



# Kent Academic Repository

**Dennis, Emily B., Fagard-Jenkin, Calliste and Morgan, Byron J. T. (2022) *rGAI: An R package for fitting the generalized abundance index to seasonal count data*. Ecology and Evolution, 12 (8).**

## Downloaded from

<https://kar.kent.ac.uk/96339/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1002/ece3.9200>

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY (Attribution)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

## RESEARCH ARTICLE

## rGAI: An R package for fitting the generalized abundance index to seasonal count data

Emily B. Dennis<sup>1,2</sup>  | Calliste Fagard-Jenkin<sup>3</sup>  | Byron J. T. Morgan<sup>2</sup> <sup>1</sup>Butterfly Conservation, Dorset, UK<sup>2</sup>School of Mathematics, Statistics and Actuarial Science, University of Kent, Kent, UK<sup>3</sup>Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, UK

## Correspondence

Emily B. Dennis, Butterfly Conservation, Dorset, UK.

Email: [edennis@butterfly-conservation.org](mailto:edennis@butterfly-conservation.org)

## Funding information

Leverhulme Trust, Grant/Award Number: EM-2017-061; University of Kent, Grant/Award Number: 43621; Joint Nature Conservation Committee; British Trust for Ornithology; Butterfly Conservation, the UK Centre for Ecology &amp; Hydrology

## Abstract

The generalized abundance index (GAI) provides a useful tool for estimating relative population sizes and trends of seasonal invertebrates from species' count data and offers potential for inferring which external factors may influence phenology and demography through parametric descriptions of seasonal variation. We provide an R package that extends previous software with the ability to include covariates when fitting parametric GAI models, where seasonal variation is described by either a mixture of Normal distributions or a stopover model which provides estimates of life span. The package also generalizes the models to allow any number of broods/generations in the target population within a defined season. The option to perform bootstrapping, either parametrically or nonparametrically, is also provided. The new package allows models to be far more flexible when describing seasonal variation, which may be dependent on site-specific environmental factors or consist of many broods/generations which may overlap, as demonstrated by two case studies. Our open-source software, available at <https://github.com/calliste-fagard-jenkin/rGAI>, makes these extensions widely and freely available, allowing the complexity of GAI models used by ecologists and applied statisticians to increase accordingly.

## KEYWORDS

flight period, generalized abundance index, multivoltine, phenology, seasonal abundance, stopover model

## TAXONOMY CLASSIFICATION

Conservation ecology; Population ecology; Spatial ecology

## 1 | INTRODUCTION

Suitable statistical models are vital for monitoring species' populations at a time of climate change, habitat degradation, and consequent major loss of biodiversity. The generalized abundance index (GAI, Dennis et al., 2016) provides a useful tool for modeling count data which exhibit variation in numbers within a season, and can take several different forms, including parametric and non-parametric alternatives. The development of the GAI was motivated by an application

to invertebrates, specifically butterflies, which have multistage life cycles, where counts are usually only made of the most visible adult stage. Counts within a season typically fluctuate as individuals emerge, reproduce and then die off, with one or more generations/broods of adults apparent per year. Long-term monitoring schemes often consist of many sites over many years, and previous modeling approaches were often time-consuming to run. The need for more efficient data analysis motivated the development of the GAI, which also offers flexibility in describing seasonal variation in count data.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Here, we outline the GAI approach; full details are provided in Dennis et al. (2016). Within a single year, suppose that counts are recorded at  $S$  sites each visited on at most  $T$  occasions. Each count  $y_{ij}$  for the  $i$ th site and  $j$ th visit is regarded as a realization of a discrete random variable, for example Poisson (alternative distributions are described later), with expectation

$$\lambda_{ij} = N_i a_{ij}$$

where  $N_i$  represents relative total abundance for site  $i$ , and  $a_{ij}$  denotes a function describing seasonal variation in counts in terms of a small set of parameters. Estimates of abundance are relative since not all individuals present during a visit will be observed, and detection is assumed to be constant (but see Matechou et al., 2014). Variation in transect length is also not accounted for, but could be by appropriately scaling  $N_i$ . The GAI encompasses three options for  $a_{ij}$  which describe seasonal variation in counts:

