

STATE OF CANADA'S BIRDS 2019 – METHODS

INTRODUCTION

This document provides a detailed technical description of the statistical and analytical methods used to generate the indicators of the population status of groups of birds in the State of Canada's Birds 2019 report. It includes technical and mathematical details intended for an audience of scientists and technical experts.

INDICATORS

In the State of Canada's Birds 2019 report, we present indicators of the national population status of groups of bird species that reflect broad biomes within Canada (e.g., forest birds, grassland birds, seabirds) and groups of species that are known to have distinct and noteworthy trends (e.g., aerial insectivores and birds of prey). Our groupings were generally the same as those used in the national overview section of the 2012 version of the report. For many of these main groups, we also presented a series of sub-indicators to represent sub-groups with particular conservation concerns or successes (e.g., long-distance migrant shorebirds and forest birds that winter in Canada).

For this report, we considered 449 extant species of birds that are native to North America and that regularly occurred in Canada in 1970. We excluded species that are not native to North America. We also excluded 9 species that are native to North America and regularly occur in Canada in 2019, but that did not regularly occur in Canada in 1970 (i.e., they have expanded their ranges into Canada since 1970) for mathematical reasons that are further explained below. Of these 449 species, population status information suitable to include in our analyses was available for 342 species (75% of the native species that regularly occur in Canada in 2016). We used trajectory information from 1970 – 2016, because 2016 was the most recent year data were available for many of the surveys.

The supplementary data on the report website includes lists of the species included in each indicator group, the data source for each species, and the annual indices of abundance (raw data for the composite indicator models, before being re-scaled to the base-year and log-transformed).

DATA SOURCES USED

The data for our statistical model are the annual estimates of each species' national population status (hereafter indices of annual abundance). For each species, we used information derived from the most appropriate monitoring survey to represent the species' national population status. Following (ECCC 2019), we considered any data source that could provide annual or periodic estimates of species' national population status (index of annual abundance or population estimate) over the long-term (i.e., > 20 years). For each species, we selected the most appropriate data source by giving priority to estimates that:

1. most clearly reflected the species' national population status,
2. provided frequent estimates (i.e., preferring annual over periodic estimates),
3. provided a measure of the estimate's precision or uncertainty (e.g., confidence interval or standard error), and
4. provided estimates across the greatest proportion of the full time period (1970 to 2016).

In some cases, we gave priority to certain data sources, if those data were thought to better represent a species' status. There were 20 data sources that provided data for at least one species in our analyses (Table 1).

Full data source name	Number of species relying on this data source in the State of Canada's Birds 2019	Analytical approach to Generalized Additive Model (GAM) smoothing of species trajectories: Smoothing conducted on raw data or published indices of abundance
North American Breeding Bird Survey (Canadian Estimates)	212	Raw Data
Christmas Bird Count (Continental Estimates)	43	Raw Data

Christmas Bird Count (Canadian Estimates)	18	Raw Data
Shorebird Migration Monitoring data	19	Published Indices
Waterfowl Breeding Population and Habitat Survey in Western Canada and the Northwestern United States	16	Published Indices
Seabird Colony Monitoring Program - Atlantic Ocean	12	Raw Data
Lincoln Estimators of Continental Population Size	5	Published Indices
Seabird Colony Monitoring Program - