Web-based Supplementary Materials for

A generalised abundance index for seasonal invertebrates

by Emily B. Dennis, Byron J.T. Morgan, Stephen N. Freeman,

Tom Brereton & David B. Roy

Web Appendix A

Comparison of the SO_B and N_B GAIs for the Poisson model

The P/N_B GAI can be viewed as an approximation to the P/SO_B GAI. In the stopover model, P/SO_1, assuming ϕ to be constant,

$$\lambda_{i,j} = N_{SO,i}(\alpha_{i,j} + \alpha_{i,j-1}\phi + \alpha_{i,j-2}\phi^2 + \cdots + \alpha_{i,0}\phi^{j-1}), \tag{1}$$

where $N_{SO,i}$ denotes the site parameter from the stopover model for a given site i, and for a given occasion, t, $\alpha_{i,t} = F(t_{i,j}) - F(t_{i,j} - 1)$. Comparatively, for the mixture model P/N_1 GAI,

$$\lambda_{i,j} = N_{G,i}\alpha_{i,j}, \tag{2}$$

where $\{N_{G,i}\}$ are the site parameters for the mixture model and $\alpha_{i,j} = f(t_{i,j})$. Since the multiplier of $N_{SO,i}$ is greater than that for $N_{G,i}$, we find that $N_{G,i} > N_{SO,i}$.

If we consider the sum of $\lambda_{i,j}$ over j, the coefficients of ϕ in the stopover model will sum approximately to unity as they form the area under a density. An approximate geometric sum for ϕ ($\phi < 1$) remains which will produce $1/(1-\phi)$. This suggests that the site estimates will differ between the two models by a scaling factor of approximately $1 - \phi$.

We compare model performance for the P/N_2 and P/SO_2 GAIs for five bivoltine butterfly species for data from a sample of 100 UKBMS sites for 2010. Different starting values for the parameters could yield different local maxima (Matechou et al., 2014; McLachlan and

Peel, 2004), therefore each model was run from five random starting values and a comparison made of each model with the highest likelihood value.

Web Figure 1 demonstrates empirically that the estimates of N differ between the P/SO_2 and P/N_2 GAIs by a scaling factor of approximately $1-\phi$. The stopover model is generally favoured in terms of AIC and overdispersion (Web Table 1). Estimates of μ_1 and μ_2 are earlier for the stopover model than the mixture model. This result could be anticipated since the brood means in the stopover model represent the entry of individuals into the population, whereas the corresponding parameters in the mixture model consist of both individuals that have entered the population and those that have survived from previous weeks. Estimates of σ from the mixture model, which relate to the length of the flight period, are greater than from the stopover model where σ relates to the length of the emergence period. The parameter ϕ from the stopover model provides additional information compared to the P/N_2 GAI, but the stopover model takes an average of seven times longer to run.