- **Splines**—seasonal variation is described by a flexible curve representing  $a_{ij}$ , for example using B-splines (Dennis et al., 2016). The  $a_{ij}$  are scaled to sum to unity, such that they describe how  $N_i$  is spread over the season. The seasonal curve is typically assumed to be the same across all  $S$  sites and the smoothness of the curve is defined by the number of knots for the spline. This option corresponds closely with the method previously developed for modeling butterfly count data (Dennis et al., 2013) and is the approach typically used for abundance trend reporting—see for example Brereton et al. (2020); Fox et al. (2021).
- **Mixture model**—seasonal variation is taken as a mixture of  $B$  Normal probability density functions (corresponding to one or more broods within a year) so that

$$a_{ij} = \sum_{b=1}^B w_{ib} \frac{1}{\sigma_{ib} \sqrt{2\pi}} \exp \left\{ -\frac{(t_{ij} - \mu_{ib})^2}{2\sigma_{ib}^2} \right\},$$

where  $t_{ij}$  denotes the  $j$ th occasion, which is the time during the season typically measured by day or week, and  $w_{ib}$ ,  $\mu_{ib}$  and  $\sigma_{ib}$  correspond to the weight, mean, and standard deviation, respectively, for the  $i$ th site and  $b$ th brood, and  $\sum_{b=1}^B w_{ib} = 1$ ,  $B \geq 1$ . The weights  $w_{ib}$  describe the relative sizes of the  $B$  broods with respect to each other. As in the spline case, the mixture model is a phenomenological model, where the  $a_{ij}$  effectively describe how  $N_i$  is spread over time, where  $a_{ij}$  integrate to unity.

- **Stopover model**—this is based upon the model proposed in Matechou et al. (2014) which incorporates parameters relating to butterfly life span, that is, lifetimes of individual adult butterflies, by estimating survival probabilities. In brief,

$$a_{ij} = \sum_{d=1}^j \beta_{i,d-1} \left( \prod_{k=d}^{j-1} \phi_{k,c} \right)$$

for  $j = 1, \dots, T$  and  $c = k - d + 1$ , where  $\beta_{i,d-1}$  are the proportions of individuals emerging at visit  $d$ , such that  $\sum_{d=1}^T \beta_{i,d-1} = 1$  for site  $i$ , which are described by appropriate areas under a mixture of  $B$  Normal densities,

$$\beta_{i,d-1} = \sum_{b=1}^B w_{ib} \{F_{i,b}(t_{i,d}) - F_{i,b}(t_{i,d} - 1)\},$$

where  $F_{i,b}(t_{i,d}) = \Pr(X \leq t_{i,d})$  for  $X \sim N(\mu_{i,b}, \sigma_{i,b}^2)$ , with mean emergence date  $\mu_{i,b}$ , standard deviation  $\sigma_{i,b}$ , and weighting  $w_{i,b}$ .  $\phi_{k,c}$  is the probability an individual present for  $c$  occasions and present at visit  $k$ , will remain until visit  $k+1$ . Since  $\phi_{k,c}$  represents apparent survival probability from one visit to the next, adult life spans may be estimated by  $1/(1-\phi_{k,c})$ . Unlike the spline and mixture models, the stopover model proposes a mechanism, of which the  $N_i$  are a part of. Hence, the model results in complex bounds on the  $a_{ij}$ , where it is the emergence parameters  $\beta_{i,d-1}$  which sum to unity. See Matechou et al. (2014) for full details of this model.

Where counts,  $y_{ij}$ , are assumed to be Poisson, efficient model fitting of the GAI is achieved by maximizing a concentrated (or profile) likelihood with respect to only the parameters associated with  $\{a_{ij}\}$  and estimating each  $N_i$  by suitably scaled site totals. An iterative approach is taken when assuming negative binomial and zero-inflated Poisson distributions, as explained in Dennis et al. (2016).

To date, the GAI has primarily been adopted as a method for producing species' abundance trends—for example, it is used annually for reporting of UK butterfly trends (Brereton et al., 2020), which contribute to UK biodiversity indicators (Department for Environment, Food and Rural Affairs, UK, 2020). The approach has also been used in status assessments for larger moths in Britain (Fox et al., 2021; Randle et al., 2019), and in analyses of European butterfly populations (Van Swaay et al., 2020). Where the specific goal is to produce abundance trends, the spline option for describing seasonal variation is used. Here, flight periods are typically assumed to be fixed over sites (as originally developed using generalized additive models, Dennis et al., 2013), or geographical subsets of them (Schmucki et al., 2016).

However, the GAI presents wider opportunities for further insights into seasonal count data, particularly through the application of the parametric descriptions of seasonal variation from the mixture and stopover models. The rGAI package therefore extends previous software to provide accessible code that can accommodate the inclusion of relevant covariates, as well as any number of broods within a defined season. The package also provides the opportunity for wider exploration of stopover models, including estimating species' life spans.

