

Estimating Breeding Bird Survey Trends and Annual Indices for Canada: How Do the New Hierarchical Bayesian Estimates Differ from Previous Estimates?

ADAM C. SMITH^{1,2}, MARIE-ANNE R. HUDSON¹, CONSTANCE DOWNES¹, and CHARLES M. FRANCIS¹

¹Canadian Wildlife Service, Environment Canada, National Wildlife Research Centre, 1125 Colonel By Drive, Ottawa, Ontario K1A 0H3 Canada

²Corresponding author: Adam.Smith@ec.gc.ca

Smith, Adam C., Marie-Anne R. Hudson, Constance Downes, and Charles 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(2): 119–134.

Canadian data from the North American Breeding Bird Survey (BBS) provide information on the population status and trends for over 300 species that regularly breed in Canada. Since the first assessments were made in the mid-1970s, both the dataset and the suite of statistical tools and techniques available to researchers have grown. As a result, Canadian BBS trend estimates have been derived from numerous statistical models. Because the BBS data are relatively complex, different statistical models can generate different trend and status estimates from the same data. In 2013, Environment Canada's Canadian Wildlife Service began producing BBS status and trend estimates using a hierarchical Bayesian model. To give users of BBS trends and annual indices of abundance a better understanding of these estimates, we demonstrate and explain some of the similarities and differences between the new hierarchical Bayesian estimates and those from the previous model; discuss the philosophical and inferential consequences of estimating trends with the new model; and describe how the hierarchical Bayesian model differs from the model currently used in the United States. Overall, trends and annual indices from the new model are generally similar to estimates from the previous model; however, they are more precise, less variable among years, better represent the spatial variation across Canada in population status, and allow for more intuitive and useful assessments of uncertainty.

Key Words: breeding bird survey; hierarchical Bayesian model; maximum likelihood model; population trend; population monitoring

Introduction

The North American Breeding Bird Survey (BBS) provides information on the population status (trends and annual indices of abundance) of almost 300 bird species in Canada. Indeed, it is the primary source of information for over 200 of those species. Due, in part, to its long history and continental coverage, the BBS is considered to be the backbone of land bird conservation in North America (Rich *et al.* 2004*). It is a vital component of a wide variety of high-level conservation documents for Canada, e.g., the State of Canada's Birds (North American Bird Conservation Initiative Canada 2012*) and the Status of Birds in Canada website (Environment Canada 2011*). Data from the survey have also been used in hundreds of scientific publications on topics ranging from assessing population trends to modelling habitat associations or impacts of climate change on range shifts (Pardieck *et al.* 2008).

The approaches used to analyze BBS data have evolved over time as new statistical methods have become available and as the dataset has grown to support increasingly complex models. The first Canadian BBS routes were conducted in 1966, and the first national trend estimates included the data from the first 10 years (Erskine 1978*). Population trends in that report were estimated using a chaining-style analysis of averaged year-to-year changes in species counts on routes run in successive years, with the chain of year-to-year changes

indexed to a value of 100 in a base year (1973). Subsequently, both the Canadian Wildlife Service (CWS) and the United States Geological Service (USGS) began to analyze the BBS data annually, but using different methods. Early analyses by the United States' agencies (at the time, the Fish and Wildlife Service) calculated annual indices using weighted average counts across years and trends using route-regression models (Robbins *et al.* 1986). Through the 1980s and 1990s, various route-regression models (Geissler and Noon 1981) were used by both the CWS and USGS, although variation in approaches to weighting routes led to different estimates (Thomas and Martin 1996). The USGS analyses continued to use route-regression models through 2008, while, from 2002 to 2009, Canadian BBS data were analyzed using a maximum likelihood (ML) model, which estimated an annual index. Starting with the 2011 BBS data, Canadian trends and annual indices have been estimated using a hierarchical Bayesian (HB) model, which is very similar to the model adopted by the USGS in 2009 and described in Link and Sauer (2002) and Sauer and Link (2011).

HB methods suit the BBS data's complex structure, because the Bayesian framework provides a coherent and flexible approach for modelling the effects of sampling variation (e.g., not all routes are surveyed each year; not all birds present on a route are detected on any given survey; and detection probabilities vary among

observers) separately from temporal and geographic variation in the underlying populations (e.g., population change due to changing weather, resource abundance, and human activity). In addition, the hierarchical structure is particularly efficient and powerful because the distributional assumptions of the random effects greatly reduce the number of parameters that need to be estimated to model observer effects and other nuisance parameters (Clark 2005). Similarly, the hierarchical structure of the model mirrors the hierarchical structure of the data and allows effects to be modelled at the appropriate scale, e.g., nuisance parameters, such as overdispersion and the effects of changes in observers over time, can be efficiently modelled across all counts, while trends and annual variation around the trend can be modelled across routes within strata (Link and Sauer 2002).

Bayesian methods also provide intuitive and direct estimates of the uncertainty around the population parameters (i.e., population trends and indices of annual abundance) that are of primary interest to most users. Bayesian methods are based on the idea that our understanding of an imperfectly known parameter, such as a population trend, has a probability distribution. That is, the mean, median, or mode of the distribution is our best estimate of the parameter, and the variance, standard error, or percentiles (i.e., the credible interval) of the distribution represent the uncertainty around our best estimate. Our initial understanding of that distribution is termed the prior probability distribution. The data are used to update the prior probability distribution, through Bayes' rule, to produce the posterior probability distribution, which expresses our final understanding of the parameter (Link and Barker, 2009). For example, with a Bayesian framework, it is relatively simple to estimate the posterior distribution of the total change in a population since 1970, and from that distribution, make concise and direct statements about the probability that a population has declined; whereas with previous analyses, uncertainty estimates were largely limited to probabilities of observing the data (or more extreme data) if the true population had not changed.

The differences between Bayesian and frequentist (i.e., traditional) estimates of uncertainty are exemplified in comparing Bayesian credible intervals with frequentist confidence intervals. Credible intervals provide a probable range of values for a parameter (e.g., in the case of a 95% credible interval, the range of values that contains, with a 95% probability, the true value of a population trend). Many readers will be more familiar with confidence intervals and likely with interpreting them in the same way. However, for confidence intervals this interpretation is incorrect, and the confusion largely stems from the unintuitive nature of the true definition of confidence interval, which relates to an infinite number of hypothetical realisations of the

data (Clark 2005). In essence, credible intervals, and Bayesian methods in general, provide intuitive and useful measures of uncertainty. For example, the current BBS results website maintained by Environment Canada (www.ec.gc.ca/ron-bbs) provides, for every trend estimate, associated estimates of the probabilities of eight thresholds of population change (e.g., probability that the population has decreased, increased by > 100%, etc.). Species-at-risk status assessments can now easily consider an explicit measure of the probability that a population has reached one of the thresholds used to categorize a species as threatened or endangered (e.g., COSEWIC 2011*).

The application of HB methods to the BBS represents the most recent development in estimating the status and trends of hundreds of bird species across North America. Three closely related HB models are described in the literature or are in use for analyzing trend and annual indices from the BBS data. The first model (HB₁) was described by Link and Sauer (2002). It was subsequently refined to account for the bias in retransforming annual index estimates from the log-scale to the scale of average counts (second equation in Sauer and Link 2011). The second model (HB₂) is currently used in annual analyses by the USGS. The third model (HB₃) includes an additional refinement that has the effect of scaling the annual indices more closely to the average observed count on routes within a stratum and improving estimates of trends for larger regions. HB₃ is the model currently used in the CWS annual analyses. The distinction between models HB₂ and HB₃ is important because trend and annual index estimates from the USGS include Canada and Canadian provinces. Therefore, anyone interested in a species trend for Canada must choose between the estimates derived by the USGS (using model HB₂) and those derived by the CWS (using model HB₃).

Because the BBS data and the statistical models used to analyze them are complex, different analytical approaches can result in different estimates of population parameters, particularly if the underlying signals in the data are relatively weak or variable or if the sampling is imbalanced. The BBS sampling method is a stratified random design that aims to generate a balanced and approximately representative sample of bird populations across North America (Erskin 1978*). Of course, as with any field survey, practical constraints introduce some variation and bias into the data: uneven spatial distribution of active routes, changes in the number and spatial coverage of routes surveyed over time, and variation among and changes over time in observers. The nested structure of the data (i.e., stops nested within routes, routes selected within degree-blocks and aggregated into analytical strata) provides added challenges. Analyses of BBS data have different ways of accounting for all or some of these sources of variation and bias, and, as a result, different analyses can

sometimes generate different estimates of population trends and annual indices (e.g., Thomas 1996; Thomas and Martin 1996; Sauer and Link 2011).

For many species and regions, the underlying signals in the data are clear and strong, the sample sizes are large, or the BBS sampling is very well balanced in space and time. In these cases, even very different models will generate similar estimates of population status and trends. In addition, although point estimates of trend for a given species and region may differ among models, most conservation or management decisions and actions are based on relatively broad overall patterns in population status. For example, the Status of Birds in Canada website classifies species based on six categories of population status, such as large decrease, moderate increase, little change, and data deficient (Environment Canada 2011*). Each of these categories encompasses a relatively wide range of trend values; thus, even when point estimates of trend or trajectory differ among models, the broad population inferences may be similar, particularly for populations with relatively extreme status (i.e., large decrease or large increase), which are likely to be the primary focus of management.

