Mixture models for multi-brooded butterfly species



Emily B. Dennis

B.J.T. Morgan, T. Brereton, S.N. Freeman, & D.B. Roy





Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL





26th June 2014

Motivation	Data	Models	Application	Discussion
•	000	000000	000000000	O
Motivation				

- Abundance indices are highly important for the monitoring and conservation of various taxa, including insects such as butterflies.
- Over a ten-year period, three-quarters of UK butterfly species have shown declines in distribution, population, or both (Fox et al., 2011).
- Butterflies are the most widely monitored invertebrate taxon and are recognised as valuable indicators of biodiversity (UK Biodiversity Action Plan).
- We develop a method that is both efficient and general for estimating relative abundance.

Motivation	Data	Models	Application	
	● 00	000000	000000000	

UK Butterfly Monitoring Scheme (UKBMS)

- Wide-scale system of ~1000 weekly transect walks (Pollard and Yates, 1993)
- Sites are self-selected but monitored intensively (up to 26 counts per year)
- Began in 1976 with 34 sites
- Assess population trends
- Biodiversity indicators



Motivation	Data	Models	Application	
	000	000000	000000000	

UK Butterfly Monitoring Scheme (UKBMS)

- Wide-scale system of ~1000 weekly transect walks (Pollard and Yates, 1993)
- Sites are self-selected but monitored intensively (up to 26 counts per year)
- Began in 1976 with 34 sites
- Assess population trends
- Biodiversity indicators



Relative abundance index for Wall Brown



Motivation	Data	Models	Application	Discussion
O	oo●	000000	000000000	O

UKBMS data



		Ja	m		Γ	Fe	ь		Γ	Ha	эг			Ap	г			Ha	y		Γ	Ju	ın		Γ	Ju	ıl		Τ	Au	g			Se	P		Γ	0c	t			No	v			De	ec	
	1	2	з	4	1	2	з	4	1	12	3	4	1	2	3	4	1	2	з	4	1	2	з	4	1	2	3	4	1	2	з	4	1	2	з	4	1	2	з	4	1	2	3	4	1	2	1	3 4
Ovun	۲			20					ł.				۲							Γ	Γ	Γ	Γ		Γ	Г	Г	Γ		0	0										0						k	
Larva			Γ	Γ	Γ				Γ	Τ	Γ		•	0	0	•	•	0	0	l,	C	C	C			Γ	Γ	Γ		Γ	Γ	Γ					Γ				Γ	Γ					Γ	Τ
Pupa	Γ		Γ	Г	Γ	Γ			Γ	Т	Г					Γ				Γ	Г		C	C	C	C	К	C	C		Γ	Γ				Γ	Γ				Γ	Γ				Γ	Γ	Τ
Inago	Г	Γ	Г	Т	Г	Г		Γ	Г	Т	Г			Γ		Г			Γ	Г	Г	Г	Γ	Γ	Г	•	•	l.	ю	×0	•	•	٠		Γ	Г	Г	Γ		Γ	Г	Г				Г	Т	Т

http://www.ukbutterflies.co.uk/species.php?species=coridon

Motivation	Data	Models	Application	Discussion
O	000	●00000	000000000	O

Generalised additive models (GAMs)

- Fit non-parametric curves to estimate a common flight period for each year (Dennis et al., 2013).
- Derive an index from fitting a regression model with site and year as predictors, offsetting for seasonal effects.
- Disadvantage: slow can take weeks to derive a bootstrap interval for a single species.



Wall Brown

Mixture models for bivoltine butterflies



Emily Dennis, University of Kent, UK

Motivation	Data	Models	Application	Discussion
O	000	000000	000000000	O

Descriptive mixture model for bivoltine butterflies

- We propose a parametric alternative to GAMs.
- Consider a Poisson likelihood, with expectation $\lambda_{i,j} = N_i a_{i,j}$, where,

$$a_{i,j} = w_i \frac{1}{\sigma_{i,1}\sqrt{2\pi}} \exp\left\{-\frac{(t_{i,j} - \mu_{i,1})^2}{2\sigma_{i,2}^2}\right\} + (1 - w_i) \frac{1}{\sigma_{i,2}\sqrt{2\pi}} \exp\left\{-\frac{(t_{i,j} - \mu_{i,2})^2}{2\sigma_{i,2}^2}\right\}.$$

Then

$$L(\mathbf{N}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}; \mathbf{y}) = \prod_{i=1}^{R} \prod_{j=1}^{T} \frac{\exp(-N_{i}a_{i,j})(N_{i}a_{i,j})^{\mathbf{y}_{i,j}}}{\mathbf{y}_{i,j}!}$$

• Parameters to optimise: *R* site parameters, as well as parameters within *a*_{*i*,*j*}.