## 2 | rGAI PACKAGE OVERVIEW

The rGAI package extends the software made available in the supplementary materials of Dennis et al. (2016). Model fitting is by maximum-likelihood, and model parameters are transformed using combinations of logarithmic (e.g., for  $\mu$  and  $\sigma$ ) or logistic (e.g., for  $w$  and  $\phi$ ) link functions. rGAI functions with simple inputs allow survey data to first be verified (for duplicate or missing entries across time, or across sites). Then, initial values can be selected for the model fitting process, accounting for the appropriate link scale. The GAI with the mixture or stopover model description for seasonal variation can be fitted with covariate inclusion for parameters of interest, and measures of

uncertainty on parameter estimates can be produced using bootstrapping methods. Table 1 details the most important functions in the rGAI package and provides a brief description of their intended use. The rGAI package also includes a markdown vignette file, with a tutorial-style presentation of all of the package's functionality, as well as installation instructions. The latest version of rGAI is made freely available at <https://github.com/calliste-fagard-jenkin/rGAI>.

Covariates can be included in the linear predictors for the mean emergence date of individuals  $\{\mu_{i,b}\}$ , the standard deviation of each component  $\{\sigma_{i,b}\}$ , and the weightings  $\{w_{i,b}\}$ . Covariate formulae can be specified individually for each brood, or be shared by all broods, for all count distributions (Poisson, negative binomial, or zero-inflated Poisson). The package flexibly allows any number of broods within a season, and includes appropriate custom link functions for the relevant parameters, for example to ensure that for the case  $B = 3$ ,  $\mu_1 < \mu_2 < \mu_3$  and  $w_1 + w_2 + w_3 = 1$ .

The package vignette, as well as two examples below, provide concrete examples of how this flexibility may be used to produce models that are more representative of underlying species biology than was possible with previous available software.

### 3 | EXAMPLES

We now demonstrate usage of the rGAI package via two examples. These examples are presented to illustrate the general capabilities of the rGAI package for modeling seasonal count data, rather than providing full statistical analyses.

**TABLE 1** Description of key functions in the rGAI package.

| Function                                    | Description                                                                                                                                                                                                                                                                               |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>extract_counts</code>                 | Extracts a table of counts across sites and occasions from an input data.frame, to facilitate data cleaning and visualization                                                                                                                                                             |
| <code>transform_starting_values</code>      | Produces a set of initial parameter values on the link scale, given user inputs on the parameter scale                                                                                                                                                                                    |
| <code>fit_GAI</code>                        | Fits GAI models, with any number of broods, with a spline, mixture model, or stopover model to describe seasonal variation. Counts can be modeled with a negative binomial, Poisson, or zero-inflated Poisson distribution                                                                |
| <code>bootstrap</code>                      | Produces bootstrap confidence intervals for all parameters by either resampling them from their asymptotically Normal distribution (parametric bootstrap), or re-fitting models by resampling sites (non-parametric bootstrap). Bootstraps can be provided on the link or parameter scale |
| <code>transform_output</code>               | Transforms parameter estimates from the link scale to the parameter scale, with custom covariate values, or those observed in the data                                                                                                                                                    |
| <code>transform_bootstrap_parameters</code> | Transforms bootstrap confidence intervals of parameter values from the link scale to the parameter scale, for custom covariate values                                                                                                                                                     |
| <code>plot</code>                           | Produces plots of fitted GAI models, with the option of scaling curves by the site total, or producing plots showing variation between sites due to covariate values                                                                                                                      |
| <code>AIC</code>                            | Extracts Akaike's information criterion for a fitted GAI model                                                                                                                                                                                                                            |
| <code>summary</code>                        | Produces a summary of a fitted GAI model, with parameter estimates and standard deviations for parameters, on the link scale                                                                                                                                                              |

Both examples feature seasonal count data for 2018 from the UK Butterfly Monitoring Scheme (UKBMS, Brereton et al., 2020). The scheme consists of a long-running network of transects at which counts of butterflies are made on a weekly basis from April to September under standardized, favorable weather conditions (Pollard & Yates, 1993).

#### 3.1 | Incorporating covariates

We illustrate the usage of covariates in the rGAI package, to allow for seasonal variation in counts to vary over space, by application to data for the Common Blue, *Polyommatus icarus*. This species is known to be bivoltine in the south of the UK, with two generations of adult butterflies per season, but univoltine further north, with a single peak in counts per season, as discussed in Asher et al. (2001, p. 47), and demonstrated by Matechou et al. (2014) using a stopover model.

