**N mixture models for site-level indices**

Luke C Evans, Tom H Oliver

**Overview and aims**

Broadly, the DRUID project aims to understand rates of declines of British insects, the drivers of any declines, and the subsequent impact on ecosystem function and services. As part of this, in collaboration with RSPB and BCT, we aim to estimate the impact of changes in key insect food resources on the population growth of selected species at higher trophic levels (i.e. birds and bat species). This requires linking changes in local insect numbers to changes in bird or bat population sizes. A key stage in the analysis, therefore, is the construction of local-scale population indices.

Beyond the DRUID project, these local-scale population indices can be highly valuable for informing new understanding in landscape ecology and to inform site-level management. For example, other standardised monitoring schemes such as the UK Butterfly Monitoring Scheme have used site level indices to explore patterns such as: i) the effects of habitat fragmentation on population dynamics (e.g. ability to withstand and recover from extreme weather events), ii) effects of habitat and topopgraphic diversity on population stability, iii) which site and landscape management characteristics are most effective in enhancing population numbers.

A difficulty with constructing these indices is that there are relatively few counts of birds and bats taken at any site within a given year and these counts are subject to imperfect detection. This produces two sources of potential error, error due to only encountering a limited proportion of the population at a given visit and error from detecting only some fraction of those that were encountered.

An approach to deal with this is the N-mixture model (Royle 2004) which explicitly accounts for these sources of error when estimating local population sizes from count data. Here we apply a series of hierarchical general additive (GAM) N-mixture models that attempt to estimate indices of local population abundance and provide insight into factors that may influence detection (e.g detector type, transect number). We view each component of this model structure as providing some benefit to the local population estimation problem: 1) the hierarchical structure regularizes local parameters (i.e. shrinkage) so that local estimates are constrained based on all the data for that species, 2) the GAM component provides high flexibility to model variability in population size - as applied when producing national level trends (e.g. Bat Conservation Trust 2021), 3) the N-mixture component accounts for the different sources of count and observation error.

Here we describe the results of applying these models to two simulation experiments, the bird and bat data from the Environmental Change Network (ECN), and finally, a sample of count data for a single bird (Blue tit) and a bat species (Common pipistrelle) from the RSPB and BCT national monitoring data respectively (i.e. the BBS survey and the NBMP). A key goal of this early report is gaining feedback for refining the models to produce high quality indices for linking local population estimates to insect populations for research and publications as part of the DRUID project. However, I have also taken the opportunity to make code user-friendly to increase the usability of the site-level indices in other research projects. To this end, all code for these experiments and analyses is provided in a private GitHub and I have taken some effort to make the code relatively easy to follow and adapt should it prove useful.

**Methods**

*The N-mixture*

The N-mixture model considers the count data to result from a mixture of an ecological process and an observational process that can be jointly estimated. An observer first encounters at visit *i*, but does not entirely detect, Ni individuals from an unobserved true population size λ.

Ni ~ Poisson(λ) ; λ >= Ni

Factors affecting changes in λ over time represent the ecological component and can be estimated in a log linear model

Log(λ) = α + β X

Where α is an intercept, X is a matrix of variables, and β a vector of parameters. We use three different model structures to estimate log(λ) which we describe in more detail below.

The model then considers that of the encountered individuals, N*i*, yi are detected (yi is the actual count data recorded) with probability of detection given by parameter p.

yi | Ni ~ Binom(Ni,p)

Similarly, to the ecological component, factors affecting detection can be modelled in a logit-linear model

Logit(p) = α + δ X

Here we use the notation δ to indicate coefficients on the observational model which can and do, use different data and parameters to the ecological model.

The parameter p can vary with visits, but λ is considered a fixed population size for that site in that year.

*Model structures*

We apply three main mixed-effects GAM structures when estimating changes in the latent abundance, a shared trend across years with a site-level offset, a hierarchical fit that has a shared trend and an individual trend and a site level offset, and an individual trend with no shared trend and a site level offset. In the code I use the shorthand *global*, *shrink,* and *free* to refer to these model structures respectively and for convenience, I shall label them as such throughout.