Our three goals in this paper are to describe in detail the HB₃ model we are now using in Canada to estimate population trends and how it differs from the HB₂ model used by the USGS; to demonstrate and explain some of the similarities and differences in the estimates provided by the HB₃ model and the previous Canadian model (the ML model); and to discuss the most important philosophical and inferential consequences of estimating trends and annual indices of abundance with the HB₃ model.

Methods

Maximum likelihood model

The model used to generate Canadian BBS trends between 2002 and 2009 is a weighted ML model (see Appendix 1). It uses a series of equations to derive ML estimates of stratum and year-specific annual indices ($\lambda_{i,j}$):

$$\lambda_{i,j} = \exp(\mu + \alpha_i + \beta_j + \gamma_{j,k})$$

The model assumes counts of birds on BBS routes are Poisson distributed, and estimates terms to account for the effects of each year (β_j), variation among routes (α_i), and variation among observers and observation conditions in each year ($\gamma_{j,k}$). The means of the estimated effects of year, route, and observer are fixed at 0, and μ is not estimated, but held constant as the observed average count of birds across all years and routes. The analysis is conducted using the sampling strata of the survey: degree blocks (regions defined by one degree latitude and one degree longitude, Figure 1 inset). The ML estimation of annual indices is weighted so that each degree block with data influences the estimate in proportion to its area, regardless

of the number of routes with data. Weighting limits the influence of degree blocks with more routes, which, because the BBS is volunteer-based, tend to occur in areas with larger human populations. The weight for each route from degree block i used in the analysis is $n_i^{-1} * Ai$, where n is the number of routes in the degree block and A is the degree block's area. The model was run through a custom C++ program.

The ML model estimates the annual indices directly, and trends are derived from log-linear regressions of the annual indices of abundance. A small constant (0.001) is added to the annual indices before the log-linear regression, because in years when the species is not observed the annual index is set to 0. The standard errors of the abundance indices and trends are estimated using a "jackknife" procedure, which discards one route at a time. A full description of the model has never been published; however further details are provided in Appendix 1.

Hierarchical Bayesian model, HB₃

The CWS is now generating Canadian BBS trends using the HB₃ model. It assumes that observed counts on a BBS route j in year t and stratum i have a Poisson distribution with mean ($\lambda_{i,j,t}$). The strata used are areas created by the intersection of provincial and territorial boundaries with those of the North American Bird Conservation Regions (BCRs; North American Bird Conservation Initiative International 2013*; Figure 1). On the log scale, the λ s are modelled by fixed effects for first-year observer effects (η , where the value of $\mathbf{I}(j,t)$ is 1 if a route-year combination represents an observer's first year on that route and 0 otherwise) and stratum-specific fixed effects for abundance (S_j), trend over time (β_j), and variance of the year effects within each stratum ($\sigma_{k,t}^2$). The model also has random effects for overdispersion ($\epsilon_{i,j,t}$) of counts and stratum-specific random terms for year effects ($\lambda_{i,t}$) and observer-route effects ($\omega_{i,j}$).

$$\log(\lambda_{i,j,t}) = S_i + \beta_i * t + \gamma_{i,t} + \omega_{i,j} + \eta \mathbf{I}(j,t) + \epsilon_{i,j} \quad [\text{equation 1}]$$

All prior distributions are standard, diffuse (non-informative) conjugate priors; that is, the priors are chosen so that the estimates are not influenced by any prior knowledge—non-informative—and the specific distributions, such as the inverse gamma for variances, ensure proper posterior distributions—conjugate priors. Specifically, all variances are assigned diffuse inverse gamma prior distributions (scale and shape parameters set to 0.001) and the parameters S , β , γ_i , and η are given diffuse normal distributions (mean 0, variance 10^6).

The annual indices of abundance for stratum i and year t are exponentiated sums of the year, stratum, and trend effects estimated in equation 1, scaled by the proportion of routes in the stratum on which the species was observed (z_i).

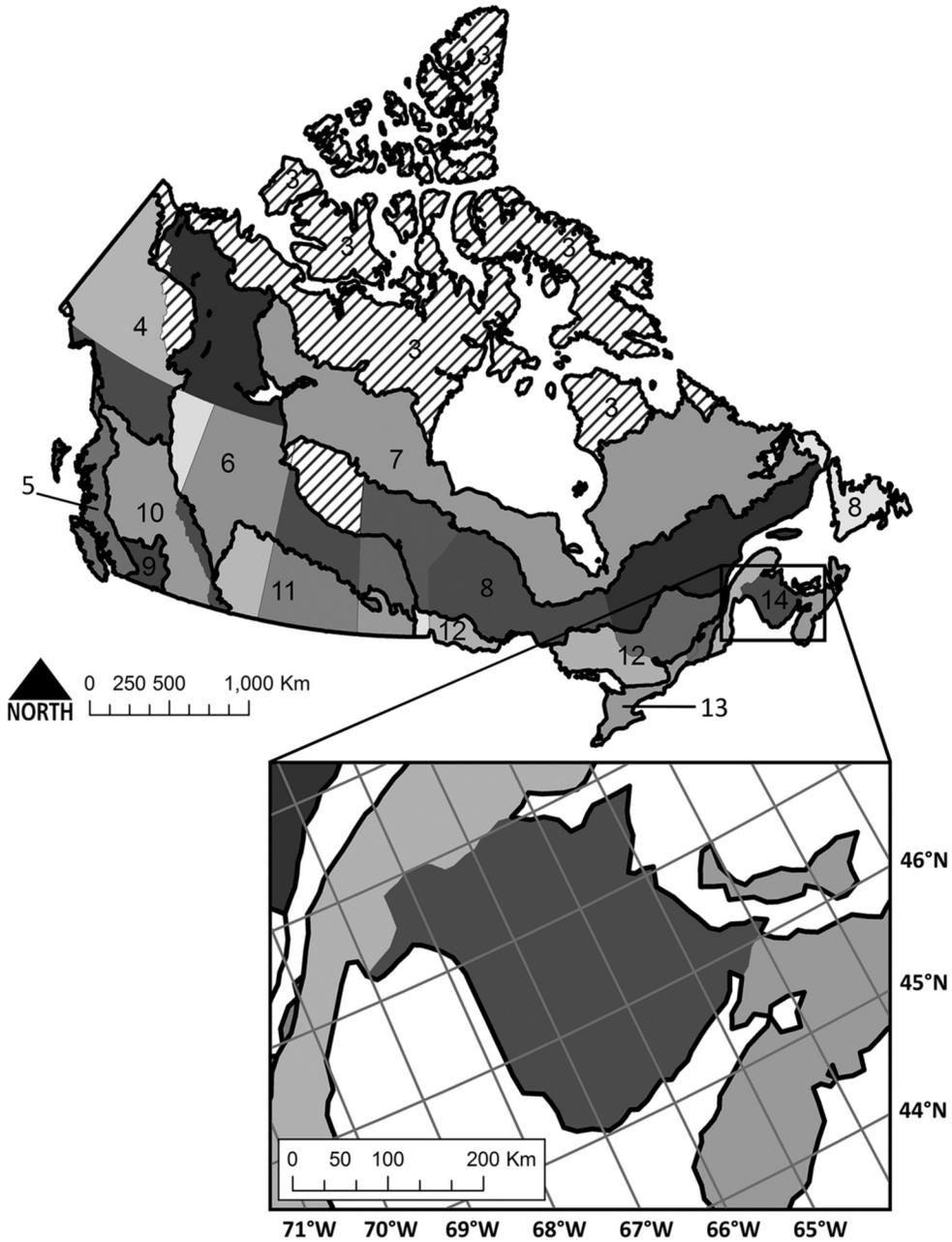


FIGURE 1. Geographic strata for analysis of Breeding Bird Survey (BBS) data using the Canadian hierarchical Bayesian model. The strata are the areas created by the intersection of provincial and territorial boundaries with those of Bird Conservation Regions (BCRs). Exceptions are the strata consisting of all of BCR 7, which crosses several provinces and territories, and all of BCR 14, which includes Nova Scotia and Prince Edward Island. BCR borders are indicated by heavy black lines: BCR 3, Arctic Plains and Mountains; BCR 4, Northwestern Interior Forest; BCR 5, Northern Pacific Rainforest; BCR 6, Boreal Taiga Plains; BCR 7, Taiga Shield and Hudson Plains; BCR 8, Boreal Softwood Shield; BCR 9, Great Basin; BCR 10, Northern Rockies; BCR 11, Prairie Potholes; BCR 12, Boreal Hardwood Transition; BCR 13, Lower Great Lakes/St. Lawrence Plain; and BCR 14, Atlantic Northern Forest. Within provinces and territories, strata are indicated by different shades of grey. Regions filled with diagonal lines have insufficient BBS data to be included in any analyses. The inset shows the original sampling strata for the BBS, i.e., degree-blocks, overlaid on the province of New Brunswick.

