Ornithological Applications

AmericanOrnithology.org

Volume 123, 2020, pp. 1–16
DOI: 10.1093/ornithapp/duaa065

RESEARCH ARTICLE

North American Breeding Bird Survey status and trend estimates to inform a wide range of conservation needs, using a flexible Bayesian hierarchical generalized additive model

Adam C. Smith[1,*] and Brandon P. M. Edwards[2]

[1] Canadian Wildlife Service, Environment Climate Change Canada, National Wildlife Research Centre, Ottawa, Canada
[2] Department of Mathematics & Statistics, University of Guelph, Guelph, Canada
*Corresponding author: adam.smith2@canada.ca

Submission Date: April 7, 2020; Editorial Acceptance Date: October 22, 2020; Published December 26, 2020

ABSTRACT

The status and trend estimates derived from the North American Breeding Bird Survey (BBS) are critical sources of information for bird conservation. However, the estimates are partly dependent on the statistical model used. Therefore, multiple models are useful because not all of the varied uses of these estimates (e.g., inferences about long-term change, annual fluctuations, population cycles, and recovery of once-declining populations) are supported equally well by a single statistical model. Here we describe Bayesian hierarchical generalized additive models (GAMs) for the BBS, which share information on the pattern of population change across a species' range. We demonstrate the models and their benefits using data from a selection of species, and we run full cross-validation of the GAMs against 2 other models to compare the predictive fit. The GAMs have a better predictive fit than the standard model for all species studied here and comparable predictive fit to an alternative first difference model. In addition, one version of the GAM described here (GAMYE) estimates a population trajectory that can be decomposed into a smooth component and the annual fluctuations around that smooth component. This decomposition allows trend estimates based only on the smooth component, which are more stable between years and are therefore particularly useful for trend-based status assessments, such as those by the International Union for the Conservation of Nature. It also allows for the easy customization of the model to incorporate covariates that influence the smooth component separately from those that influence annual fluctuations (e.g., climate cycles vs. annual precipitation). For these reasons and more, this GAMYE model is a particularly useful model for the BBS-based status and trend estimates.

Keywords: Bayesian, Breeding Bird Survey, cross-validation, generalized additive model, population change, status and trend estimates

LAY SUMMARY

- The status and trend estimates derived from the North American Breeding Bird Survey are critical sources of information for bird conservation, but they are partly dependent on the statistical model used.
- We describe a model to estimate population status and trends from the North American Breeding Bird Survey data, using a Bayesian hierarchical generalized additive mixed model that allows for flexible population trajectories and shares information on population change across a species' range.
- The model generates estimates that are broadly useful for a wide range of common conservation applications, such as International Union for the Conservation of Nature status assessments based on trends or changes in the rates of decline for species of concern, and the estimates have better or similar predictive accuracy to other models.

Estimations de l'état et des tendances du Relevé des oiseaux nicheurs de l'Amérique du Nord pour documenter un large éventail de besoins de conservation, en utilisant un modèle additif généralisé hiérarchique bayésien flexible

RÉSUMÉ

Les estimations de l'état et des tendances dérivées du Relevé des oiseaux nicheurs de l'Amérique du Nord (BBS) sont des sources d'information essentielles pour la conservation des oiseaux. Toutefois, les estimations dépendent en partie du modèle statistique utilisé. Par conséquent, les modèles multiples sont utiles car les utilisations variées de ces estimations (p. ex., les inférences sur les changements à long terme, les fluctuations annuelles, les cycles des populations, le rétablissement de populations autrefois en déclin) ne sont pas toutes soutenues de manière égale par un seul modèle statistique. Nous décrivons ici des modèles additifs généralisés hiérarchiques bayésiens (GAM) pour le BBS, qui partagent des informations sur le patron de changement des populations dans l'aire de répartition d'une espèce. Nous faisons la démonstration des modèles et de leurs avantages en utilisant des données provenant d'une sélection d'espèces, et nous effectuons une validation croisée complète des GAM par rapport à deux autres modèles pour comparer l'adéquation prédictive. Les GAM ont une meilleure adéquation prédictive que le modèle standard pour toutes les espèces étudiées, et une adéquation prédictive comparable à un modèle de première différence alternatif. De plus, une version du GAM décrite ici (GAMYE) estime une trajectoire de population qui peut être décomposée en une composante lisse et en fluctuations annuelles autour de cette composante lisse. Cette décomposition permet d'obtenir des estimations de tendances basées seulement sur la composante lisse qui sont plus stables d'une année à l'autre et qui sont donc particulièrement utiles pour les évaluations de l'état des populations basées sur les tendances, telles que celles de l'Union internationale pour la conservation de la nature. Elle permet également d'adapter facilement le modèle afin d'incorporer des covariables qui influencent la composante lisse séparément de celles qui influencent les fluctuations annuelles (p. ex., les cycles climatiques vs les précipitations annuelles). Pour ces raisons et d'autres encore, ce modèle GAMYE est particulièrement utile pour les estimations de l'état et des tendances basées sur le BBS.

Mots-clés: bayésien, Relevé des oiseaux nicheurs, validation croisée, modèle additif généralisé, changement dans la population, estimations de l'état et des tendances

INTRODUCTION

Estimates of population change derived from the North American Breeding Bird Survey (BBS) are a keystone of avian conservation in North America. Using these data, the Canadian Wildlife Service (CWS, a branch of Environment and Climate Change Canada) and the United States Geological Survey (USGS) produce national and regional status and trend estimates (estimates of annual relative abundance and rates of change in abundance, respectively) for 300–500 species of birds (Sauer et al. 2014, Smith et al. 2019). 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 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 they are used by many conservation organizations and researchers to visualize, analyze, and assess the population status of many North American bird species (Rosenberg et al. 2017, North American Bird Conservation Initiative [NABCI] Canada 2019, Rosenberg et al. 2019).

While the estimates of status and trend from the BBS serve many different purposes, not all uses are equally well supported by the standard models, and so there is a need for alternative models and for 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 for overall status (Partners in Flight 2019), short-term trends to assess extinction risk (International Union for the Conservation of Nature [IUCN] 2019), changes in population trends to assess species recovery (Environment Climate Change Canada 2016), or annual fluctuations (Wilson et al. 2018). Each one of these uses relies on different parameters or spatial and temporal variations in those parameters, and no single model can estimate all parameters equally well. This is not a criticism; it is true of any single model. For example, the standard model used between 2011 and 2017 in the United States and 2011 and 2016 in Canada is essentially a Poisson regression model, which estimates population change using random year effects around a continuous slope in a Bayesian hierarchical framework (Sauer and Link 2011, Smith et al. 2014). These slope and year effects are well suited to estimating annual fluctuations around a continuous long-term change, but the model tends to be conservative when it comes to estimating changes in a species' population trend (e.g., population recovery after a decline) or population cycles (Fewster et al. 2000, Smith et al. 2015). Similarly, short-term trends (e.g., the last 10 yr of the time series) derived from this standard model incorporate information from the entire time series (i.e. the slope component of the model). For many purposes, this is a reasonable and useful assumption, which guards against extreme and imprecise fluctuations in short-term trends. However, this feature of the model is problematic for assessing changes in trends of a once-declining species, such as the recovery of a species at risk (Environment and Climate Change Canada 2016).

Generalized additive models (GAMs, Wood 2017) provide a flexible framework for tracking changes in populations over time, without any assumptions about a particular temporal pattern in population change (Fewster et al. 2000, Knape 2016). The semi-parametric smooths can fit almost any shape of population trajectory, including stable populations, constant rates of increase or decrease, cycles of varying frequency and amplitude, or change points in population trends (Wood 2017). Furthermore, the addition of new data in subsequent years has relatively little influence on estimates of population change in the earlier portions of the time series. In contrast, the slope parameter in the standard models effectively assumes that there is some consistent rate of change. As a result, to the extent that the slope parameter influences the estimated trajectory, estimates of the rate of a species population change in the early portion of the time series (e.g., during the 1970s or 1980s) can change in response to the addition of contemporary data and recent rates of population change.

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 the population trajectory across a species range. In the terminology of Pedersen et al. 2019, this hierarchical structure on the GAM parameters would make our model a "Hierarchical Generalized Additive Model" (HGAM). However, it also includes random effects for parameters not included in the smooth and could therefore be referred to as a Generalized Additive Mixed Model (GAMM), in the terminology of Wood 2017. Similarly in the standard model, the slope parameters can be estimated as random effects and share information among strata, which improves estimates of the 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. 2020), 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. Of course, for some conservation uses, this independent estimation of population trajectories might be critical (e.g., if one were interested specifically in estimating the differences in trends among provinces or states), and in these situations, the sharing of information could be problematic.

Trend estimates (interval-specific rates of mean annual population change, Sauer and Link 2011, Link et al. 2020) derived from 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 change in the trends over time. For example, the population trend criteria of the IUCN (IUCN 2019) or Canada's national assessments by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) are based on rates of change over 3 generations. For most bird species monitored by the BBS, this 3-generation time is approximately the same as the 10-yr, short-term trends produced by the CWS and USGS analyses. Because of the inclusion of year effects in the standard model, these short-term trends fluctuate from year to year, complicating the quantitative assessment of a species trend in comparison to the thresholds. Species trends may surpass the threshold in 1 yr, 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 is assessed.

In this article, we describe a status and trend model that uses a hierarchical GAM to estimate the relative abundance trajectory of bird populations, 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 GAM version that includes random year effects (conceptually similar to Knape et al. 2016), to the fit of 2 alternative models commonly applied to BBS data (Sauer and Link 2011, Smith et al. 2015, Link et al. 2020).