The structures for site *j* are as follows:

Global:

Log(λj) = αj +

Log(β1:k)~ N(0, 0.05)

Log(αj) ~ N(2, 1)

Where K is the number of splines covering the sampling years, βk a vector of coefficients, and bk basis functions (here we use B spline basis functions).

Shrink:

Log(λj) = αj +

Log(βjk)~ N(k , σ)

Log(1:k)~ N(0, 0.05)

σ ~ Exponential(5)

Log(αj) ~ N(2, 1)

Where symbols are as above but now include hierarchical parameters on the coefficients and individual coefficients per site.

Free:

Log(λj) = αj +

Log(βjk)~ N(0, 0.05)

σ ~ Exponential(5)

Log(αj) ~ N(2, 1)

As we fit the GAM models in a Bayesian context we pre-select the number of basis functions using Fewster recommendation of 0.3 \* number of years (Fewster et al. 2000) and in certain experiments, we add one more and one less knot and compare models using approximations of cross-validation scores (WAIC; Watanabe and Opper 2010).

For the observational process, we focus only on three models, one simple intercept model for all sites and visits that we apply to birds and bats, while the second and third observation models are bird and bat specific with p varying with site and visit.

Shared:

Logit(p) = δ

Logit(δ) = N(0,1.6)

Here we use δ rather than β to distinguish observational and ecological parameters as is applied in the code.

Bird observation:

Logit(pi) = α + δ1 \* Visiti + δ2 \* Distance bandi

Logit(δ1:2) = N(0,1.6)

Bat observation:

Logit(pi) = α + δ1 [detector]i + δ2 \* Spots coveredi

Logit(δ1) = N(0,1.6)

Logit(δdetectors) = N(0,1.6)

Note, the observational components are currently quite simplistic and almost certainly require modification as I have only selected observational factors that intuitively seemed like they might have an effect for a demonstration. For the bird observation, we treat both visit (the number of visits over the course of the year) and distance band as linear effects assuming detectability will (presumably) drop linearly with distance and detectability increase or decrease with visits (i.e. you see more/fewer birds later in the year due to vegetation or some other factor). For bats, we assume detectability will increase linearly with the number of survey spots covered and that each make of bat detector could either increase or decrease detectability.