$$n_{i,t} = z_i * e^{S_i + \beta_i * t + \gamma_{i,t} + 0.5 * \sigma_{\omega_i}^2 + 0.5 * \sigma_{\varepsilon}^2}$$

[equation 2]

The variance components ($0.5 * \sigma_{\omega_i}^2 + 0.5 * \sigma_{\varepsilon}^2$) are added to correct for retransformation bias: the exponent of the average of a normal distribution — the sum of $S_i + \beta_i * t + \gamma_{i,t}$ — underestimates the average of the log-normal distribution (Newman 1993). The variance components ensure that the indices are scaled to an average number of birds per route. Annual indices for larger regions, i.e., provinces, territories, BCRs, or country, are area-weighted sums of the stratum-level estimates:

$$N_t = \frac{\sum A_i * n_{i,t}}{A}$$

[equation 3]

Trends in the HB₃ model are estimated as geometric mean annual changes in population size over specific intervals, expressed as a percentage. That is, the trend \bar{B} from year t_a to year t_b for a given region is:

$$\bar{B} = 100 * \left(\left(\frac{N_{t_b}}{N_{t_a}} \right)^{\frac{1}{t_b - t_a}} - 1 \right)$$

[equation 4]

The HB₃ model described here is very similar to the HB₂ model used by the USGS (described in Sauer and Link 2011) with one modification: the HB₂ model estimates a common observer–route variance across all strata (σ_{ω}^2 instead of $\sigma_{\omega_i}^2$). In the HB₃ model, the stratum-specific observer–route variances ($\sigma_{\omega_i}^2$) are drawn from a hyperdistribution that allows the variance of observer–route effects to differ among strata and yet be estimated relatively efficiently, even for strata with relatively few observer–route combinations. The hyperdistribution is defined such that the precision (the inverse of the variance) of stratum-specific observer–route effects ($1/\sigma_{\omega_i}^2$), is assumed to come from a log-normal distribution with a common mean and variance, i.e., $\log(1/\sigma_{\omega_i}^2) = N(\mu_{\omega}, \sigma_{\omega}^2)$. In comparison, the HB₂ model estimates a single observer–route variance term (σ_{ω}^2) across all observer–route combinations, which is given a diffuse inverse gamma prior distribution (scale = shape = 0.001). The WinBUGS code for the HB₃ model is available in Appendix 2A.

Comparisons between models

We used the HB₃ and ML models to estimate trends and annual indices for all species using a common dataset. The dataset included all observations from Canadian BBS routes, run from 1970 through 2011, under acceptable observation conditions (i.e., coded as RunType = 1; USGS 2012*). Although the original datasets were identical, the models have slightly different inclusion criteria for a given route–year combination. For example, the ML model estimates a route–

specific trend parameter that requires a minimum of 2 years of observations on each route, whereas the HB₃ model estimates a stratum-specific trend parameter and, therefore, can include data from routes with only 1 year of data. These differences were relevant to very few route–year combinations (< 1%), so they resulted in only very small differences in the final estimates. All species names (English, French, and scientific), as well as all trend estimates for each species, region, and trend period, are available in Appendix 3.

For most comparisons, we contrasted estimates from the two models at three scales. First, the national scale provides a comprehensive comparison that has relevance to many conservation and management decisions. Second, the BCR-scale comparisons highlight some of the important spatial characteristics of each model, e.g., differences in area weights, spatial scope of inference, and minimum data criteria (Canada's 11 BCRs are shown in Figure 1). Finally, a comparison for one province (New Brunswick; Figure 1 inset), which is an identical stratum in both analyses, provides the most direct comparison. This last comparison highlights differences between the overall frameworks of the ML and HB₃ models, because the data, spatial extent, and relative area weights are as similar as possible.

We made quantitative comparisons of five characteristics of the two models: (1) the number of species in each region with trend estimates, (2) the magnitude of trends, both long term (> 40 years) and short term (10 years), (3) the precision of trend estimates, (4) the inter-annual variability of short-term trends, and (5) the scale of the annual indices. We did not include a comparison of the number of statistically significant trends in the two models (i.e., a trend estimate for which the confidence or credible interval excludes zero). Statistical significance is a synthetic metric that confounds trend magnitude and precision; therefore, making such a comparison would add nothing new to the comparisons made here.

Magnitude of trends: To compare the magnitude of the long- and short-term trends, we used an HB model analogous to a weighted paired *t*-test that accounts for the uncertainty of each individual trend estimate. Extreme trend estimates (i.e., estimates that are larger in absolute magnitude) are more likely to be imprecise and would strongly influence a direct comparison of trends if not weighted by precision. Our comparison approach was derived from one originally described in Sauer and Link (2002), which assumes that estimated trends are part of a collection of trends that share a common mean and variance. However, we did not assume a common average trend for each model (as suggested in Link and Barker 2009) and did not, therefore, shrink imprecise estimates toward a common mean. The comparison model uses trend estimates ($\hat{\beta}_{s,m}$, trend estimates for species *s* from model *m*) as data and accounts for uncertainty in the estimation of those trends using estimates of the variance of each

trend ($\hat{\sigma}_{s,m}^2$). It makes pairwise, within-species comparisons between the posterior estimates of trend from each model

$$(\beta_{s,HB} - \beta_{s,ML})$$

and compares the overall average species-specific differences

$$\left(\frac{\sum_1^S \beta_{s,HB} - \beta_{s,ML}}{S}\right)$$

The comparison model estimates the average, species-level difference (i.e., the result of subtracting one estimate from another) in trends between the two models, weighted by the relative precision of each estimate. The average of the species-level differences is an estimate of the relative bias in the two models: positive differences indicate that the HB₃ trend estimates tend to be more positive (less negative) than the ML trend estimates for a given species. In addition to evaluating bias, we also estimated the correlation coefficient of species-level posterior trend estimates from this model, as a measure of overall agreement between trends estimated by the two models. The WinBUGS code for the comparison model is provided in Appendix 2B.

To compare the tendency for each model to produce extreme trend estimates (i.e., larger in absolute magnitude), we estimated the slope of a regression of ML trends on HB trends, using major axis, model II regression methods (Legendre and Legendre 1998). Slope values > 1 would indicate that, on average, trends from the ML model are more extreme for a given species. Values < 1 indicate that HB trends are more extreme. We used an unweighted regression, because precision-weighted comparisons would greatly reduce the influence of the extreme trends, which are often imprecisely estimated.

Precision of trends: To compare the precision of trend estimates between the two models, we treated credible intervals for the HB₃ model and confidence intervals for the ML model as comparable estimates of variance (Sauer and Link 2011). Despite their fundamental mathematical and philosophical differences, 95% credible intervals from simple Bayesian models using diffuse priors are generally very similar to 95% confidence intervals from comparable frequentist models and, in practice, they are used in similar ways to assess uncertainty in model estimates (Clark 2005). We used the model described for comparing the magnitude of trends to simultaneously compare the precision of trend estimates

$$\left(\frac{1}{\sigma_{s,m}^2}\right)$$

This model accounts for both the imprecision of each trend estimate (through the estimated variance of trends) and the imprecision around the estimate of the variance. Specifically, the model assumes that the estimated variances ($\hat{\sigma}_{s,m}^2$) of each trend estimate are mutu-

ally independent (both within species [*s*] and within models [*m*]) and independent of their associated trend estimates (Sauer and Link 2002). The true variance of each trend ($\sigma_{s,m}^2$) was estimated using a chi-squared distribution with *n* degrees of freedom, such that

$$\frac{n \cdot \hat{\sigma}_{s,m}^2}{\sigma_{s,m}^2} \sim \chi_n^2$$

where *n* is the number of routes on which the species was observed (i.e., the number of routes contributing data to the trend analysis). We then estimated the proportion of species for which the HB₃ trend was more precise (i.e., $\hat{\sigma}_{s,HB}^2 < \hat{\sigma}_{s,ML}^2$) and the average pairwise differences in the standard errors of each species' trend

$$\left(\frac{\sum_1^S \sigma_{s,HB}^2 - \sigma_{s,ML}^2}{S}\right)$$

where *S* is the number of species.

Inter-annual variability in short-term trends: The two models have very different definitions of trend (Sauer and Link, 2011): the HB₃ model estimates trend as an endpoint comparison of the annual indices in the first and last years of a given period; the ML model estimates a series of annual indices, then derives trend as the slope of a log-linear regression of the indices against time. Because the HB₃ model trend estimate is an end-point comparison, which does not directly include any of the indices in the intervening years, trend estimates from the HB₃ model could be more variable among consecutive years (i.e., greater inter-annual variation). For example, consecutive 10-year trends estimated using the HB₃ model (e.g., 2000–2010 and 2001–2011) are based on completely different pairs of annual indices. Variability of this sort is particularly important because these short-term, 10-year trends are the basic estimates of population change used for species assessments by COSEWIC. Undue inter-annual variation in trends could introduce some uncertainty and instability into important and potentially costly conservation decisions.