METHODS

Overview

We designed a Bayesian hierarchical model for estimating status and trends from the North American BBS that uses a GAM smooth to estimate the medium- and long-term temporal components of a species population trajectory (i.e. changes occurring over time periods ranging from 3 to 53 yr). In the model, the parameters of the GAM smooths are treated as random effects within the geographic strata (the spatial units of the predictions, intersections of Bird Conservation Regions, and province/state/territory boundaries), so that information is shared on the shape of the population trajectory across the species' range. In comparison to the non-Bayesian HGAMs in the work of Pedersen et al. 2019, our model is most similar to the "GS" model, which has a global smooth in addition to group-level smooths with a similar degree of flexibility. We applied 2 versions of the GAM: one in which the GAM smooth was the only component modeling changes in abundance over time (GAM) and another in which random year effects were also estimated to allow for single-year departures from the GAM smooth (GAMYE, which is conceptually similar to the model described in the work of Knape 2016).

For a selection of species, we compared estimates and predictive accuracy of our 2 models using the GAM smooth, against 2 alternative models that have been used to analyze the BBS data. We chose the main comparison species (Barn Swallow, *Hirundo rustica*) because of the striking differences between trajectories from the SLOPE model and a number of nonlinear models (Smith et al. 2015, Sauer and Link 2017). We added a selection of other species to represent a range of anticipated patterns of population change, including species with known change points in their population trajectories (Chimney Swift, *Chaetura pelagica*; Smith et al. 2015), and species with relatively more data and known large and long-term trends (Wood Thrush, *Hylocichla mustelina* and Ruby-throated Hummingbird, *Archilochus colubris*) and species with relatively fewer data and long-term changes (Canada Warbler, *Cardellina canadensis*, Cooper's Hawk, *Accipiter cooperii*, and Chestnut-collared Longspur, *Calcarius ornatus*). Finally, we also added a few species with strong annual fluctuations and/or abrupt step-changes in abundance (Pine Siskin, *Spinus pinus* and Carolina Wren, *Thryothorus ludovicianus*).

The BBS data are collected along roadside survey routes that include 50 stops at which a 3-min point count is conducted, once annually, during the peak of the breeding season (Robbins et al. 1986, Hudson et al. 2017, Sauer et al. 2017). All of the models here use the count of individual birds observed on each BBS route (summed across all 50 stops) in a given year by a particular observer. The 4 statistical models differed only in the parameters used to model changes in species relative abundance over time. We used 15-fold cross-validation (Burman 1989) to estimate the observation-level, out-of-sample predictive accuracy of all 4 models (Vehtari et al. 2017, Link et al. 2020). We compared the overall predictive accuracy among the models, and we explored the spatial and temporal variation in predictive accuracy in depth.

We compared 4 alternative BBS models, all of which have the same basic structure:

$$\log\left(\lambda_{s,j,t}\right) = \theta_s + \ \Delta\ _s\left(t\right) + \eta I\left[j,t\right] + \omega_j + \ \varepsilon_{s,j,t}$$

The models treat the observed BBS counts as overdispersed Poisson random variables, with mean $\lambda_{s,j,t}$ (i.e. geographic stratum s, observer and route combination j, and year t). The means are log-linear functions of stratum-specific intercepts (θ_s, estimated as fixed effects and with the same priors following Smith et al. 2014), observer-route effects (ω_j, estimated as random effects and with the same priors following Sauer and Link 2011), first-year startup effects for an observer(η, estimated as fixed effects and with the same priors following Sauer and Link 2011), a count-level random effect to model overdispersion ($\varepsilon_{s,j,t}$, estimated using heavy-tailed, t-distribution and with the same priors following Link et al. 2020), and a temporal component estimated using a function of year, which varies across the 4 models ($\Delta_s\left(t\right)$). The models here only varied in their temporal components ($\Delta_s\left(t\right)$).

Bayesian Hierarchical GAMs

Generalized additive model. The main temporal component $\Delta_s\left(t\right)$ in the GAM was modeled with a semi-parametric smooth, estimated following Crainiceanu et al. (2005) as

$$\Delta_s^{\mathrm{GAM}}\left(t\right) = \sum_{k=1}^{K} \beta_{s,k}\chi_{t,k}$$

where K is the number of knots, $\chi_{t,k}$ is the year t and k th entry in the design matrix X(defined below), and $\beta_{s,k}$is the K-length vector of parameters that control the shape of the trajectory in stratum s. Each $\beta_{s,k}$ is estimated as a random effect, centered on a mean across all strata (a hyperparameter B_k)

$$\beta_{s,k} \quad \sim \quad \mathrm{Normal}\left(\mathrm{B}_k, \sigma_\beta^2\right)$$

and

$$\mathrm{B}_K \quad \sim \quad \mathrm{Normal}\left(0, \sigma_\mathrm{B}^2\right)$$

where the variance σ_B^2 acts as the complexity penalty, shrinking the complexity and the overall change of the mean trajectory toward a flat line). It would be possible to add an additional slope parameter, as was done in the work of Crainiceanu et al. 2005, but we have found that the BBS data for most species are insufficient to allow for the separate estimation of the linear component to population change and the additive smooth. In addition, we see little benefit to including a linear component because the assumptions required to include a constant linear slope for a 53-yr time series are unlikely to be met for any continental-scale population. In combination, these variance parameters ($\sigma_\beta^2, \sigma_\mathrm{B}^2$) control the complexity penalty of the species trajectories and the variation in pattern and complexity among strata and were given the following priors, following advice in the work of Crainiceanu et al (2005):

$$\sigma_\beta^2 \quad \sim \quad \frac{1}{\mathrm{gamma}\left(2, 0.2\right)}$$

$$\sigma_\mathrm{B}^2 \quad \sim \quad \frac{1}{\mathrm{gamma}\left(10^{-2}, 10^{-4}\right)}$$

These prior parameters were chosen to ensure that the priors are sufficiently vague that they are overwhelmed by the data, particularly for σ_B^2 that controls the shape of the survey-wide trajectory (Crainiceanu et al 2005). We have so far had good results across a wide range of species using these priors, and in tests of alternative priors, there is no effect on posterior estimates (Supplementary Material Figure S9). For example, estimates of B_K and σ_B for Chestnut-collared Longspur (a relatively data-poor species) are unchanged even if using a much more restrictive prior on σ_B that places 99% of the prior density for σ_B below 1.2 ($\sigma_\mathrm{B}^2 \sim \frac{1}{\mathrm{gamma}(2,0.2)}$). However, these variance priors are an area of ongoing research, aimed at improving the efficiency of the Markov Chain Monte Carlo (MCMC) sampling.

The design matrix for the smoothing function (X) has a row for each year and a column for each of K knots. The GAM smooth represented a third-degree polynomial spline:$\chi_{t,k} = |t' - t'_k|^3$, and was calculated in R, following the work of Crainiceanu et al (2005). We centered and re-scaled the year values to improve convergence, so that $t' = {}^{(t-\mathrm{midyear})}\!/_{T}$, where midyear is the middle year of the time series and T is the number of years in the time series. Here, we have used 13 knots (K = 13), across the 53-yr time series of the BBS (1966–2018), which results in approximately one knot for every 4 yr in the time series. With this number of knots, we have found that the 53-yr trajectories are sufficiently flexible to capture all but the shortest-term variation (i.e. variation on the scale of 3–53 yr, but not annual fluctuations). Models with more knots are possible but in the case of a penalized smooth, the overall patterns in the trajectory will be similar, as long as a sufficient number of knots is allowed (Wood 2017). The number of knots could be customized in a species-specific approach; however, because we are looking for a general model structure that can be applied similarly across the >500 species in the BBS, we have fixed the number of knots at 13. Our approach relies on the shrinkage of the smoothing parameters (Β, β) to ensure that the trajectories are only as complex as the data support, and the limited number of knots constrains the complexity of the additive function (Fewster et al. 2000, Wood 2017).

GAMYE. The GAMYE was identical to the GAM, with the addition of random year effects ($\gamma_{t,s}$) estimated independently among strata, following the works of Sauer and Link (2011) and Smith et al. (2015), as

$$\gamma_{t,s} \sim \text{Normal}\left(0, \sigma^2_{\gamma,s}\right)$$

where $\sigma^2_{\gamma,s}$ are stratum-specific variances. Thus, the temporal component for the GAMYE is given by

$$\Delta_s^{\text{GAMYE}}(t) = \sum_{k=1}^{K} \beta_{s,k}\chi_{t,k} + \gamma_{t,s}$$

The GAMYE trajectories are therefore an additive combination of the smooth and random annual fluctuations. The smooth components of the trajectory in the GAMYE are generally similar to those from the GAM, but tend to be slightly less variable (i.e. less complex) because the year effects components can account for single-year deviations from the longer-term patterns of population change. The full trajectories from the GAMYE (smooth plus the year effects) generally follow the same overall pattern as the GAM estimates and include abrupt single-year changes in abundance, which increases the capacity to model step-changes in abundance.

Alternative Models

For a selection of species, we compared the predictions and predictive accuracy of the 2 GAMs against 2 alternative models previously used for the BBS.

