
Matechou, Eleni; Dennis, Emily B.; Freeman, Stephen N.; Brereton, Tom. 2014. **Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model**. *Journal of Applied Ecology*, 51 (3). 766-775. 10.1111/1365-2664.12208

Monitoring abundance and phenology in (multivoltine) butterfly species; a novel mixture model

Eleni Matechou[*], Emily B. Dennis[†], Stephen N. Freeman [‡], Tom Brereton [§]

November 8, 2013

Key words: Common blue butterfly, normal mixtures, parameter redundancy, sample counts, stopover duration, "super-population", UKBMS

Article type: Standard

Number of words in the summary: 361, **Number of words in the main text**: 4808, **Number of words in references**: 950, **Number of words in figure captions**: 91, **Number of words in table captions**: 43, **Number of tables**: 2, **Number of figures**: 4, **Number of references**: 28

---

[*] Department of Statistics, University of Oxford, 1 South Parks Road, OX1 3TG, UK, matechou@stats.ox.ac.uk

[†] School of Mathematics, Statistics and Actuarial Science, University of Kent, ed234@kent.ac.uk

[‡] Centre for Ecology and Hydrology, Wallingford, sfreeman@ceh.ac.uk

[§] Butterfly Conservation, tbrereton@butterfly-conservation.org

Summary

1. Data from ’citizen science’ surveys are increasingly valuable in identifying declines in widespread species, but require special attention in the case of invertebrates, with considerable variation in number, seasonal flight patterns and, potentially, voltinism. There is a need for reliable and more informative methods of inference in such cases.

2. We focus on data consisting of sample counts of individuals that are not uniquely identifiable, collected at one or more sites. Arrival/emergence and departure/death of the individuals take place during the study. We introduce a new modelling approach, which borrows ideas from the “stopover” capture-recapture literature, that permits the estimation of parameters of interest, such as mean arrival times and relative abundance or, in some cases, absolute abundance, and the comparison of these between sites.

3. The model is evaluated using an extensive simulation study which demonstrates that the estimates for the parameters of interest obtained by the model are reliable, even when the data sets are sparse, as is often the case in reality.

4. When applied to data for the Common blue butterfly *Polyommatus icarus* at a large number of sites, the results suggest that the mean emergence times, as well as the relative sizes of the broods, are linked to site Northing, and confirm field experience that the species is bivoltine in the south of the UK but practically univoltine in the north.

Synthesis and applications

5. Our proposed “stopover” model is parameterised with biologically informative con-

stituents; times of emergence, survival rate and relative brood sizes. Estimates of absolute or relative abundance, that can be obtained alongside these underlying variables, are robust to the presence of missing observations, and can be compared in a statistically rigorous framework. These estimates are direct indices of abundance, rather than "sightings", implicitly adjusted for the possible presence of repeat sightings during a season. At the same time they provide indices of change in demographic and phenological parameters that may be of use in identifying the factors underlying population change. The model is widely applicable and this will increase the utility of already valuable and influential long-standing surveys in monitoring the effects of environmental change on phenology or abundance.

1 Introduction

The sizes of wildlife populations, and particularly of many insect populations, at a study site can change daily, with new individuals being added through birth and/or immigration and removed through death and/or emigration. Studying these changes is of great interest in conservation and monitoring, but a complete census is rarely possible and a sampling survey is all that can be achieved. For widespread species, “citizen science” schemes are increasingly adopted, with potentially large numbers of volunteers able to cover large geographical areas over a longer period of time than is practical by any other means. Trends from such data are now available from across many (though by no means all) taxonomic groups. Modelling invertebrate data poses both problems and opportunities, with seasonal occurrence and individuals, possibly of several generations, having lifespans considerably shorter than the sampling period.

Such data usually comprise a number of simple series of counts collected to a standardised protocol. Where individuals can be uniquely marked or identified by unique physical characteristics, a number of capture-recapture (or mark-recapture) modelling approaches have been developed for estimating the size of a population. Collectively, these have become known as “Jolly-Seber-type” models (Jolly, 1965; Seber, 1965). Furthermore, recently developed “stopover” models (Pledger et al., 2009; Matechou et al., 2013), that build on the Schwarz & Arnason (1996) mark-recapture approach (SA), explicitly model the unknown times of arrival and departure of the individuals and provide estimates of the arrival rates at the site during the sampling period, as well as indirect estimates of the average duration of stay of the individuals at the site, referred to as mean stopover duration. However, these models do not directly apply to data where the individuals are not uniquely identifiable. In these cases, the data commonly consist only of sample counts obtained on each of a number of occasions within a survey period, and the number of times an individual is detected is unknown.

In this paper, we introduce a new approach for data of this type. The merits of this new model are that underlying biological processes are explicitly parameterised. This permits the flexible estimation of parameters of interest such as absolute (or relative) abundance, mean arrival times and probabilities of retention, and the formal comparison of these across a number of sites and/or years. Whereas existing models essentially estimate the numbers of sightings, we estimate relative or absolute abundance, with an implicit adjustment made for multiple encounters while estimating the relative sizes of the arrival groups, their mean arrival times, and the mean stopover duration in a season. These additional parameters are potentially useful in explaining changes in abundance or phenology and can be functionally related to environmental covariates.

In section 2 we introduce the model, with particular attention to issues of model-fitting and parameter identifiability. In section 3 we explore the performance of the model via a large-scale simulation study. In section 4 we present an illustrative example based on data from the UK Butterfly Monitoring scheme (UKBMS) (Botham et al., 2011) where we consider the case of seasonally emerging butterflies which are only countable - in adult form - during a fixed period. Specifically, we consider multivoltine species in which more than one generation appear during each season, with potentially considerable overlap in their times of flight. The paper then concludes with a discussion.

2 Material and Methods

2.1 The model

The model presented in this section borrows ideas from models for capture-recapture data, and specifically from the SA parameterisation of the Jolly-Seber model. SA use the idea of a "super-population", N, first introduced by Crosbie & Manly (1985), which is the total number of unique individuals that were present at the site during the study. For a study assumed to consist of T occasions, eg. days, weeks etc, they explicitly model the arrival of the individuals at the site using the entry parameters, $\beta_{j-1}, j = 1, \ldots, T$, which are the proportions of N that were new arrivals at occasion j with $\sum_{j=1}^{T} \beta_{j-1} = 1$.

We assume here, without loss of generality, that the T sampling occasions are equally spaced. The data set, $\mathbf{y}$, is a vector of sample counts of length T collected at the site with $K \leq T$ non-missing entries. For an individual animal to contribute to a count obtained on occasion j, y_j, it has to have arrived before occasion j, to have remained until j and to be detected given that it is present on occasion j. The probability of remaining at the site until the next occasion, referred to as retention probability, can be time- or age-dependent, where age is defined to be the unknown time since entry to the site. We denote by ϕ_{ja} the probability that individuals that have been at the site for a occasions, and are present on occasion j, will remain until occasion $j+1$. The probability of detecting an individual that is present on occasion j, referred to as detection probability, is denoted by p_j.

It is natural to treat entry j in $\mathbf{y}$ as the realisation of a Poisson distribution with expectation λ_j. Each λ_j is a function of the "super-population" size and of the entry, retention and detection probabilities. Specifically, the expected number of individuals counted at the site on occasion j is equal to $\lambda_j = N\left[\sum_{b=1}^{j} \beta_{j-1}\left\{\prod_{k=b}^{j-1} \phi_{ka}\right\}\right] p_j,\ j = 1, \ldots, T,\ a = k - b + 1$, where $b = 1, \ldots, j$ are the possible times of entry to the population for an individual detected on occasion j. For example, $\lambda_3 = N\left(\beta_0\phi_{11}\phi_{22} + \beta_1\phi_{21} + \beta_2\right)p_3$, that is individuals detected on occasion 3 entered the population either before occasion 1 but remained until occasion 3, or between occasions 1 and 2 and remained until occasion 3, or between occasions 2 and 3.

