Supplementary material: efficient statistical inference methods for assessing changes in species' populations using citizen science data

Emily B. Dennis, Alex Diana, Eleni Matechou & Byron J.T. Morgan

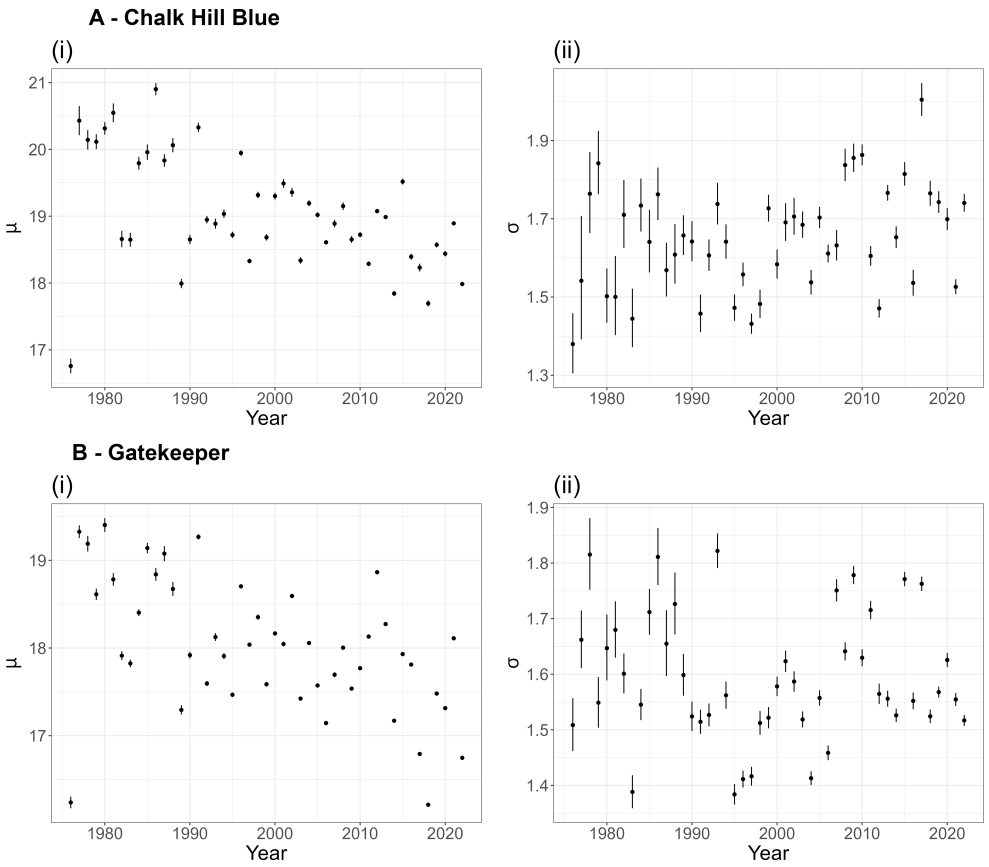

Figure S1: Estimates of (i) μ_r, and (ii) σ_r, the mean and standard deviation of the Normal flight period curve, for each year r, from the extended GAI applied to Chalk Hill Blue (A) and Gatekeeper (B). Plots show the point estimates with 95% confidence intervals (CI).