SLOPE. The SLOPE model includes a slope parameter and random year effects to model species trajectories. It is a linear year-effects model currently used by both the CWS (Smith et al. 2014) and the USGS (Sauer et al. 2017) as an omnibus model to supply status and trend estimates from the BBS (essentially the same as model SH, the SLOPE model with heavy-tailed error in the work of Link et al. 2017). The temporal component in the SLOPE model is

$$\Delta_s^{\text{SLOPE}}(t) = \beta_s*(t - t_{\text{mid}}) + \gamma_{t,s}$$

DIFFERENCE. The first difference model (DIFFERENCE) is based on a model described in the work of Link and Sauer (2016) and models the temporal component as

$$\Delta_s^{\text{DIFFERENCE}}(t) = \gamma_{t,s} = N\left(\gamma_{t-1,s}, \sigma^2_{\gamma_s}\right)$$

The DIFFERENCE model includes year effects that follow a random walk prior 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 analyses in this article 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 2020) to access the BBS data and run all of the models used here. We used the same number of burn-in iterations (10,000), thinning rate (1/10), chains (3), and number of saved samples from the posterior (3,000) to estimate trends and trajectories for all models. We examined trace plots and the Rhat statistic to assess convergence. 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 (see Data availability in Acknowledgements).

Cross-Validation

We used a temporally and spatially stratified v-fold cross-validation (Burman 1989, often termed “k-fold,” but here we use Berman’s original “v-fold” to distinguish it from “k” which is often used to describe the number of knots in a 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 almost every combination of strata and years. We did this by randomly allocating each count within a given stratum and year to one of the 15 folds. We chose this approach over a leave-one-out (loo) cross-validation approach using a random subset of counts (Link et al. 2017) because we wanted to assess the predictive success across all counts in the dataset, explore the temporal and spatial patterns in predictive success. Also, although full loo cross-validation minimizes bias and variance of the estimate of predictive accuracy (Zhang and Yang 2015), a full loo is not practical for computational reasons and cross-validation with

k >10 is a reasonable approximation of loo (Kohavi 1995, Vehtari et al. 2017). We could also have chosen to conduct structured cross-validation (Roberts et al. 2017), but cross-validation in a Bayesian context has particularly large computational requirements; there are multiple levels of dependencies in the BBS data (dependences in time, space, and across observers); and models being compared vary in the way they treat some of those dependencies (models that share information differently in space and/or time). Therefore, we chose a relatively simple non-structured approach where the folds are balanced in time and space, and for a given species were identical across all models compared. We followed a similar procedure to that outlined in the work of Link et al. (2017) to implement the cross-validation in a parallel computing environment, using the R-package for each (Wallig and Weston 2020). We used the end-values from the model-run using the full dataset as initial values in each of the 15 cross-validation runs, ran a short burn-in of 1,000 samples, then used a draw of 3,000 samples of the posterior with a thinning rate of 1/10 spread across 3 chains. We did not calculate the widely applicable information criterion (WAIC), because previous work has shown that WAIC does not approximate loo well for the BBS data (Link et al. 2017, 2020).

We used the estimated log predictive density ($\text{elpd}_{i,M}$) to compare the observation-level, out-of-sample predictive success of all 4 models (Vehtari et al. 2017, Link et al. 2020). For a given model M, elpd is the estimated log posterior density for each observation i, for the model M fit to all data except those in the set v that includes i ($Y_{-v,\ i\in v}$). That is,

$$\text{elpd}_{i,M} = \quad \log\left(f_M\left(Y_i | Y_{-v, \quad i\in v}, X_i\right)\right)$$

Larger values of elpd indicate better predictive success, that is a higher probability of the observed data given the model M, the estimated parameters, the vector of covariates for observation i, such as the year, observer-route, etc. (X_i), and all of the data used to fit the model ($Y_{-v,\ i\in v}$).

We have not summed elpd values to generate Bayesian predictive information criterion (BPIC) values; rather, we have compared model-based estimates of mean difference inelpd between pairs of models. We modeled the elpd values so that we could account for the imbalances in the BBS data in time and space (i.e. the variation in number of counts among strata and years). The raw sum of the elpd values would give greater weight to the regions with more data and to the recent years in the time series, which have more counts. Therefore, expanding on the approach in the work of Link et al. 2020 that used a z-score to estimate the significance of the difference in fit between 2 models, we used a hierarchical model to estimate the mean of the differences in predictive fit (δ_i^{elpd}). We first calculated the difference in the elpd of each observed count (Y_i) under models 1 and 2, as

$$\delta_{i,M1-M2}^{\text{elpd}} = \quad \log\left(f_1\left(Y_i | Y_{-v,\ i\in v}, X_i\right)\right) - \log\left(f_2\left(Y_i | Y_{-v,\ i\in v}, X_i\right)\right)$$

, so that positive values of $\delta_{i,M1-M2}^{\text{elpd}}$ indicate more support for model 1. We then analyzed these δ_i^{elpd}values using an additional Bayesian hierarchical model, with random effects for year and strata to account for the variation in sampling effort in time and space. These random effects account for the imbalances in the BBS data among years and regions, and the inherent uncertainty associated with any cross-validation statistic (Link et al. 2017, Vehtari et al. 2017). This model treated the elpd differences for a count from a given year t and stratum s ($\delta_{i,s,t}^{\text{elpd}}$) as having a t-distribution with an estimated variance (σ_δ^2) and degrees of freedom (ν). That is,

$$\delta_{i,s,t}^{\text{elpd}} = t\left(\mu_i, {\sigma^2}_\delta\ , \nu\right)$$

$$\mu_i = \phi + \psi_s + \psi_t$$

From the model, ϕ was our estimate of the overall comparison of the mean difference in predictive fit for Model 1 – Model 2 ($\delta_{M1-M2}^{\text{elpd}} = \ \phi$), $\phi + \psi_s$ was the estimate of the mean difference in stratum s, and $\phi + \psi_t$ was the estimated difference in year t. The year and stratum effects ($\psi_s + \psi_t$) were estimated as random effects with a mean of zero and estimated variances given uninformative inverse gamma priors. We used this t-distribution as a robust estimation approach, instead of the z-score approach used by Link et al. (2020) because of the heavy tails in the distribution of the δ_i^{elpd} values (Supplementary Material Figure S7). Given these heavy tails, a statistical analysis assuming a normal distribution in the differences would give an inappropriately large weight to a few counts where the prediction differences were large in magnitude (Gelman et al. 2014). In essence, our model is simply a "robust" version of the z-score approach (Lange et al. 1989) with the added hierarchical parameters to account for the spatial and temporal imbalance in the BBS data.

Trends and Population Trajectories

For all models, we used the same definition of trend following Sauer and Link (2011) and Smith et al. (2015); that is, an interval-specific geometric mean of proportional changes in population size, expressed as a percentage. Thus, the trend estimate for the interval from year a (t_a) through year b (t_b) is given by

$$R_{a:b} = 100 \times \left(\left(\frac{N_{t_a}}{N_{t_b}}\right)^{\frac{1}{t_a - t_b}} - 1\right)$$

where $N_.$ 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 2 kinds of trend estimates: $R_{a:b}$ that include all aspects of the trajectory, and $R'_{a:b}$ that removes the annual fluctuations, including only the GAM smooth components.

Population trajectories are the collection of annual indices of relative abundance across the time series. These indices approximate the mean count on an average BBS route, conducted by an average observer, in a given stratum and year. For all the models here, we calculated these annual indices for each year t and stratum s following the work of Smith et al. (2019) as

$$N_{s,t} = z_s \times \frac{\sum_{j\in J_s} e^{\mathrm{A}_{s,t}+\omega_j+0.5\times\sigma^2{}_{\varepsilon}}}{n_{Js}}$$

where each $N_{s,t}$ is an exponentiated sum of the relevant components of the model ($\mathrm{A}_{s,t}$), observer-route effects (ω_j), and count-level extra-Poisson variance ($0.5 \times \sigma^2{}_{\varepsilon}$), averaged over count-scale predictions across all of the n_{Js} observer-routes j in the set of observer-route combinations in stratum s (J_s), and then multiplied by a correction factor for the proportion of routes in the stratum on which the species has been observed (z_s, i.e. the proportion of routes on which the species has been observed; on all other routes species abundance is assumed to equal zero and they are excluded from the model, see Sauer and Link 2011). This is slightly different from the approach described in the works of Sauer and Link (2011) and Smith et al. (2015) and an area of ongoing research. We have found that this different annual index calculation ensures that the annual indices are scaled more similarly to the observed mean counts, which can affect the relative weight of different strata in trends estimated for broader regions (e.g., continental and national trends), but it has no effect on the trends estimated within each stratum and no effect on the cross-validation results presented here. For a discussion on the differences between these 2 ways of calculating annual indices, refer to Supplementary Material Appendix A.

For the GAMYE model, we calculated 2 versions of the species trajectory (N_s): one that included the annual variation in the trajectory,

$$N_{s,t} = z_s \times \frac{\sum_{j\in J_s} e^{\mathrm{A}_{s,t}+\omega_j+0.5\times\sigma^2{}_{\varepsilon}}}{n_{Js}}$$

$$\mathrm{A}_{s,t} = \theta_s + f_s(t) + \quad \gamma_{s,t}$$

and a second that excluded the annual variations, including only the smoothing components of the GAM to estimate the time series,

$$Ng_{s,t} = z_s \times \frac{\sum_{j\in J_s} e^{\mathrm{Ag}_{s,t}+\omega_j+0.5\times\sigma^2{}_{\varepsilon}}}{n_{Js}}$$

$$\mathrm{Ag}_{s,t} = \theta_s + f_s(t)$$

We calculated population trajectories and trends from the GAMYE model using both sets of annual indices ($N_{s,t}$ and $Ng_{s,t}$). When comparing predictions against the other models, we used the $N_{s,t}$ values to plot and compare the population trajectories (i.e. including the year effects), and the $Ng_{s,t}$ values to calculate the trends (i.e. removing the year-effect fluctuations).

RESULTS

Model Predictions

Population trajectories from the GAM, GAMYE, and DIFFERENCE are similar. All 3 of these models suggest that Barn Swallow populations increased from the start of the survey to approximately the early 1980s, compared to the SLOPE model predictions that show a relatively steady rate of decline (Figure 1). The trajectories for all species from both GAMs and the DIFFERENCE model were less linear overall than the SLOPE model trajectories and tended to better track nonlinear patterns, particularly in the early years of the survey and often in more recent years as well (Figure 1, Supplementary Material Figures S1 and S6). GAM and GAMYE trajectories vary a great deal among the strata, particularly in the magnitude and direction of the long-term change (Figure 2 for Barn Swallow). However, there are also many similarities among the strata, in the nonlinear patterns that are evident in the continental mean trajectory (e.g., the downward inflection in the early 1980s in Figure 2 and Supplementary Material Figure S2). Figure 3 shows the estimated trajectories for Barn Swallow in the 6 strata that make up BCR 23 from the GAMYE, DIFFERENCE, and SLOPE models. The GAMYE estimates suggest that the species' populations increased in the early portion of the time series in all of the strata, and this is a pattern shared with the continental mean trajectory for the species (Figure 2). In contrast, the estimates from the SLOPE model only show an increase in the stratum with the most data (i.e. the most stacked gray dots along the x-axis indicating the number of BBS routes contributing data in each year, US-WI-23), the DIFFERENCE model shows more of the early increase in many strata, except those with the fewest data. In the other strata with fewer data, the SLOPE trajectories are strongly linear and the DIFFERENCE trajectories are

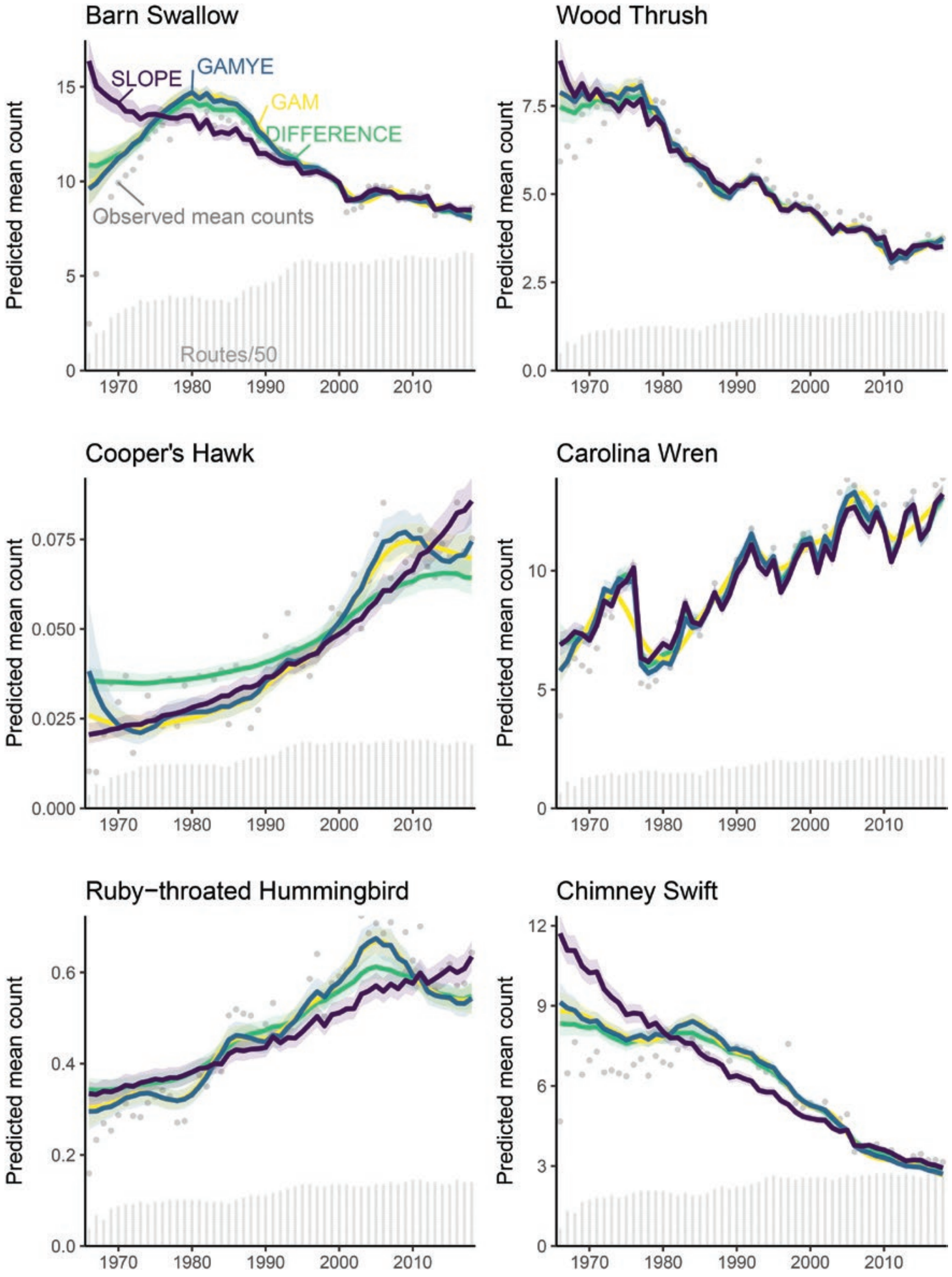

FIGURE 1. Survey-wide population trajectories for Barn Swallow (*Hirundo rustica*), Wood Thrush (*Hylocichla mustelina*), Cooper's Hawk (*Accipiter cooperii*), Carolina Wren (*Thryothorus ludovicianus*), Ruby-throated Hummingbird (*Archilochus colubris*), and Chimney Swift (*Chaetura pelagica*), estimated from the BBS using 2 models described here that include a GAM smoothing function to model change over time (GAM and GAMYE) the standard regression-based model used for BBS status and trend assessments since 2011 (SLOPE), and a first difference time series model (DIFFERENCE). The stacked dots along the *x*-axis indicate the approximate number of BBS counts used in the model in each year; each dot represents 50 counts.

particularly flat in the early years with particularly few data. The cross-validation results suggest that for Barn Swallow, the GAMYE is preferred over the SLOPE model, and generally preferred (some overlap with 0) to the DIFFERENCE model (Figure 4), particularly in the early years of the survey (pre-1975, Supplementary Material Figure S6). Finally, the general benefits of sharing information among strata on the shape of the population trajectory are evident for the GAM, GAMYE, and the SLOPE models in Figure 5, where there is no relationship between the sample size and the absolute value of the long-term trend for Cooper's Hawk (more below).

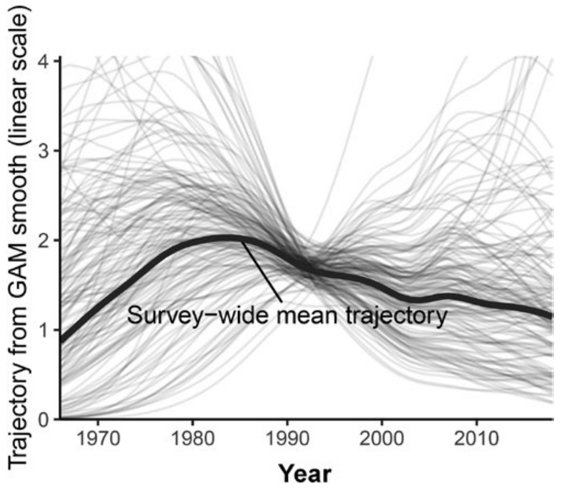

FIGURE 2. Variation among the spatial strata in the random-effect smooth components of the GAMYE model applied to Barn Swallow data from the BBS. Gray lines show the strata-level random-effect smooths, and the black lines show the survey-wide mean trajectory.

For most species here, the GAMs or the DIFFERENCE model generally were preferred over the SLOPE model (Figure 4, Supplementary Material Figure S3). For the 2 species with population trajectories that are known to include strong year effects (Carolina Wren and Pine Siskin), the GAM model that does not include year effects performed poorly (Figure 4). For Carolina Wren, the DIFFERENCE model was preferred clearly over the GAMYE (Figure 4), and yet the predicted trajectories from the 2 models are similar (Figure 1). In contrast, for Pine Siskin, the DIFFERENCE and GAMYE were similar in their predictive accuracy (Figure 4) and yet the predicted trajectories are noticeably different in the first 10 yr of the survey (Supplementary Material Figure S1). For Cooper's Hawk, the GAMYE model was generally preferred over the DIFFERENCE model, although there was some overlap with zero (Figure 4), but in this case, the predicted trajectories are different. The DIFFERENCE trajectory for Cooper's Hawk suggests much less change in the species' population over time than the GAM or GAMYE (Figure 1).

Cooper's Hawk provides an example of a species with sparse data, for which the sharing of information in space may be particularly relevant. In a single stratum, the model has relatively a few data with which to estimate changes in populations through time. For example, the mean counts for the species indicate that on average one bird was observed for every 40 BBS routes run in the 1970s, and since the species population has increased, it still requires more than 10 routes to observe a single bird. For this species, the models that share information among strata on population change (GAM, GAMYE, and SLOPE) suggest a greater change in populations than the DIFFERENCE. For these models, where the stratum-level population change parameters are able to share information across the species' range, the absolute change in the population does not depend on the sample size in the region. In addition, for each of these models, there is still large variability in the trends estimated for data-sparse regions, demonstrating that while the estimates benefit from the sharing of information among strata, the local trends are still influenced by the local data. In contrast, there is a strong relationship between the magnitude of the trend and the number of routes contributing data to the analysis for the DIFFERENCE model (Figure 5, Supplementary Material Figure S4). In strata with fewer than 10 routes contributing data, the DIFFERENCE trends are almost all close to zero. In these relatively data-sparse strata, the DIFFERENCE model has little information available to estimate population change, and so the prior is more relevant and the population changes are shrunk toward zero. In contrast, the other models can integrate data from the local stratum with information on changes in the species' population across the rest of its range.