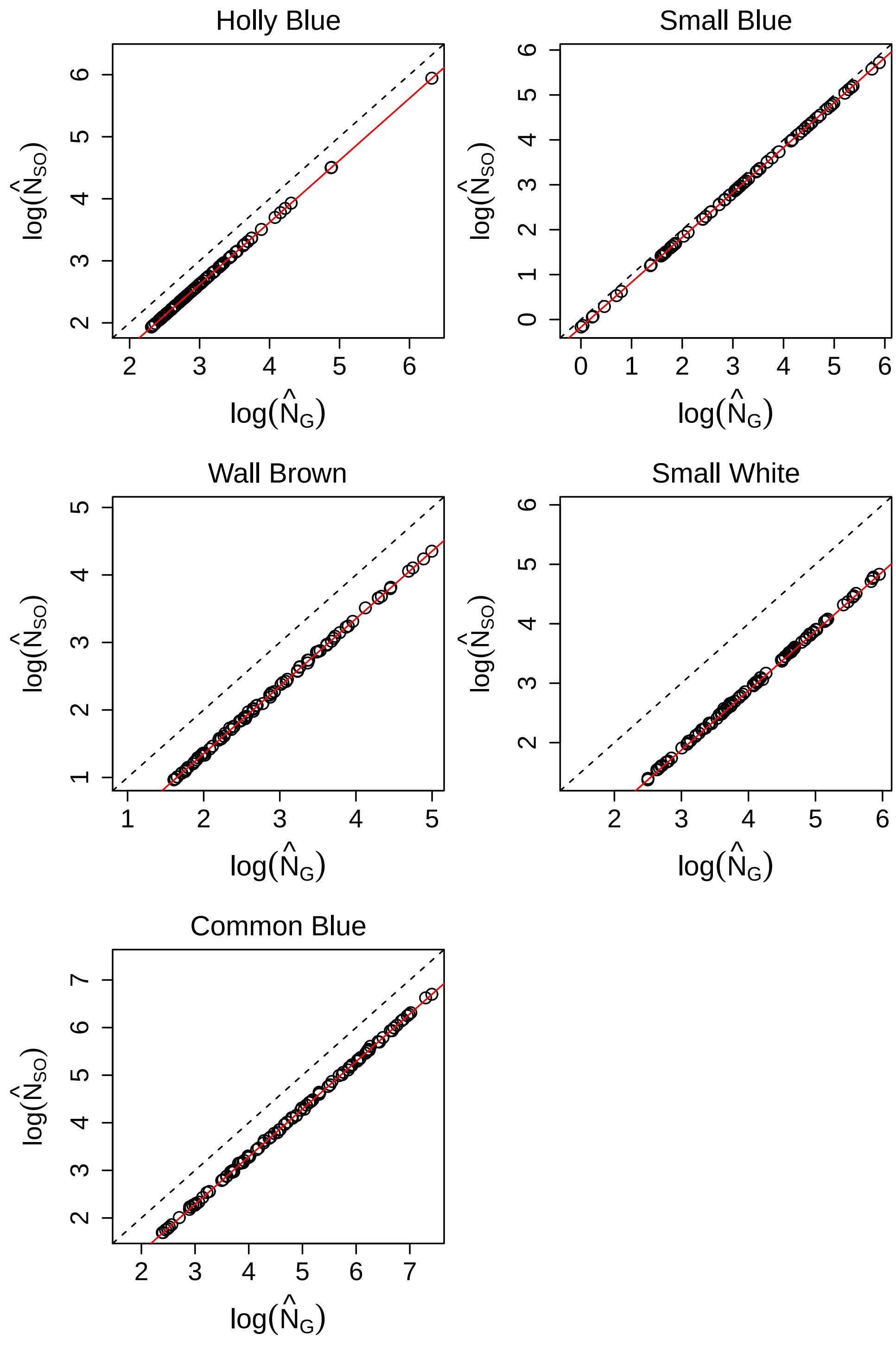
References

Matechou, E., Dennis, E. B., Freeman, S. N., and Brereton, T. (2014). Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *Journal of Applied Ecology*, 51(3):766–775.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. Wiley, New York.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1):108–115.



Web Figure 1: Comparison of estimated site parameters, $\hat{N}_G$ from the P/N_2 GAI and $\hat{N}_{SO}$ from the P/SO_2 GAI. Both axes are displayed on the log scale. The dashed line indicates the 1-1 line and the red line indicates the line with offset $\log(1-\hat{\phi})$.

Web Table 1: Parameter estimates from P/SO_2 and P/N_2 GAI for five illustrative species. Estimates are shown for the best model from five different starting values, in terms of AIC. The computation time is given in seconds, n is the number of parameters and D is the dispersion estimate (residual deviance/degrees of freedom).

a) P/SO_2 GAI															
Species	n	Time	Log(L)	AIC	w		μ_1		μ_d		σ		D	ϕ	
Holly Blue	5	2.14	-2114.3	4238.6	0.287	(0.011)	6.373	(0.264)	11.379	(0.123)	2.102	(0.110)	1.325	0.375	(0.108)
Small Blue	5	2.65	-2262.6	4535.2	0.767	(0.008)	4.577	(0.120)	7.755	(0.071)	1.478	(0.055)	3.144	0.148	(0.089)
Wall Brown	5	2.20	-2500.3	5010.7	0.372	(0.010)	5.951	(0.085)	10.887	(0.08)	1.286	(0.051)	1.844	0.507	(0.021)
Small White	5	3.79	-4343.0	8696.1	0.120	(0.005)	6.205	(0.130)	10.286	(0.112)	1.824	(0.056)	3.005	0.653	(0.015)
Common Blue	5	2.14	-6677.3	13364.6	0.260	(0.004)	4.948	(0.032)	8.858	(0.029)	1.189	(0.019)	5.958	0.447	(0.009)
b) P/N_2 GAI															
Species	n	Time	Log(L)	AIC	w		μ_1		μ_d		σ		D		
Holly Blue	4	0.39	-2115.0	4238.0	0.286	(0.011)	7.424	(0.107)	11.382	(0.123)	2.307	(0.042)	1.325		
Small Blue	4	0.23	-2263.3	4534.6	0.766	(0.008)	5.242	(0.033)	7.743	(0.069)	1.573	(0.022)	3.142		
Wall Brown	4	0.42	-2532.3	5072.7	0.363	(0.010)	7.251	(0.070)	10.839	(0.084)	1.856	(0.030)	1.877		
Small White	4	0.43	-4421.1	8850.1	0.110	(0.004)	8.047	(0.121)	10.551	(0.123)	2.626	(0.028)	3.085		
Common Blue	4	0.42	-6924.6	13857.2	0.253	(0.004)	6.103	(0.028)	8.943	(0.031)	1.665	(0.010)	6.244		

Web Appendix B

An hierarchical model approach

An alternative approach to optimising a concentrated likelihood involves treating the individual site effects as random effects. Using an hierarchical approach, we assume the site parameters, N_i, to be independent random variables with a particular distribution function $f(N_i, \theta)$.