Here, the rGAI package was used to fit a GAI with Poisson distribution and a stopover model to describe seasonal variation, where mean emergence dates were regressed linearly on northing for both broods, and the weighting parameter was described as a quadratic function of northing. Parameter estimates are given in Table 2, with constant survival probability,  $\phi$ , and constant standard deviations for each brood ( $\sigma_1$  and  $\sigma_2$ ). The `transform_output` function was used to produce parameter estimates on the parameter scale for specified covariate values (northing), which are shown in Figure 1, as well as estimates of seasonal pattern,

which are presented in Figure 2. Figures 1 and 2 illustrate the gradual delay in mean emergence date with increasing northing, along with the increasing closeness of the two broods, until the species' seasonal pattern shows only one generation in the most northern parts of the UK.

This brief example is based upon data for a single year, but application to multiple years offers the potential to assess analytically

**TABLE 2** Parameter estimates (with standard errors, SE) for the GAI, fitted with Poisson distribution and stopover model, applied to UK count data for the Common Blue butterfly in 2018, where  $\mu_1$  is the mean emergence for the first brood,  $\mu_d$  is the difference between mean emergence times  $\mu_1$  and  $\mu_2$ , and  $w_1$  is the weighting of the size of the first brood with respect to the second brood, such that  $w_1 + w_2 = 1$ .

| Parameter                      | Estimate | SE    |
|--------------------------------|----------|-------|
| $\mu_1$ (intercept)            | 2.334    | 0.003 |
| $\mu_1$ (northing)             | 0.142    | 0.002 |
| $\mu_d$ (intercept)            | 1.964    | 0.004 |
| $\mu_d$ (northing)             | -0.168   | 0.003 |
| $w_1$ (intercept)              | -0.438   | 0.019 |
| $w_1$ (northing)               | 0.790    | 0.020 |
| $w_1$ (northing <sup>2</sup> ) | 0.241    | 0.018 |
| $\sigma_1$                     | 1.572    | 0.017 |
| $\sigma_2$                     | 1.292    | 0.026 |
| $\phi$                         | 0.468    | 0.009 |

Note: As they vary with northing, estimates for  $\mu_1$ ,  $\mu_d$  and  $w_1$  are shown on the link scale (log link for  $\mu$  and logistic link for  $w_1$ ). See estimates on the parameter scale in Figure 1. Estimates for the standard deviation of the emergence period for each brood,  $\sigma_1$  and  $\sigma_2$ , and weekly survival probability,  $\phi$ , are constant, and therefore shown on the parameter scale.

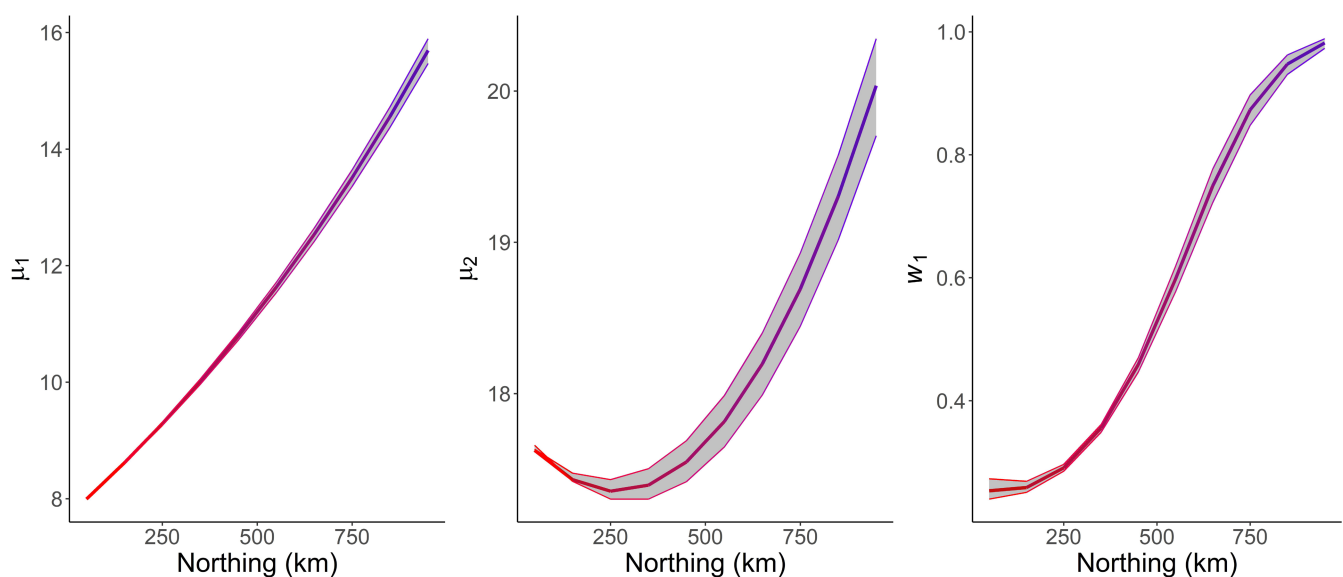
how the change from two to one broods might vary over time, for example due to climatic factors, as well as consider whether changes in abundance for different broods are in synchrony (Asher et al., 2001, p. 162).