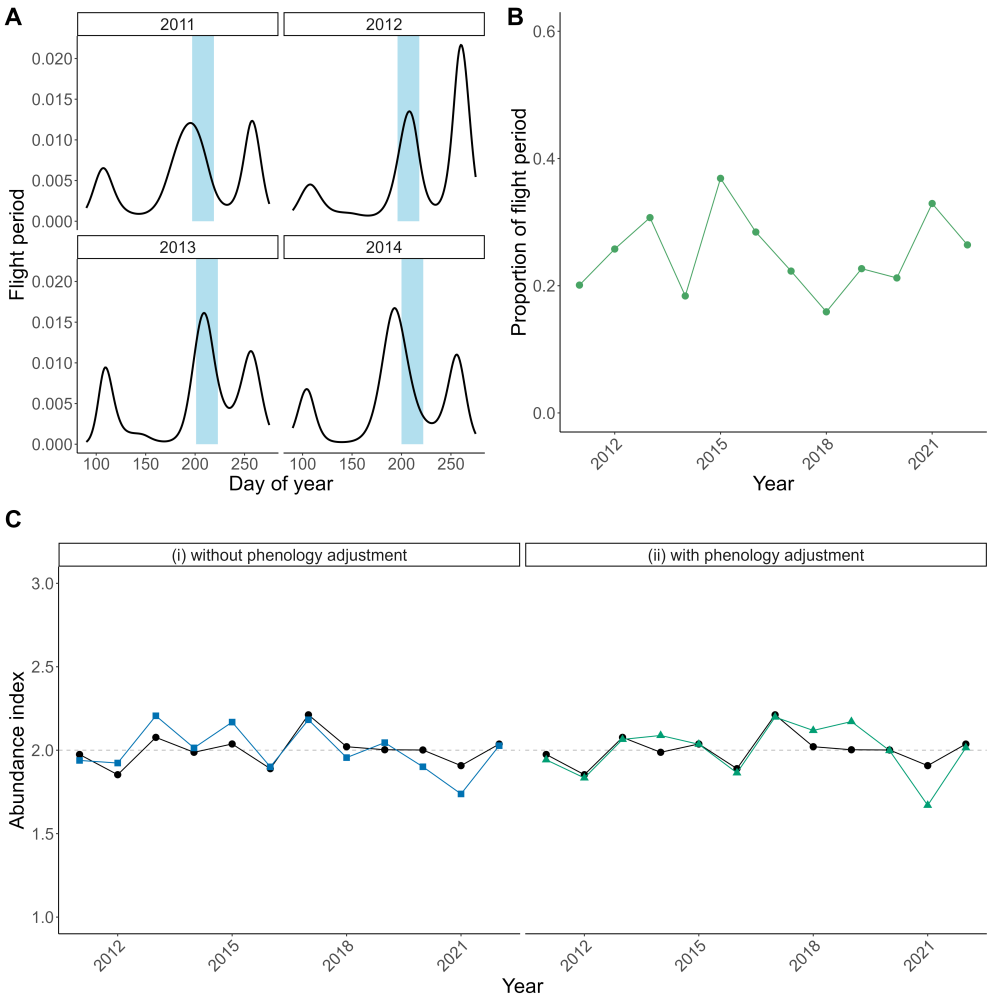

Figure S2: Demonstration of the phenology adjustment approach for the Comma *Polygonia c-album* butterfly: A) flight period curves estimated from UKBMS data for four years. The blue shaded areas represents the BBC sampling period each year. B) the proportion of the Comma flight period covered by the BBC sampling period each year. C) relative abundance indices produced from the GAI applied to UKBMS data (black), from BBC data without phenology adjustment (i, blue squares), and from BBC data with phenology adjustment (ii, green triangles). Indices are on the $\log_{10}$ scale with a mean value of 2 (indicated by the horizontal dashed lines).

S1 Extended GAI

Here we describe the concentrated likelihood approach for the extended GAI in more detail. We suppose that counts of adults are recorded at S sites, each visited on up to V occasions, in each of Y successive years. In any particular year r, the count $y_{s,v,r}$ can be treated as the realisation of an appropriate discrete random variable. For example, if this is taken as Poisson, with expectation $\lambda_{s,v,r}$ for site s, visit v and year r, the likelihood has the form

$$\mathrm{L}(\boldsymbol{N}, \boldsymbol{\theta}; \boldsymbol{y}) = \prod_{s=1}^{S} \prod_{v=1}^{V} \prod_{r=1}^{Y} \frac{\exp(-\lambda_{s,v,r}) \lambda_{s,v,r}^{y_{s,v,r}}}{y_{s,v,r}!}, \tag{1}$$

where $\boldsymbol{\theta}$ are the model parameters which determine the forms of the functions $\{a_{s,v,r}\}$, specified below. Using the same structural form for the Poisson means as for the standard GAI, results in the likelihood:

$$\mathrm{L}(\boldsymbol{N}, \boldsymbol{\theta}; \boldsymbol{y}) \propto \prod_{r=1}^{Y} \prod_{s=1}^{S} \prod_{v=1}^{V} \exp(-N_{s,r} a_{s,v,r}) (N_{s,r} a_{s,v,r})^{y_{s,v,r}}.$$

We now incorporate the expression for the annual model $N_{s,r} = e^{\alpha_s + \beta_r}$ - see ter Braak et al. (1994) - which results in the following expression for the log-likelihood, ignoring an additive constant.