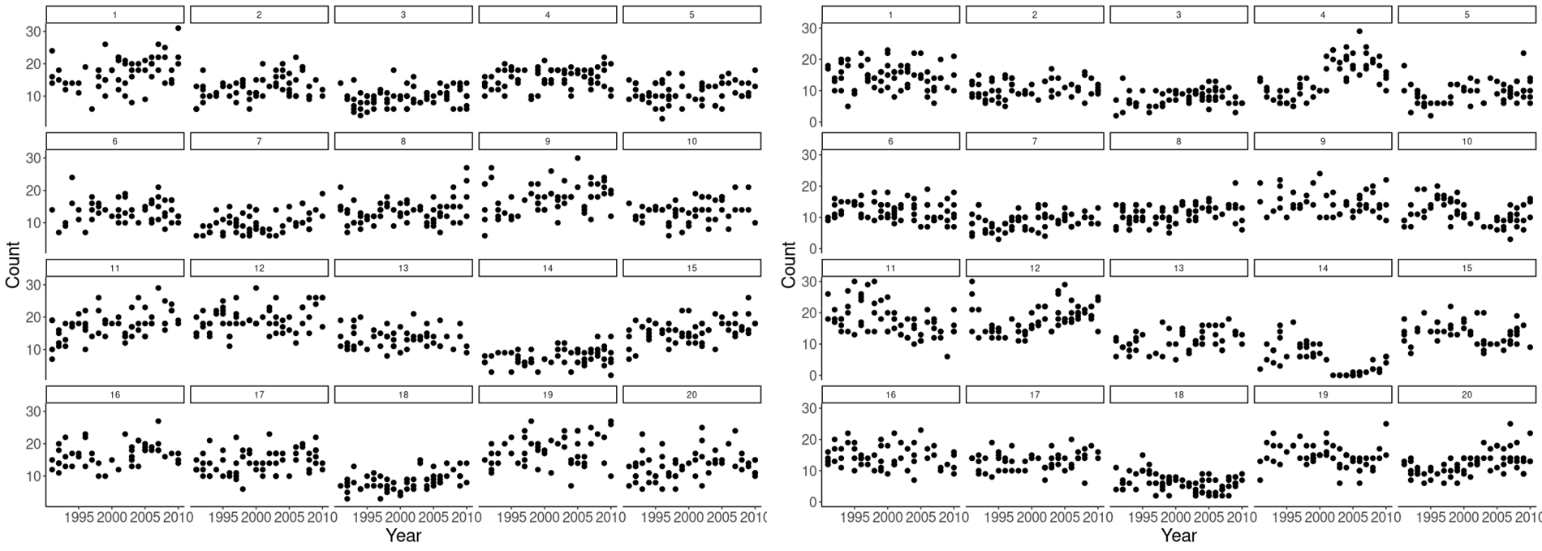
*Samplers*

The code is replicated in two different Bayesian languages: *Nimble* (de Valpine et al. 2017) and *stan* (Stan Development Team 2019). Nimble has similar syntax to BUGS, but models are written in R taking advantage of some syntax familiar to R users. It uses a range of MCMC samplers to obtain posterior estimates of parameters. An advantage is that Nimble appears to have a growing user base in ecology and has a specialised package for fitting some common ecological models (Goldstein et al. 2019; Code in the github shows how the Meehan approximation for the N-mixture is calculated). The stan models are written as a probabilistic program in stan syntax which is a modified form of c++. Stan uses the NUTS sampler which offers efficient sampling of high-dimensional posteriors but cannot estimate discrete parameters and thus requires a step integrating out certain parameters. Stan is highly expressive and very efficient when drawing samples but tends to run slightly more slowly that Nimble, but given its higher efficiency it is difficult to make definite comparisons at this stage. Note, that in this code we take a ‘profiling’ approach whereby we apply limited MCMC chains and/or iterations in order to explore the various model structures without large computation times. To provide robust results for publications will typically require longer MCMC chains and more concern around diagnostics and adjustment of the MCMC hyper parameters (particularly for Nimble). However, we have made sure that fatal inferential issues that would entirely invalidate model comparisons (such as divergent transitions in NUTS) are avoided.

*Simulation experiments 1 and 2*

For the first simulation experiment, we replicated, in a virtual ecologist approach, five visits per year across 20 sites and for 20 years of observations (Fig 1). Sites were randomly sampled from an autocorrelated landscape where position determined the starting population size. Populations then increased linearly at a fixed rate for the 20 years. Detectability was set at 0.8 across all years and sites. In the code, we apply the three model structures, with intercept only detection, but replace the GAM component with a simple linear effect of year. We use this experiment to demonstrate the impact of observational error at the local scale and to introduce the Nimble and stan models prior to further development.

In the second simulation, we used the same approach to generate starting populations, but population change then followed a sinusoidal trajectory where the starting position on the sinusoidal curve was dependent on position. Consequently, we have variation in the intercepts and trends across sites (Fig 1). We use this experiment to introduce the GAM versions of the models.



**Figure 1.** Simulated counts for 20 sites across 20 years for a) a linear increasing trend, and b) a space-varying sinusoidal trend. The noise in the data shows the impact of count and observational error for even simple trends.

*ECN data*

For the ECN trial we downloaded count data for bats and birds from the Environmental Information Data Centre (Rennie et al. 2017a, 2017b) and accompanying data on site identification and location. The ECN data provides counts in a similar structure to the RSPB and BCT data with counts at each of the 10 sites (Figure 2) collected from multiple visits over ~18 years. This data was used to test how the various model structures impact predictions of population size for the blue tit and the common pipistrelle and uses model selection to suggest the select a model with the lowest WAIC. This is similar to the analysis of the BCT and RSPB data but is a much smaller sample size and uses the intercept only detection model, fitting only one estimate of detectability across all sites and visits.