To compare inter-annual variation in short-term trends, we calculated all possible 10-year trends ($\beta_{\text{year } x - \text{year } x+10}$) using each model from the entire time-series of BBS data, i.e., $\beta_{1970-1980}$, $\beta_{1971-1981}$, ..., $\beta_{2001-2011}$. We then calculated the absolute value of the change in estimates between all 32 consecutive pairs of trends, e.g., $\Delta\beta_1 = |\beta_{1970-1980} - \beta_{1971-1981}|$, $\Delta\beta_2 = |\beta_{1971-1981} - \beta_{1972-1982}|$, ..., $\Delta\beta_{32} = |\beta_{2000-2010} - \beta_{2001-2011}|$. We averaged these estimates of consecutive change in trend across the full time series

$$\overline{(\Delta\beta)} = \sum_{i=1 \dots 32} \Delta\beta_i / 32$$

and compared the averages between the ML and HB₃ models for each species at the national scale.

Scale of annual indices: The annual indices from the BBS reflect the relative abundance of a species

among regions and influence the relative weight of each region's population trend in composite trends for larger regions. In addition, the indices are used in many conservation documents as a scale of the relative abundance among species. In regions that are identical in the two analyses, the two models produce annual indices of abundance on the same scale — the average expected count on an average route, conducted by an average observer. However, annual indices from the two models are derived in fundamentally different ways. The ML model estimates the annual indices directly, and trends are derived from log-linear regressions of the indices on year. The HB₃ model estimates trends directly and annual indices are derived estimates of departures from a long-term trend line. We compared the scaling of annual indices from the two models for a region that is identical in the two analyses — New Brunswick. For each species, we calculated the average of the annual indices from each model for New Brunswick for 1985–1995, which roughly represents the middle of the time series for the BBS in Canada. We limited our quantitative comparison to this one region and window of time because it provided the most direct comparison. It was largely unaffected by differences among strata in area weights, the inclusion or exclusion of particular routes, or in the indices at either end of the time series, which are more likely to be influenced by the way in which the ML model handles years with no observations or by strong underlying trends estimated using the HB₃ model.

To quantify differences in the scale of annual indices, we used a generalized linear mixed model with a Gaussian error distribution, a log-link, a fixed effect of model type (HB₃ vs. ML), and a random effect for species. We used the log-link because the models that produce the annual indices estimate the variances of the indices on the log scale and so that we could provide each species with an approximately equal weight in the analysis (i.e., comparing estimates on the linear scale would give abundant species undue weight). We report the retransformed parameter estimate and confidence intervals for the fixed effect of model type as the average difference in annual indices between the two models.

The HB₃ model used here produces annual indices on a slightly different scale than the HB₂ model, which estimates a single observer–route variance term for all strata (σ_0^2). Therefore, we also compared annual indices from the HB₂ model with indices from the ML model. Observer–route variance influences the scaling of annual indices in both HB models in two ways. First, the structure of the observer–route variance (σ_0^2 vs. σ_{0i}^2) affects the estimates of the average stratum-level abundance (S_i). Second, the estimates of observer–route variance are used as a retransformation factor that scales the annual indices to an estimate of average counts per route.

If a common observer–route variance across all strata is estimated, as in the HB₂ model, some of the ecologically relevant variation in abundance among strata is modelled as observer–route error and not incorporated into the estimates of S_i , which, as a result, become more similar among strata, i.e., they approach the survey-wide average abundance. The observer–route variance estimate is composed of the variation in abundance among individual BBS routes as well as the variation among observers. Although the variance among observers is entirely a source of error, the variance among routes represents, at least partly, ecologically-relevant spatial variation in abundance. This spatial variation in abundance among routes is also nested within the variation in abundance among strata — routes are uniquely associated with a given strata. As a result, when a common observer–route variance is estimated (σ_0^2 in the HB₂ model), some of the variation in abundance among routes is modelled as error. By contrast, the stratum-specific estimates of observer–route variance (σ_{0i}^2 in the HB₃ model) allocate more of the variation in abundance among strata to the estimates of S_i .

The stratum-specific variance estimates in the HB₃ model also retain greater variation in abundance among strata, because they provide stratum-specific retransformation factors to rescale the annual indices. The observer–route variance is one of two variance components ($0.5 * \sigma_{0i}^2$ in equation 2) that are added to the parameter estimates on the log-scale to adjust for the bias in transforming from the log scale to the linear scale (i.e., count of birds). Annual indices in the HB₃ model are scaled by a multiplicative function of $e^{0.5 * \sigma_{0i}^2}$, which varies among strata; the analogous function ($e^{0.5 * \sigma_0^2}$) is the same for all strata in the HB₂ model.

All figures that include comparisons of the ML and HB₃ models at the BCR scale have been arranged so that the BCRs are sorted in descending order of an approximate estimate of data quality, i.e., BCRs are sorted in descending order of

$$\frac{\sqrt{n_b}}{A_b}$$

for each BCR, b , where n_b is the number of routes and A_b is the area of the BCR. Some additional comparisons are included in Appendix 4.

Results

Magnitude of trends

Overall, trends estimated from the ML and HB₃ models were generally similar. Estimates from the two models were correlated for all regions and time periods (Figure 2), with higher correlations at the level of individual strata and BCRs than at the national level (compare plots for Canada and New Brunswick in Figure 2 and Figure 3A). Long-term trends were generally more correlated than short-term trends (Figure 3A). Trends

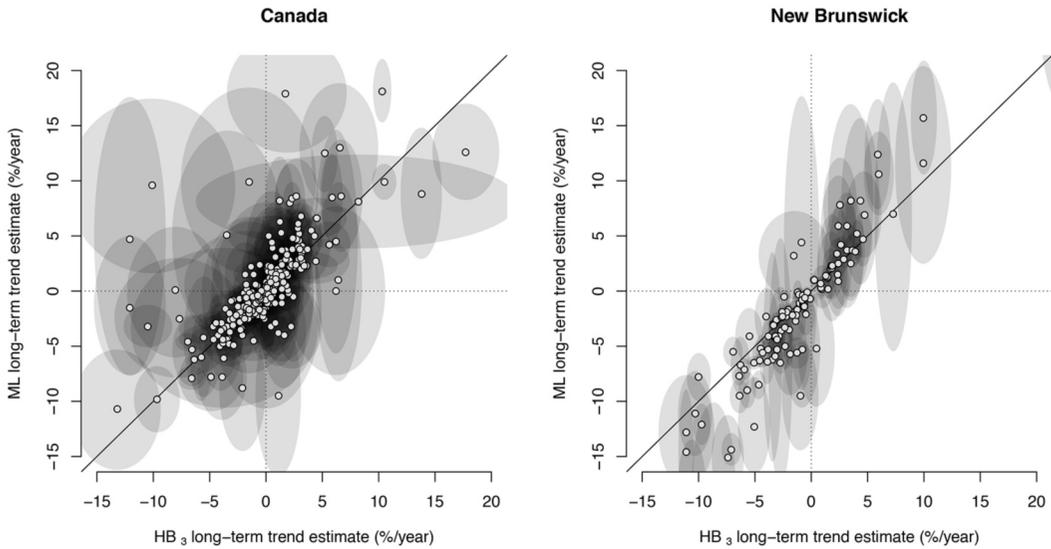


FIGURE 2. Point estimates (circles) and associated uncertainty (semi-transparent ellipses) of long-term Breeding Bird Survey trends for Canada and New Brunswick calculated from the previously used maximum likelihood (ML) model and the new hierarchical Bayesian (HB_3) model. The diameters of the ellipses represent the central 67th percentile of the posterior distribution of trend estimates from the Bayesian trend comparison model (vertical diameter = ML trend estimates and horizontal diameter = HB_3 trend estimates). The diagonal line indicates a 1:1 relationship. The dotted black lines divide the plot into quadrants representing combinations of trend-signs from the two models (e.g., points and posterior mass plotted in the upper right quadrant represent species with positive trend estimates from both models). Differences in trends for Canada incorporate differences between the two models in terms of model structure and assumptions, as well as differences in area weights, routes, and regions of the country included in the estimates. Differences in trend estimates for New Brunswick are almost entirely due to differences in model structure and assumptions. One species was omitted from both graphs because its trend estimates were extreme ($> 25\%/year$) and including it changed the scale such that it became difficult to discern relations among the remaining species.

from the two models also tended to be more correlated in regions with more data (i.e., BCRs closer to the left side of Figure 3A, which have more routes in relation to their area).