It is natural in this instance for $f(N_i, \theta)$ to be a continuous distribution, where N_i can take any non-negative value. The gamma distribution is a sensible choice, since the Poisson-gamma mixture is well known to produce a negative-binomial distribution. Here we explore the gamma distribution with shape parameter β and rate parameter α. For a given site i and visit j, $\lambda_{i,j} = a_{i,j} N_i$. If we drop subscripts for simplicity then the likelihood will be based upon

$$\Pr(Y = y) = \int_0^\infty \frac{e^{-aN}(aN)^y}{y!} \frac{\alpha^\beta}{\Gamma(\beta)} N^{\beta-1} e^{-\alpha N} \,\mathrm{d}N,$$

which simplifies to

$$\Pr(Y = y) = \binom{y + \beta - 1}{y} \left(\frac{a}{a + \alpha}\right)^y \left(\frac{\alpha}{a + \alpha}\right)^\beta.$$

Hence, a Poisson-gamma mixture where the Poisson expectation is the scalar product, aN, is a negative-binomial distribution parameterised by $r = \beta$ and $p = \frac{a}{a+\alpha}$.

Consequently, the likelihood over S sites and T visits for the Poisson-gamma model is

$$L(\alpha, \beta, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}; \boldsymbol{y}) = \prod_{i=1}^{S} \prod_{j=1}^{T} \binom{y_{i,j} + \beta - 1}{y_{i,j}} \left(\frac{a_{i,j}}{a_{i,j} + \alpha}\right)^{y_{i,j}} \left(\frac{\alpha}{a_{i,j} + \alpha}\right)^\beta. \qquad (3)$$

Incorporating the hierarchical aspect into the model increases the number of parameters relative to the GAI with a concentrated likelihood, by the addition of parameters for the gamma distribution.

The density of N_i is given by Bayes theorem as

$$f_{N_i}(n_i|y_{i,j}, a_{i,j}, \beta, \alpha) \propto n_i^{y_{i,j}+\beta-1} e^{-n_i(a_{i,j}+\alpha)},$$

which is a gamma distribution with shape parameter $y_{i,j} + \beta$ and rate parameter $a_{i,j} + \alpha$. Hence, averaging over j, we can estimate each N_i by

$$E(N_i) = \frac{y_{i,\cdot} + \beta}{a_{i,\cdot} + \alpha}. \tag{4}$$

This expression generalises (2) in the main paper, and as $\alpha, \beta \to 0$, keeping the ratio constant results in (2).

In other scenarios, a discrete distribution for N_i may be more appropriate. For example in Royle (2004), the Poisson distribution is mixed with the Binomial distribution.

Negative-binomial-gamma model

As for the concentrated likelihood model, the negative-binomial provides an alternative to the Poisson model. Parameterising the negative-binomial in terms of $(r, a_{i,j}N)$, where $a_{i,j}N$ is the mean, the negative-binomial-gamma likelihood is

$$L(\alpha, \beta, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}; \boldsymbol{y}) = \prod_{i=1}^{S} \prod_{j=1}^{T} \int_0^\infty \frac{\Gamma(r + y_{i,j})}{y_{i,j}!\Gamma(r)} \left(\frac{r}{r + a_{i,j}N}\right)^r \left(\frac{a_{i,j}N}{r + a_{i,j}N}\right)^{y_{i,j}} \frac{\alpha^\beta}{\Gamma(\beta)} N^{\beta-1} e^{-\alpha N} \, \mathrm{d}N. \tag{5}$$

The integral in (5) does not have a simple solution as in the Poisson-gamma case, hence evaluation of the likelihood requires numerical integration. In `R`, we use the standard `integrate` function (with a tolerance of 1e-4). Due to this need for numerical integration, fitting the negative-binomial-gamma model is difficult and only limited results have been obtained. The negative-binomial-gamma model is also much more time-consuming to fit compared to the Poisson-gamma.

Comparison with GAI

We compare model performance for the P/N_2 GAI, the analogous hierarchical Poisson-gamma model, and the NB/N_2 GAI, for five bivoltine species for UKBMS data from 2010. Since the focus here was on model comparison, all parameters in $a_{i,j}$ were assumed to be constant spatially (w, μ_1, μ_d and σ^2). This resulted in four, five and six model parameters for the P/N_2 GAI, NB/N_2 GAI and Poisson-gamma model, respectively.