The model likelihood is $L(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{p}|\mathbf{y}) = \prod_{j=1}^{T}\left[\frac{\exp(-\lambda_j)\lambda_j^{y_j}}{y_j!}\right]$. If sample j was missed then $p_j = 0$ and hence $\lambda_j = y_j = 0$ and observation j does not contribute to the likelihood calculation.

As will be explained in section 2.2, the total number of parameters or combinations of parameters that can be estimated by the model is equal to K. Allowing the entry probabilities to freely vary by time introduces $T - 1$ parameters to the model, a number of which can be practically equal to 0 for data sets of the type considered in this paper, since the period during which individuals arrive at the site may be much shorter than, and is assumed to be encompassed by, the sampling period. Therefore, we suggest that the entry probabilities are modelled/constrained using a mixture of M normal distributions instead. Each of these distributions relates to one arrival group, eg. one distinct brood, and has its own (relative) weight w_m, $m = 1, \ldots, M$ and mean, μ_m, and possibly its own variance σ_m^2. Consequently, the proportion of N that were new arrivals on occasion $j,\ j = 1, \ldots, T$ is equal to: $\beta_{j-1} = \sum_{m=1}^{M} w_m\left[F_m(j) - F_m(j-1)\right]$ where, $F_m(j) = P(X \leq j)$ when $X \sim N(\mu_m, \sigma_m^2)$. By the definition of the β parameters, $F_m(0) = 0$ and $F_m(T) = 1\ \forall m$, therefore $\beta_0 = \sum_{m=1}^{M} w_m\left[F_m(1)\right]$ and $\beta_{T-1} = 1 - \sum_{m=1}^{M} w_m\left[F_m(T-1)\right]$. A demonstrating example when $M = 2$ with overlapping times between the two arrival groups is given in Fig.

1. Hence, regardless of the size of T, the model requires the estimation of $2M$ parameters in the case of homoscedastic mixture distributions and $3M-1$ parameters in the heteroscedastic. This modelling approach for the entry parameters provides a smooth representation of the arrival pattern of the individuals to the site as well as estimates of the mean arrival times and relative sizes of the different arrival groups.

[Figure 1 about here.]

Similarly, we propose to model retention probabilities using parametric curves. An example is the logistic curve where $\text{logit}(\phi_{ja}) = \alpha_\phi + \beta_\phi \cdot x_j, \forall a$ where x_j can correspond to the value of a time-varying environmental covariate such as temperature or simply to calendar time. Another option is to use the more flexible quadratic function where $\text{logit}(\phi_{ja}) = \alpha_\phi + \beta_\phi \cdot x_j + \gamma_\phi \cdot x_j^2, \forall a$. Alternatively, as mentioned above, ϕ can be modelled as a function of age, $\text{logit}(\phi_{ja}) = \alpha_\phi + \beta_\phi \cdot a, \forall j$.

Finally, detection probabilities can be modelled either as constant over time, as appropriate, or as dependent on a time-varying covariate v, such as temperature or sampling effort at the time of sampling, with $\text{logit}(p_j) = \alpha_p + \beta_p \cdot v_j$. As will be shown in section 2.2, in the first case, parameters N and p are only estimable as a product and therefore the model provides estimates of the total number of individuals that were detected at least once but not estimates of the "super-population" size. In the latter case, simulations presented in section 3 suggest that the model becomes more "data-hungry" and richer data sets with higher counts from more sites and maybe better separated groups are required for its estimates to be reliable.

If data sets from multiple sites are available, then one can use an integrated modelling approach to analyse them simultaneously. The data set now consists of matrix $\mathbf{Y}$ with entry y_{ij} equal to the count obtained at site i on sampling occasion j. Specifically, for data sets collected at S sites the likelihood becomes $L(\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{p}|\mathbf{Y}) = \prod_{i=1}^{S} \prod_{j=1}^{T} \left[\frac{\exp(-\lambda_{ij})\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right]$, where λ_{ij} is the Poisson mean for site i on occasion j.

The number of estimable parameter combinations now increases, with a maximum of ST if there are no empty cells in $\mathbf{Y}$, which allows for some of the assumptions for the model parameters to be relaxed. For example, the mean arrival times can be modelled in terms of a site-specific covariate, z: $\log(\mu_{im}) = \alpha_m + \beta_\mu \cdot z_i$, where α_m, $m = 1, \ldots M$ is the log mean arrival time of group m with $z_i = 0$ and β_μ is the shift of all means to a direction indicated by its sign when z_i changes. A similar approach can be employed for modelling the relative weights of the arrival groups. The precision of the estimates is expected to increase as the number of sites increases if the expectations of the model for each site have a number of parameters in common. However, parameters $N_1, \ldots, N_S$ are again only estimated each as a product with p when the latter is assumed constant for all sites and sampling occasions.

All of the simulation and data analysis results presented in this paper were obtained using R Core Team (2013) and computer code to perform the model-fitting is available from the first author upon request.

2.2 Parameter Redundancy

If one or more parameters in a model cannot be estimated, then the model is termed parameter redundant and is non-identifiable. Catchpole & Morgan (1997) showed that for a model from the exponential family of distributions, such as the Poisson model described in section

2.1, one can identify whether it will be parameter redundant for all data sets by calculating the rank of a matrix of first derivatives. Specifically, the number of estimable parameter combinations in the model for data from a single site is equal to the rank of matrix $\mathbf{D}$ with entries

$$D_{lk} = \frac{\partial \lambda_k}{\partial \theta_l},\ k = 1, \ldots, T,\ l = 1, \ldots, q$$

where $\boldsymbol{\lambda}$ is the vector of means of the model, in this case of length T, and $\boldsymbol{\theta}$ is the vector of parameters, of length q. The rank of $\mathbf{D}$, r, is less than or equal to T if there are no missing data and less than or equal to K otherwise. If r is less than q then the model is parameter redundant and it is not possible to identify unique maximum likelihood estimators for at least some of its parameters. If on the other hand $r = q$ then the model is termed full rank.

Symbolic algebra packages, such as Maple, can be used to calculate the entries of $\mathbf{D}$ as well as its rank. However, if the model structure is too complex, Maple can run out of memory when calculating r symbolically. To deal with this limitation, Choquet & Cole (2012) proposed a hybrid symbolic-numerical approach where the entries of $\mathbf{D}$ are found symbolically but its rank is calculated numerically for values of the parameters randomly chosen from the parameter space. If r is calculated less than q then any zero entries in the numerical estimation of the left kernel of $\mathbf{D}$ suggest the parameters that are estimable. As Choquet & Cole (2012) point out, a point chosen at random from the parameter space, especially if this choice is poor, can result in r being estimated as smaller than the actual model rank, and they therefore suggest choosing around 5 sets of random values and repeating the procedure for each set. The model rank is equal to the largest value for r obtained from these 5 repetitions.

For data collected at S sites, the number of estimable parameter combinations cannot exceed ST. If detection probability is assumed constant across all sites and all sampling occasions then $N_1, \ldots, N_S$ only appear as a product with p and therefore are not estimable separately from it. This result is also verified by adopting the aforementioned symbolic-numerical methods for all specifications of the other model parameters introduced in section 2.1.

Parameters $N_1, \ldots, N_S$ and p do however become separately estimable if p is modelled using a time-varying covariate. This finding is similar to that shown by Cole & Choquet (2013) who incorporated random effects to separate confounded parameters in capture-recapture models. Similarly to Cole & Choquet (2013), we have found that if either the covariate used to model p does not vary considerably across the samples, or its effect on p is not statistically significant, then the model becomes near-parameter redundant, which means that even though it is in theory full rank it actually behaves like a parameter redundant model in practice.

