beta}, \quad (9)$$

then

$$dr = -\frac{e^{(t-w-\mu)/\beta}}{\beta} dw$$

$$dw = -\frac{\beta dr}{r}.$$

Rewriting Eq. (8) in terms of r and dr, I obtain

$$f_T(t) = -\alpha\delta \int_{e^{(t-\mu)/\beta}}^{0} e^{-\alpha w} \frac{1}{(1+r)^{\delta+1}} dr. \quad (10)$$

I then rearrange Eq. (9) to obtain

$$w = t - \mu - \beta \mathrm{Ln}(r)$$

and substitute this result into the exponential term in Eq. (10) yielding

$$e^{-\alpha[t-\mu-\beta \mathrm{Ln}(r)]} = e^{-\alpha(t-\mu)} e^{\alpha\beta \mathrm{Ln}(r)} = e^{-\alpha(t-\mu)} r^{\alpha\beta}$$

which leads to

$$f_T(t) = -\alpha\delta e^{-\alpha(t-\mu)} \int_{e^{(t-\mu)/\beta}}^{0} \frac{r^{\alpha\beta}}{(1+r)^{\delta+1}} dr.$$

Finally, reversing the limits of integration, I obtain the PDF of the sum of an exponential and generalized logistic in the same form as the normalized Zonneveld model:

$$f_T(t) = \alpha\delta e^{-\alpha(t-\mu)} \int_0^{e^{(t-\mu)/\beta}} \frac{r^{\alpha\beta}}{(1+r)^{\delta+1}} dr. \quad (11)$$

As the integral above is the incomplete Beta function, Eq. (11) can be written

$$f_T(t) = \alpha\delta e^{-\alpha(t-\mu)} B_z(1+\alpha\beta, \delta - \alpha\beta)$$

where $z = e^{(t-\mu)/\beta} / (1 + e^{(t-\mu)/\beta})$.

When $\delta = 1$ the generalized logistic reduces to the classical (symmetrical) logistic distribution and $f_T(t)$ simplifies to

$$f_T(t) = \alpha e^{-\alpha(t-\mu)} \int_0^{e^{(t-\mu)/\beta}} \frac{r^{\alpha\beta}}{(1+r)^2} dr$$

$$= \alpha e^{-\alpha(t-\mu)} B_z(1+\alpha\beta, 1-\alpha\beta)$$

which is the normalized abundance time course of the classical Zonneveld model, Eq. (6).

Appendix B: Cumulants of the EZ and CZ

The cumulants of the distribution of the sum of two random variables are simply the sums of the cumulants of the

component distributions. The first three cumulants (the mean, the variance, and the third central moment) of the normalized EZ model can thus be obtained from the cumulants of the generalized logistic emergence (Wu et al. 2000) and exponential lifespan (Evans et al. 2000) distributions, yielding

$$\begin{aligned}\kappa_{ez}^{(1)} &= \mu - \beta[\gamma + \psi(\delta)] + \frac{1}{\alpha}\\ \kappa_{ez}^{(2)} &= \beta^2\left[\frac{\pi^2}{6} + \psi^{(1)}(\delta)\right] + \frac{1}{\alpha^2}\\ \kappa_{ez}^{(3)} &= \beta^3[\psi^{(2)}(1) - \psi^{(2)}(\delta)] + \frac{2}{\alpha^3}\end{aligned} \tag{12}$$

where $\gamma \approx 0.577216$ is Euler's constant, $\psi(\bullet)$ is the digamma function, and $\psi^{(i)}(\bullet)$ is the polygamma function of order i. For the normalized CZ, these expressions simplify to

$$\begin{aligned}\kappa_{cz}^{(1)} &= \mu + \frac{1}{\alpha}\\ \kappa_{cz}^{(2)} &= \frac{\beta^2\pi^2}{3} + \frac{1}{\alpha^2}\\ \kappa_{cz}^{(3)} &= \frac{2}{\alpha^3}.\end{aligned} \tag{13}$$

Appendix C: Moments estimators

The method of moments is one approach to parameter estimation and can be more convenient to work with than maximum likelihood (Clark 2007). Solving Eqs. (13) for the three parameters (μ, β, and α) of the normalized CZ produces the moment estimators given in the main text as Eqs. (5).

When fitting the CZ to the EZ, it is possible to solve directly for the moment estimators of the CZ in terms of the parameters of the EZ by substituting Eqs. (12) into the right hand side of Eqs. (5), yielding

$$\begin{aligned}\hat{\mu}_{cz} &= \mu_{ez} + \frac{1}{\alpha_{ez}} - \beta_{ez}[\gamma + \psi(\delta_{ez})] - \left(\frac{1}{\alpha_{ez}^3} - \frac{\beta_{ez}^3}{2}\left[\psi^{(2)}(\delta_{ez}) + 2\zeta(3)\right]\right)^{1/3}\\ \hat{\beta}_{cz} &= \frac{\sqrt{3}}{\pi}\sqrt{\frac{1}{\alpha_{ez}^2} + \beta_{ez}^2\left[\frac{\pi^2}{6} + \psi^{(1)}(\delta_{ez})\right] - \left(\frac{1}{\alpha_{ez}^3} - \frac{\beta_{ez}^3}{2}\left[\psi^{(2)}(\delta_{ez}) + 2\zeta(3)\right]\right)^{2/3}}\\ \hat{\alpha}_{cz} &= \left(\frac{1}{\alpha_{ez}^3} + \frac{\beta_{ez}^3}{2}\left[\psi^{(2)}(1) - \psi^{(2)}(\delta_{ez})\right]\right)^{-1/3}\end{aligned} \tag{14}$$

where $\zeta(\bullet)$ is the Riemann zeta function.