**Figure 2** ECN sites with concurrent Bird and Bat observations

*RSPB and BCT data*

For the RSPB and BCT data we randomly selected 100 sites from each respective dataset and then fitted the models for the blue tit and the common pipistrelle. For the common pipistrelle, we used the field transect data of the NBMP. We apply a similar approach to the ECN data by fitting all three model structures and comparing the WAIC scores to indicate the best fitting model for the sample. The key difference in these models relative to the ECN fits, however, is that we apply the bird and bat specific observation models and estimate a detection probability for each visit.

**Results**

We provide here a brief overview of the preliminary results for each of the experiments, though with the caveats as stated in the section on samplers. We plot only the results from Nimble as results from the stan models were generally very similar, though there are some exceptions e.g., Nimble estimates lower detection probabilities, and sometimes produced a more varying fit than stan. These subtitles are detailed more thoroughly in the code, but to save space here I only provide a general picture. Overall, however, trends in the simulated data were recovered by the models, and models for both samplers converged when estimating population size and produced similar predictions.

*Simulation experiments 1 and 2*

The N mixture model could extract the underlying linear trend in the true site abundances in experiment 1 (Figure 3) and when compared against an alternative index, such as the mean across counts within a year, resulted in fewer false population changes driven by observational and count errors.

Chart, line chart

Description automatically generated

**Figure 3.** Model comparison for population estimates from count data simulated at two sites. True values of population size are shown in grey, mean model estimates shown in red for the standard N mixture version and pink when using the Meehan approximation. An alternative index, the mean of counts at a site and year, is shown in black.

Results were similar for the second experiment where there were non-linear and site varying trends in abundance.

Chart, scatter chart

Description automatically generated

**Figure 4.** Modelled population estimates from the *free* model structure applied to count data simulated at two sites. True values of population size are shown in grey, and mean model estimates are shown in red for the N mixture using the Meehan approximation. The mean of counts at a site and year, are shown in black.

*ECN*

When fitting to the ECN count data, models from both languages produced qualitatively similar predictions for both the blue tit (Figure 6) and the common pipistrelle (Figure 5) and the shrink model resulted in the lowest WAIC scores in all cases. A noticeable difference between the languages was again estimates of detection, this resulted in the same pattern in abundance being shifted up in the Nimble fits relative to the stan fits. Effectively the Nimble models predicted two-times or greater abundance than the stan models.

A note when interpreting these plots; the means look very low relative to the fits. In fact, the fits look like they are fitting nothing! This is less pronounced in the stan fits as the population estimates are uniformly lower so it’s easier to see the comparison between the data and the model. But it also results from the model structure where it assumes the population size is at least as large as the biggest count for that year at that site. In the code I show fits compared to the max values observed from any visit in a year and this gives a better indication of what the model is using for estimation and below I discuss this assumption in more detail.

*Graphical user interface

Description automatically generated*

**Figure 5.** Modelled population estimates of the common pipistrelle from all model structures applied to count data at ECN sites. Panels show different sites. The global model structure is shown in red, shrink in light blue and free in dark blue. The mean of counts at a site and year, are shown in black.

*Graphical user interface, timeline

Description automatically generated*

**Figure 6.** Modelled population estimates of the blue tit from all model structures applied to count data at ECN sites. The global model structure is shown in red, shrink in light blue and free in dark blue. The mean of counts at a site and year, are shown in black.

*National monitoring schemes*

For the samples of 100 sites from the BCT and RSPB data, results from stan and Nimble models were again similar, but there was variation in the best supported model with the shrink model being better supported for the blue tit (Figure 8) and the free model (Figure 7) for the common pipistrelle.

A picture containing scatter chart

Description automatically generated

**Figure 7.** Modelled population estimates of the common pipistrelle from all model structures applied to selected count data from the field transects of the NBMP. Panels show different sites. The global model structure is shown in red, shrink in light blue and free in dark blue. The mean of counts at a site and year, are shown in black.

A picture containing diagram

Description automatically generated

**Figure 8.** Modelled population estimates of the blue tit from all model structures applied to selected count data from the BBS. The global model structure is shown in red, shrink in light blue and free in dark blue. The mean of counts at a site and year, are shown in black.