Although these results do not need to be reproduced when the models are fitted to data, the Maple (Maplesoft, Waterloo, Canada) code used to derive these results is available upon request from the first author. Simulations shown in section 3 that explore the different model specifications suggest that as the model becomes more complicated, and especially when N and p are separately estimated, the data set needs to be richer for the model to perform adequately. As is usually the case with sparse data sets, results that hold in theory might not be true in practice in terms of the estimable parameter combinations in the model and results obtained by analysing sparse data sets should be treated with caution.

2.3 Model-fitting considerations

Different starting values for the parameters in mixture models can yield different local maxima since the surface of a mixture model likelihood may be multimodal. This implies that in this case different starting values for the mean arrival times of the groups could lead to different results. It is recommended that the optimisation algorithm is started from a number of different values to ensure a wide search and to obtain a number of different local maxima from which to choose the best i.e. the one that results in the highest likelihood value.

The starting value for each arrival mean can be randomly sampled from the possible arrival times, which are all values between 1 and T. Those for the standard deviations of the arrival groups can be chosen to be large, eg. 5-6 depending on the length of the study, to eliminate as much as possible the appearance of spurious maximisers which may result from the fact that the likelihood for mixtures of heteroscedastic normal distributions does not have a global maximum value and continues to increase when one, or more, of the values of the variances of the groups decrease. These maximisers often lead to singularities in the variance-covariance matrix.

For a detailed description of the issues of multimodality and spurious maximisers see McLachlan & Peel (2000) and in particular sections 2.12 and 3.10.

2.4 Goodness of fit

The expected number of individuals detected at site i on occasion j, $\widehat{y}_{ij}$ is equal to $\widehat{y}_{ij} = \widehat{N}_i \widehat{p}_{ij} \sum_{b=1}^{j} \left[\widehat{\beta}_{i,b-1} \left\{\prod_{j=b}^{j-1} \widehat{\phi}_{i,ja}\right\}\right]$, $a = j - b + 1$, where $\widehat{\beta}_{i,b-1}$ and $\widehat{\phi}_{i,ja}$ are the estimated entry and retention probabilities for site i, respectively.

The residual deviance of the fitted Poisson model can be used to assess its fit. However, when a number of cell counts in $\mathbf{Y}$ are low then the asymptotic distribution of the residual deviance may not be χ^2 anymore, but the fit can be assessed less formally by plotting the observed and fitted values against or alongside one another.

2.5 Stopover duration

The mean stopover duration at site i, MSD_i, is equal to $MSD_i = \sum_{b=1}^{T}\sum_{d=b}^{T}(d-b+1)\beta_{i,b-1}\left(\prod_{j=b}^{d-1}\phi_{i,ja}\right)(1-\phi_{i,d(d-b+1)})$, $a = j-b+1$, where $d = b,\ldots,T$ are the possible exit times from the population for an individual that entered on occasion b. In the case of bivoltine insect species, this is the average duration across both broods. In cases where the population is closed to migration, this will generally be the average (adult) lifespan of an individual, although for the few species that overwinter as adults individuals emerge from, and may leave the study into, a state of diapause.

3 Simulations

This section presents an extensive simulation study which examines the performance of the model for a wide range of assumptions for the parameters. The simulations are divided in two sections: in section A, the fitted models have $p_{ij} = p \; \forall i, j$ and therefore parameters $N_1, \ldots, N_S$ are only estimated as a product with p, while in section B, detection probabilities are logistically regressed on an artificial covariate, generated from a Unif[5,15] for all sites/occasions, and estimates of $N_1, \ldots, N_S$, separate from detection probabilities, are obtained.

3.1 Section A

Simulation A1 sets $S = 10$, $T = K = 15$, $N = (609, 869, 659, 848, 553, 346, 871, 875, 227, 545)$, $M = 2$ with $\mu_{i,1} = 2$, $\mu_{i,2} = 7$, $w_{i,1} = 0.4$, $w_{i,2} = 0.6$ and $\sigma_{i,1} = \sigma_{i,2} = 1\ \forall i$, $p_{i,j} = 0.2$ and $\phi_{i,j} = 0.6\ \forall i, j$. Fig. 2 (a) shows the counts obtained in one simulation run for all sites. Figs 2 (b) and (c) demonstrate that the model provides satisfactory estimates of both the numbers of individuals detected once at each site, as well as of the entry parameters.

[Figure 2 about here.]

This simulation is used as a baseline for evaluating several extensions of the model. The results are shown as supplementary material. In simulation A2, observations are deleted at random and hence at any site $K \neq T$, resulting in around 20% of the data being missing (Fig. S1). In simulations A3, A4 and A5, retention probabilities are, respectively, a function of calendar time, a function of age and a function of the square of calendar time (Figs S2, S3, S4). Note that in simulation A4 T is set equal to 20 since the second group remains for longer than in simulation A3. In simulation A6, w_1 is logistically regressed on a fictitious covariate (Fig. S5) while in simulation A7, the logarithms of μ_1 and μ_2 are regressed on a fictitious covariate (Fig. S6). The case of heteroscedastic arrival groups is examined in simulation A8 (Fig. S7) while, finally, simulation A9 sets $M = 3$ with $\mu = (2, 6, 10)$ and $w = (0.4, 0.5, 0.1)$ (Fig. S8).

The results suggest that the model performs well in all of these cases. When the relationship between $\text{logit}(\phi)$ and calendar time is quadratic, there is greater uncertainty for the part of the curve that corresponds to the early sampling occasions, compared to the case when $\text{logit}(\phi)$ is linearly dependent on time. Similarly, when $\text{logit}(\phi)$ depends linearly on age, there is more uncertainty in the part of the curve that corresponds to the older individuals.

The validity of model selection criteria, such as the Akaike Information Criterion (AIC) (Akaike, 1973), in choosing the number of mixture components is doubtful because of violation of regularity conditions (McLachlan & Peel, 2000, Chapter 6). However, their use has gained support in the literature, for example in Cubaynes et al. (2012). We performed a small simulation study to examine the performance of AIC in choosing the right value for M when M is set equal to 1, 2 or 3. Specifically, we simulated data with $M = 1$, $\mu = 5$, $\sigma = 1$ or with $M = 2$, $w_1 = 0.4$, $w_2 = 0.6$, $\mu_1 = 3$, $\mu_2 = 7$ and $\sigma_1 = \sigma_2 = 1$, or with $M = 3$, $w_1 = 0.4$, $w_2 = w_3 = 0.3$, $\mu_1 = 3$, $\mu_2 = 7$, $\mu_3 = 11$ and $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and all other parameters set as in the baseline simulation. For each set, we fitted models with $M = 1, 2, 3$ and used AIC to choose between them. The number of times each model is selected for each set, out of 100 simulations, given below, suggests that in this case AIC successfully selects the right value for M in the majority of cases:

[Table 1 about here.]

3.2 Section B

If detection probabilities are allowed to vary according to a fictitious site- and time-varying covariate, then the model also requires richer data sets with higher counts in order to perform adequately. For example, if the average detection probability is 0.2 and all other parameters are as in simulation A1, then the median relative bias (MRB) in the estimates for N is around -9% for all sites. If detection probabilities are set on average equal to 0.7, then the MRB falls to around 5% for all sites. If N doubles for all sites compared to simulation A1, then

MRB is around 3.5% when average detection probability is 0.2, and only 0.1% when average detection probability is 0.7. The results of the latter simulation (B1), are presented in Fig. 3 together with the counts obtained in one of the simulation runs at all sites.