The decomposed trajectories from the GAMYE allow us to calculate trends from the smooth but also plot trajectories that show the annual fluctuations (Supplementary Material Figure S5). For example, the smooth trajectory for the Carolina Wren captures the general patterns of increases and decreases well, while the full trajectory also shows the sharp population crash associated with the extreme winter in 1976 (Figure 6). Calculating trends from the smooth component generates short-term estimates that vary less from year to year for species with relatively strong annual fluctuations (Figure 7). For example, Figure 8 shows the series of short-term (10-yr) trend estimates for Wood Thrush in Canada, from the smooth component of the GAMYE, the GAMYE including the year effects, the DIFFERENCE model, and the SLOPE model used since 2011. In this example, the 10-yr trend estimate from the GAMYE with the year effects and the SLOPE model both cross the IUCN trend threshold criterion for Threatened (IUCN 2019) at least once in the last 12 yr, including 2011, when the species' status was assessed in Canada (COSEWIC 2012). In contrast, a trend calculated from the decomposed GAMYE model using only the smooth component (GAMYE—Smooth Only in Figure 8) fluctuates much less between years.

Cross-Validation Varies in Time and Space

The preferred model from the pairwise predictive fit comparisons varied in time and space (Figures 4, 9, and 10 and Supplementary Material Figure S6). The contrast between GAMYE and DIFFERENCE for Barn Swallow provides a useful example: depending on the year or the region of the continent, either the GAMYE or the DIFFERENCE model

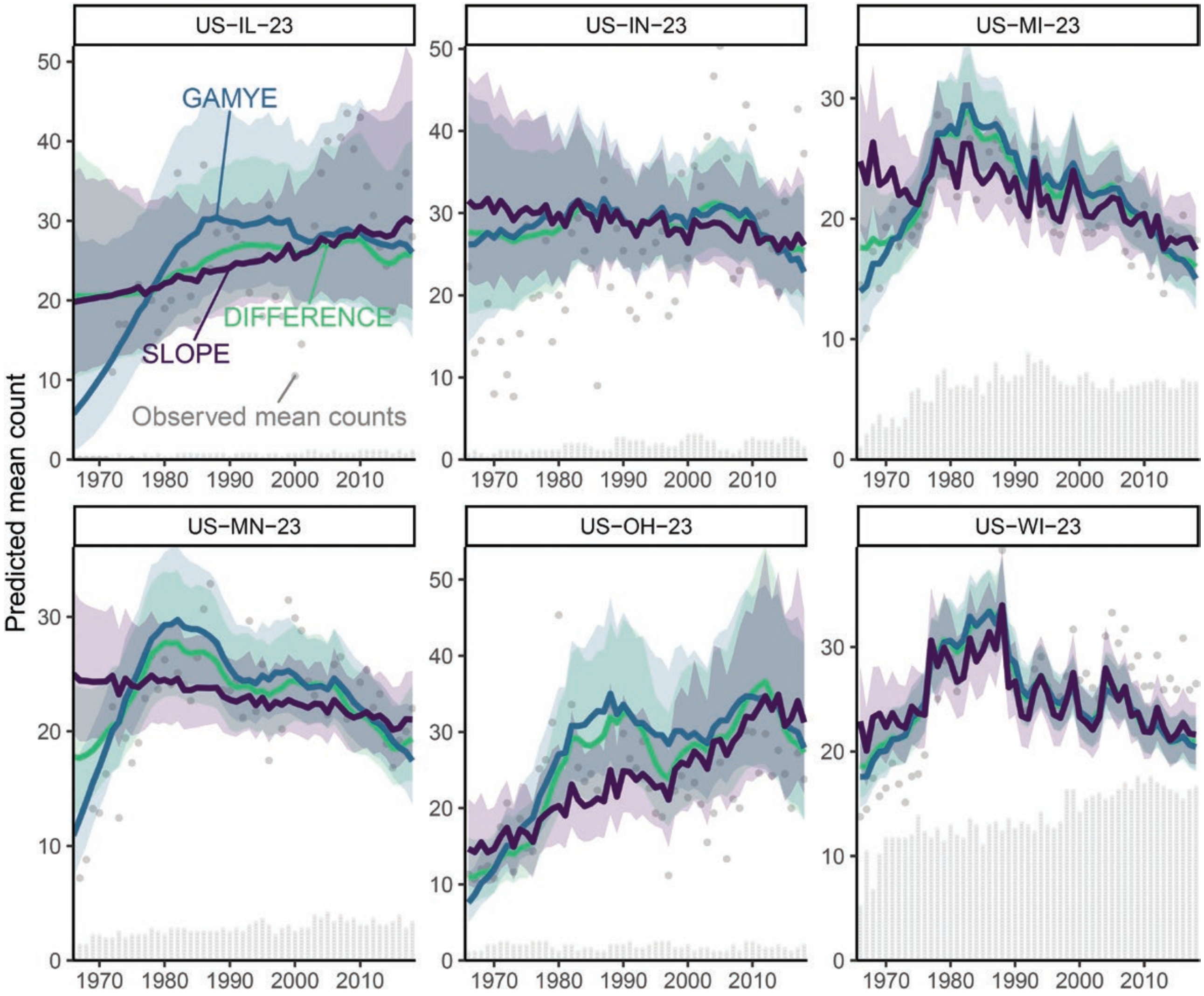

FIGURE 3. Stratum-level predictions for Barn Swallow population trajectories in BCR 23 from GAMYE against the predictions from the SLOPE and DIFFERENCE model. The similarity of the overall patterns in the GAMYE as compared to the SLOPE estimates demonstrates the inferential benefits of the sharing of information among regions on the nonlinear shape of the trajectory. In most strata, the similar patterns of observed mean counts and the GAMYE trajectories suggest a steep increase in Barn Swallows across all of BCR 23 during the first 10 yr of the survey. The GAMYE estimates show this steep increase in almost all of the strata, whereas the SLOPE predictions only show this pattern in the most data-rich stratum, US-WI-23. The DIFFERENCE trajectories model the nonlinear shapes well in all but the most data-sparse region (US-IL-23) and years (<1972). The facet strip labels indicate the country and state-level division of BCR 23 that makes up each stratum. The first 2 letters indicate all strata are within the United States, and the second 2 letters indicate the state. The stacked dots along the *x*-axis indicate the number of BBS counts in each year and stratum; each dot represents one count.

was the preferred model, but overall, and in almost all regions and years, the 95% CI of the mean difference in fit between GAMYE and DIFFERENCE overlapped 0 (Figures 4, 9, and 10). For Barn Swallow, the GAMYE model has generally higher predictive fit during the first 5 yr of the time series, but then the DIFFERENCE model is preferred between approximately 1975 and 1983. The geographic variation in predictive fit is similarly complex. In the eastern parts of the Barn Swallow's range, the GAMYE model generally out-performed the DIFFERENCE model, whereas the reverse is generally true in the remainder of the species' range (Figure 10). Although the mapped colors only represent the point-estimates, they suggest an interesting spatial pattern in the predictive fit of these 2 models for this species. Many of the species considered here show similarly complex temporal and spatial patterns in the preferred model based on predictive fit (Supplementary Material Figure S6).

DISCUSSION

Using Bayesian hierarchical semi-parametric GAM smooths to model time series of population abundance with the North American BBS generates useful estimates

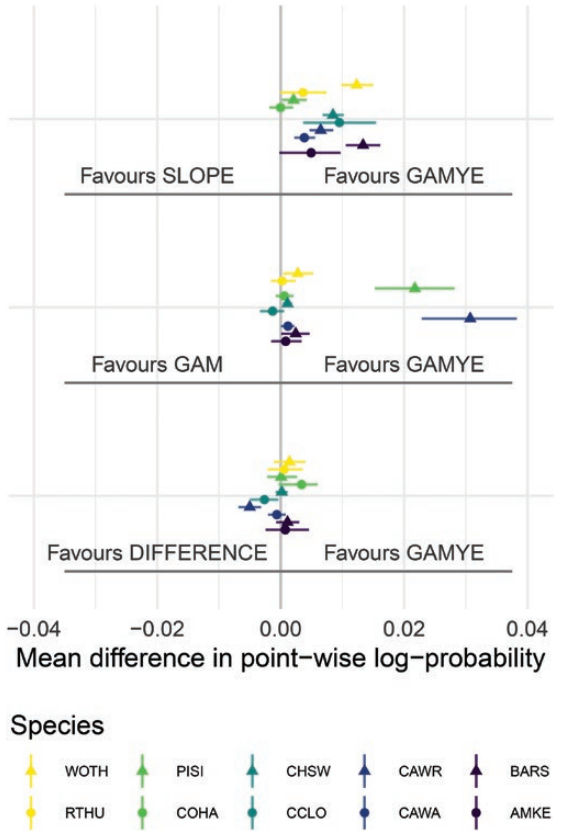

FIGURE 4. Overall differences in predictive fit between the GAMYE and SLOPE and GAMYE and GAM for Barn Swallow and 9 other selected species. Species short forms are WOTH = Wood Thrush (*Hylocichla mustelina*), RTHU = Ruby-throated Hummingbird (*Archilochus colubris*), PISI = Pine Siskin (*Spinus pinus*), Cooper's Hawk (*Accipiter cooperii*), CHSW = Chimney Swift (*Chaetura pelagica*), CCLO = Chestnut-collared Longspur (*Calcarius ornatus*), CAWR = Carolina Wren (*Thryothorus ludovicianus*), CAWA = Canada Warbler (*Cardellina canadensis*), AMKE = American Kestrel (*Falco sparverius*).

of population trajectories and trends and has better or comparable out-of-sample predictive accuracy, in comparison to the SLOPE or DIFFERENCE model. The flexibility of the GAM smoothing structure to model long- and medium-term temporal patterns, and the optional addition of random year effects to model annual fluctuations, allow it to model a wide range of temporal patterns within a single base model (Fewster et al. 2000, Wood 2017). We fit the smoothing parameters as random effects to share information across geographic strata within a species' range and to improve the estimates of population trajectories for data-sparse regions (Pedersen et al. 2019). For almost all species included here, the 2 GAM-based models clearly out-performed the standard model (SLOPE) used for the CWS and USGS analyses since 2011 (Sauer and Link 2011, Smith et al. 2014) and showed similar out-of-sample predictive accuracy as a first difference, random-walk trajectory model (Link et al. 2020). On a practical note, the GAM-based models required approximately 40% more time than the SLOPE or DIFFERENCE model to generate a similar number of posterior samples but given the 53 yr of effort to collect the data, we suggest this is a small price to pay for useful status and trend estimates.