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = \sum_r \sum_s \sum_v \{-e^{(\alpha_s + \beta_r)} a_{s,v,r} + y_{s,v,r}(\alpha_s + \beta_r) + y_{s,v,r} \log(a_{s,v,r})\}. \tag{2}$$

Here $\{\alpha_s\}$ and $\{\beta_r\}$ are respectively site and year effects to be estimated. In order to form maximum-likelihood parameter estimates efficiently we use concentrated likelihood as above. We start by concentrating out the parameters $\boldsymbol{\alpha}$, by analogy with what is done in the GAI, when there are just data from one year.

Differentiating with respect to α_s then gives:

$$\frac{\partial \ell}{\partial \alpha_s} = \sum_r \sum_v \{-e^{(\alpha_s + \beta_r)} a_{s,v,r} + y_{s,v,r}\} + \text{ other terms},$$

where the other terms lack parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Next we set $\frac{\partial \ell}{\partial \alpha_s} = 0$ to give

$$e^{\alpha_s} \sum_r e^{\beta_r} a_{s,.,r} = y_{s,.,.}.$$

Thus

$$e^{\alpha_s} = \frac{y_{s,.,.}}{\sum_r e^{\beta_r} a_{s,.,r}}. \tag{3}$$

We now substitute for α_s in Equation (2), which after some cancellation gives

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = \sum_r \sum_s \sum_v \left[-\frac{y_{s,.,.} e^{\beta_r} a_{s,v,r}}{\sum_j e^{\beta_j} a_{s,.,j}} + y_{s,v,r} \{\beta_r - \log(\sum_j e^{\beta_j} a_{s,.,j})\} \right]. \tag{4}$$

We can now maximise efficiently with respect to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

S2 Variational inference

The transition from complete to observed data likelihoods for the simplest single-season occupancy model and then using Gibbs sampling is given by Royle and Dorazio (2008, p.107) and the corresponding results for the general single-season occupancy model are given by Dorazio and Rodriguez (2012).

Thus, in our case the joint probability of the data y and the latent variables z given (β^ψ, β^p) can be written as

$$p(y, z|\beta^\psi, \beta^p) = p(y|z, \beta^p) p(z|\beta^\psi) = \prod_{j:z_j=1} \left\{ \prod_{i:k_i=j} p_i^{y_i} (1-p_i)^{1-y_i} \right\} \left\{ \prod_{j=1}^{n} z_j^{\psi_j} (1-z_j)^{1-\psi_j} \right\}.$$

The observed data likelihood for(β^ψ, β^p) given the data y can then be written as

$$\mathrm{L}(\beta^\psi, \beta^p; y) = \prod_{j:o_j=1} \left\{ \psi_j \prod_{i:k_i=j} p_i^{y_i} (1-p_i)^{1-y_i} \right\} \prod_{j:o_j=0} \left[\psi_j \left\{ \prod_{i:k_i=j} (1-p_i) \right\} + (1-\psi_j) \right],$$

where $\psi_j = \mathrm{logistic}(X_j^\psi \beta^\psi)$, $p_i = \mathrm{logistic}(X_i^p \beta^p)$ and o_i is a latent variable which is equal to 1 if sampling unit i is confirmed occupied (that is, if at least one detection was recorded) and 0 otherwise.