[Figure 3 about here.]

Further simulation results are shown as supplementary material. In particular, simulation B2 explores the case when 20% of the counts are missing (Fig. S9), simulation B3 has ϕ logistically regressed on time (Fig. S10) while simulation B4 has ϕ logistically regressed on age (Fig. S11). Note that in simulations B3 and B4 the groups are better separated for the model to perform well with $\mu_1 = 2$ and $\mu_2 = 9$ and also in simulation B4 $T = 20$, as was the case in simulation A4.

4 Application to UKBMS count data

Butterfly counts are characterised by their high variability throughout the season, representing the different patterns of emergence for each species. Different species of butterfly exhibit varying levels of voltinism, with one, two or more generations per year.

The UKBMS consists of counts made weekly from the beginning of April until the end of September using the transect method, which is described in depth in Pollard & Yates (1993). Transects are typically 2-4 km long and walked within specified periods of the day and when weather conditions are suitable for butterfly activity. The scheme design allows for counts to be made throughout the season for butterfly activity, during which abundance will vary according to different seasonal patterns of emergence.

We apply the model introduced in section 2.1 to UKBMS data for the Common blue *Polyommatus icarus*, collected in 2010. This species is known to exhibit bivoltine populations in the south of the UK, whilst populations become single brooded in the north. However, a precise latitude at which this occurs or knowledge of how this boundary may have changed over time are both unknown (Asher et al., 2001).

We considered $M = 2$ homoscedastic normal mixture distributions for the arrival of the butterflies at the sites. For computational efficiency, data were limited to a random sample of 50 monitored sites, excluding sites where more than 6 counts were missing from the season or the sum of the counts made was less than 10. Common Blue overwinters as a caterpillar, and is therefore not seen in flight until late spring. The start of the season was defined as the week with the first positive count, with season length totalling 21 weeks.

Model comparison was made for varying parameter assumptions using AIC. The mixture means and weights were estimated as either constant, or as a function of site Northing, as described in section 2.1. Retention probabilities were modelled as constant, or as logistically dependent on calendar time, age or calendar time squared. Detection probabilities were set either as constant and common across sites or as logistically dependent on temperature at the site on the day of sampling, which is also recorded by the data collector. Missing temperature records were replaced by the average of neighbouring sites. Each model was started from ten different random starts for the parameters to determine the optimal local maximum, as discussed in section 2.3 and all covariates were standardised.

Table 2 provides the AIC values and the number of parameters of the models considered. The two models with the greatest support have the weights and mixture means dependent on Northing and the logits of retention probabilities as dependent on the square of calendar time. The most favoured model also has the logits of detection probabilities dependent on temperature and an AIC value considerably lower than that of the second most favoured model, which has a constant detection probability across sites and time.

[Table 2 about here.]

Parameter estimates and associated standard errors for the preferred model are given in Table 3 and they are similar to those derived from the second best model, shown in Table S1.

[Table 3 about here.]

The residual deviance of the selected model is D=3952, with $(n-p)$=924-61=863 which implies a moderate lack of fit and dispersion estimated as approximately 4.58. However, comparison of the observed counts with estimated fitted counts from the model show reasonable correspondence for most sites, implying overdispersion rather than a failure in the model structure (Figure S12) and all standard errors have been adjusted for overdispersion.

The estimated retention probabilities, shown in Fig. 5(a), peak around week 11 of 21, before dropping off towards the ends of the season. They are estimated as approximately zero for the initial weeks of the season. The 95% confidence intervals constructed around the logit of retention probabilities demonstrate the greater uncertainty for the part of the curve that corresponds to the start of the season (Fig. S13). This is because the obtained counts are considerably low in the first few weeks and the first few columns of the data set are very sparse. A similar result was observed for simulated data, as mentioned in section 3.

The weighting of the first normal distribution increases with Northing, with the second brood almost disappearing in the North (Fig. 5(b)). The means of the two normal distributions suggest a later time of emergence in the North (Fig. 5(c)). This is also demonstrated by the entry parameters; two relatively even broods at southern sites, with the first brood dominating at high Northing, in addition to a later emergence (Figure 4). The 95% confidence intervals shown in Fig. 5(b) and Fig. 5(c) are constructed using the Delta method in R package *msm* Jackson (2011).

[Figure 4 about here.]

Finally, the estimated "super-population" sizes for all 50 sites are shown in Fig. 5(d), together with their asymptotic 95% confidence intervals back-transformed from the log-scale. The black dots are the estimates of Np, that is the product of the "super-population" sizes and detection probability, when that is assumed constant and common across all sites, derived by the second best model, as shown in Table 2. As expected, and especially for sites with higher estimated N, these two point estimates are far apart, with N greater than Np.

[Figure 5 about here.]

5 Discussion

Simple series of annually replicated counts, made to a standardised protocol, are essential for conservation monitoring. Through butterfly counts of this kind, for example, have been identified not only widespread declines in some species, but also several marked successful responses to targeted management (Thomas, Simcox & Hovestadt, 2011). Optimal methods of analysing such data would also shed light on the factors behind the population changes, by estimating demographic and phenological changes alongside the population trends.

Many approaches have been adopted for comparing relative abundances at different locations, or different points in time. Simple Poisson models are often applied to data in the form of multiple, often short and incomplete, time series arising from standardised survey protocols. A specific problem with butterfly data, such as those of the UKBMS, is the need to account for the seasonal patterns - inevitable in the counts - which mean that expected counts, even at an individual site, vary greatly within a season. To date, this seasonal pattern has usually been estimated via a Generalized Additive Model (GAM: Rothery & Roy (2010); Dennis et al. (2013)). Such models however rely on interpolating any missing values, and as such estimate the total numbers of sightings, rather than individuals.

A GAM approach is clearly non-parametric and empirical. Our proposed method is still Poisson-based (although other distributions might be considered if preferred), but seeks to explain the changing counts within a season via models which are both biologically realistic, and of considerable value in conservation management. Missed visits within the season are easily encompassed due to the use of parametric functions to constrain the model parameters, and we can now estimate more flexibly a number of quantities, in addition to abundance or relative abundance. Additionally, we can ensure that indices of abundance are not biased due to differences in seasonal flight periods or multiple sightings of individuals within a season. The retention/survival probability ϕ is a demographic variable that can be converted to the estimated flight-period (adult life-expectancy) of an individual; through the arrival parameter β alone we are informed about phenological change (the average time of arrival/emergence) and, importantly, the relative strengths of two (or more) broods in multivoltine species. Extensive simulations have shown that the model can be expected to perform well with data of a scale readily achievable in practice.

Note that the method includes the pioneering work of Zonneveld (1991) as a special case, in which data from single sites are modelled individually with a constant rate of survival. The number of parameters estimated restricts the capacity for useful inference at a single site and season (see Calabrese, 2012). Extension to simultaneous analyses at multiple, potentially many, sites which may be expected to share at least some parameters allows "borrowing strength" and improved inference on key ecological parameters - the assumption of constant survival can be tested in a robust framework, for example. Although we have chosen normal distributions for the arrival times, alternatives are readily adopted: Zonneveld (1991) considered logistic, and Cornulier et al. (2009) also used asymmetric distributions within mixture models to permit a degree of skewness in the hatching dates of birds from monitored nests. Clearly the model is also straightforward to apply to univoltine species, though we have concentrated on the bivoltine case here due to its special interest and difficulties, and to illustrate the connection with the widely-used stopover models in other contexts. In the latter, arrival and departure of (marked) individuals from a location are usually considered as immigration and emigration, though the mathematics is clearly analogous. The butterflies in our data are of course not individually identifiable, and this prohibits the estimation of genuine abundance when detection probability is assumed constant, something which is possible in conventional stopover models. Nonetheless, in this case the model confounds abundance and detectability, so if the latter can be assumed constant comparable measures of relative abundance arise, and are of use in management and monitoring.