At the national level, 54% of species had HB_3 trends that were more negative than their ML trends, and HB_3 trends were on average slightly more negative: the average species-level difference in long-term trends was $-0.26\%/year$, although the 95% credible interval—the interval between the 2.5 and 97.5 percentiles of the posterior distribution—included zero (-0.68 to 0.16 ; Figure 3B). In contrast, national short-term trends from the HB_3 model were on average more positive: 64% of species had HB_3 trends that were more positive, and the average difference was $1.1\%/year$, (95% credible interval -0.41 to 1.7 ; Figure 3B). For most BCRs, the average differences between trend estimates from the two models followed the same pattern observed in the national estimates, i.e., the differences in long-term trends were more often negative and differences in short-term trends were more often positive (open circles below the dotted line and closed circles above in Figure 3B).

All trends from the HB_3 model tended to be less extreme (i.e., less in absolute magnitude, points above

the dotted line in Figure 4), except for the national long-term trends, where the opposite was true.

Precision of trends

Trend estimates from the HB_3 model, both long- and short-term, were more precise than estimates from the ML model in all regions. For national trends, 63% of species had more precise trends with the HB_3 model, and the standard errors of HB_3 trends averaged approximately $0.5\%/year$ smaller than the standard errors of the ML trends (Figure 5). Within the BCRs, HB_3 trends were more precise for almost all species (73–98%) and their standard errors averaged approximately $2\%/year$ smaller (Figure 5).

Inter-annual variability of short-term trends

Ten-year trend estimates from the HB_3 model were generally less variable among years than those from the ML model (Figure 6). Across species, the mean absolute year-to-year change in 10-year trend estimates was $1.12\%/year$ smaller using the HB_3 model than the ML model for a given species. The average species-level difference ($HB_3 - ML$) in the mean absolute year-to-year change in national 10-year BBS trends was -1.12 (SE 0.17); in Figure 6A, most points are to the right of the 1:1 diagonal line. In general, 10-year trends

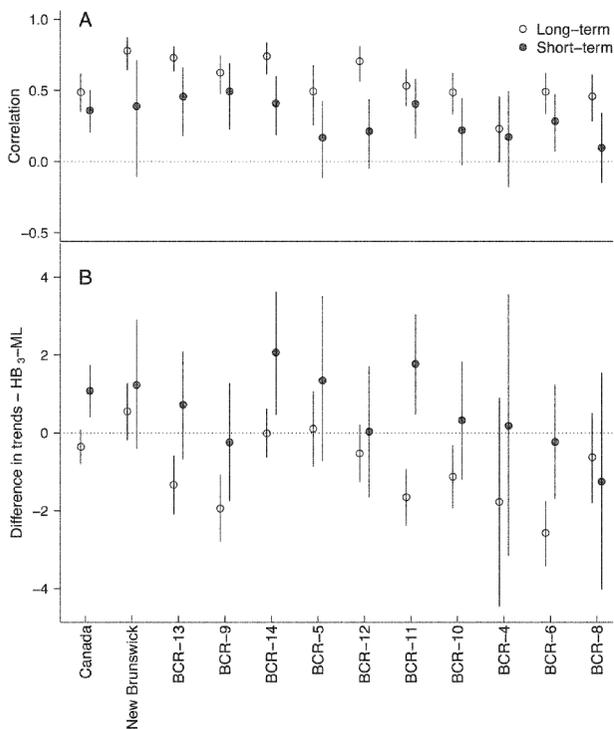


FIGURE 3. Correlation and comparison of the magnitude of long- and short-term Canadian Breeding Bird Survey trends estimated using the earlier maximum likelihood (ML) model and the new hierarchical Bayesian (HB₃) model. Bird Conservation Regions (BCRs) are arranged in descending order of the ratio of the square root of sample size to area. The correlations plotted in the upper graph show the overall similarity of trends estimated using the two models. The differences plotted in the lower graph show the relative bias in the two models: negative values indicate that the HB₃ trend estimates were smaller than the ML estimates (i.e., the population change was more negative and/or less positive); positive values indicate the opposite. All comparisons were made on the posterior estimates of trends, output from a model that accounts for the relative precision of the original estimates. The circles represent the means of the posterior distribution, with error bars indicating 95% credible intervals. See Figure 1 for BCR definitions.

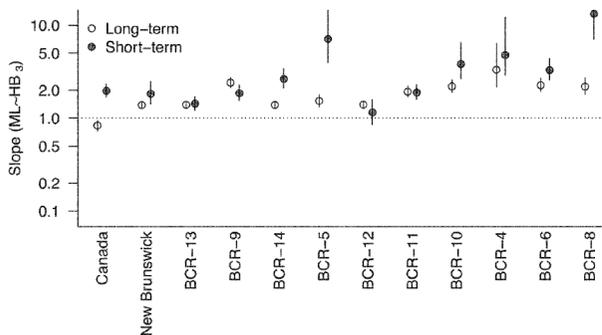


FIGURE 4. Slopes of major axis, model II regressions of Breeding Bird Survey trend estimates from the previously used maximum likelihood (ML) model regressed on estimates from the new hierarchical Bayesian (HB₃) model. Bird Conservation Regions (BCRs) are arranged in descending order of the ratio of the square root of sample size to area. These comparisons do not consider the precision of the estimated trends, for reasons explained in Methods. Values > 1.0 indicate that the absolute values of the ML trends are, on average, more extreme than the HB₃ trends for each species (i.e., the ML trends are larger in absolute magnitude) and values < 1.0 indicate that the HB₃ trends are more extreme. Error bars indicate 95% confidence intervals for the slope estimates. Slope values are plotted on a logarithmic scale so that equal deviations from the 1:1 expectation appear equal above and below the dotted line (e.g., a slope of 2.0 represents the same magnitude of deviation from the expectation as a slope 0.5). See Figure 1 for BCR definitions.

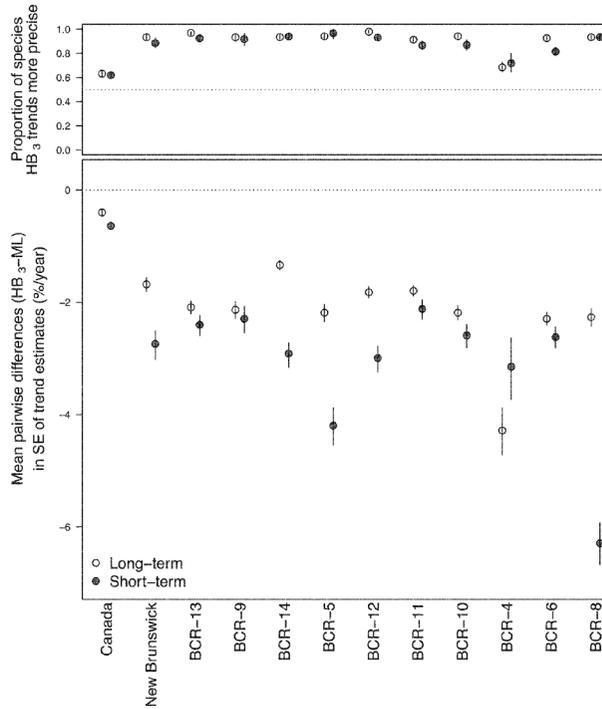


FIGURE 5. Comparisons of precision of Breeding Bird Survey trend estimates between the previously used maximum likelihood (ML) model and the new hierarchical Bayesian model (HB_3) in various regions of Canada. Bird Conservation Regions (BCRs) are arranged in descending order of the ratio of the square root of sample size to area. The upper graph shows the proportion of species in each region for which the HB_3 trend estimates were more precise than the ML trend estimates. The lower graph shows the mean pairwise differences in trend estimates' standard errors (SE); negative values indicate the HB_3 estimates were more precise, i.e., have smaller SEs. Circles and associated error bars represent the means and 95% credible intervals of the posterior distribution of estimates from a trend comparison analysis that accounts for the uncertainty in both the estimates of trend and the estimates of the variance of the trends. See Figure 1 for BCR definitions.

from the ML model were much more variable for species with relatively imprecise annual index estimates (Figure 6B, lower row). For some species, 10-year trends from the HB_3 model were more variable than those from the ML model (points to the left of the diagonal line in Figure 6A); these are species with relatively precise estimates of highly variable annual indices (Figure 6B, upper row).

Scale of annual indices

On average, annual indices from the HB_3 model were somewhat larger than those from the ML model (mean difference = 1.29, 95% credible interval 1.23–1.36), however the log–log relationship closely follows the 1:1 line (Figure 7). Mean estimates from the HB_3 model were more similar to the annual indices from the ML model than those generated by the HB_2 model with a common observer–route effect (mean difference 2.51, 95% credible interval 2.32–2.71). In addition, the differences between the HB_2 and ML model depended on the species' abundance; almost all species with ML indices > 2 were above the 1:1 line and the slope of the

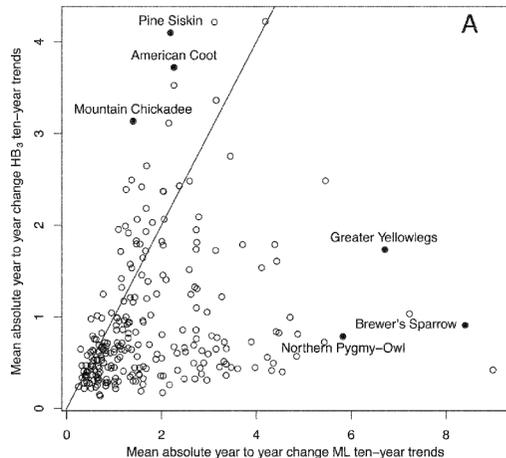


FIGURE 6A.