By fitting the GAI with a stopover model, an estimate of survival was also obtained ( $\hat{\phi} = 0.47$ ), with a 95% confidence interval of (0.45, 0.49), produced by a parametric bootstrap. Application to multiple years would allow for assessment of potential changes in species' lifespans over time, as demonstrated for two butterfly species in Dennis et al. (2016).

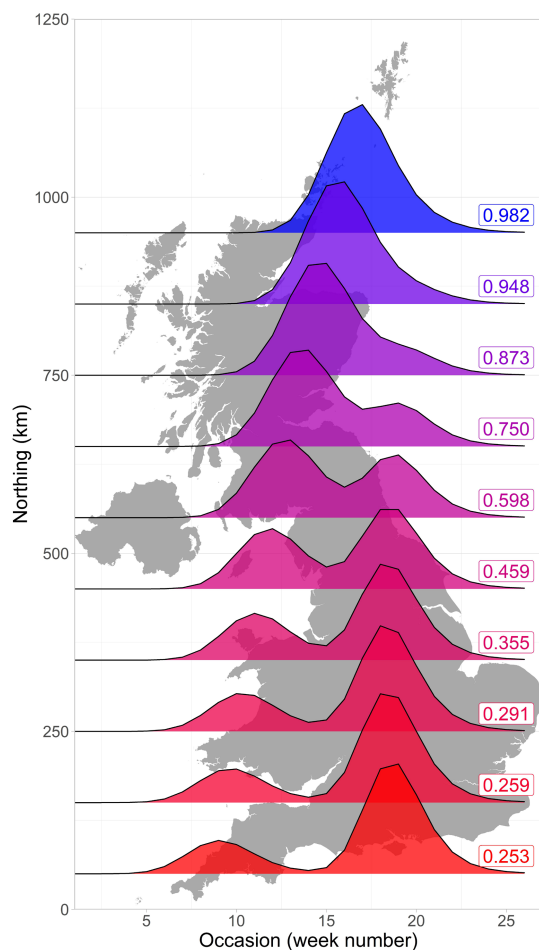
### 3.2 | Modeling multiple broods

The mixture and stopover model formulations of the GAI are described in Dennis et al. (2016) in terms of any general number of broods/generations,  $B$ , within a year/season, but to date, code for a general number of broods has not been widely available; hence, applications have been limited to a maximum of  $B = 2$ , with the exception of a small simulation example for the stopover model in Matechou et al. (2014) which considered values up to  $B = 3$ . The rGAI package extends existing code to allow for any number of broods.

Here, we demonstrate application of the package to data for Small Copper, *Lycaena phlaeas*, for which the overall seasonal pattern suggests three broods within a year, with the third brood at the end of the monitoring season (Figure 3). Using the mixture model formulation, we fit and compare several models for varying  $B = 1, 2, 3$ , as well as different distributions (Poisson, zero-inflated Poisson and negative binomial), and fixed or individual standard deviations,  $\sigma$ , which describe the variation in flight period lengths. Note that this was not an exhaustive model comparison.



**FIGURE 1** Parameter estimates of mean emergence times,  $\mu_1$  and  $\mu_2$ , and mixing probability,  $w_1$ , from fitting the GAI with Poisson distribution and a stopover model to counts of the Common Blue butterfly in 2018, with varying northing. For  $\mu_1$  and  $\mu_2$ , week 1 corresponds to the start of April. 95% confidence intervals derived by parametric bootstrap are shown.

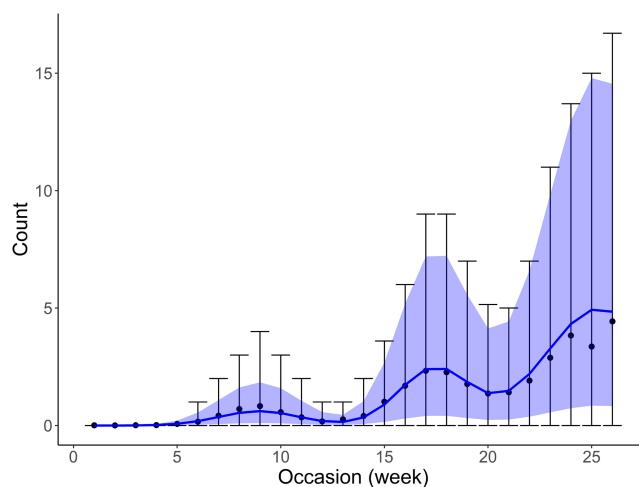


**FIGURE 2** Estimated seasonal pattern for a sample of northing values (each 100 km, from 50 to 950 km) from fitting the GAI with Poisson distribution and a stopover model to counts of the Common Blue butterfly in 2018. The area under the curve is the same for each northing value. The estimate of the mixing probability,  $w_1$ , which describes the size of the first brood relative to the second, is given for each northing value at 100 km intervals. Week 1 corresponds to the start of April.

