BBS GAM draft

Abstract:

Introduction

Estimates of population change derived from the North American Breeding Bird Survey form the foundation of avian conservation in North America. The Canadian Wildlife Service (CWS, a division of Environment and Climate Change Canada) and the United States Geological Survey (USGS) produce national and regional status and trend estimates for 300-400 species of birds (Smith et al. 2019, Sauer et al. 2015). These estimates are derived from models designed to account for some of the sampling imperfections inherent to an international, long-term field survey, such as variations in which sites or routes are surveyed in a given year and variability among observers (Sauer and Link 2011, Smith et al. 2014). Producing these estimates requires significant analytical expertise, time, and computing resources, but the estimates are an invaluable resource that are used by many conservation organizations and researchers to visualize, analyze, and assess the population status of over 400 species of birds (e.g., Saracco et al. 2008, NABCI Canada 2019, NABCI U.S. 2014, Rosenberg et al. 2019, Stanton et al. 2019, Downes et al. 2017).

While the estimates of status and trend produced by the USGS and CWS are used for many different purposes, not all uses of the estimates are equally well supported by the standard models, and so there is a need for alternative models and for a continual evolution of the modeling. Different conservation-based uses of the BBS status and trend estimates relate to different aspects of population change, including long-term trends, short-term trends, changes in population trends, or annual fluctuations, and no single model can estimate all parameters equally well (). This is not a criticism of the standard model; it is true of any single model. For example, the standard models used since approximately 2011, estimate population change using slope-parameters and random year-effects in a hierarchical Bayesian framework (Sauer and Link 2011, Smith et al. 2014). These slope and year-effects are well suited to estimating annual fluctuations around a long-term change, but the model tends to be conservative when it comes to estimating changes in a species’ population trend, or population cycles (Fewster et al. 2000, Smith et al. 2015, Amano et al. 2012). Similarly, short-term trends (e.g., the last 10-years of the time-series) derived from the standard models incorporate information from the entire time-series (i.e., the slope component of the model). This is a reasonable and useful assumption for data-sparse species and regions, which guards against extreme and imprecise fluctuations in short-term trends. However, for assessing changes in trends of a once-declining species (e.g., recovery of a species at risk), this feature of the model is problematic.

GAMs provide a flexible framework for tracking changes in populations over time, without any assumptions about a particular pattern in population change (Fewster et al., 2000, Knape 2016). The semi-parametric smooths are able to fit almost any shape of population trajectory, including stable populations, constant rates of increase or decrease, cycles of varying frequency and amplitude, or abrupt change points in population trends (Wood 2017). In addition, the addition of contemporary data has little or no influence on estimates of population change in the earlier portions of the time-series. By contrast, the slope parameter in the standard models estimates a constant rate of population change across the entire time-series, effectively assuming that there is some consistent rate of change. As a result, estimates of the rate of a species population change in the early portion of the time series (e.g., during the 1970s or 80s) will change as new data are added in response to contemporary data and recent rates of population change. The random walk structure of a first-difference model (Link et al. 2017) assumes that the population in each year is similar to the population in the previous year—a biologically reasonable assumption—but if data are sparse in a given region and year, the population trends are shrunk towards zero (i.e., effectively assuming a stable population).

GAMs also provide a useful framework for sharing information on the shape and rate of population change across a species range. The GAM smoothing parameters can be estimated as random effects within geographic strata, thus allowing the model to share information on the shape of a species population trajectory across sub-regions of a species range. In the terminology of Pedersen et al. 2019, this hierarchical structure on the GAM parameters would make our model a “HGAM” (Hierarchical Generalized Additive Model). However, it is also includes random effects for parameters not included in the smooth and could therefore be referred to as a GAMM (Generalized Additive Mixed Model), in the terminology of Wood 2017. In the standard model, the slope parameters can be estimated as random effects and share information among strata, which improves estimates of trend for relatively data-sparse regions (Link et al. 2017, Smith et al. 2019). Although recent work has shown that the standard model is, for many species, out-performed by a first-difference model (Link et al. 2019), the population change components of the first-difference model (Link et al. 2017), include no way to share information on population change in space and so population trajectories are estimated independently among strata.

The inherently smooth temporal patterns generated by GAMs are well suited to particularly common conservation uses, such as assessments of trends in populations from any portion of a time-series as well as assessments of the rate of change in the trends over time. For example, the population trend criteria of the IUCN or COSEWIC assessments for endangered or threatened species, are based on rates of change over 3 generations. For most bird species monitored by the BBS, this 3-generation time is approximately 10-years and so short-term trends are generally used. Because of the annual fluctuations estimated by the standard model, these short-term trends can fluctuate from year to year, greatly complicating the quantitative assessment of a species trend in comparison to the regulatory thresholds. Species trends may surpass the threshold in one year, but not in the next. The same end-point comparisons on estimates from a GAM will change much more gradually over time, and be much less dependent on the particular year in which a species was assessed.