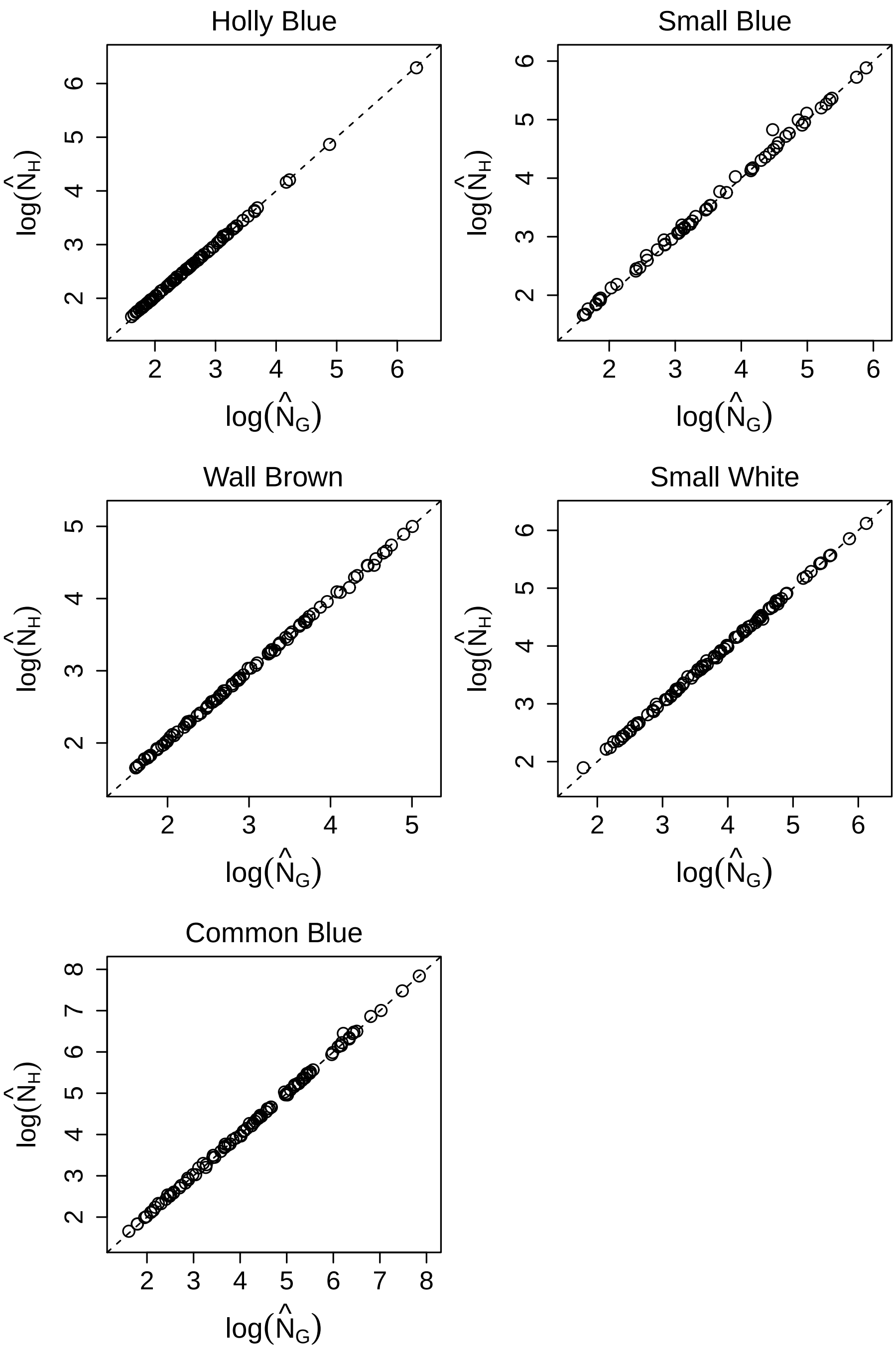
The Poisson-gamma model has lower AIC values than the P/N_2 GAI for four out of the five species, but the NB/N_2 GAI consistently has AIC values that are the lowest (Web Table 2). Given that the models are applied to large, noisy data sets, there are often large differences in AIC as each model describes the data, particularly in terms of overdispersion, differently. The Poisson-gamma model is an intermediate option between the two GAIs: it allows for variation in $\{N_i\}$, whereas the NB/N_2 GAI also estimates the appreciable additional variation in the raw data with respect to the Poisson.

Estimates of the four parameters associated with the mixture components show minimal differences between the three models. The associated standard errors are consistently smallest for the P/N_2 GAI, and are larger from the NB/N_2 GAI and Poisson-gamma model, which may be anticipated as a consequence of accounting for overdispersion. Estimates of the average abundance, $\hat{G}$, which were estimated by the expression in (7) of the main paper, are similar for the different methods, as well as the associated 95% confidence intervals, which were estimated via a bootstrapping approach. For the hierarchical Poisson-gamma model, $\hat{G}$ could also be estimated simply by $\hat{G} = \hat{\alpha}/\hat{\beta}$. Individually, comparison of the $\{\hat{N}_i\}$ from the P/N_2 GAI, estimated from (2) of the main paper, and from the Poisson-gamma model, derived from (4) of this web appendix, also correspond well (Web Figure 2).

The computation times for the P/N_2 GAI are lower than for the hierarchical Poisson-gamma model and NB/N_2 GAI. Computation times for the NB/N_2 GAI are longer than for the Poisson case due to the iterative concentrated likelihood approach. The differences in computation time for the hierarchical model compared to the GAIs would be more significant for the negative-binomial-gamma models, which are not straightforward to fit. We conclude that the GAI is preferable to the hierarchical model as it is simpler and more efficient, whilst producing similar results, and the negative-binomial GAI performs best.