The non-exponential lifespan distribution of the normalized LZ means it cannot be derived as the sum of the emergence and lifespan distributions, so closed-form expressions for its cumulants are not available. Instead, the first three cumulants of the normalized LZ can, for any given parameter set, be calculated numerically as

$$\begin{aligned}\kappa_{lz}^{(1)} &= \int_{-\infty}^{\infty} t\dot{x}_{lz}(t)dt\\ \kappa_{lz}^{(2)} &= \int_{-\infty}^{\infty} (t - \kappa_{lz}^{(1)})^2\dot{x}_{lz}(t)dt\\ \kappa_{lz}^{(3)} &= \int_{-\infty}^{\infty} (t - \kappa_{lz}^{(1)})^3\dot{x}_{lz}(t)dt\end{aligned} \tag{15}$$

and these values can then be substituted into the moment estimators to obtain the parameters of the CZ that produce the best fit to the LZ. These numerical integrations are much quicker to calculate than randomly generating datasets from the LZ and fitting the CZ to it via the maximum likelihood approach described in the next section.

Appendix D: Large sample maximum likelihood estimation

The moment estimators in the main text allow the CZ to be fit directly to the generalized models without (for a given set of parameters) first generating many random transect count datasets from the focal generalized model, fitting the CZ to each realization, and then averaging the CZ's parameter estimates over the realizations. The moment approach is computationally efficient and it allows me to directly assess the ability of the CZ to approximate the shapes the generalized models can produce. As the moments of each generalized model that are used to fit the CZ are the "true values" of these quantities and not estimates of them calculated from a random sample, it is an asymptotic approach in the sense that it is what would be obtained as sample size became infinitely large. The method of moments often produces similar (or in some cases exactly the same) estimators as the more standard method of maximum likelihood (Clark 2007), but can be more biased and less reliable than maximum likelihood (Bolker 2008). To check the moment-based results, I therefore use a large sample maximum likelihood approach that approximates the asymptotic situation represented by the moment estimators. This approach is too computationally intensive to be feasible to thoroughly explore a broad range of parameter space, but does provide a way to spot check the moment-based results.

For a given set of generalized model parameters, I calculated the time course in abundance $x_{gz}(t)$ as described in the main text. For each day t for which $x_{gz}(t) > 0.005$, I randomly drew an observed count from a Poisson distribution with mean $x_{gz}(t)$. This procedure was repeated 50 times for each parameter set yielding 50 random datasets

for that parameter set with complete coverage of the flight season. I used INCA 1.53 (http://www.urbanwildlands.org/INCA/) to fit the CZ to each of these random datasets via maximum likelihood. For each parameter of interest (i.e., N and $\langle l \rangle$), I then calculated its average across the 50 random datasets and used this value as the "asymptotic" maximum likelihood estimate of that parameter. Finally, I calculated the percent error between the asymptotic MLE of that parameter and its true value. The asymptotic MLE results are plotted as points along the percent error curves in Fig. 2. The two methods agree closely in all cases, indicating that the moment estimators produce reasonable results.

References

Auckland JN, Debinski DM, Clark WR (2004) Survival, movement, and resource use of the butterfly *Parnassius clodius*. Ecol Entomol 29:139–149

Ban Y, Kiritani K, Miyai S, Nozato K (1990) Studies on ecology and behavior of Japanese black swallowtail butterflies: VIII. Survivorship curves of adult male populations in *Papilio helenus nicconicolens* Butler and *P. protenor demetrius* Cramer (Lepidoptera: Papilionidae). Appl Entomol Zool 25:409–414

Bolker BM (2008) Ecological models and data in R. Princeton University Press, Princeton

Bouyer J, Sana Y, Samandoulgou Y, Cesar J, Guerrini L, Kabore-Zoungrana C, Dulieu D (2007) Identification of ecological indicators for monitoring ecosystem health in the trans-boundary W Regional park: a pilot study. Biol Conserv 138:73–88

Brakefield PM (1982) Ecological studies on the butterfly *Maniola jurtina* in Britain. II. Population dynamics: the present position. J Anim Ecol 51:727–738

Calabrese JM, Fagan WF (2004) Lost in time, lonely, and single: reproductive asynchrony and the Allee effect. Am Nat 164:25–37