In this paper, we describe a status and trend model that uses a hierarchical GAM, smoothing function to estimate the annual population status of birds using data from the BBS. This model allows for the sharing of information about a species’ population trajectory among geographic strata and for the decomposition of long- and medium-term population changes from annual fluctuations. We also compare the fit of the GAM, and a version that also includes random year-effects, to the fit of the standard BBS status and trend model (Sauer and Link 2011, Smith et al. 2015, and Smith et al. 2019) as well as the fit of a first-difference model where trajectory is modeled as a random walk from year-1 (Link and Sauer 2015, Link and Sauer 2019).

Methods

Overview

We designed a hierarchical Bayesian model for estimating status and trends from the North American Breeding Bird Survey (BBS) that uses a Generalized Additive Model (GAM) smooth to estimate the medium- and long-term temporal components of a species population trajectory. In the model, the stratum-level GAM smooths are treated as random-effects, so that information is shared on the shape of a species’ population trajectory among sub-regions of the species’ range (an alternative parameterization to the “GS” model in Pedersen et al. 2019). We applied two versions of the GAM: one in which the GAM smooth provided the only parameters modeling changes in abundance over time (GAM), and another in which random year effects were also estimated to allow for single-year deviations from the GAM smooth (GAMYE).

Using the BBS data for a selection of species, we compared estimates and predictive accuracy of our two models using the GAM smooth, against two alternative models that have been used to analyze the BBS. The four statistical models differed only in the parameters used to model changes in species relative abundance over time. We used k-fold cross validation to estimate the Bayesian Predictive Information Criterion (BPIC, a measure of out-of-sample predictive success; Link and Sauer 2015) for each of the models. We compared overall BPIC values among the models, and we explored the spatial and temporal variation in predictive accuracy in depth.

Bayesian hierarchical GAM

We based the GAM, and all of the other models used here, on the model described in Sauer and Link (2011), varying only the components that model change in populations over time. All of the models here have the same basic structure:

The models treat the observed BBS counts as overdispersed Poisson random variables, with mean (i.e., geographic stratum i, observer and route combination j, and year t). The means are log-linear functions of stratum-specific intercepts (), observer-route effects (), first-year startup effects for a given observer (), a count-level random effect to model overdispersion (), and a temporal component estimated using some function of year (). The models here only varied in their temporal components. Priors were set following Link et al. (2017), and using the heavy-tailed, t-distribution to model the count-level extra-Poisson dispersion (Link et al. 2019).

For the simpler of the two models that use the GAM smooth, the temporal component was estimated as a semi-parametric smooth, estimated following Crainiceanu et al (2005).

Where is the number of knots, is the year-t and k-th entry in the design matrix (defined below), and is the -length vector of parameters that control the shape of the trajectory in stratum-i. Each is estimated as a random effect, centered on a Hyperparameter: a mean across all strata ().

The mean smoothing parameters () are themselves estimated as normally distributed random effects with a mean of zero, and variance controlling the shrinkage towards a first-degree polynomial (i.e., a line). These variance parameters () control the complexity penalty of the smooths and the variation in complexity among strata (Crainiceanu et al. 2005).

The design matrix for the smoothing function ( ) has a row for each year, and a column for each of knots. The GAM smooth represented a 3rd-degree polynomial spline: , and it was calculated in R, following Crainiceanu et al (2005). We have explored an alternative parameterization, following Wood et al. 2017 using the jagam function in the R-package mgcv (i.e., model GS in Pedersen et al. 2019). For the species here, the overall trajectories obtained with our model parameterization were very similar to results from a non-Bayesian analysis of the same data using mgcv. However, when we tried to implement a fully Bayesian version of the mgcv parameterization, parameter convergence was significantly less efficient (> 5-fold increase in computation time).

Here, we have used 13 knots ( = 13), across the 53 year time-series of the BBS (1966-2018), the default setting in the R-package bbsBayes is to add one knot for every 4 years in the time-series. With this number of knots, we have found that the 53-year trajectories are sufficiently flexible to capture all but the shortest-term variation (i.e., long- and medium-term variation but not annual fluctuations). We have explored models with more knots (up to 20), but the estimated trajectories from those models were almost indistinguishable from those presented here. We set the knots at even intervals along the time-series.