Web Table 2: Model comparison for a) the P/N_2 GAI, b) the hierarchical Poisson-gamma model and c) the NB/N_2 GAI. The computation time is given in seconds. $\hat{G}$ is the index of abundance from the expression of (7) of the main paper, with a 95% confidence interval estimated via bootstrapping. Parameter estimates are given with the associated standard errors in brackets.

	Species	Time	Log(L)	AIC	$\hat{G}$	$\hat{w}$	$\hat{\mu}_1$	$\hat{\mu}_d$	$\hat{\sigma}$		
	Holly Blue	0.34	-2115	4238	21.8 (14.4, 32.9)	0.29 (0.011)	7.42 (0.107)	11.38 (0.123)	2.31 (0.042)		
	Small Blue	0.31	-2263	4535	60.3 (47.5, 77.7)	0.77 (0.008)	5.24 (0.033)	7.74 (0.069)	1.57 (0.022)		
a)	Wall Brown	0.39	-2532	5073	28.5 (23.4, 32.8)	0.36 (0.010)	7.25 (0.070)	10.84 (0.084)	1.86 (0.030)		
	Small White	0.57	-4421	8850	73.5 (60.3, 87.9)	0.11 (0.004)	8.05 (0.121)	10.55 (0.123)	2.63 (0.028)		
	Common Blue	0.37	-6925	13857	190.6 (138, 233.8)	0.25 (0.004)	6.10 (0.028)	8.94 (0.031)	1.67 (0.010)		
	Species	Time	Log(L)	AIC	$\hat{G}$	$\hat{w}$	$\hat{\mu}_1$	$\hat{\mu}_d$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$
	Holly Blue	4.61	-2113	4238	21.7 (14.3, 33.2)	0.31 (0.023)	7.40 (0.162)	11.50 (0.197)	2.37 (0.071)	0.28 (0.019)	0.014 (0.001)
	Small Blue	0.80	-1664	3340	61.7 (47.6, 78.7)	0.68 (0.049)	5.22 (0.071)	8.13 (0.136)	1.32 (0.038)	0.29 (0.024)	0.003 (5e-04)
b)	Wall Brown	1.25	-2175	4362	28.4 (23.3, 32.9)	0.36 (0.027)	7.44 (0.153)	10.68 (0.175)	2.05 (0.053)	0.29 (0.019)	0.011 (0.001)
	Small White	4.44	-3431	6874	73.9 (60.6, 88.2)	0.11 (0.009)	8.10 (0.179)	10.88 (0.195)	2.62 (0.060)	0.46 (0.023)	0.006 (4e-04)
	Common Blue	1.96	-3979	7969	192.7 (137.2, 233.3)	0.23 (0.019)	6.43 (0.112)	9.02 (0.128)	1.79 (0.038)	0.25 (0.011)	0.001 (1e-04)
	Species	Time	Log(L)	AIC	$\hat{G}$	$\hat{w}$	$\hat{\mu}_1$	$\hat{\mu}_d$	$\hat{\sigma}$	$\hat{r}$	
	Holly Blue	2.51	-1826	3661	21.8 (14.4, 33.3)	0.27 (0.018)	6.90 (0.145)	11.62 (0.168)	2.22 (0.053)	0.81 (0.077)	
	Small Blue	1.56	-1475	2961	60.5 (48.0, 78.3)	0.75 (0.021)	5.30 (0.079)	7.94 (0.135)	1.54 (0.040)	0.64 (0.057)	
c)	Wall Brown	3.31	-1965	3940	28.6 (23.5, 32.8)	0.31 (0.020)	7.28 (0.117)	10.93 (0.141)	1.87 (0.039)	0.55 (0.042)	
	Small White	3.22	-3144	6298	73.9 (60.4, 88.1)	0.12 (0.008)	8.24 (0.164)	10.72 (0.176)	2.62 (0.050)	0.88 (0.054)	
	Common Blue	3.79	-3429	6869	192.6 (139.8, 237.5)	0.22 (0.012)	6.53 (0.082)	8.78 (0.096)	1.76 (0.026)	0.75 (0.042)	



Web Figure 2: Comparison of estimated site parameters, $\hat{N}_G$, from the P/N$_2$ GAI and $\hat{N}_H$ from the hierarchical Poisson-gamma model. Both axes are displayed on the log scale and the dashed line indicates the 1-1 line.

Web Appendix C

Efficiency of the concentrated likelihood approach

We compare the performance of optimising a full versus a concentrated likelihood for simulated data for Poisson, negative-binomial and zero-inflated Poisson GAI, for both mixture and stopover models. Data were simulated from the relevant fitted model, based on a single year for $S = 50$ sites and $T = 26$ visits, where for illustration the parameter values used were based upon reasonable values that might be applicable for data for a real species. For the negative-binomial and zero-inflated Poisson cases, we set $r = 0.75$ and $\psi = 0.75$, respectively. For the stopover models, we set $\phi = 0.5$. We assume a univoltine species where the counts arise from a Normal distribution with $\mu = 10$ and $\sigma = 2.5$, and N_i for each site was drawn from a Poisson distribution with an expectation of 150.