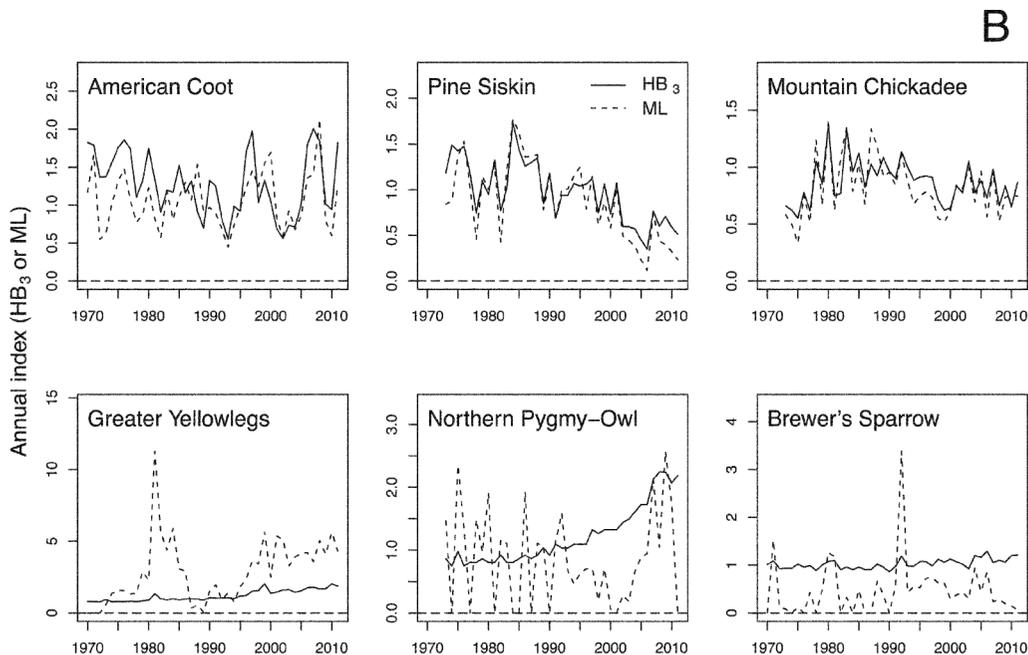


FIGURE 6. Comparison of the year-to-year variability in 10-year Breeding Bird Survey trends for Canada estimated using the maximum likelihood (ML) model and the new hierarchical Bayesian (HB_3) model. A shows the species-level estimates of the absolute year-to-year change in trend values for the two models, from a series of consecutive 10-year trends from $t_{1970-1980}$, $t_{1971-1981}$, ..., $t_{2001-2011}$. The filled circles indicate the species trends plotted in B. B shows overlays of annual indices from the HB_3 model (solid lines) and the ML model (dashed lines). American Coot (*Fulica americana*), Pine Siskin (*Spinus pinus*) and Mountain Chickadee (*Poecile gambeli*) are examples of species for which the HB_3 trends are more variable, i.e., points above the diagonal in A. Greater Yellowlegs (*Tringa melanoleuca*), Northern Pygmy-Owl (*Glaucidium gnoma*), and Brewer's Sparrow (*Spizella breweri*) are species for which the ML trends are more variable, i.e., points below the diagonal in A.

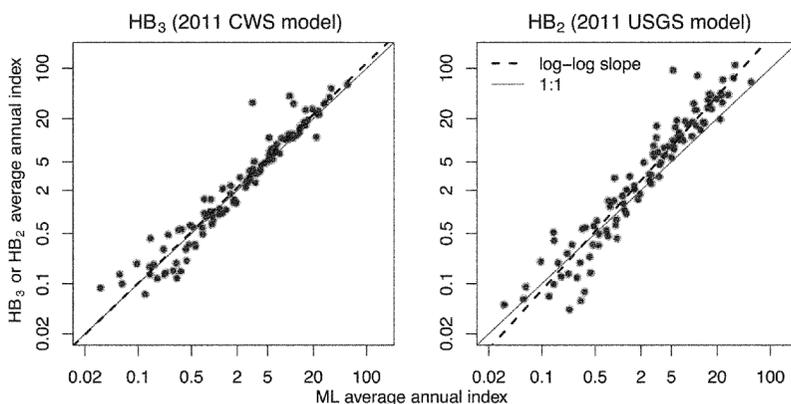


FIGURE 7. Log-scale comparisons of the average of Breeding Bird Survey annual indices for New Brunswick (for 1985–1995) from the previously used maximum likelihood (ML) model, plotted against annual indices from two hierarchical Bayesian models: the new Canadian model (HB_3) and the HB_2 model described in Sauer and Link (2011). Each dot represents one species, and the black dashed lines are predictions from linear models regressing the log of the average annual indices from the two HB models on the log of the average annual indices from the ML model. The solid grey line represents a 1:1 relationship. Note: CWS = Canadian Wildlife Service; USGS = United States Geological Survey

log–log relationship diverged from the 1:1 line (Figure 7).

Discussion

Canadian BBS trends estimated using the HB_3 model are generally similar in magnitude to those estimated from the ML model that was in use between 2002 and 2009. This suggests that for most species, our understanding of population status has not been drastically changed by the adoption of the HB_3 model.

Overall, the trends from the new HB_3 model are more precise, less likely to have extreme values, and less variable among years than those from the previous ML model. A similar hierarchical model (HB_2) showed similar improvements compared with the route-regression model previously used to estimate BBS trends in the United States (Sauer and Link 2011). Although no single model is ideal for all species in all situations, the differences demonstrated here suggest that the HB_3 model provides better information on the population status of approximately 300 bird species breeding in Canada. The broad similarities in trend estimates from the two models are not surprising; the BBS represents a large dataset, and many temporal trends are clear and strong regardless of the analytical models used to estimate them (as noted in similar cross-model comparisons in Thomas and Martin 1996 and Sauer and Link 2011). For some species, the two models generate strikingly different trend esti-

mates, but, in most cases, the estimates are extremely imprecise for both models. For example, in Figure 2, the points that fall far from the 1:1 line are generally very imprecise (i.e., have large credibility ellipses), and the credibility ellipses overlap the 1:1 line in at least one dimension for most of these species.

Magnitude

In some BCRs, long-term trends derived from the HB_3 model are generally more negative/less positive than those from the ML model. This difference is partially the result of a statistical artefact in the ML model; thus, the HB_3 trend estimates are less biased. In the ML model, for years in which a species was not observed in a particular region, the annual index is set to a small arbitrary value. That is, the annual index itself is estimated at zero, but when the trend is calculated, a small constant is added (0.001) so that the log-linear regression can be calculated. For species and regions with these zero counts near either end of a time series, any estimated long-term trend will be strongly influenced by this arbitrary value (O'Hara and Kotze 2010). Because of the smaller number of BBS routes in earlier years, species are more likely to have annual index estimates of zero for those years, and, therefore, their abundance is more likely to be represented by the arbitrary value leading to positive bias in the ML model (Figure 8). In contrast, in the HB_3 model, the annual index for those years is estimated to be a non-zero val-

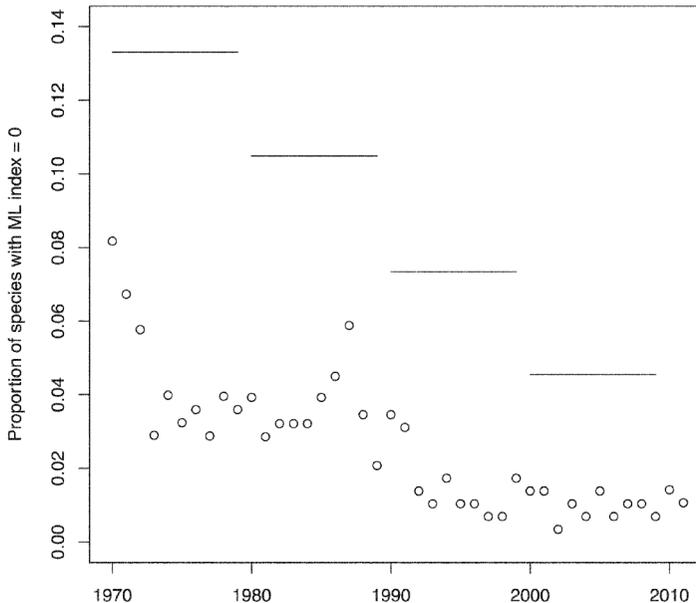


FIGURE 8. Proportion of species with annual index estimates equal to zero when calculated with the maximum likelihood model, previously used to estimate Breeding Bird Survey trends in Canada. Circles indicate annual proportions; horizontal lines indicate proportions of species with at least one annual index equal to zero in each of four decades (1970s, 1980s, 1990s, and 2000s).

ue that reflects the best estimate of the abundance in that year, given the number of routes with zero counts in that year, and the estimates of all other parameters in the model.