Calabrese JM, Ries L, Matter SF, Debinski DM, Auckland JN, Roland J, Fagan WF (2008) Reproductive asynchrony in natural butterfly populations and its consequences for female matelessness. J Anim Ecol 77:746–756

Casella G, Berger RL (2002) Statistical inference. Duxbury Press, Pacific Grove

Clark JS (2007) Models for ecological data: an introduction. Princeton University Press, Princeton

Crone EE, Pickering D, Schultz CB (2007) Can captive rearing promote recovery of endangered butterflies? An assessment in the face of uncertainty. Biol Conserv 139:103–112

Cushman JH, Boggs CL, Weiss SB, Murphy DD, Harvey AW, Ehrlich PR (1994) Estimating female reproductive success of a threatened butterfly: influence of emergence time and host plant phenology. Oecologia 99:194–200

Evans M, Hastings N, Peacock B (2000) Statistical distributions. Wiley-Interscience, New York

Fagan WF, Cosner C, Larsen EA, Calabrese JM (2010) Reproductive asynchrony in spatial population models: how mating behavior can modulate Allee effects arising from isolation in both space and time. Am Nat 175:362–373

Gross K, Kalendra EJ, Hudgens BR, Haddad NM (2007) Robustness and uncertainty in estimates of butterfly abundance from transect counts. Popul Ecol 49:191–200

Haddad NM, Hudgens B, Damiani C, Gross K, Kuefler D, Pollock K (2008) Determining optimal population monitoring for rare butterflies. Conserv Biol 22:929–940

Iwasa Y, Odendaal FJ, Murphy DD, Ehrlich PR, Launer AE (1983) Emergence patterns in male butterflies: a hypothesis and a test. Theor Popul Biol 23:363–379

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley, New York

Kawagoe T, Suzuki N, Matsumoto K (2001) Multiple mating reduces longevity of females of the windmill butterfly *Atrophaneura alcinous*. Ecol Entomol 26:258–262

Lederhouse RC (1983) Population structure, residency and weather related mortality in the black swallowtail butterfly, *Papilio polyxenes*. Oecologia 59:307–311

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: the case of the El Segundo blue butterfly, *Euphilotes bernardino allyni*. J Insect Conserv 5:197–206

Parmesan C (2007) Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Glob Change Biol 13:1860–1872

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42

Pollard E (1977) A method for assessing changes in the abundance of butterflies. Biol Conserv 12:115–134

Pollard E, Yates TJ (1993) Monitoring butterflies for ecology and conservation: the British butterfly monitoring scheme. Chapman & Hall, London

Rothery P, Roy DB (2001) Application of generalized additive models to butterfly transect count data. J Appl Stat 28:897–909

Roy DB, Sparks TH (2000) Phenology of British butterflies and climate change. Glob Change Biol 6:407–416

Roy DB, Rothery P, Moss D, Pollard E, Thomas JA (2001) Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. J Anim Ecol 70:201–217

Schtickzelle N, Le Boulengé E, Baguette M (2002) Metapopulation dynamics of the bog fritillary butterfly: demographic processes in a patchy population. Oikos 97:349–360

Sims SR, Shapiro AM (1983) Seasonal phenology of *Battus philenor* (L.) (Papilionidae) in California. J Lep Soc 37:281–288

Sisk TD, Launer AE, Switky KR, Ehrlich PR (1994) Identifying extinction threats. Bioscience 44:592–604

Thomas JA (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. Philos Trans R Soc B Biol Sci 360:339–357

van Swaay CAM, Nowicki P, Settele J, van Strien AJ (2008) Butterfly monitoring in Europe: methods, applications and perspectives. Biodivers Conserv 17:3455–3469

Warren MS, Hill JK, Asher TJ, Fox R, Huntley B, Roy BD, Telfer MG, Jeffcoate S, Harding P, Jeffcoate G, Willis SG, Greatorex-Davies JN, Moss D, Thomas CD (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. Nature 414:65–69

Wu JW, Hung WL, Lee HM (2000) Some moments and limit behaviors of the generalized logistic distribution with applications. Proc Natl Sci Counc ROC (A) 24:7–14

Xue FS, Kallenborn HG, Wei HY (1997) Summer and winter diapause in pupae of the cabbage butterfly, *Pieris melete* Ménétriés. J Insect Physiol 43:701–707

Zheng C, Ovaskainen O, Saastamoinen M, Hanski I (2007) Age-dependent survival analyzed with Bayesian models of mark–recapture data. Ecology 88:1970–1976

Zonneveld C (1991) Estimating death rates from transect counts. Ecol Entomol 16:115–121

Zonneveld C (1992) Polyandry and protandry in butterflies. Bull Math Biol 54:957–976

Zonneveld C (1996a) Sperm competition cannot eliminate protandry. J Theor Biol 178:105–111

Zonneveld C (1996b) Being big or emerging early? Polyandry and the trade-off between size and emergence in male butterflies. Am Nat 147:946–965

Zonneveld C, Metz JAJ (1991) Models on butterfly protandry: virgin females are at risk to die. Theor Popul Biol 40:308–321