For the simplest P/N_1 GAI, the concentrated likelihood has just two parameters to estimate, and for the full likelihood, with the addition of a parameter for each site, there are 52 parameters to estimate. The negative-binomial and zero-inflated Poisson mixture models each required one additional parameter to be estimated. Similarly where the stopover model formulation was used, an additional parameter, ϕ, was estimated.

The concentrated likelihoods were maximised using the `optim` function in the `R` software package (R Core Team, 2015) with the default Nelder-Mead algorithm, as were all of the GAI analyses in this paper. The full likelihoods were maximised using the BFGS algorithm, since the Nelder-Mead algorithm did not always optimise. Iterative likelihood optimisation for the negative-binomial and zero-inflated Poisson cases was performed until the difference in the current and previous log-likelihood value was < 0.001.

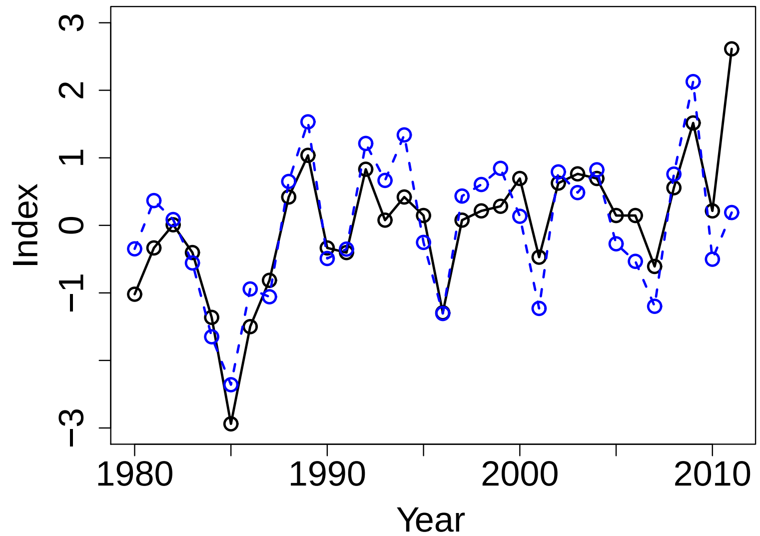
Based on the average time taken to fit each model to one simulated dataset, using a concentrated likelihood approach showed very large reductions in computation time (Web Table 3). In particular for the Poisson case, fitting the full parameter model took over 100 times longer than fitting the concentrated likelihood model for both the mixture and stopover models. Despite requiring iterative likelihood optimisation, the concentrated approach was also faster than optimising the full likelihood in the zero-inflated Poisson and negative-binomial cases. The zero-inflated Poisson and negative-binomial mixture models always each converged within 3 and 5 iterations through steps (ii)-(iv) of Section 2.3, respectively, whereas for the stopover model formulation the zero-inflated Poisson model took a maximum of 23 iterations, and hence took the longest time to fit. In all cases the stopover model took longer than the mixture model to fit, which would be anticipated given the greater complexity of the model, which also has an additional parameter to estimate.

Web Table 3: Average computation times (in seconds) from 20 simulated datasets, fitting the full and concentrated likelihood approach for the mixture and stopover models. The mean and maximum number of iterations are given for the ZIP and NB iterative concentrated likelihood approach.