Based on AIC, the model correctly identifies the species as having three broods (Table 3) and indicates the negative binomial distribution to be the most suitable, as well as individual estimates of  $\sigma$ , for each brood, suggesting that the time period for each brood varies in length. Although we use AIC for illustration here, alternative model selection approaches could be used, such as other information criteria or tests based directly on likelihood values.

Transformed parameter estimates for the best-fitting model are given in Table 4, along with 95% confidence intervals produced using a parametric bootstrap. Using a bootstrap also allows for the production of a confidence interval for other quantities of interest, for example, the estimated seasonal pattern,  $a_{ij}$ , estimates of relative site abundance,  $N_i$ , and predicted counts, avoiding the use of the delta method which is likely to be complex in these cases.

This example demonstrates the potential of the GAI for modeling species with more than two generations within a season, and wider



**FIGURE 3** Observed mean count per week (black circles), averaged over sites, with 5% and 95% quantiles of all observed counts shown as error bars, for the Small Copper butterfly in 2018. The predicted mean count per week, averaged over sites, is shown in blue, along with predicted 5% and 95% quantiles for comparison. Predicted values are estimates from the best-fitting model from Table 3, for which parameter estimates are given in Table 4. Week 1 corresponds to the start of April.

application could again involve extension to analysis over multiple years, as well as the incorporation of relevant covariates to account for spatial variation as in the previous example.

## 4 | DISCUSSION/FUTURE AVENUES

The rGAI package has been designed to provide easy-to-use software for fitting the GAI, particularly with parametric descriptions of seasonal variation through mixture and stopover models. The ability to include covariates flexibly into parameters of interest offers the potential for further studies and improved understanding of spatio-temporal variation in species' phenology. Hodgson et al. (2011) considered variation in phenology over space and time using generalized additive models, but through parametric descriptions of seasonal variation, the GAI can provide simple phenological summaries from parameters of interest, as well as separately for each brood, thus offering opportunities beyond many previous phenology studies which have been limited to species exhibiting a single peak in abundance or to the first generation only (Bell et al., 2019; Macgregor et al., 2019; Roy et al., 2015).

Generalization to accommodate any number of broods/generations within a season provides the opportunity for application of the GAI to species which are known to exhibit more than two broods per year, as well as to species with a less predefined number of broods, which may vary over space and time when species show phenotypic plasticity in voltinism and phenology (Macgregor et al., 2019). The rGAI package provides opportunities to better test for and identify such variation. Although this is applicable for several multivoltine butterfly species in the UK, there is even greater potential/relevance

**TABLE 3** Model comparison for selected GAI fitted with mixture models applied to counts for the Small Copper butterfly, where  $n$  is the number of model parameters.

| Model                                   | $n$ | AIC    | $\Delta$ AIC |
|-----------------------------------------|-----|--------|--------------|
| P, $B = 1$                              | 2   | 42,075 | 11,431       |
| P, $B = 2$ , $\sigma_{\text{shared}}$   | 4   | 42,081 | 11,437       |
| P, $B = 3$ , $\sigma_{\text{shared}}$   | 6   | 39,488 | 8844         |
| ZIP, $B = 3$ , $\sigma_{\text{shared}}$ | 7   | 35,731 | 5087         |
| NB, $B = 3$ , $\sigma_{\text{shared}}$  | 7   | 30,665 | 22           |
| NB, $B = 3$ , $\sigma_{1,2,3}$          | 9   | 30,644 | 0            |

Note: Models are defined by the distribution used (P = Poisson, ZIP = zero-inflated Poisson, NB = negative binomial), the number of broods  $B$ , and, for  $B > 1$ , whether  $\sigma$ , the standard deviation for the flight period curves, are shared across broods or estimated per brood. AIC denotes the Akaike information criterion and  $\Delta$ AIC denotes the difference for each model between its AIC value and the smallest AIC value in the set of fitted models. The best model corresponds to  $\Delta$ AIC = 0.