For these data, we also fit the bat and bird specific observation models. Here we attempted to account for the detector used and the spots surveyed for the bats, and for the birds, number of visits and the distance band. The models find that these factors may influence detection - though at this stage I think they should be interpreted cautiously. For the bats, some bat detectors increased or decreased detection, and observations increased with the number of spots surveyed. For birds, later visits were associated with slightly higher detectability and increasing distance bands were associated with lower detectability. The coefficients on the observational models (named delta in the plots) are described in more detail in Figure 9 for the common pipistrelle and in Figure 10 for the blue tit.

Chart

Description automatically generated

**Figure 9.** Coefficients on the observation model for the common pipistrelle with the means and standard deviations of the posteriors – note these are on the logit scale. The top of the plot shows δ2 the parameter showing the positive effect of the number of spots surveyed, and at the bottom, δ0 the intercept(i.e. α). The middle points show the effect on detectability for a particular bat detector. The blue dots show detectors not fitted in this sample of 100 sites. δs for these detectors can be matched to the detector brand from a list provided in the code – e.g. this first go shows Pettersson models are associated with increased detectability and Magenta models with lower detectability.

Chart, box and whisker chart

Description automatically generated

**Figure 9.** Coefficients on the observation model for the blue tit with means and standard deviations of the posteriors – note these are on the logit scale. At the top δ2 shows the reduction in detectability with the distance band, then δ1 the possible increase in detectability with later visits and then finally the intercept.

**Discussion and next stages**

Here I will raise some points worth a quick discussion around the models and indices they produce. Firstly, the N mixture models explicitly attempt to estimate the true fixed population at a location in a given year. I think it is questionable at this local scale that such a fixed population really exists due to movement and survivorship etc. The two modelling languages also vary (so far during development) in the absolute value they predict for the populations even though patterns of population change are very similar. Consequently, I think it is sensible to view outputs as relative indices rather than an absolute estimate of abundance and/or detectability.

I have so far used a set number of GAM based models to estimate abundances. I suggest some advantages for their use; however, other structures could be considered. For example, simpler linear models accounting for population growth in a year and density dependence could be used. Alternatively, more sophisticated models could also be considered, I provide an example for stan in the code where I use a distance-based gaussian process to estimate the parameters on the GAM so that sites close to each other have more similar estimates than those further away (this model has low WAIC but is computationally expensive). Deciding between these alternatives is probably dependent on the goal of the research project and for the DRUID project I think the GAM probably strikes the right balance, but it is worth consideration. I also only use Poisson error structures on λ, and negative binomial models may provide a better fit if there are large numbers of zero counts at sites. nimbleEcology provides fast approximations of negative binomials and so adjusting the Nimble models is trivial, stan provides support for negative binomials but this would take a little more work to implement.

A decision with the GAM models is the number of splines to use. This can impact the variability in the population trend estimated across the years. When using a lower number of sites, I have used WAIC to guide model choice and the models are fast enough with this amount of data for this to be not too time-consuming. However, fits might also be guided by experience with these taxa about either too little or too much population variation at a site.

Between the models we have developed (global, shrink, and free), I think shrink may have some advantages by allowing flexibility at the site level while also constraining this variation based on estimation from all the data. The shrink model can also produce national trends through the hyper parameters and this might provide good opportunities to validate the approach against other methods for national level indices.

A key limitation with the models currently is the observational components. Here it would be helpful to have feedback on what might be important for detection (or interesting to test) for factors that are either provided with the datasets or otherwise available. It would also be helpful to consider the form the effect might take e.g. temperature – linear increase.

Finally, when comparing the different modelling languages, it is good to consider computation time. Overall, I have found that stan is superficially slower, but it samples more efficiently, whereas Nimble completes the actual sampling much faster. For a smaller sample of sites, say for matching with moth indices from Rothamsted traps, either language is fine to use, and they are fast enough for some model selection through WAIC. However, if we were to fit the whole dataset speed comparisons would probably need to be much more of a focus.

*Next stages*

It would be helpful to have this document and the code shared with modellers at BCT and RSPB who work on national trends, or similar indices, so that they could provide feedback on the approach.