The ability to estimate changes in phenology, demography and voltinism along with abundance provides a rigorous statistical basis for comparisons of these, and their relationships with environmental covariates or with one another. Thus, for example, in a species such as the Common Blue we are able not only to estimate abundance/relative abundance at each site, but to apportion this between two broods. Use of a covariate (site Northing) confirms, and quantifies, long-standing field experience that the species is effectively single-brooded in the far North of its range although the two broods are comparable in the South, and that the broods emerge later with increasing Northing, presumably as a consequence of later spring/summer conditions. Given that, for simplicity, survival is often assumed constant in modelling butterfly populations, e.g. Soulsby & Thomas (2012), and in the absence of evidence for senescence, it is interesting that for Common Blue models with age-dependence in survival fare poorly. However, the evidence for variation with time across the season is pronounced.

Butterfly data are regarded as especially useful environmental indicators. They are, for example, relatively easy to collect via 'citizen science' schemes as they are visible, popular with the public and, in the UK at least, species are few and largely easy to locate and identify, compared to many invertebrate groups. The method is, however, equally applicable to other seasonal invertebrate species.

The sensitivity of butterflies to climatic or land-use changes makes them useful indicators of the effects on wider biodiversity, as shown by the adoption of the UKBMS into governmental indicators of biodiversity trends in the UK (Defra, 2011) and beyond (van Swaay, Nowicki, Settele & van Strien, 2008). Given the demonstrated utility of butterfly surveys in studying, for example, climate change (Roy & Sparks, 2000; Roy et al., 2001) or consequences of agricultural practice (Woodcock et al., 2012; Jonason et al., 2011) and, for specialist species, habitat fragmentation (Brückmann, Krauss & Steffan-Dewenter, 2010) we believe that the greater flexibility and robustness of the models described here will greatly increase the value of such surveys in the future.

The method has important applications in conservation biology as it enables absolute abundance of an insect species seasonal population to be estimated from strip transects. This is without the need for intrusive and labour intensive marking techniques or technically demanding distance sampling that involves counting butterflies in distance bands. Phenology and abundance are modelled simultaneously, and mean date of emergence, which is a new statistic for butterflies, can be used to monitor species responses to climate change. Altitude and aspect are also thought to affect butterfly phenology and it is straightforward to include these in the model in the future. Detectability was modelled as a function of temperature at the site on the day of sampling, but other covariates such as habitat type, recorder effort, experience or age can be also incorporated if available.

Acknowledgments

We are grateful to Diana Cole for providing us with Maple code for the Choquet & Cole (2012) hybrid symbolic-numerical method and for all the useful discussions on parameter identifiability and to Steve Buckland and Byron Morgan for all their insightful suggestions.

The UKBMS is operated by the Centre for Ecology & Hydrology and Butterfly Conservation and funded by a multi-agency consortium including the Countryside Council for Wales, Defra, the Joint Nature Conservation Committee, Forestry Commission, Natural England, the Natural Environment Research Council, the Northern Ireland Environment Agency and Scottish Natural Heritage. The UKBMS is indebted to all volunteers who contribute data to the scheme.

Supporting Information

Additional Supporting Information may be found in the online version of this article (Supplementary material for "Monitoring abundance and phenology in (multivoltine) butterfly species; a novel mixture model"):

Section 1. Simulation results

Section 2. UKBMS data results

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. B. N. Petrov, and F. Caski, (eds.) Proceeding of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest., 267–281.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G., & Jeffcoate, S. (2001). The Millennium Atlas of Butterflies in Britain and Ireland. Oxford University Press.

Botham, M. S., Brereton, T. M., Middlebrook, I., Randle, Z. & Roy, D. B. (2011). United kingdom butterfly monitoring scheme report for 2010. Tech. rep., CEH, Wallingford.

Brückmann, S. V., Krauss, J. & Steffan-Dewenter, I. (2010). Butterfly and plant specialists suffer from reduced connectivity in fragmented landscapes. Journal of Applied Ecology, 47, 799–809.

Calabrese, J. M. (2012). How emergence and death assumptions affect count-based estimates of butterfly abundance and lifespan. Population Ecology, 54, 431–442.

Catchpole, E. A. & Morgan, B. J. T. (1997). Detecting parameter redundancy. Biometrika, 84, 187–196.

Choquet, R. & Cole, D. J. (2012). A hybrid symbolic-numerical method for determining model structure. Mathematical Biosciences, 236, 117–125.

Cole, D. J. & Choquet, R. (2013). Parameter redundancy in models with individual random effects. Tech. rep., School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury.

Cornulier, T., Elston, D. A., Arcese, P., Benton, T. G., Douglas, D. J. T., Lambin, X., Reid, J., Robinson, R. A. & Sutherland, W. J. (2009). Estimating the annual number of breeding attempts from breeding dates using mixture models. Ecology Letters, 12, 1184–1193.

Crosbie, S. F. & Manly, B. F. (1985). Parsimonious modelling of capture-mark-recapture studies. Biometrics, 41, 385–398.

Cubaynes, S., Lavergne, C., Marboutin, E. & Gimenez, O. (2012). Assessing individual heterogeneity using model selection criteria: how many mixture components in capture-recapture models? Methods in Ecology and Evolution, 3, 564–573.

Defra (2011). UK biodiversity indicators in your pocket 2011. Published by Defra on behalf of the UK Biodiversity Partnership, Defra, London.

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

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. Journal of Statistical Software, 38, 1–29.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-Stochastic model. Biometrika, 52, 225–247.

Jonason, D., Andersson, G. K. S., Ockinger, E., Rundlof, M., Smith, H. G. & Bengtsson,

J. (2011). Assessing the effect of the time since transition to organic farming on plants and butterflies. Journal of Applied Ecology, 48, 543–550.

Matechou, E., Morgan, B. J. T., Pledger, S., Collazo, J. E. & Lyons, J. A. (2013). Integrated analysis of capture-recapture-resighting data and counts of unmarked birds at stop-over sites. JABES, 18, 120–135.

McLachlan, G. & Peel, D. (2000). Finite Mixture Models. Wiley.

Pledger, S., Efford, M., Pollock, K. H., Collazo, J. A. & Lyons, J. E. (2009). Stopover duration analysis with departure probability dependent on unknown time since arrival. Environmental and Ecological Statistics (Edited by D.L.Thomson, E.G.Cooch and M.J. Conroy), 3, 349–363.

Pollard, E. & Yates, T. J. (1993). Monitoring butterflies for ecology and conservation: the British Butterfly Monitoring Scheme. Chapman & Hall, London, UK.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rothery, P. & Roy, D. B. (2010). Application of generalized additive models to butterfly transect count data. Journal of Applied Statistics, 28, 897–909.

Roy, D. B., Rothery, P., Moss, D., Pollard, E. & Thomas, J. A. (2001). Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. Journal of Animal Ecology, 70, 201–217.

Roy, D. B. & Sparks, T. H. (2000). Phenology of british butterflies and climate change. Global Change Biology, 6, 407–416.

Schwarz, C. J. & Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. Biometrics, 52, 860–873.

Seber, G. A. F. (1965). A note on the multiple-recapture census. Biometrika, 52, 249–259.