The decomposition of the estimated population trajectory into the smooth and year-effect components is a feature of the GAMYE that is particularly useful for conservation applications. It allows the user to estimate and visualize separate trends and trajectories that include or exclude the annual fluctuations (Knape 2016). This allows the estimates to suit a range of conservation and management applications that rely on visualizing and estimating multiple aspects of population change. For example, the smoothed population trajectories capture the medium- and long-term changes in populations that are most relevant to broad-scale, multi-species assessments like the "State of the Birds" reports (NABCI Canada 2019) where the annual fluctuations of a given species are effectively noise against the signal of community-level change over the past 50 yr (Rosenberg et al. 2019). Similarly, estimates of population trends (interval-specific, rates of annual change) derived from the smooth component are responsive to medium-term changes and so can be used to identify change points in trends such as the recovery of species at risk (Environment Climate Change Canada 2016).

Trends derived from the smooth component of the GAMYE are responsive to medium-term changes, but also much less likely to fluctuate from year to year and therefore more reliable for use in species at risk status assessments (James et al. 1996). In many status assessments, such as those by IUCN and COSEWIC, population declines beyond a particular threshold rate can trigger large investments of resources related to policy and conservation actions. For example, in both the IUCN red-listing and Canada's federal species at risk assessments (IUCN 2019) estimated population declines greater than 30% over 3 generations is one criterion that results in a "Threatened" designation. If the estimated rate of population decline fluctuates from year to year, and is therefore strongly dependent on the particular year in which a species is assessed, there is an increased risk of inaccurate assessments. These inaccuracies could result in failures to protect species or inefficient investments of conservation resources. Of course, the full assessments of species' status are sophisticated processes that consider more than just a single-trend estimate. However, the example of Wood Thrush trends for Canada in Figure 8 shows that trends used to assess the species were below the threshold for "Threatened" status in 2011, but not in either year adjacent to 2011. The smooth-only trend never dips below the threshold (Figure 8) and raises the question of whether Wood Thrush would have been assessed

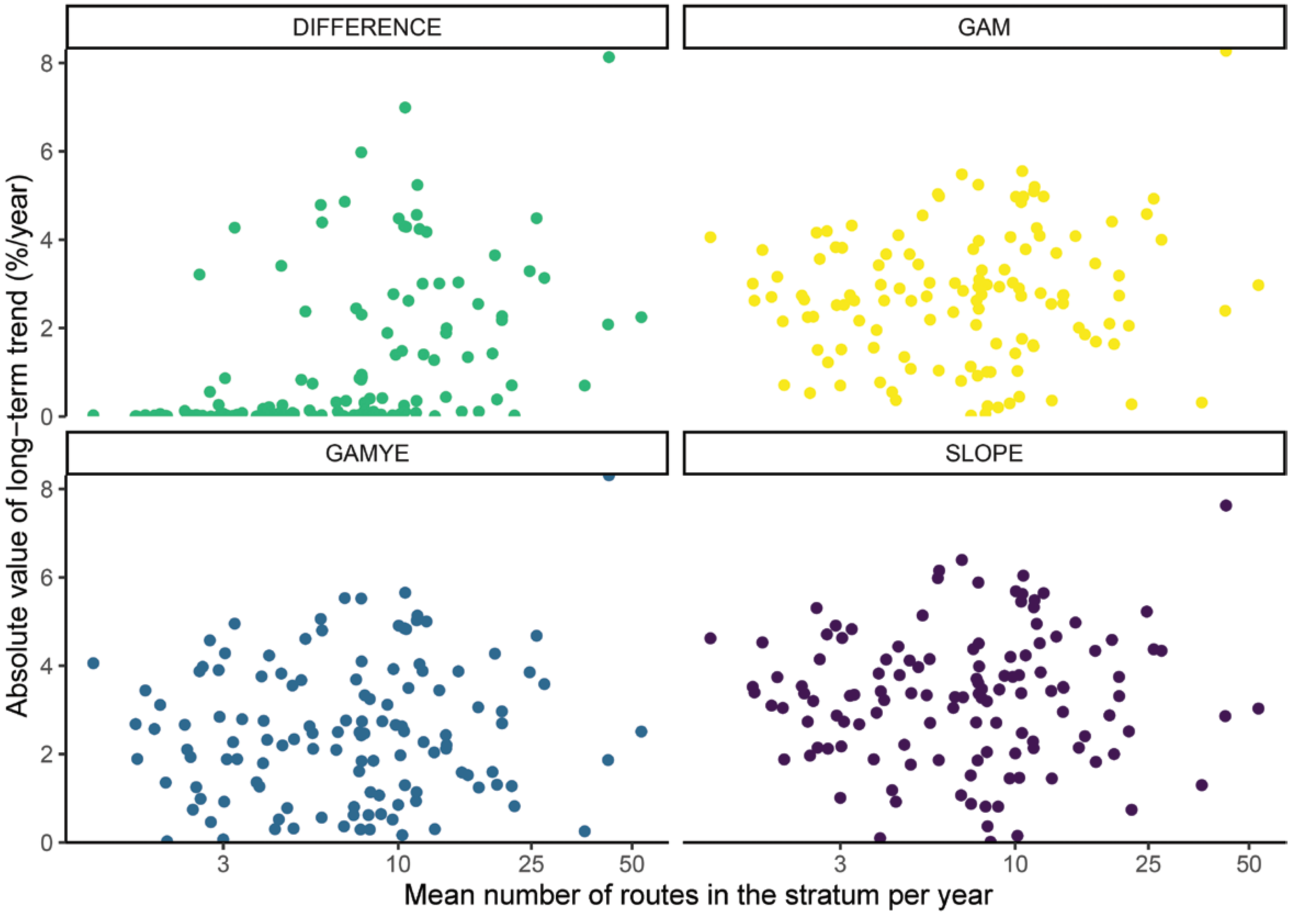

FIGURE 5. Relationship between the absolute value of estimated long-term trends (1966–2018) and the amount of data in each stratum, from the 4 models compared here for Cooper's Hawk, a species with relatively sparse data in each individual stratum. More of the trends estimated with the DIFFERENCE model are close to zero, suggesting a stable population, and particularly where there are relatively few routes contributing data in each year. This relationship is not evident for the same data modeled with one of the 3 models that are able to share some information among strata on population change (GAM, GAMYE, and SLOPE).

as Threatened in Canada if the relevant trend had not been estimated in 2011 or had been estimated using a different model (COSEWIC 2012).

Alternative metrics of population trends that remove the annual fluctuations have been used with for the BBS, such as LOESS smooths (James et al. 1996) or slopes of log-linear regression lines calculated as part of the underlying model (Link and Sauer 1994) or as derived parameters from series of estimated annual indices (Sauer and Link 2011). Trend estimates that remove the effect of the annual fluctuations are generally a common approach to summarizing average rates of change in other monitoring programs (Fewster et al. 2000 for U.K. breeding birds and Bogaart et al. 2020 for European breeding birds). Many alternative definitions of trend could be calculated using the annual indices derived from any one of the models compared here (Supplementary Material Figure S8). However, for the last decade, both national agencies have supplied authoritative trend estimates based on end-point comparisons of annual indices, which include the annual fluctuations (Sauer and Link 2011, Smith et al. 2015). Similarly, calculating alternative metrics of trend from the annual indices in a way that propagates uncertainty would be done best using information from the full posterior distribution of each annual index. Given that these full posterior distributions are challenging for users to manipulate and summarize, we suggest that providing authoritative trends based on the smooth component from the GAMYE is a practical and simple solution. These smooth-based trends are responsive to cycles or changes in rates of population change (discussed in the works of James et al. 1996 and Sauer and Link 2011) while they also limit the annual fluctuations that might otherwise undermine the utility and credibility of BBS trends for species status assessments (see also Smith et al. 2015).