**TABLE 4** Transformed parameter estimates for the best-fitting GAI model (based on the AIC values given in Table 3) applied to counts for the Small Copper butterfly.

| Parameter  | Estimate | Lower  | Upper  |
|------------|----------|--------|--------|
| $\mu_1$    | 8.970    | 8.839  | 9.102  |
| $\mu_2$    | 17.437   | 17.307 | 17.564 |
| $\mu_3$    | 25.388   | 25.222 | 25.550 |
| $\sigma_1$ | 1.911    | 1.814  | 2.016  |
| $\sigma_2$ | 1.711    | 1.615  | 1.810  |
| $\sigma_3$ | 2.596    | 2.414  | 2.808  |
| $w_1$      | 0.064    | 0.060  | 0.068  |
| $w_2$      | 0.229    | 0.214  | 0.244  |
| $r$        | 0.830    | 0.786  | 0.880  |

Note: 95% confidence intervals are provided based on a parametric bootstrap. The means and standard deviations of the flight period are denoted by  $\mu_b$  and  $\sigma_b$ , for each brood  $b$ .  $w_1$  and  $w_2$  describe the weighting of the size of the first and second brood, where  $w_1 + w_2 + w_3 = 1$ .  $r$  is the dispersion parameter for the negative binomial distribution.

beyond the UK, for example in Europe where species may be multi-voltine in warmer parts of their range. Models for multivoltine species may also have increasing relevance as climate warming may lead to increases in species' voltinism (Altermatt, 2010).

In future releases of the rGAI package, we intend to allow survival  $\phi$  to vary with respect to spatial covariates, or within the season in terms of time or age (Matechou et al., 2014). There is also the potential to account for variation in detection probability, to reduce bias in estimates of relative abundance, using relevant covariates (Matechou et al., 2014). The package can also be extended for multi-year fits and trend estimation; see for example Dennis et al. (2016). We also hope that researchers may contribute new developments to the package; for example accounting for skewness in patterns of seasonal variation/emergence would be of interest (Calabrese, 2012).

The GAI is also relevant for other taxa, for example birds on migration—see Matechou et al. (2013), beetles—see Dennis et al. (2021) who model the two-year life cycle of fuliginators, *Iberodorcadion fuliginator*—and bees—see Matechou et al. (2018) who use a dynamic stopover model to analyze citizen science data on bumblebees, from the *BeeWalk* scheme. We anticipate that the rGAI package will enhance and enable further research by ecologists and applied statisticians, which can improve our understanding of changes in species' populations and phenology.

## AUTHOR CONTRIBUTIONS

**Emily B. Dennis:** Conceptualization (lead); data curation (lead); formal analysis (equal); software (equal); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Calliste Fagard-Jenkin:** Software (equal); writing – review and editing (equal). **Byron J. T. Morgan:** Conceptualization (equal); writing – original draft (equal); writing – review and editing (equal).

## ACKNOWLEDGMENTS

The UK Butterfly Monitoring Scheme is organized and funded by Butterfly Conservation, the UK Centre for Ecology & Hydrology, British Trust for Ornithology, and the Joint Nature Conservation Committee. The UKBMS is indebted to all volunteers who contribute data to the scheme. We are thankful to Alessandro Mari and Eleni Matechou for spotting a minor error in the package. We thank the associate editor and two anonymous referees for their insightful comments that improved this manuscript. Byron Morgan was supported by a Leverhulme research fellowship. Funding was provided by the University of Kent research impact grant 43621.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The butterfly count data that feature in the examples are available in the rGAI package from <https://github.com/calliste-fagard-jenkin/rGAI>.

## ORCID

Emily B. Dennis  <https://orcid.org/0000-0003-2692-4415>

Calliste Fagard-Jenkin  <https://orcid.org/0000-0001-6786-4936>

Byron J. T. Morgan  <https://orcid.org/0000-0002-5465-8006>

## REFERENCES

- Altermatt, F. (2010). Climatic warming increases voltinism in European butterflies and moths. *Proceedings of the Royal Society B: Biological Sciences*, 277(1685), 1281–1287.
- 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.
- Bell, J. R., Botham, M. S., Henrys, P. A., Leech, D. I., Pearce-Higgins, J. W., Shortall, C. R., Brereton, T. M., Pickup, J., & Thackeray, S. J. (2019). Spatial and habitat variation in aphid, butterfly, moth and bird phenologies over the last half century. *Global Change Biology*, 25(6), 1982–1994.