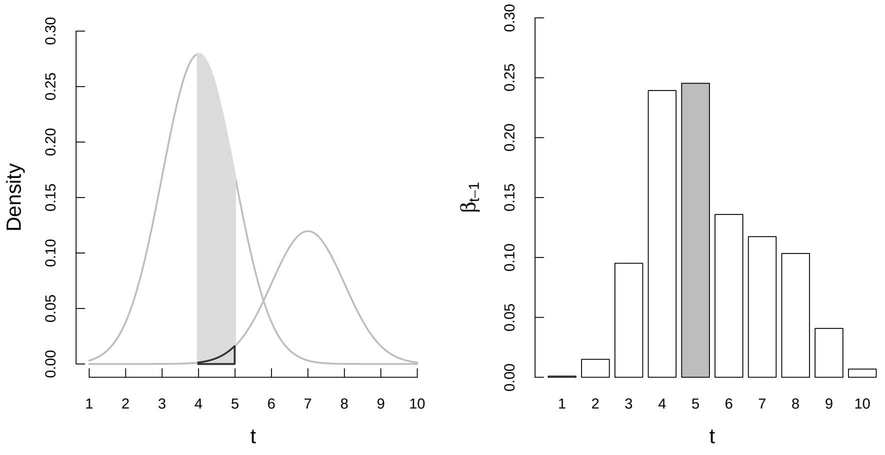
Soulsby, R. L. & Thomas, J. A. (2012). Insect population curves: modelling and application to butterfly transect data. Methods in Ecology and Evolution, 3, 832–841.

Thomas, J. A., Simcox, D. J. & Hovestadt, T. (2011). Evidence-based conservation of butterflies. Journal of Insect Conservation, 15, 241–258.

van Swaay, C. A. M., Nowicki, P., Settele, J. & van Strien, A. J. (2008). Butterfly monitoring in Europe: methods, applications and perspectives. Biodiversity and Conservation, 17, 3455–3469.

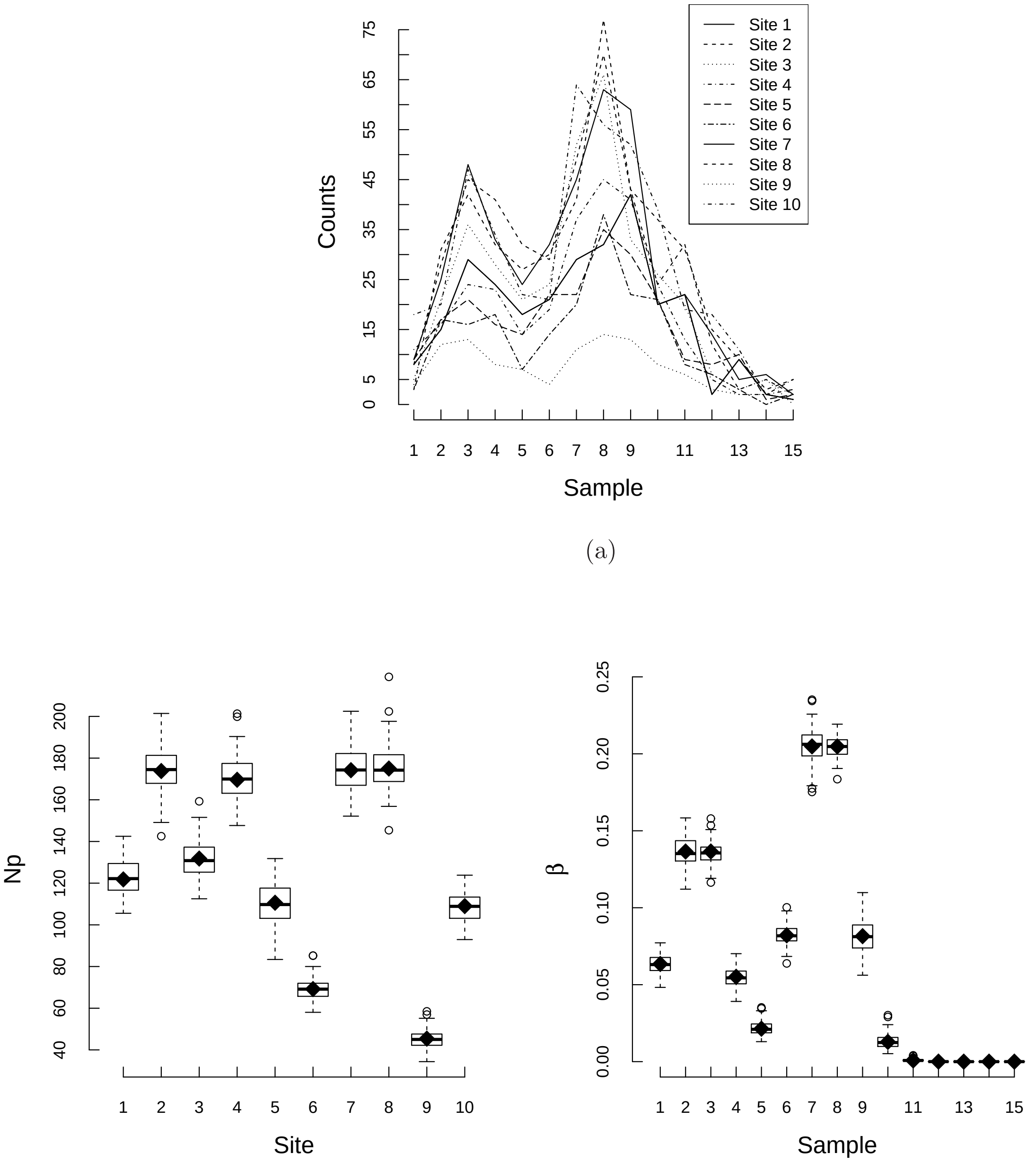
Woodcock, B. A., Bullock, J. M., Mortimer, S. R., Brereton, T., Redhead, J. W., Thomas, J. A. & Pywell, R. F. (2012). Identifying time lags in the restoration of grassland butterfly communities: A multi-site assessment. Biological Conservation, 155, 50–58.

Zonneveld, C. (1991). Estimating death rates from transect counts. Ecological Entomology, 16, 115–121.



(a) (b)