The gradient of the ELBO, $\nabla_\lambda \mathbf{E}_{\theta \sim q_\lambda(\theta)} [\log p(y, \theta) - \log q_\lambda(\theta)]$, is not straightforward to compute since the variational parameter λ appears in the expectation. To overcome this issue, Kingma and Welling (2013) propose writing the variational distribution $q_\lambda(\theta)$ as a deterministic function of the variational parameter λ and a noise term ϵ independent of λ, that is, $\theta = g(\lambda, \epsilon)$. Once this is done, the ELBO can be rewritten as $\mathbf{E}_\epsilon [\log p(y, g(\lambda, \epsilon)) - \log q_\lambda(g(\lambda, \epsilon))]$ and the gradient operator can be brought inside the expectation. We note that expressing θ

in the form $g(\lambda, \epsilon)$ is straightforward since having chosen q to be a normal distribution θ is simply $\mu + L\epsilon$, where $\epsilon \sim \mathrm{N}(0, I)$.

Next, we can perform a Monte Carlo approximation to compute the gradient as

$$\frac{1}{M}\sum_{m=1}^{M} \nabla_\lambda \left(\log(p(y|\beta^\psi, \beta^p)) - \log(q_\lambda(g(\lambda, \epsilon_i)\right)$$

where $\epsilon_i \sim f(\cdot)$. The gradient of the second term does not pose problems. The gradient of the first term $\nabla_\lambda \log(p(y|\beta^\psi, \beta^p))$ can be decomposed as $\sum_{i=1}^{n} \log(p(y_i|\beta^\psi, \beta^p))$.

Using the chain rule, we can compute the derivative $\frac{\partial l_i}{\partial \lambda_i}$ as $\frac{\partial l_i}{\partial \beta} \times \frac{\partial \beta}{\partial \lambda_i}$, where $\beta = (\beta^\psi, \beta^p)$.

The gradient $\frac{\partial l_i}{\partial \beta}$ can be computed as

$$\frac{\partial l_i}{\partial \beta^\psi} = \sum_{o_j=1} X_j^\psi \frac{1}{1+\exp(X_j^\psi \beta^\psi)} + \sum_{o_j=0} X_j^\psi \frac{\exp\left(-X_j^\psi \beta_\psi\right)(\hat{p}_j - 1)}{(\psi_j \hat{p}_j + (1-\psi_j))(1+\exp -X_j^\psi \beta_\psi)^2}$$

$$\frac{\partial l_i}{\partial \beta^p} = \sum_{o_j=1}\sum_{i:k_i=j}\left(y_i \frac{X_i^p}{1+\exp(X_i^p\beta^p)} - (1-y_i)X_i^p \frac{\exp(X_i^p\beta^p)}{1+\exp(X_i^p\beta^p)}\right) + \sum_{o_j=0}\sum_{i:k_i=j} \frac{q_j \psi_j \hat{p}_j}{\psi_j \hat{p}_j + (1-\psi_j)}$$

where $\hat{p}_j = \prod_{i:k_i=j}(1-p_i)$ and $q_j = -\sum_{i:k_i=j} X_i^p \frac{\exp\left(-X_i^p\beta_p\right)}{(1+\exp\left(-X_i^p\beta_p\right))^2}$

Expressions for the gradient $\frac{\partial \beta}{\partial \lambda_i}$ can be found in Tan and Nott (2018).

We have set L such that the the intercept of the detection and occupancy probability are dependent a-posteriori, while all the other parameters are independent a-posteriori.

References

ter Braak, C. J. F., van Strien, A. J., Meijer, R. and Verstrael, T. J. (1994) Analysis of monitoring data with many missing values; which method? In *Bird Numbers 1992 Distribution, Monitoring and Ecological Aspects* (eds. E. J. M. Hagemeijer and T. J. Verstrael).

Dorazio, R. M. and Rodriguez, D. T. (2012) A Gibbs sampler for Bayesian analysis of site-occupancy data. *Methods in Ecology and Evolution*, **3**, 1093–1098.

Kingma, D. P. and Welling, M. (2013) Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*.

Royle, J. A. and Dorazio, R. M. (2008) *Hierarchical Modeling and Inference in Ecology*. Academic Press, Amsterdam.

Tan, L. S. L. and Nott, D. J. (2018) Gaussian variational approximation with sparse precision matrices. *Statistics and Computing*, **28**, 259–275.