In some conservation or scientific uses of the BBS-based population trajectories, the annual fluctuations may be important components of the trajectory (e.g., winter-related mortality of Carolina Wrens), and in these

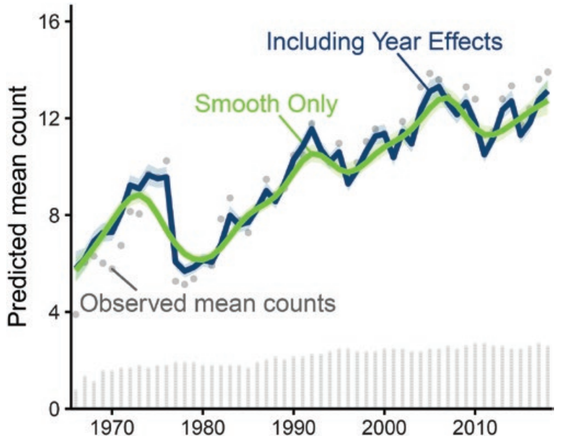

FIGURE 6. Decomposition of the survey-wide population trajectory for Carolina Wren (*Thryothorus ludovicianus*), from the GAMYE, showing the full trajectory ("Including Year Effects," $N_{s,t}$) and the isolated smooth component ("Smooth Only," $Ng_{s,t}$), which can be used to estimate population trends that are less sensitive to the particular year in which they are estimated. The stacked dots along the *x*-axis indicate the approximate number of BBS counts used in the model; each dot represents 50 counts.

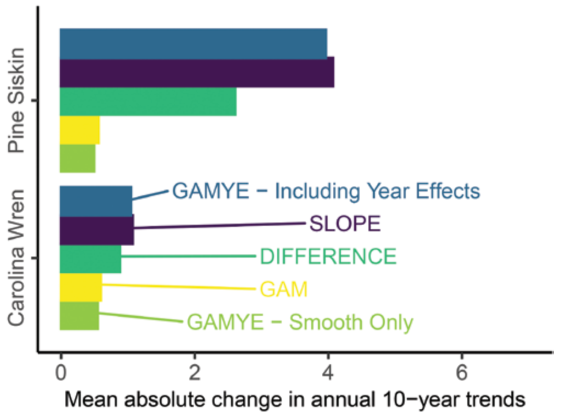

FIGURE 7. Interannual variability of 10-year trend estimates for 2 species with large annual fluctuations (% yr^{-1}). Trends from the GAM, which does not model annual fluctuations, and from the GAMYE using only the smooth component, which removes the effect of the annual fluctuations, are less variable between subsequent years (i.e. more stable) than trends from the GAMYE including the year effects or the other 2 models that include the annual fluctuations.

situations, both components from the GAMYE can be presented. This comprehensive estimate of a species' population trajectory is likely the best approach for the official presentation of a time series. At a glance, managers, conservation professionals, and researchers can glean information about fluctuations that might relate to annual covariates such as precipitation, wintering ground conditions, or cone-crop cycles. The GAMYE structure allows an agency like the CWS to provide estimates in multiple versions (e.g., full trajectories and smoothed trajectories in the same presentation, such as Figure 6), drawn from a coherent model, to suit a wide range of conservation applications, and to produce them in an efficient way. For example, there are situations where the ability for a user to access a ready-made separation of the yearly fluctuations from the underlying smooth could be helpful in the initial formulation of an ecological hypothesis. In addition, for custom analyses (Edwards and Smith 2020), a researcher can modify the basic GAMYE model to include annual

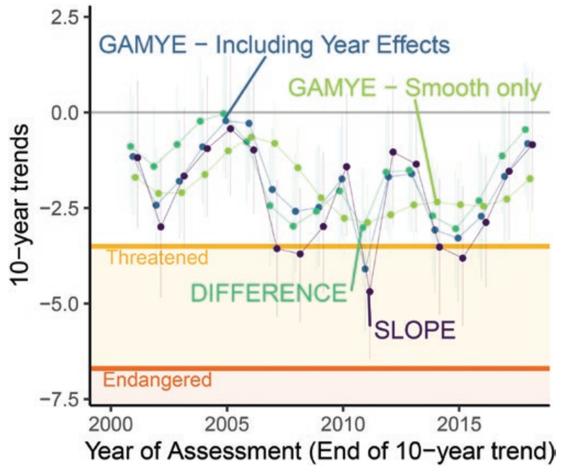

FIGURE 8. Sequential, short-term trend estimates for Wood Thrush (*Hylocichla mustelina*) in Canada from 3 alternative modeling approaches, and their comparison to the IUCN trend criteria for "Threatened" (in orange) and "Endangered" (in red). Trends estimated from the decomposed trajectory of the GAMYE that include only the smooth component (in blue) are more stable between sequential years than trends from the other models that include annual fluctuations.

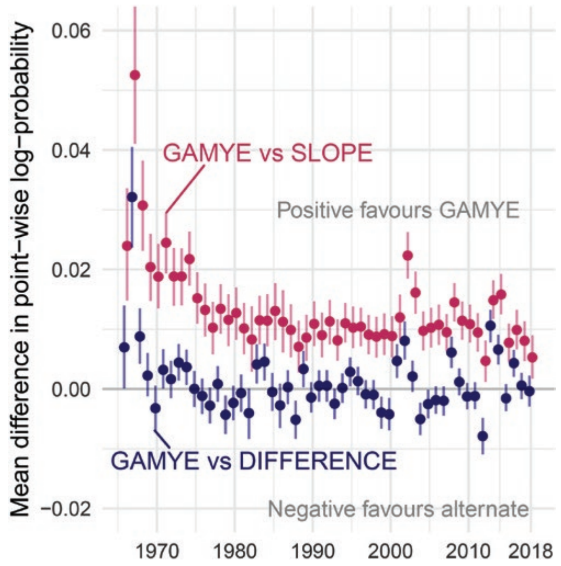

FIGURE 9. Annual differences in predictive fit between the GAMYE and SLOPE (blue) and the GAMYE and DIFFERENCE model (red) for Barn Swallow.

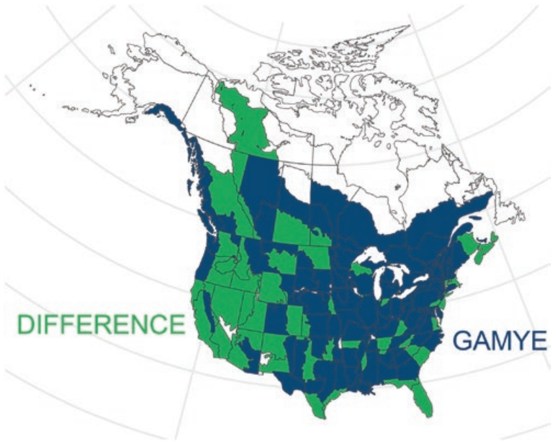

FIGURE 10. Geographic distribution of the preferred model for Barn Swallow, according to the point estimate of the mean difference in predictive fit between GAMYE and DIFFERENCE. The GAMYE is generally preferred in the Eastern part of the species' range, but the DIFFERENCE is preferred in many regions in the Western part of the species' range. Note: in most regions, the differences in predictive fit were variable and neither model was clearly preferred (i.e. the 95% CI of the mean difference included 0).

covariates on the yearly fluctuations (e.g., extreme weather during migration, or spruce cone mast-years) and other covariates on the smooth component (e.g., climate cycles).

Predictive Accuracy

Overall, the cross-validation comparisons generally support the GAMYE, GAM, or DIFFERENCE model over the SLOPE model for the species considered here, in agreement with Link et al. (2020). For Barn Swallow, the overall difference in predictive fit, and particularly the increasing predictive error of the SLOPE model in the earliest years, strongly suggests that in the period between the start of the BBS (1966) and ~1983 (Smith et al. 2015), Barn Swallow populations increased. All models agree, however, that since the mid-1980s populations have decreased.

Using all data in our cross-validations allowed us to explore the spatial and temporal variation in fit and to compare the fit across all data used in the model. We have not reported absolute values of predictive fit because estimates of fit from a random selection of BBS counts, or simple summaries of predictive fit from the full dataset, are biased by the strong spatial and temporal dependencies in the BBS data (Roberts et al. 2017). However, because our folds were identical across models, and we modeled the differences in fit with an additional hierarchical model that accounted for repeated measures among strata and years, we are reasonably confident that relative-fit assessments are unbiased within a species and among models. Alternative approaches, such as blocked cross-validation (Roberts et al. 2017) to assess the predictive success of models in time and space, and targeted cross-validation (Link et al. 2017) to explore the predictive success in relation to particular inferences (e.g., predictive accuracy in the end-point years used for short- and long-term trend assessments) are an area of ongoing research.

The overall predictive fit assessments provided some guidance on model selection for the species here, but not in all cases. The SLOPE model compared poorly against most other models in the overall assessment, similar to Link et al. 2020. However, among the other 3 models, many of the overall comparisons failed to clearly support one model, even in cases where the predicted population trajectories suggested different patterns of population change (e.g., Cooper's Hawk). For a given species, the best model varied among years and strata. These temporal and spatial patterns in predictive fit complicate the selection among models, given the varied uses of the BBS status and trend estimates (Rosenberg et al. 2017).

In general, estimates of predictive accuracy are one aspect of a thoughtful model building and assessment process, but are insufficient on their own (Gelman et al. 2013 p. 180, Burnham and Anderson 2002 p. 16). This is particularly true if there is little or no clear difference in overall predictive accuracy, but important differences in model predictions. For example, the overall cross-validation results do not clearly distinguish among the SLOPE, DIFFERENCE, and GAMYE for Cooper's Hawk, and yet predictions are different between the DIFFERENCE model and the others (Figures 1, 4, and 5). Interestingly, the cross-validation approach in the work of Link et al. 2020 suggested that the DIFFERENCE model was preferred over the SLOPE for Cooper's Hawk, but we did not find that here (Supplemental Material Figure S3). The important differences in trend estimates and the equivocal cross-validation results suggest that further research is needed into the criteria for, and consequences of, model selection for BBS status and trend estimates. Model selection is also complicated when overall predictive accuracy appears to clearly support one model and yet the important parameters (trends and trajectories) are not noticeably different. For example, the overall cross-validation results for Carolina Wren suggest that the DIFFERENCE model is preferred over the GAMYE, and yet the trajectories are almost identical (Figures 1 and 4). Predictive accuracy is also complicated when robust predictions are required for years or regions with relatively few data against which predictions can be assessed (e.g., the earlier years of the BBS, or data-sparse strata that still have an important influence on the range-wide trend). Model selection is complicated, and predictive accuracy would never be the only criterion used to select a model for the BBS analyses. Limits to computational capacity and a desire to avoid a data-dredging all-possible-models approach ensure that some thoughtful process to select the candidate models is necessary.