Motivation	Data	Models	Application	Discussion
O	000	000000	000000000	O

General formulation for a concentrated likelihood

For a Poisson model with $\lambda_{i,j} = N_i a_{i,j}$, for a given site *i*,

$$\ell = \text{Log}(L) = -N_i \sum_{j=1}^{T} a_{i,j} + \log(N_i) \sum_{j=1}^{T} y_{i,j} + \sum_{j=1}^{T} y_{i,j} \log(a_{i,j}) - \sum_{j=1}^{T} \log(y_{i,j}!).$$

Differentiating ℓ with respect to N_i and equating to zero, we find that

$$N_i = \frac{\sum_{j=1}^T y_{i,j}}{\sum_{j=1}^T a_{i,j}}.$$

Hence, we can substitute for N_i and simply optimize the concentrated Poisson likelihood, with $\lambda_{i,j} = a_{i,j} \frac{\sum_{j=1}^{J} y_{i,j}}{\sum_{j=1}^{T} a_{i,j}}$.

Motivation	Data	Models	Application	Discussion
O	000	000000	000000000	O

General formulation for a concentrated likelihood

- This significantly reduces the number of parameters leading to large improvements in computation time.
- For a simple mixture model, the reduction is from *R* + 4 parameters to just 4 parameters.
- The *a*_{*i,j*} are general mixture of Normal distributions, spline, stopover model formulation.
- The approach also applies to a negative-binomial or zero-inflated Poisson and negative-binomial models, each incorporating an additional parameter.

Motivation		Models	Application	
0	000	000000	000000000	0

Stopover models for butterfly data (Matechou et al., 2014)

• Assuming a Poisson model and a closed super-population size

$$L(\boldsymbol{N},\boldsymbol{\beta},\boldsymbol{\phi};\boldsymbol{y}) = \prod_{i=1}^{R} \prod_{j=1}^{T} \frac{\exp(-\lambda_{i,j})\lambda_{i,j}^{y_{i,j}}}{y_{i,j}!}, \quad \text{where} \quad \lambda_{i,j} = N_i \left\{ \sum_{b=1}^{j} \beta_{i,b-1} \left(\prod_{k=b}^{j-1} \phi_{k,a} \right) \right\}$$

- For example if j = 3, then $\lambda_{i,3} = N_i(\beta_{i,0}\phi_{1,1}\phi_{2,2} + \beta_{i,1}\phi_{2,1} + \beta_{i,2})$.
- Arrival parameters, β , are modelled using a mixture of normal distributions.

Motivation		Models	Application	
0	000	000000	•00000000	0

Comparison of the stopover and descriptive mixture model

Species	Log(L)	AIC	W	μ_1	μ_2	σ	D	ϕ
Holly Blue	-2601.1	5212.1	0.26	7.24	18.90	2.10	1.42	0.32
Small Blue	-2348.1	4706.3	0.77	9.56	17.32	1.47	1.30	0.16
Wall Brown	-2183.0	4376.0	0.38	7.94	18.77	1.26	1.25	0.48
Small White	-4555.0	9119.9	0.14	5.55	16.52	1.84	2.44	0.68
Common Blue	-8978.5	17967.1	0.26	8.69	17.79	1.26	5.34	0.51

a) Stopover model n = 5

b) Descriptive mixture model n = 4

Species	Log(L)	AIC	W	μ_1	μ_2	σ	D	
Holly Blue	-2601.0	5210.0	0.26	8.19	19.83	2.26	1.63	
Small Blue	-2349.0	4706.1	0.77	10.24	17.98	1.57	1.72	
Wall Brown	-2220.6	4449.3	0.38	9.24	20.05	1.78	1.54	
Small White	-4639.8	9287.7	0.13	7.51	18.76	2.70	3.03	
Common Blue	-9336.3	18680.5	0.25	10.04	19.22	1.84	6.70	

0 000 000000	00000000 0000000000000000000000000000	

Application to UKBMS data for a single year

a) Poisson n = 4

Species	Log(L)	AIC	D	W	μ_1	μ_{d}	σ	
Holly Blue	-2601.0	5210.0	1.63	0.26	8.19	11.64	2.26	
Small Blue	-2349.0	4706.1	1.72	0.77	10.24	7.74	1.57	
Wall Brown	-2220.6	4449.3	1.54	0.38	9.24	10.81	1.78	
Small White	-4639.8	9287.7	3.03	0.13	7.51	11.26	2.70	
Common Blue	-9336.3	18680.5	6.70	0.25	10.04	9.18	1.84	

b) Negative-binomial n = 5

Species	Log(L)	AIC	D	W	μ_1	μ_{d}	σ	r
Holly Blue	-2207.7	4425.4	0.62	0.28	7.78	11.88	2.24	0.83
Small Blue	-1565.4	3140.8	0.48	0.71	10.30	7.93	1.56	0.63
Wall Brown	-1813.8	3637.5	0.52	0.36	9.43	10.75	1.94	0.62
Small White	-3443.9	6897.9	1.32	0.14	7.63	11.45	2.63	1.11
Common Blue	-4490.9	8991.7	1.54	0.25	10.75	8.63	2.00	0.76

	Discussion
0 000 000000 00000000	0

Timing comparison for the concentrated likelihood

Comparison of efficiency for the descriptive mixture model with and without the concentrated likelihood approach.