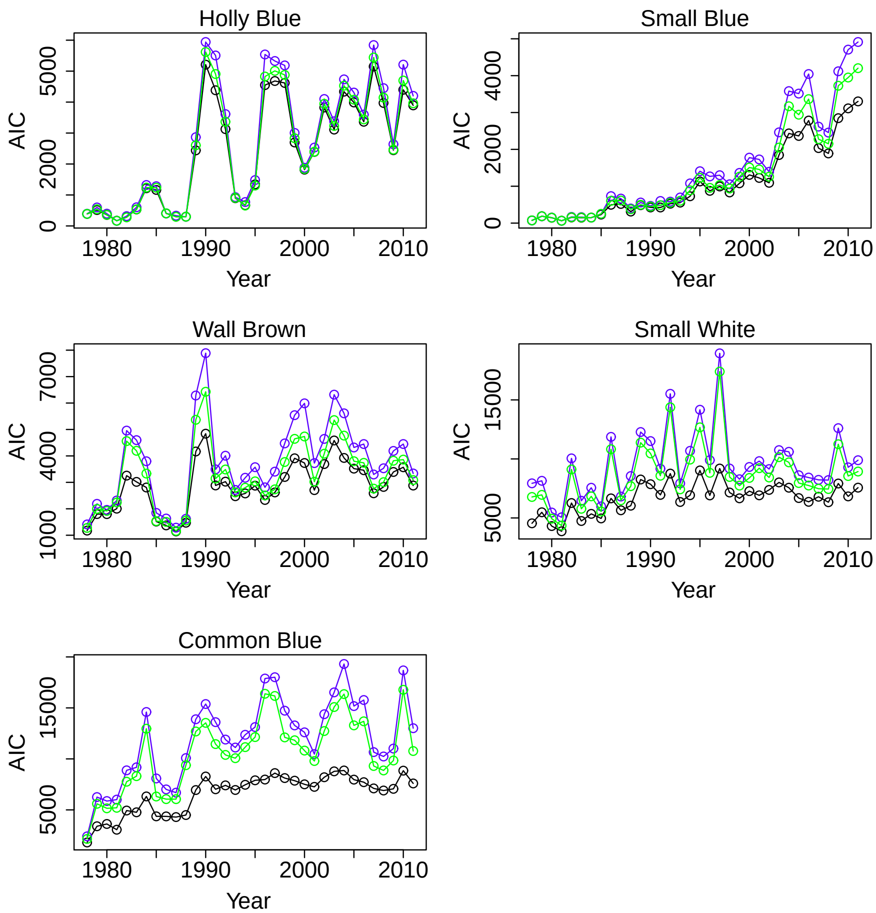
	Computation time		No. of iterations	
Model	Full	Concentrated	Mean	Max
P/N_1	8.6	0.1	-	-
ZIP/N_1	18.3	0.7	3	3
NB/N_1	20.3	0.7	4	5
P/SO_1	66.9	0.6	-	-
ZIP/SO_1	101.5	9.8	11	23
NB/SO_1	93.9	5.2	6	7