With improvements in the indices ongoing, the next stage is to start to link changes in population size (Nt– Nt-1) to changes in the abundance of food resources. I have done a pilot of this for the blue tit BBS data and the Rothamsted trap data and found some promising results suggesting variation in the *Erannis defoliaria* influences population change in the blue tit. For the other species, as part of our collaborations between DRUID, RSPB and BCT we are identifying key insect food resources for selected bird and bat species and then using the DRUID derived indices to evaluate the role of these food sources on population change.

As part of understanding the role of insect food, I also plan to evaluate the effects of weather and land cover on population change to account for these factors in our analysis. For our selected species if you have any prior knowledge about what would be sensible to consider (or something that would be interesting to test) then please let me know. Otherwise, I was planning to model seasonal mean temperatures and precipitation interacting with dominant landcover.

**References**

Bat Conservation Trust. 2021. The National Bat Monitoring Programme. Annual Report 2021.

Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson. 2000. Analysis of Population Trends for Farmland Birds Using Generalized Additive Models. Ecology 81:1970–1984.

Goldstein, B., D. Turek, L. Ponisio, and P. de Valpine. 2019. nimbleEcology: Distributions for Ecological Models in’nimble’. R package version 0.1. 0.[Google Scholar].

Rennie, S., J. Adamson, R. Anderson, C. Andrews, J. Bater, N. Bayfield, K. Beaton, D. Beaumont, S. Benham, V. Bowmaker, C. Britt, R. Brooker, D. Brooks, A. Brunt, J. Brunt, S. Clawson, G. Common, R. Cooper, S. Corbett, N. Critchley, P. Dennis, J. Dick, B. Dodd, N. Dodd, N. Donovan, J. Easter, M. Flexen, A. Gardiner, D. Hamilton, P. Hargreaves, M. Hatton-Ellis, M. Howe, J. Kahl, M. Lane, S. Langan, D. Lloyd, Y. McElarney, C. McKenna, S. McMillan, F. Milne, L. Milne, M. Morecroft, M. Murphy, A. Nelson, H. Nicholson, D. Pallett, D. Parry, I. Pearce, G. Pozsgai, R. Rose, S. Schafer, T. Scott, L. Sherrin, C. Shortall, R. Smith, P. Smith, R. Tait, C. Taylor, M. Taylor, M. Thurlow, A. Turner, K. Tyson, H. Watson, M. Whittaker, and C. Wood. 2017a. UK Environmental Change Network (ECN) bird data: 1995-2015. NERC Environmental Information Data Centre.

Rennie, S., J. Adamson, R. Anderson, C. Andrews, J. Bater, N. Bayfield, K. Beaton, D. Beaumont, S. Benham, V. Bowmaker, C. Britt, R. Brooker, D. Brooks, J. Brunt, G. Common, R. Cooper, S. Corbett, N. Critchley, P. Dennis, J. Dick, B. Dodd, N. Dodd, N. Donovan, J. Easter, M. Flexen, A. Gardiner, D. Hamilton, P. Hargreaves, M. Hatton-Ellis, M. Howe, J. Kahl, M. Lane, S. Langan, D. Lloyd, M. Lundy, Y. McElarney, C. McKenna, S. McMillan, F. Milne, L. Milne, M. Morecroft, M. Murphy, A. Nelson, H. Nicholson, D. Pallett, D. Parry, I. Pearce, G. Pozsgai, R. Rose, S. Schafer, T. Scott, L. Sherrin, C. Shortall, R. Smith, P. Smith, R. Tait, C. Taylor, M. Taylor, M. Thurlow, A. Turner, K. Tyson, H. Watson, M. Whittaker, and C. Wood. 2017b. UK Environmental Change Network (ECN) bat data: 1993-2015. NERC Environmental Information Data Centre.

Royle, J. A. 2004. N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. Biometrics 60:108–115.

Stan Development Team. 2019. Stan Modeling Language Users Guide and Reference Manual, 2.29.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics 26:403–413.

Watanabe, S., and M. Opper. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research 11.