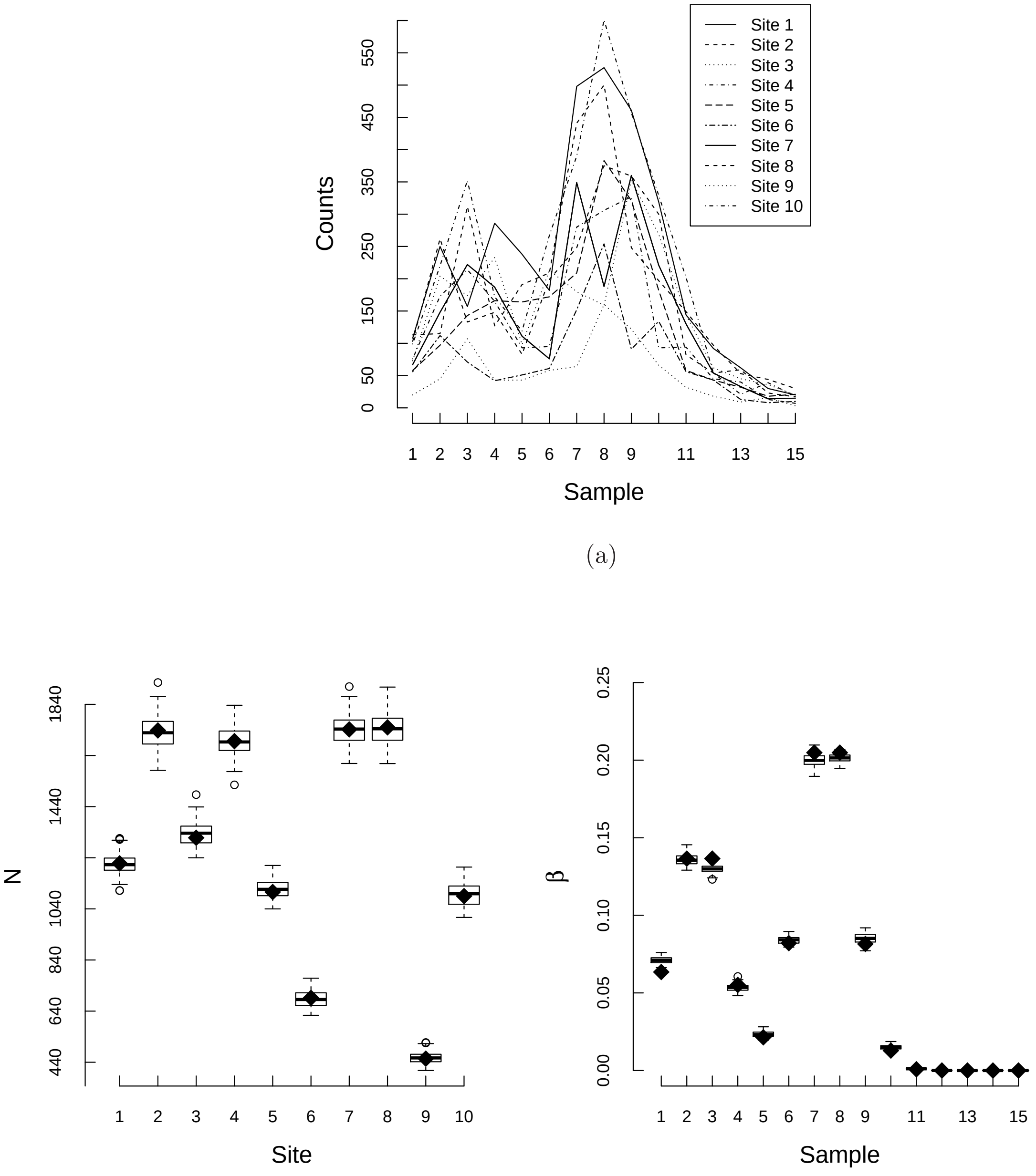
Figure 1: (a): Two normal densities with $\mu_1 = 4$, $\mu_2 = 7$ and $\sigma_1 = \sigma_2 = 1$ with corresponding weights equal to 0.7 and 0.3. If an observation, x_1, is drawn randomly from a $N(4, 1)$ distribution then $0.7 \times P(4 \leq x_1 \leq 5)$ is given by the gray shaded area (0.7×0.34) while if an observation x_2 is drawn from a $N(7, 1)$ distribution then $0.3 \times P(4 \leq x_2 \leq 5)$ is given by the area inside the black lines (0.3×0.02). (b): The resulting β parameters. For example, β_4, which is shaded in gray, is given by $0.7 \times 0.34 + 0.3 \times 0.02 = 0.244$.



(a)

(b) (c)

Figure 2: Simulation A1. (a): Obtained counts from one simulation run for all sites. (b) and (c): Box-plots of derived estimates for Np and β from 100 replications and true values, indicated by the black diamonds.



(a)

(b)

(c)

Figure 3: Simulation B1. Derived estimates for N and β from 100 replications. The true values are indicated by the black diamonds.

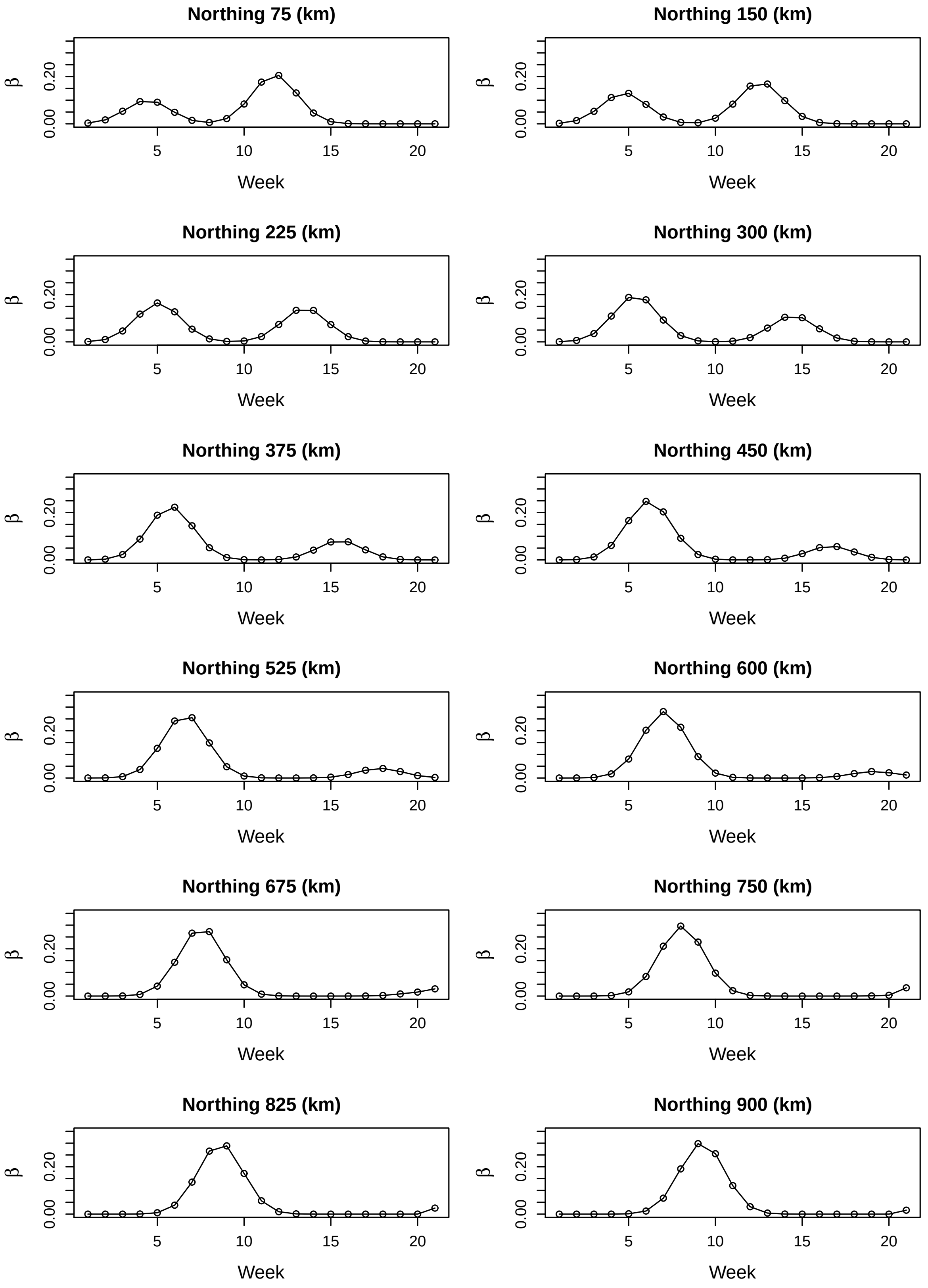
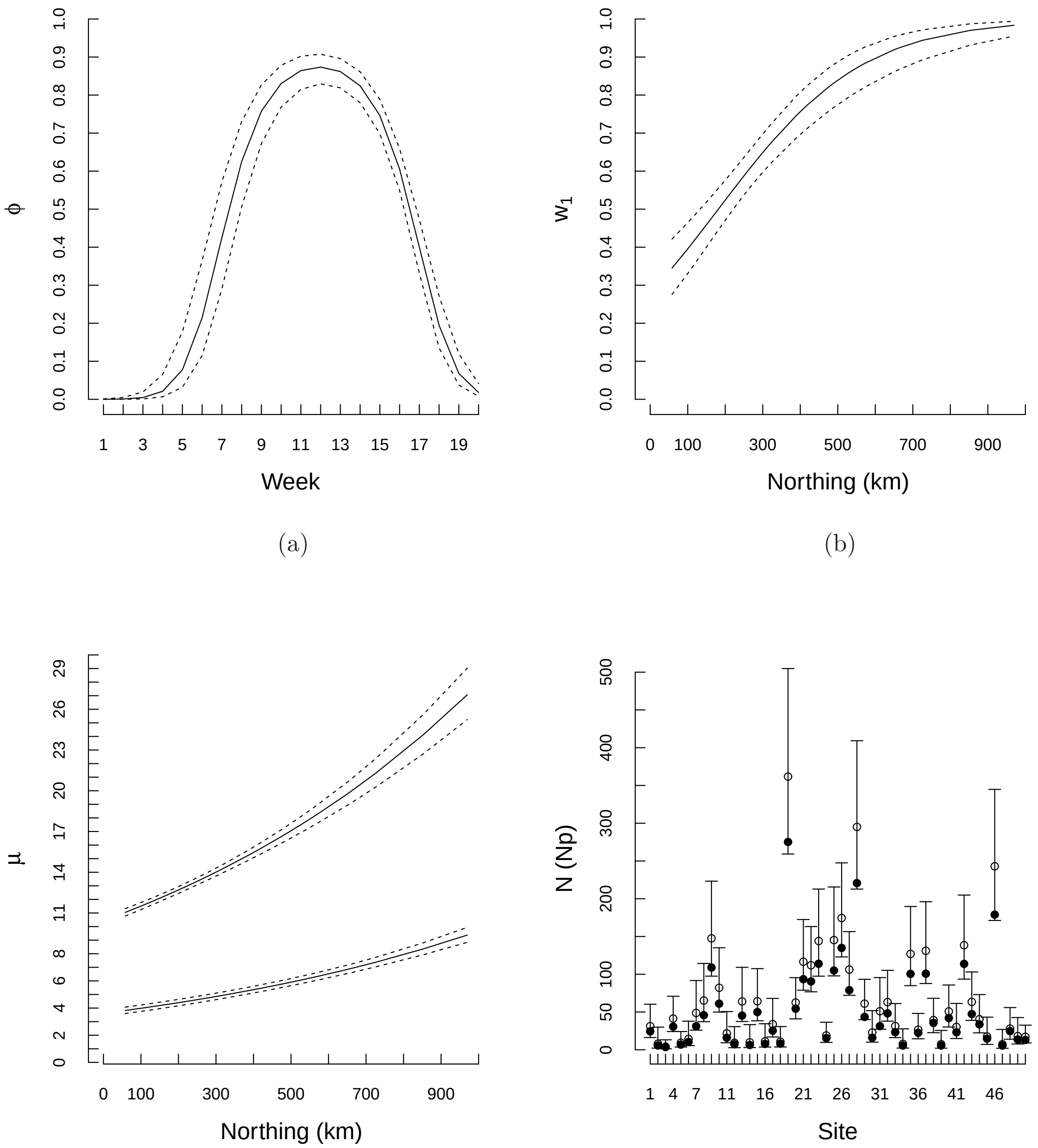