Precision

Canadian BBS trends estimated using the HB_3 model were more precise than those from the ML model, and this difference was more pronounced for smaller regions (i.e., New Brunswick and the BCRs). At the national level, the HB_3 model's hierarchical structure makes efficient use of the data and gives moderately more precise estimates than the ML model. For smaller regions, the hierarchical structure of the model results in additional efficiencies. Because the stratum-level parameters are estimated within a single national model, estimates for the error, over-dispersion, first-year observer effects, and the observer–route variance (i.e., the nuisance parameters) share information among strata. This sharing of information makes for more precise stratum-level estimates of the nuisance parameters than would be possible if they were estimated solely from the much smaller set of routes within each stratum. In contrast, the ML model accounts for the observer and route-level nuisance parameters separately within each region.

The HB_3 model's increased precision has important benefits for conservation in that it decreases uncertainty around the assessment of species' status. Steep rates of change or abrupt changes in populations (e.g., 30% decreases in populations that may warrant species at risk status; COSEWIC 2011*) can be identified with greater certainty and over shorter periods of time. Similarly, species can be more confidently classified into status categories, such as those used in Environment Canada's Status of Birds in Canada website (www.ec.gc.ca/soc-sbc).

Variability among years

For most species, the HB_3 short-term trend estimates are less variable among years than the ML short-term trend estimates, because the hierarchical structure of the year-effect parameters makes annual fluctuations much less sensitive to sampling error and annual route coverage. For a few species, short-term HB_3 trend estimates are more variable among years than ML trends, but only when there is relatively strong evidence supporting large annual fluctuations in the population's status. In essence, although short-term HB_3 trend estimates will fluctuate more for populations whose status is well estimated, those fluctuations are likely appropriate because they are more likely to reflect real changes in the populations and not sampling error. In these cases, management decisions that rely on short-term trend estimates, such as COSEWIC status assessments, should also examine the population's recent trajectory to give some context to the trend estimate in any given year. For data-poor populations, the HB_3 model's short-term trends will be more stable across years, so man-

agement decisions can be made with some confidence that a species' status assessment will not strongly depend on the year in which it was assessed.

Annual indices

At the stratum level, annual indices from the HB_3 model are very similar to those from the ML model. They are directly interpretable as the expected count by an average observer on an average route in the stratum. They are also more similar in scale to the indices from the ML model than annual indices from the HB_2 model. Estimating observer–route variance at the stratum level in the HB_3 model brings the scaling of the annual indices closer to a scale familiar to users of the ML model results, which is approximately equal to the observed average count across the routes run in any given year.

The improvement in scaling of the annual indices in the HB_3 model over the HB_2 model also has an influence on regional trends, because it better reflects the variation in abundance among strata. Because regional annual indices are sums of the stratum-level indices, the relative contribution of each stratum's trend to a regional trend is partly mediated by the relative abundance of birds in each stratum. In relatively rare but potentially important cases, regional trends from the HB_3 and HB_2 models can be strikingly different: where the average abundance, trend, and observer–route variance are highly variable among strata. For example, the HB_2 model applied to Canadian data for Wood Thrush (*Hylocichla mustelina*) estimates the national trend at -1.8 (95% credible interval -2.6 to -0.9), while the HB_3 model estimates the national trend at -4.4 (95% credible interval -5.3 to -3.6). This large difference — the credible intervals do not overlap — is due to the relative weight in the two analyses of the largely stable Wood Thrush population in southern Ontario (Ontario BCR 13). In this region, the observer–route variance is much lower than in any other stratum in the analysis. As a result, the observer–route variance retransformation factors are very different in the two models ($\sigma_0^2 \gg \sigma_{\text{BCR13}}^2$). The larger retransformation factor in the HB_2 model creates annual indices that overestimate the observed abundance of Wood Thrush in southern Ontario by a factor of 3 and, similarly, overestimate the proportion of the national population that occurs in the region and, therefore, its influence on the national trend.

Users of published BBS estimates should be aware that the scale of annual indices from the HB_3 model and the ML model may differ because the models account differently for routes in regions not included in each species analysis. The HB_3 model estimates an annual index scaled to the expected count averaged across routes within the strata included in the analysis, i.e., it excludes routes where the species data are too sparse or that are outside the species' breeding range. The ML model's annual indices are scaled to the expected count averaged across all routes within the re-

gion being analyzed, e.g., for national trends, all routes run in the country are included, regardless of whether they fall within the species' breeding range. For broadly distributed and common species, for which all strata and all routes are included in the HB_3 analysis, the scales of the national estimates from each model will be equal. For more locally distributed species, for which the HB_3 analysis only includes a subset of the strata (e.g., any species that breeds exclusively west of the Rockies), the HB_3 annual indices will be scaled to a much higher abundance because they ignore all of the routes and strata outside of the species' breeding range, where the average counts are zero. It is important to note that this difference does not affect our quantitative comparisons of annual indices among the HB_3 , ML, and HB_2 models (i.e., Figure 7), because we compared them within a single stratum, nor does it affect trend estimates in the two models; it only affects the scale of the annual indices.

Improved population inference from BBS

Beyond the quantitative comparisons, some philosophical arguments suggest that the HB_3 model represents an improvement over the ML model. First, the Bayesian framework provides much more intuitive and flexible inference than the frequentist framework of the ML model. The correct inferential interpretation of Bayesian credible intervals is almost certainly a more natural expression of the type of inference that users of BBS trends desire, i.e., there is a 95% probability that the trend is greater than the lower bound and less than the upper bound. In contrast, the correct interpretation from frequentist analyses (usually confined to tests of significance) do not relate to the estimates of population change, but to the data that were sampled, e.g., if the true rate of population change is zero, there is less than a 5% chance of observing the data that were collected. In addition, the ability to estimate the full posterior distribution of a broad suite of derived parameters, such as the probability that the population has declined more than 50% in 10 years (i.e., one of the COSEWIC criteria for "Endangered" status), provides a practical and flexible approach to assessing the uncertainty around particular conclusions one might draw from a BBS trend.

Second, the HB framework and the HB_3 model in particular provide an efficient, flexible, coherent and complete framework for the analysis, which the ML model lacks. The hierarchical structure of the model makes efficient use of the data and is less sensitive to annual variation in sampling error. In addition, the HB_3 model can be expanded easily to include covariables and explicit spatial structure (e.g., Thogmartin *et al.* 2004; Nielson *et al.* 2008), and the HB framework lends itself well to composite and comparative analyses of the BBS and other surveys (e.g., Link and Sauer 2007 and the method used here to compare the magnitude and precision of trends from the two models). Finally, the HB_3 model estimates trends and annual in-

dice within a single coherent model structure which the ML model lacks. For example, the ML model requires the addition of an arbitrarily chosen constant to some years' annual indices to estimate trends.

A third example of improved inference from results of the HB_3 model is that the trends and estimates of uncertainty around trends for multi-strata regions (e.g., national trends), are a much clearer reflection of the full uncertainty of the population status estimates within the region. That is, regional and national estimates are combinations of stratum-level estimates, appropriately weighted by the proportion of the species' breeding population in each stratum (by summing indices of relative abundance, weighted by the stratum's area). In addition, the strata in the HB_3 model are more likely to reflect spatial variation in population trends and trajectories than the degree-block strata in the ML analysis. The HB_3 strata are structured on BCRs, which are relevant to population processes, and political units, which are relevant to management and human activity. In contrast, inferences regarding trends and uncertainty from the ML model only apply to the degree blocks included in the analysis, i.e., degree-blocks with data. Furthermore, the degree-block weighting of the ML model means that the relative influence of populations with potentially disparate trends and trajectories among different regions (i.e., BCRs and political units) depends on the relative sampling intensity of the BBS in those regions (i.e., the number of degree-blocks with data). For example, the White-throated Sparrow (*Zonotrichia albicollis*) occurs on BBS routes in the southern BCRs (12, 13, and 14), where there are many degree-blocks with data and where trends are strongly negative. However, the bulk of its population occurs in the more northern BCRs (6, 7, and 8), where trends are relatively stable and there are few degree-blocks with data. The degree-block weighting of the ML analysis puts relatively more weight on the decreasing southern populations (national short-term ML trend $-0.9\%/year$ [95% confidence interval -1.7% to -0.1%]), whereas the stratum area weights of the HB_3 analysis put more weight on the more positive and much less precisely estimated trends from the northern populations (national short-term HB_3 trend $0.59\%/year$ [95% credible interval -1.64% to 3.87%]). The estimated average trend from the HB_3 model better reflects the status of the bulk of the Canadian population of the White-throated Sparrow in the northern BCRs and, given that the northern BCRs are relatively poorly surveyed by the BBS, the wider credible interval of the national HB_3 trend is a more appropriate measure of the uncertainty around the population's status. This more appropriate weighting of imprecisely measured trends from regions that are relatively large but poorly surveyed also explains the modest improvements in precision at the national level, relative to the greater improvements within individual BCRs (Figure 3).