- Brereton, T. M., Botham, M. S., Middlebrook, I., Randle, Z., Noble, D., Harris, S., Dennis, E. B., Robinson, A., Peck, K., & Roy, D. B. (2020). *United Kingdom Butterfly Monitoring Scheme report for 2019*. Technical report, UK Centre for Ecology & Hydrology, Butterfly Conservation, British Trust for Ornithology and Joint Nature Conservation Committee.
- Calabrese, J. M. (2012). How emergence and death assumptions affect count-based estimates of butterfly abundance and lifespan. *Population Ecology*, 54(3), 431–442.
- 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(7), 637–645.
- Dennis, E. B., Kéry, M., Morgan, B. J. T., Coray, A., Schaub, M., & Baur, B. (2021). Integrated modelling of insect population dynamics at two temporal scales. *Ecological Modelling*, 441, 109408.
- Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T. M., & Roy, D. B. (2016). A generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4), 1305–1314.
- Department for Environment, Food and Rural Affairs, UK. (2020). UK Biodiversity Indicators 2020.
- Fox, R., Dennis, E. B., Harrower, C. A., Blumgart, D., Bell, J. R., Cook, P., Davis, A. M., Evans-Hill, L. J., Hanes, F., Hill, D., Isaac, N. J. B., Parsons, M. S., Pocock, M. J. O., Prescott, T., Randle, Z., Shortall, C. R., Tordoff, G. M., Tuson, D., & Bourn, N. A. D. (2021). *The State of Britain's Larger Moths 2021*. Butterfly Conservation, Rothamsted Research and UK Centre for Ecology & Hydrology.
- Hodgson, J. A., Thomas, C. D., Oliver, T. H., Anderson, B. J., Brereton, T., & Crone, E. (2011). Predicting insect phenology across space and time. *Global Change Biology*, 17(3), 1289–1300.
- Macgregor, C. J., Thomas, C. D., Roy, D. B., Beaumont, M. A., Bell, J. R., Brereton, T., Bridle, J. R., Dytham, C., Fox, R., Gotthard, K., Middlebrook, I., Nylin, S., Platts, P. J., Rasteiro, R., Saccheri, I. J., Villoutreix, R., Wheat, C. W., & Hill, J. K. (2019). Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year. *Nature Communications*, 10(1), 1–10.
- Matechou, E., Dennis, E. B., Freeman, S. N., & Brereton, T. (2014). Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *Journal of Applied Ecology*, 51(3), 766–775.
- Matechou, E., Freeman, S. N., & Comont, R. (2018). Caste-specific demography and phenology in bumblebees: modelling *BeeWalk* data. *Journal of Agricultural, Biological, and Environmental Statistics*, 23, 427–445.
- Matechou, E., Morgan, B. J. T., Pledger, S., Collazo, J. A., & Lyons, J. E. (2013). Integrated analysis of capture–recapture–resighting data and counts of unmarked birds at stop-over sites. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(1), 120–135.
- Pollard, E., & Yates, T. J. (1993). *Monitoring butterflies for ecology and conservation: the British Butterfly Monitoring Scheme*. Chapman & Hall.
- Randle, Z., Evans-Hill, L. J., Parsons, M. S., Tyner, A., Bourn, N. A. D., Davis, A. M., Dennis, E. B., O'Donnell, M., Prescott, T., Tordoff, G. M., & Fox, R. (2019). *Atlas of Britain & Ireland's Larger Moths*. Pisces Publications.
- Roy, D. B., Oliver, T. H., Botham, M. S., Beckmann, B., Brereton, T., Dennis, R. L., Harrower, C., Phillimore, A. B., & Thomas, J. A. (2015). Similarities in butterfly emergence dates among populations suggest local adaptation to climate. *Global Change Biology*, 21(9), 3313–3322.
- Schmucki, R., Pe'er, G., Roy, D. B., Stefanescu, C., Van Swaay, C. A., Oliver, T. H., Kuussaari, M., Van Strien, A. J., Ries, L., Settele, J., Musche, M., Carnicer, J., Schweiger, O., Brereton, T. M., Harpke, A., Settele, J., Kün, E., & Juilliard, R. (2016). A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes. *Journal of Applied Ecology*, 53, 501–510.
- Van Swaay, C. A. M., Dennis, E. B., Schmucki, R., Sevillega, C. G., Aghababayan, K., Åström, S., Balalaikins, M., Bonelli, S., Botham, M., Bourn, N., Brereton, T., Cancela, J. P., Carlisle, B., Collins, S., Dopagne, C., Dziekanska, I., Escobés, R., Faltynnek Fric, Z., Feldmann, R., ... Roy, D. B. (2020). *Assessing Butterflies in Europe - Butterfly Indicators 1990-2018*. Technical report, Butterfly Conservation Europe & ABLE/eBMS ([www.butterfly-monitoring.net](http://www.butterfly-monitoring.net)).

**How to cite this article:** Dennis, E. B., Fagard-Jenkin, C., & Morgan, B. J. T. (2022). rGAI: An R package for fitting the generalized abundance index to seasonal count data. *Ecology and Evolution*, 12, e9200. <https://doi.org/10.1002/ece3.9200>