Web Appendix D

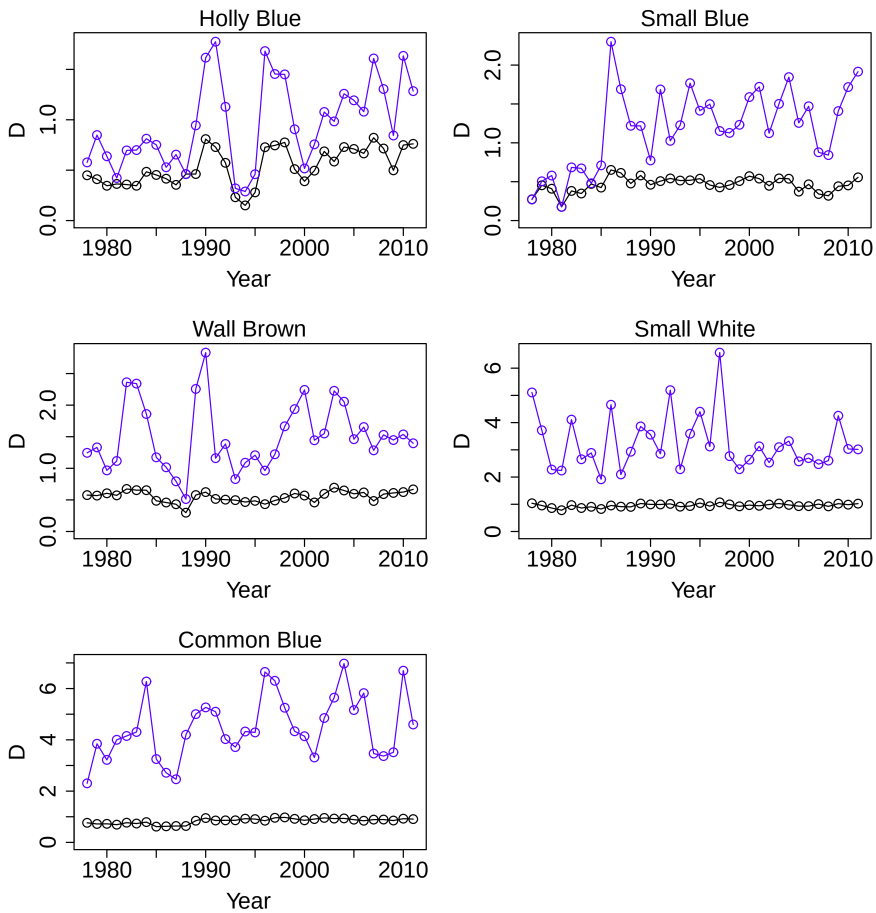
Supplementary tables and figures



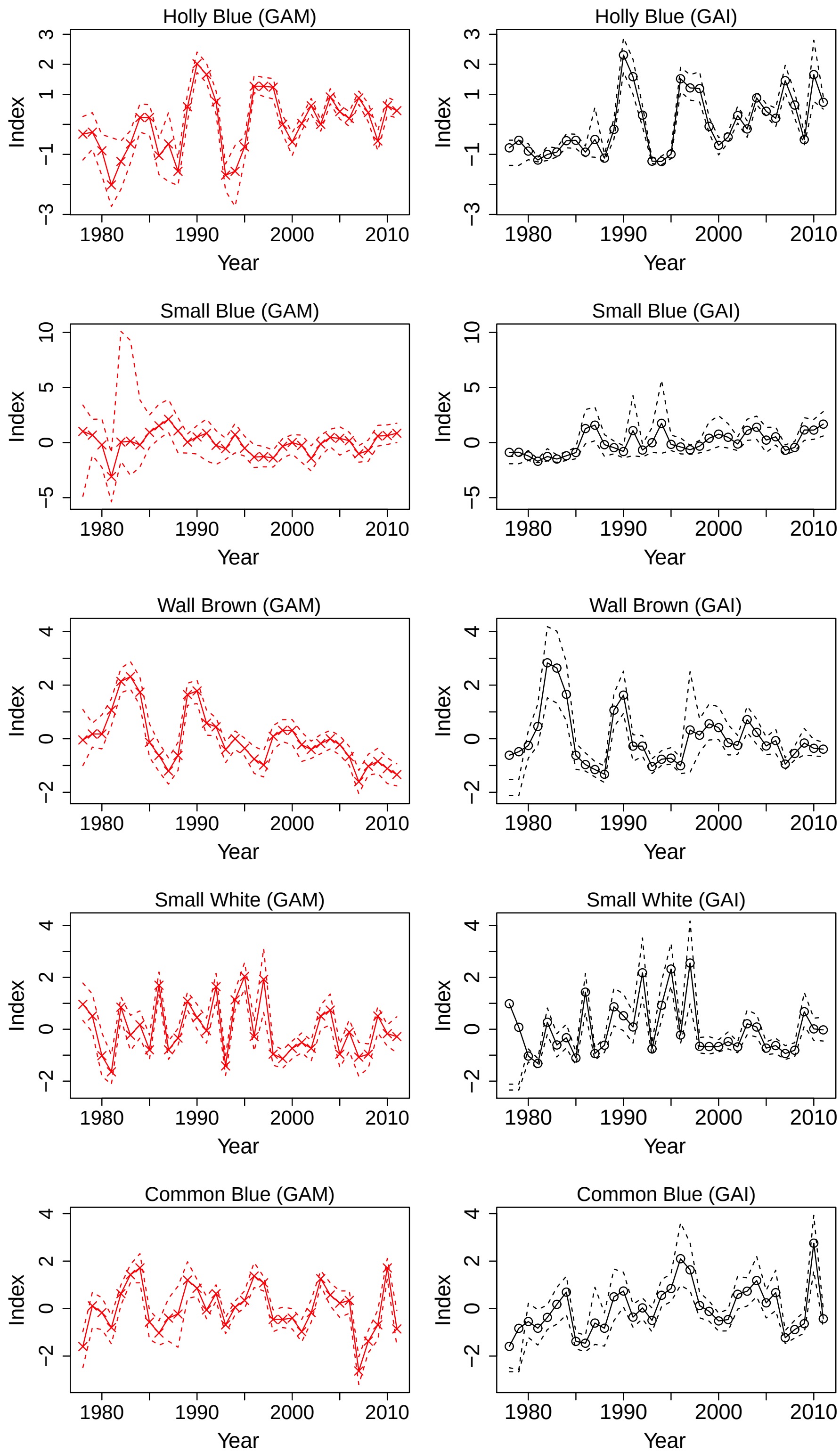
Web Figure 3: Relative abundance indices for the GAM approach (black solid) and P/C GAI (blue dashed) for Speckled Wood.



Web Figure 4: AIC values from the P/N_2 (blue), ZIP/N_2 (green) and NB/N_2 (black) GAIs.



Web Figure 5: Dispersion values (residual deviance/degrees of freedom) from the P/N_2 (blue), and NB/N_2 (black) GAIs.



Web Figure 6: Comparison of indices with bootstrapped intervals derived from the GAM (red) and NB/N_2 GAI (black).

Web Table 4: Latin names of the sample of butterfly species considered.

Species	Latin name
Common Blue	*Polyommatus icarus*
Dark Green Fritillary	*Argynnis aglaja*
Holly Blue	*Celastrina argiolus*
Small Blue	*Cupido minimus*
Small White	*Pieris rapae*
Speckled Wood	*Pararge aegeria*
Wall Brown	*Lasiommata megera*
White Admiral	*Limenitis camilla*

Web Table 5: Parameter estimates (and asymptotic standard errors) for the best (in terms of AIC) multi-year P/N_2 GAI for Wall Brown.

Parameter	Estimate	Std. error
Logit of w		
Intercept	-0.899	0.002
Slope for north	-0.027	0.002
Slope for year	0.229	0.002
Slope for year.north	-0.123	0.002
Log of μ_1		
Intercept	2.135	0.001
Slope for north	0.056	0.002
Slope for year	-0.088	0.002
Slope for year.north	0.016	0.002
Log of μ_d		
Intercept	2.463	0.003
Slope for north	-0.002	0.003
Slope for year	0.037	0.009
Slope for year.north	-0.006	0.011
Log of σ		
Intercept	0.613	0.010
Slope for year	0.020	0.010