Because of the flexibility of the HB_3 model framework, the best short-term trend and the best long-term trend do not necessarily include data from the same spatial areas. Since 1966, spatial coverage by Canadian BBS routes has expanded gradually into the more northerly strata. The earliest BBS routes in Canada were run in the Maritimes and Quebec in 1966. By the early 1970s, much of southern Canada had adequate coverage. However, some more northerly regions have only had adequate coverage for the last 20 years. Because of the paucity of data from the Northwest Territories and much of the Yukon before the late 1980s or later, trends in these areas cannot reasonably be extrapolated back to 1970. Long-term trend estimates for large regions (e.g., national trends) can be estimated, but they must exclude some strata, for which we have only recent data. In contrast, short-term estimates for those regions should include all of the strata with data. In the HB_3 model, deriving separate short- and long-term trend estimates that include different strata is a straightforward process of summarizing stratum-level annual estimates, and it retains the entire posterior distribution of all derived estimates. To get similar separate trends from the ML model would require multiple runs using different subsets of the data, and each subset would have a reduced sample size and, therefore, reduced precision.

Limitations of the HB_3 model and future evolution

Along with the benefits of a complex hierarchical model structure come complex consequences for population inference. For example, the trend term in the HB_3 model combined with the distributional assumptions of the random year-effects terms means that, for years with relatively sparse or highly variable data, annual indices will tend to track a fitted estimate of the species' long-term trend. These assumptions and model structure are reasonable, given that the long-term trend is estimated from the data. However, in some cases, they could lead to an over-smoothing of the annual indices. For example, the model does not account for autocorrelation in the sequence of successive year-effects and could, therefore, be insensitive to cyclical population patterns, unless each individual year contained relatively strong evidence of a departure from the long-term trend. Similarly, the number of BBS routes in Canada has more than doubled since 1980, and, therefore, sparser data in the early years means that those early annual indices are less likely to depart from the long-term trend than annual indices in later years. Although these assumptions greatly reduce the influence of sampling errors on trend estimates (Link and Sauer 2002), as evidenced by the reduced inter-annual variability of HB_3 trend estimates (Figure 6), there are species for which the smoothing of annual indices could lead to bias in both the individual indices and the trends. Future refinements of the model may include terms to model the temporal autocorrelation in the year-effects, so that the model will more closely

reflect population fluctuations when successive years show similar departures from the long-term trend.

Conclusion

Overall, the use of the HB_3 model to estimate trends and annual indices of abundance provides improved information on the population status of birds in Canada. Estimates from the HB_3 model are generally similar in direction and magnitude to estimates from the previous model, but they offer greater precision, less variation among years, a better representation of the spatial variation in population status across Canada, and a more intuitive and flexible assessment of uncertainty. As statistical science evolves, the analysis of BBS data will follow suit, continuing to improve as it has in the past. This evolution and improvement honours the contributions of the thousands of volunteers who have participated in the BBS since its inception, and ensures that this survey will remain at the forefront of our understanding of the status of Canada's and North America's birds.

Acknowledgements

We thank the thousands of skilled volunteers who have contributed to the Breeding Bird Survey over the years, as well as those who have served as provincial and territorial coordinators. Our thanks also go to Pete Blancher, Brian Collins, Bill Link, David A. W. Miller, John Sauer, and Joe Veech for their insightful comments and review of the paper, model development or both.

Documents Cited (marked * in text)

- COSEWIC (Committee on the Status of Endangered Wildlife in Canada)**. 2011. COSEWIC's assessment process and criteria. COSEWIC, Ottawa, Ontario, Canada. Accessed December 2013. www.cosewic.gc.ca/eng/sct0/assessment_process_e.cfm
- Environment Canada**. 2011. Status of Birds in Canada – 2011. Environment Canada, Ottawa, Ontario, Canada. Accessed 11 April 2014. www.ec.gc.ca/soc-sbc
- Erskine, A. J.** 1978. The first ten years of the co-operative Breeding Bird Survey in Canada. Canadian Wildlife Service Report Series number 42. Minister of Fisheries and the Environment, Ottawa, Ontario, Canada. 61 pages.
- North American Bird Conservation Initiative Canada**. 2012. The State of Canada's Birds, 2012. Environment Canada, Ottawa, Ontario, Canada. 36 pages. Accessed 11 April 2014. www.stateofcanadasbirds.org/State_of_Canada's_birds_2012.pdf
- North American Bird Conservation Initiative International**. 2013. Bird conservation regions. North American Bird Conservation Initiative, Ottawa, Ontario, Canada. Accessed July 2013. www.nabci.net/International/English/bird_conservation_regions.html#
- Rich, T. D., C. J. Beardmore, H. Berlanga, P. J. Blancher, M. S. W. Bradstreet, G. S. Butcher, D. W. Demarest, E. H. Dunn, W. C. Hunter, E. E. Inigo-Elias, J. A. Kennedy, A. M. Martell, A. O. Panjabi, D. N. Pashley, K. V. Rosenberg, C. M. Rustay, J. S. Wendt, and T. C. Will.** 2004. Partners in Flight North American Landbird

Conservation Plan. Cornell Lab of Ornithology. Ithaca, New York, USA. Accessed July 2013. www.partnersinflight.org/cont_plan.

USGS (United States Geological Survey). 2012. North American Breeding Bird Survey data retrieval site. Patuxent Wildlife Research Center, USGS, Laurel, Maryland, USA. Accessed July 2013. www.pwrc.usgs.gov/bbs/RawData.

Literature Cited

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2–14.

Geissler, P. H., and B. R. Noon. 1981. Estimates of avian population trends from the North American Breeding Bird Survey. Pages 42–51 in *Estimating Numbers of Terrestrial Birds*. Edited by C. J. Ralph and J. M. Scott. Cooper Ornithological Society, Studies in Avian Biology 6.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Number 20 in *Developments in Environmental Modelling*. Elsevier, Amsterdam.

Link, W. A., and R. J. Barker. 2009. *Bayesian Inference: With Ecological Applications*. Academic Press, Burlington, Massachusetts, USA. xiii, 339 pages.

Link, W. A., and J. R. Sauer. 2002. A hierarchical model of population change with application to Cerulean Warblers. *Ecology* 83: 2832–2840.

Nielson, R. M., L. L. McDonald, J. P. Sullivan, C. Burgess, D. S. Johnson, D. H. Johnson, S. Bucholtz, S. Hyberg, and S. Howlin. 2008. Estimating the response of Ring-necked Pheasants (*Phasianus colchicus*) to the Conservation Reserve Program. *Auk* 125: 434–444.

Newman, M. C. 1993. Regression analysis of log-transformed data: statistical bias and its correction. *Environmental Toxicology and Chemistry* 12: 1129–1133.

O'Hara, R. B., and D. J. Kotze. 2010. Do not log-transform count data. *Methods in Ecology and Evolution* 1: 118–122.

Pardieck, K. L., D. J. Ziolkowski, and B. G. Peterjohn. 2008. *A Bibliography for the North American Breeding Bird Survey*. Version 2008. USGS Patuxent Wildlife Research Center, Laurel, Maryland, USA.

Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. *The breeding bird survey: its first fifteen years, 1965–1979*. Resource publication 157. United States Fish and Wildlife Service, Washington, D.C., USA. 196 pages.

Sauer, J. R., and W. A. Link. 2002. Hierarchical modeling of population stability and species group attributes from survey data. *Ecology* 83: 1743–1751.

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

Thogmartin, W. E., J. R. Sauer, and M. G. Knutson. 2004. A hierarchical spatial model of avian abundance with application to Cerulean Warblers. *Ecological Applications* 14: 1766–1779.

Thomas, L., and K. Martin. 1996. The importance of analysis method for Breeding Bird Survey population trend estimates. *Conservation Biology* 10: 479–490.

Thomas, L. 1996. Monitoring long-term population change: Why are there so many analysis methods? *Ecology* 77: 49–51.

Supplementary material available at:
<http://www.canadianfieldnaturalist.ca>

Received 23 September 2013
Accepted 22 November 2013

SUPPLEMENTARY MATERIAL:

APPENDIX 1. Detailed description of the maximum likelihood (ML) model used to generate Canadian Breeding Bird Survey trends before 2010.

APPENDIX 2. WinBUGS (Bayesian inference Using Gibbs Sampling) language descriptions of the hierarchical Bayesian models.

APPENDIX 3. Microsoft Excel workbook showing differences in trend estimates between the maximum likelihood (ML) and the hierarchical Bayesian (HB₃) models, with separate worksheets for short- and long-term trends and for each Bird Conservation Region, Canada, and New Brunswick.

APPENDIX 4. Additional comparisons of the hierarchical Bayesian (HB₃) and maximum likelihood (ML) models, including comparisons of the number of species with trend estimates and the classification of species trends into the categorical assessments of population status used in Environment Canada's Status of Birds in Canada website (Environment Canada 2011).