Figure 4: Estimated entry parameters for Common Blue in the UK, summer 2010, at a sample of Northing values.



(a) (b)

(c) (d)

Figure 5: Point estimates obtained by the selected model together with 95% confidence intervals : (a) estimated retention probabilities (common across sites), (b) relative size of the first brood, (c) mean arrival times of the two broods, by Northing, and (d) estimated "super-population" sizes, N, of the 50 sites. The black dots in (d) are the point estimates of Np obtained by the second best model in terms of AIC.

Table 1: Each cell corresponds to the number of times each value of M was chosen by AIC, out of 100 simulations, when the true value of M is the one indicated in the first column

	Chosen M		
True M	**1**	**2**	**3**
1	83	8	9
2	0	85	15
3	0	27	73

Table 2: Model comparisons from fitting a range of models to the Common blue data. Here ℓ denotes the value of the log likelihood evaluated at the maximum likelihood estimate of the parameters. *north*: Northing, *temp*: temperature.

Model	ℓ	No .of parameters	AIC
$w_m(north)\mu_m(north)\sigma(.)\phi(t+t^2)p(temp)$	-2880.01	61	5882.03
$w_m(north)\mu_m(north)\sigma(.)\phi(t+t^2)p(.)$	-2923.42	59	5964.85
$w_m(.)\mu_m(north)\sigma(.)\phi(t+t^2)p(.)$	-3161.16	58	6438.31
$w_m(north)\mu_m(north)\sigma(.)\phi(a+a^2)p(temp)$	-3232.41	61	6586.81
$w_m(north)\mu_m(north)\sigma(.)\phi(t)p(.)$	-3260.33	58	6636.65
$w_m(north)\mu_m(north)\sigma(.)\phi(t)p(temp)$	-3258.37	60	6636.73
$w_m(north)\mu_m(north)\sigma(.)\phi(a+a^2)p(.)$	-3280.17	59	6678.34
$w_m(north)\mu_m(north)\sigma(.)\phi(a)p(temp)$	-3306.02	60	6732.04
$w_m(north)\mu_m(north)\sigma(.)\phi(.)p(temp)$	-3316.60	59	6751.19
$w_m(north)\mu_m(north)\sigma(.)\phi(a)p(.)$	-3337.4	58	6790.80
$w_m(north)\mu_m(north)\sigma(.)\phi(.)p(.)$	-3343.42	57	6800.84
$w_m(.)\mu_m(north)\sigma(.)\phi(t+t^2)p(temp)$	-3356.25	60	6832.50
$w_m(north)\mu_m(.)\sigma(.)\phi(t+t^2)p(temp)$	-3374.32	60	6868.63
$w_m(.)\mu_m(north)\sigma(.)\phi(.)p(temp)$	-3473.19	58	7062.38
$w_m(north)\mu_m(.)\sigma(.)\phi(t+t^2)p(.)$	-3559.7	58	7235.39
$w_m(.)\mu_m(north)\sigma(.)\phi(.)p(.)$	-3566.72	56	7245.43
$w_m(north)\mu_m(.)\sigma(.)\phi(.)p(temp)$	-3741.07	58	7598.15
$w_m(.)\mu_m(.)\sigma(.)\phi(t+t^2)p(temp)$	-3837.59	59	7793.17
$w_m(north)\mu_m(.)\sigma(.)\phi(.)p(.)$	-3849.68	56	7811.36
$w_m(.)\mu_m(.)\sigma(.)\phi(.)p(temp)$	-3971.42	57	8056.84
$w_m(.)\mu_m(.)\sigma(.)\phi(t+t^2)p(.)$	-4000.97	57	8115.94
$w_m(.)\mu_m(.)\sigma(.)\phi(.)p(.)$	-4045.26	55	8200.51

Table 3: Parameter estimates from the most favoured model in terms of AIC.

Parameter	Estimate	Standard error
Logit of detection probabilities, as a function of temperature		
Intercept	-4.600	1.769
Slope	5.522	2.170
Logit of relative weight of the first brood, as a function of Northing(standardised)		
Intercept	1.186	0.160
Slope	1.101	0.150
Log of mean emergence times of the two broods, as a function of Northing(standardised)		
Intercept (1)	1.688	0.023
Intercept (2)	2.748	0.013
Slope	0.209	0.010
σ	1.251	0.055
Logit of retention probabilities, as a function of time (t) and time squared (t^2)		
Intercept	1.743	0.193
Slope for t	1.562	0.266
Slope for t^2	-3.200	0.297