We agree with Link et al. (2020) that we should not select models based on a particular pattern in the results. In fact, the necessary subjective process occurs before any quantitative analyses (Burnham and Anderson 2002) and relies on "careful thinking" to balance the objectives, the model, and the data (Chatfield 1995). The careful thinking required to select a BBS model or to interpret the BBS status and trend estimates is to consider the consequences of the potential conflicts between the model structures ("constraints on the model parameters" Chatfield 1995) and the objectives of the use of the modeled estimates. For example, one of the models that shares information on population change among strata is likely preferable to the DIFFERENCE model for species with relatively sparse data in any given stratum, because the prior of the DIFFERENCE model (stable population) will be more influential when the data are sparse. This prior dependency of the results may not be identified by the lower predictive accuracy of the estimates, as the results for Cooper's Hawk demonstrate (Figure 5). Similarly, a user of estimates from the DIFFERENCE model should carefully consider the conservation-relevant consequences of the prior and model structure when assessing potential changes in the population trends of declining and relatively rare species. These species' short-term rates of decline could appear to decrease, suggesting a stabilizing population, simply due to the increasing influence of the prior, if the species observations decline to a point where it is not observed in some years. In contrast, if a user wished to explicitly compare estimates of population change among political jurisdictions or ecological units, the sharing of information among those units in the GAM-based models here might be problematic. We suggest that the GAMYE's strong cross-validation performance, its sharing of information across a species range, its decomposition of the population trajectory, and its broad utility that suits the most common uses of the BBS status and trends estimates make it a particularly useful model for the sort of omnibus analyses conducted by the CWS and other agencies.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Ornithological Applications* online.

ACKNOWLEDGMENTS

We sincerely thank the thousands of U.S. and Canadian participants who annually perform and coordinate the North American Breeding Bird Survey. We also wish to acknowledge Courtney Amundson for sharing some code on similar models, and John Sauer and Bill Link for sharing code that helped with the cross-validations and for many spirited, collegial discussions that have informed this work. We also thank the many biologists within the Canadian Wildlife Service and other users of the BBS status and trend estimates whose insightful questions and suggestions motived much of this work, including Charles Francis, Marie-Anne Hudson, Veronica Aponte, Marcel Gahbauer, Pete Blancher, and Ken Rosenberg.

Funding statement: This work was supported by operating funds from Environment and Climate Change Canada.

Ethics statement: This research was conducted in compliance with the Environment and Climate Change Canada Values and Ethics Code.

Author contributions: A.C.S. conceived the ideas and designed methodology; B.P.M.E. and A.C.S. analyzed the data; A.C.S. led the writing of the manuscript. A.C.S. and B.P.M.E. contributed critically to the drafts and gave final approval for publication.

Data availability: R scripts to download the BBS data and to perform the analyses in this article and are archived at https://zenodo.org/badge/latestdoi/228419725 (Edwards and Smith 2020).

LITERATURE CITED

Bogaart, P., M. van der Loo, and J. Pannekoek (2020). rtrim: Trends and Indices for Monitoring Data. R package version 2.1.1. https://rdrr.io/cran/rtrim/

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika 76:503–514.

Burnham, K. P., and D. R. Anderson (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second edition. Springer-Verlag, New York, NY, USA.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference (with discussion). Journal of the Royal Statistical Society (London), Series A 158:419–466.

COSEWIC (2012). COSEWIC Assessment and Status Report on the Wood Thrush *Hylocichla mustelina* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, Canada. http://www.registrelep-sararegistry.gc.ca/sar/assessment/status_e.cfm

Crainiceanu, C. M., D. Ruppert, and M. P. Wand (2005). Bayesian analysis for penalized spline regression using WinBUGS. Journal of Statistical Software 14:1–24.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. Journal of the American Statistical Association 78:605–610.

Edwards, B. P. M., and A. C. Smith (2020). bbsBayes v2.1.0, version 2.1.0. Zenodo. doi:10.5281/zenodo.3727279

Environment Climate Change Canada (2016). Recovery Strategy for the Canada Warbler (*Cardellina canadensis*) in Canada. Species at Risk Act Recovery Strategy Series. Environment Canada, Ottawa, Canada. http://www.registrelep-sararegistry.gc.ca

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.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). Bayesian Data Analysis. Chapman and Hall/CRC Press, Boca Raton, FL, USA.

Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016.

Hudson, M.-A. R., C. M. Francis, K. J. Campbell, C. M. Downes, A. C. Smith, and K. L. Pardieck (2017). The role of the North American Breeding Bird Survey in conservation. The Condor: Ornithological Applications 119:526–545.

IUCN Standards and Petitions Committee (2019). Guidelines for Using the IUCN Red List Categories and Criteria, version 14. Prepared by the Standards and Petitions Committee. http://www.iucnredlist.org/documents/RedListGuidelines.pdf

James, F. C., C. E. McCulloch, and D. A. Wiedenfeld (1996). New approaches to the analysis of population trends in land birds. Ecology 77:13–27.

Knape, J. (2016). Decomposing trends in Swedish bird populations using generalized additive mixed models. Journal of Applied Ecology 53:1852–1861.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence 2:1137–1143.

Lange, K. L., R. J. A. Little, and J. M. G. Taylor (1989). Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84:881–896.

Link, W. A., and J. R. Sauer (1994). Estimating equations estimators of trend. Bird Populations 2:23–32.

Link, W. A., and J. R. Sauer (2016). Bayesian cross-validation for model evaluation and selection, with application to the North American Breeding Bird Survey. Ecology 97:1746–1758.

Link, W. A., J. R. Sauer, and D. K. Niven (2017). Model selection for the North American Breeding Bird Survey: A comparison of methods. The Condor: Ornithological Applications 119:546–556.

Link, W. A., J. R. Sauer, and D. K. Niven (2020). Model selection for the North American Breeding Bird Survey. Ecological Applications 30:e02137.

North American Bird Conservation Initiative Canada (2019). The State of Canada's Birds, 2019. Environment and Climate Change Canada, Ottawa, Canada. www.stateofcanadasbirds.org

Partners in Flight (2019). Avian Conservation Assessment Database, version 2019. http://pif.birdconservancy.org/ACAD

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. PeerJ 7:e6876.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf

R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robbins, C. S., D. Bystrak, and P. H. Geissler (1986). The Breeding Bird Survey: Its First Fifteen Years, 1965–1979. U.S. Fish and Wildlife Service Resource Publication 157, U.S. Fish and Wildlife Service, Washington, DC, USA. https://pubs.er.usgs.gov/publication/5230189

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, et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical or phylogenetic structure. Ecography 40:913–929.

Rosenberg, K. V., P. J. Blancher, J. C. Stanton, and A. O. Panjabi (2017). Use of North American Breeding Bird Survey Data in avian conservation assessments. The Condor: Ornithological Applications 119:594–606.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, et al. (2019). Decline of the North American avifauna. Science (New York, N.Y.) 366:120–124.

Sauer, J. R., and W. A. Link (2011). Analysis of the North American Breeding Bird Survey using hierarchical models. The Auk 128:87–98.

Sauer, J. R., J. E. Hines, J. E. Fallon, K. L. Pardieck, D. J. Ziolkowski, Jr., and W. A. Link (2014). The North American Breeding Bird Survey, Results and Analysis 1966–2013, version 01.30.2015. USGS Patuxent Wildlife Research Center, Laurel, MD, USA.

Sauer, J. R., K. L. Pardieck, D. J. Ziolkowski, A. C. Smith, M.-A. R. Hudson, V. Rodriguez, H. Berlanga, D. K. Niven, and W. A. Link (2017). The first 50 years of the North American Breeding Bird Survey. The Condor: Ornithological Applications 119:576–593.

Smith, A. C., M-A. R. Hudson, V. Aponte, and C. M. Francis (2019). North American Breeding Bird Survey–Canadian Trends Website, Data-version 2017. Environment and Climate Change Canada, Gatineau, Quebec, Canada.

Smith, A. C., M.-A. R. Hudson, C. Downes, and C. M. Francis (2014). Estimating breeding bird survey trends and annual indices for Canada: How do the new hierarchical Bayesian estimates differ from previous estimates. Canadian Field-Naturalist 128:119–134.

Smith, A. C., M. A. Hudson, C. M. Downes, and C. M. Francis (2015). Change points in the population trends of aerial-insectivorous birds in North America: Synchronized in time across species and regions. PLoS One 10:e0130768.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432.

Wallig, M., and S. Weston (2020). foreach: Provides Foreach Looping Construct. R package version 1.5.1. https://cran.r-project.org/package=foreach

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

Wilson, S., A. C. Smith, and I. Naujokaitis-Lewis (2018). Opposing responses to drought shape spatial population dynamics of declining grassland birds. Diversity and Distributions 24:1687–1698.

Wood, S. N. (2017). Generalized Additive Models: An Introduction with R, second edition. CRC Press, Portland, OR, USA.

Zhang, Y., and Y. Yang (2015). Cross-validation for selecting a model selection procedure. Journal of Econometrics 187:95–112.