Species	Poiss	on	Negative-binomial		
	<i>n</i> = 104	<i>n</i> = 4	<i>n</i> = 105	<i>n</i> = 5	
Holly Blue	4.9m	0.27s	6.9m	0.70s	
Small Blue	1.9m	0.45s	3.6m	0.86s	
Wall Brown	5.2m	0.29s	6.3m	0.83s	
Small White	11.2m	0.44s	22.4m	1.36s	
Common Blue	15.8m	0.35s	10.6m	1.22s	

	000	000000	000000000				
Motivation		Models	Application				

Goodness of fit



Motivation		Models	Application	
	000	000000	000000000	

Covariate dependence



Emily Dennis, University of Kent, UK

Motivation	Data	Models	Application	Discussion
O	000	000000		O
Multi-year	modelling			

• We use the average of the site parameter estimates, \widehat{N}_i , as an index of abundance in a given year.

$$\widehat{G} = \frac{1}{R} \sum_{i=1}^{R} \widehat{N}_i = \frac{1}{R} \sum_{i=1}^{R} \frac{\sum_{j=1}^{T} y_{i,j}}{\sum_{i=1}^{T} \widehat{a}_{i,j}}$$

- Hence to obtain an index, the mixture model is fitted separately to data for each year and \hat{G} then plotted against time.
- Standard errors are calculated using a bootstrap.

Motivation	Data	Models	Application	Discussion
O	000	000000	○○○○○●○○○	O
Comparison o	of indices			GAM

GAM Mixture model



Motivation	Data	Models	Application	Discussion
O	000	000000	0000000000	O

Timing comparison for a single run

Comparison of efficiency and accuracy for the GAM and descriptive mixture model, based on a single model run.

Species	Time			
	GAM	Mixture		
		Negbin.	Poisson	
Holly Blue	9m	59s	14s	
Small Blue	32m	22s	10s	
Wall Brown	39m	56s	18s	
Small White	23m	102s	25s	
Common Blue	22m	63s	18s	

Motivation	Data	Models	Application	Discussion
O	000	000000	0000000000	O

Timing and accuracy comparison

Comparison of efficiency and accuracy for the GAM and descriptive mixture model, based on 100 bootstraps. The mixture model was fitted for five random starts for each year and each bootstrap.

Species	Time			Mean CI width		
	GAM	Mixture		GAM	Mixture	
		Negbin.	Poisson		Negbin.	Poisson
Holly Blue	15h29m	10h22m	3h01m	0.862	0.639	0.648
Small Blue	53h12m	4h13m	2h05m	3.091	1.934	1.949
Wall Brown	64h59m	9h01m	3h10m	0.860	1.062	1.089
Small White	39h23m	12h11m	4h07m	0.998	0.960	0.969
Common Blue	37h42m	9h52m	3h27m	1.066	1.295	1.298

Motivation	Data	Models	Application	Discussion
O	000	000000		O

Multi-year modelling - regressing parameters on year & northing



Motivation	Data	Models	Application	Discussion
O	000	000000		•
Discussion				

- Development of the concentrated likelihood approach leads to significant improvements in computation time.
 - Important for schemes such as the UKBMS with large-scale data.
- Parametric models are flexible for linking parameters between years and incorporating covariate dependence.
 - Benefit the study of changes in phenology and voltinism, e.g. due to climate change.
- The *a_{i,j}* can be any general function describing variation in counts over the monitoring period.
 - Splines for direct comparison to GAM and species with complex seasonal patterns.
 - Stopover models (Matechou et al., 2014).
- Wider applications for other scenarios and taxa e.g. migrant bird populations.



References



- Dennis, E. B., Freeman, S. N., Brereton, T., and Roy, D. B. (2013). Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods in Ecology and Evolution*, 4(7):637–645.
- Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T., and Roy, D. B. (2014). A general method for constructing animal abundance indices. Technical report, University of Kent.
- Fox, R., Brereton, T. M., Botham, M. S., Middlebrook, I., Roy, D. B., and Warren, M. S. (2011). The state of the UK's Butterflies 2011. Technical report, Butterfly Conservation and the Centre for Ecology & Hydrology, Dorset.
- 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.
- Pollard, E. and Yates, T. J. (1993). Monitoring butterflies for ecology and conservation: the British Butterfly Monitoring Scheme. Chapman & Hall.

Bootstrap intervals



Emily Dennis, University of Kent, UK

Mixture models for butterflies