Prior distributions on the variance components of the GAM-coefficients, were set following advice in Crainiceanu et al (2005) and Gelman (2006), as follows:

We also compared a version of the GAM that included year effects to model single-year departures from the smoothed trajectory (conceptually similar to the model described in Knape 2016, and hereafter referred to as the GAMYE). The GAMYE was identical to the GAM, with the addition of random year effects () estimated following Sauer and Link (2011) and Smith et al. (2015), as normally distributed, random effects with mean 0 and stratum-specific variances.

Adding the year-effects to the GAM structure, allowed us to decompose the estimated population trajectory into long- and medium-term changes using the GAM smooths and local annual fluctuations using the year-effects. We suggest that this decomposition is a useful feature of this model for many conservation-applications, because it allows us to estimate trends and trajectories that either include the effects of the annual fluctuations or that remove these annual fluctuations to highlight the medium- and long-term changes.

Alternative models

For a selection of species, we compared the predictions and predictive accuracy of the two GAMs against two alternative models previously used for the BBS. The SLOPE model includes a log-linear slope parameter and random year-effects to model species trajectories. It is a linear-trend model similar to the ones that are currently used by both the Canadian Wildlife Service (CWS, Smith et al. 2014) and the United States Geological Survey (USGS, Sauer et al. 2017) as an omnibus model to supply status and trend estimates from the BBS (essentially the same as model SH in Link et al 2017)

Secondly, the first-difference year effect model (DIFFERENCE) based on a model described in Link and Sauer (2015) and referred to as model DH in Link and Sauer (2017)

The DIFFERENCE model includes year-effects that follow a random walk from the first-year of the time-series, by modeling the first-order differences between years as random effects with mean zero and an estimated variance.

All four of our models share the same suite of parameters that account for variations in relative abundance among strata and among observer-route combinations, as well as the parameters that account for overdispersion and the effect of an observer’s first year of survey on a route. These effects were modeled using the same parameterization and with the same prior distributions as in Sauer and Link 2017, and for Smith et al. 2019.

All analyses in this paper were conducted in R (R Core Team, 2019), using JAGS to implement the Bayesian analyses (Plummer 2003), and an R-package bbsBayes (Edwards and Smith unpublished) to access the BBS data and run all of the models used here. The graphs relied heavily on the package ggplot2 (Wickham 2016). BUGS-language descriptions of the GAM and GAMYE, as well as all the code and data used to produce the analyses in this study are archived online (). In addition, all of the models used here can be applied to the BBS data using the R-package “bbsBayes” currently available on GitHub.

Cross-validation

We used a temporally and spatially stratified, v-fold cross-validation (Burman 1983, often termed “k-fold”, but here we use v to distinguish it from “k” knots in the GAM) with V = 15, where we held-out random sets of counts, stratified across all years and strata so that each of the V-folds included some observations from every combination of strata and years. We chose this approach over a leave-one-out approach using a random subset of counts (e.g., Link et al. 2019), because we wanted to assess the predictive success across all counts in the dataset, and because we wanted to explore the temporal and spatial patterns in predictive success. We followed a similar procedure to that outlined in Link et al. 2017 to implement the cross-validation in a parallel computing environment, using the R-packages parallel and foreach. We did not calculate WAIC because previous work has shown that WAIC does not approximate loocv well for the BBS data (Link et al. 2017).

We used the Bayesian Predictive Information Criterion (BPIC) to compare the out-of-sample predictive success of all four models (Link et al. 2019). BPIC for a given model-*M*, is the sum of the estimated log posterior probabilities of each observation-i, for the model-*M* fit to all data except those in the set-v that includes i ().

Following Link et al. 2019, we report BPIC values directly (i.e., we have not multiplied them by -2), so that larger values indicate better predictive success, that is a higher probability of the observed data given the model-*M*, the covariate vector (), and all of the data used to fit the model ().

We calculated the observation-level estimates of predictive fit for every count in the dataset, and summarized these fit values to explore the temporal and spatial patterns in support of different models (Table X).

To compare the prediction error between pairs of models, we calculated the difference in the conditional predictive ordinates of each observed count () under models 1 and 2, as , and so positive values of indicate more support for model 1. We analysed these values using an additional hierarchical Bayesian model to account for the imbalances in the BBS-data among years and regions. This model treated the values as having a t-distribution with an estimated variance and degrees of freedom. We used this robust estimation approach, instead of the z-score approach used by Link and Sauer (2019) because of the extremely heavy tails in their distribution of the values (Fig X).

We also analysed the count-level BPIC values using additional hierarchical Bayesian models that accounted for the posterior variance of each BPIC value to compare the predictive success among models, years, and geographic strata. These models treated the posterior median BPIC for each count as random variates drawn from a normal distribution with a mean and a precision given by the inverse of the posterior variance of the BPIC values. This model therefore accounted for the precision of the BPIC calculation for each count in the dataset, which we suggest is an improvement over the Z-score, paired t-test approach of Link and Sauer 2019, because it will downweight the influence of a handful of extremely poorly predicted counts. For example, Figure X shows the extremely non-normal distribution of the BPIC scores for Barn Swallow.

|  |  |  |
| --- | --- | --- |
| Summary level | Description | Rationale |
| Overall | Summarized across all data-points | Overall comparisons of predictive error among models |
| Strata | Strata-level summaries | Spatial comparisons, looking for variation in predictive error related to variation in sample sizes among strata, variation in species mean abundance, and variation across species’ range |
| Yearly | Summaries for:  each year;  first-five years;  last-five years; and  across the first- and last-five years. | Temporal comparisons, looking for variation in predictive error across the time-series, as well as variation specifically related to the end-points of the time-series, which are particularly relevant for long-term trend estimates. |

Trend estimates and their variability in space and time

We compared trends estimated from all models at the survey-wide and stratum-levels for their magnitude, precision, variation among strata, and variation among years. For all models, we used the same definition of trend following Sauer and Link (2011) and Smith et al. 2015, an interval-specific geometric mean of proportional changes in population size, expressed as a percentage. So the trend estimate for the interval from year a () through year b () is

where represents the annual index of abundance in a given year (see below). Because this estimate of trend only considers the annual abundance estimates in the years at either end of the trend period, we refer to this estimate as an end-point trend. For the GAMYE model, we decomposed the trajectory (i.e., the series of annual indices of abundance) into long- and medium-term components represented by the GAM smooth and annual fluctuations represented by the random year-effects. This decomposition allowed us to estimate two kinds of trend estimates: that include all aspects of the trajectory, and that removes the annual fluctuations, including only the GAM smooth components.

Annual indices for all the models here were calculated following Smith et al. 2019, which is conceptually similar to the approach described in Sauer and Link 2011 and Smith et al. 2015, with one small variation. Instead of using a retransformation factor that assumes a normal distribution of observer-route effects to re-scale the annual indices, we generated count scale predictions for every observer-route in a given stratum and averaged across the collection of predictions. More precisely, in the standard approach described in Sauer and Link 2011, the annual indices in a given year t and stratum i () are exponentiated sums of the relevant components of the model (). These components include the stratum-level intercepts and all of the parameters that estimate the time-series (i.e., the slopes, year-effects and GAM smooths), plus two variance components that account for the asymmetric retransformation from the log-scale parameters to the count-scale annual indices (). The variance component associated with the observer-route effects () is problematic, because it assumes that a global estimate of variance among observers and routes represents the true observer-route variance within each stratum equally well, and it assumes that the distribution of the estimated observer-route effects is approximately normal (Duan 1983). For many species, one or both of these two assumptions are not well supported and as a result, annual indices for some species and regions are over-estimated (Smith et al. 2015).

we calculated the annual indices as follows:

So that instead of relying on the half-variance, log-normal re-scaling factor (), we averaged count-scale predictions across all of the observer-routes j in the set of observer-route combinations in stratum i (). The conceptual difference is that values represent mean expected counts from among the existing collection of observer routes in a given stratum. In contrast, the values (i.e.,the standard approach) represent the mean expected count from a hypothetical new observer-route combination. However, because the variance of the observer-route effects () is not specific to the stratum, the hypothetical new observer-route is not necessarily a route from within the relevant stratum. The practical effect of this difference is that the annual indices calculated here, more closely reflect the observed average counts on BBS routes in a given stratum. Therefore, the stratum’s contribution to the overall trajectory and trend estimate better reflects the observed relative abundance of the species in that stratum. In the R-package bbsBayes, both versions of N are available for all of the models, but the approach is the default. To our knowledge, the variance component that relates to the count-level extra-Poisson variance () better meets the necessary assumptions, although the specific re-scaling factor used to reflect the t-distribution of the error is an area of ongoing research (Link et al. 2019).

For the GAMYE model, we calculated two versions of the species trajectory (): one which included the annual variation in the trajectory,

and a second which excluded the annual variations and only included the smoothing components of the GAM to estimate the time-series.

We calculated population trajectories and trends from the GAMYE model using both sets of annual indices ( and ). When comparing predictions against the other models, we use the values to plot and compare the population trajectories (i.e., including the year-effects), and the values to calculate the trends (i.e., removing the year-effect fluctuations). In the R-package bbsBayes, both versions of the population trajectories are available for the GAMYE model, but the approach is the default.

Results

Model predictions:

Population trajectories from the GAM and GAMYE are very similar. Both models suggest that BARS populations increased from the start of the survey through to approximately the early 1980s. In contrast to the standard model predictions that show a relatively steady rate of decline (Figure 1).

Trajectories vary a great deal among the strata, but are generally similar to the continental means. For data-sparse species, there is more similarity among the strata in the trajectories (graph with trajectories grouped by some arbitrary data-quality threshold).

The trajectories follow the pattern in observed counts reasonably well, but depart from it more in years with sparse data.

The decomposed trajectories from the GAMYE allow us to calculate trends from the smooth and plot trajectories that show the annual fluctuations. The smooth trajectories are very similar between the GAM and GAMYE.

Comparison to other models - Cross-validation

Both GAMs clearly outperform the standard model. Out-perform the standard model in all cases here.

Similar to the difference model. The survey-wide population trajectories for the GAMYE models are very similar to trajectories from the DIFFERENCE model. Trajectories from the GAM, follow the overall pattern of the GAMYE and are generally similar to the DIFFERENCE model, but show effectively no annual fluctuations. Trajectories from all four models are very similar in the latter half of the time-series when there are more data, but less similar in the earlier portion of the time-series when there are fewer data (Figure 1 [panel of 4 species plots with continental trajectories]). The overall cross-validation results suggest that the GAMYE and the DIFFERENCE models are comparable in their prediction accuracy, at least for these species, and outperformed both of the other models.

Do better in years and regions with sparse data?

Best model varies by region and year.

Figures

GAM trajectories – continental index, continental mean Betas… some regional estimates, plotted against the observed mean counts and the number of counts.

Annual variation in trend estimates from the different models…

Model comparisons

Trajectories for all 4 models

Trajectories of BPIC or Vj

Maps of best model

Plots of BPIC vs count-value

Discussion

This model provides useful estimates of population trajectories and trends with comparable out of sample predictive accuracy as other models.

The smoothed GAM components are particularly useful for many conservation-relevant uses. Intermediate trends, short-term trends that don’t fluctuate from year-year, short-term trends that are not influenced by the long-term trend.

We recommend the GAM-ye model for status and trend analyses, where the trend components are derived from the smooth and two estimates of trajectories can be used to examine the intermediate and long-term patterns, and the year-effects can be added to examine annual fluctuations.

There are alternative measures of trend that integrate information from all annual indices in a given trend period, such as the slope of a log-linear regression through the estimated annual indices (e.g., PIF trend estimates). However this end-point trend estimate is useful because it can be derived for any two points within the time-series and it is a derived parameter that is simple to calculate using estimates from any of the models presented here.

This decomposition serves a practical purpose in that it allows for estimates of population change and population status that exclude the annual fluctuations (Knape 2016). Excluding the annual fluctuations makes short- and medium-term trend estimates more stable between successive years and is therefore particularly useful for conservation status assessments such as population-trend criteria used in the IUCN red-listing and COSEWIC species at risk assesments. For both of these assessments, population declines beyond a particular threshold trigger large investments of resources related to policy and conservation actions, and so if the rate of population decline is strongly dependent on the particular year in which a species is assessed there is an increased risk of inefficient investments of those resources.

Predictive accuracy varies in space and time.

Predictive accuracy is not the only consideration, given that even among models with comparable BPIC, estimates of trend and trajectory were different in ways with implications for conservation.

The model can be accessed

References cited:

Crainiceanu CM, Ruppert D, Wand MP (2005). “Bayesian Analysis for Penalized Spline Re-gression Using WinBUGS.” Journal of Statistical Software, 14 (14)

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 1:515-533.

Microsoft and Steve Weston (2019). foreach: Provides Foreach Looping Construct. R package version 1.4.7. <https://CRAN.R-project.org/package=foreach>

Plummer, Martyn. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Pedersen EJ, Miller DL, Simpson GL, Ross N. 2018. Hierarchical generalized additive models: an introduction with mgcv. PeerJ Preprints 6:e27320v1 <https://doi.org/10.7287/peerj.preprints.27320v1>

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schroder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical or phylogenetic structure. - Ecography doi: 10.1111/ecog.02881.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Figures

Figure 1. Survey-wide population trajectories for Barn Swallow (Hirundo rustica) estimated from North American Breeding Bird Survey using two models described here that include a GAM smoothing function to model change over time (GAM and GAMYE) and a third trajectory estimated using the standard slope-based model used for BBS status and trend assessments since 2011 (SLOPE).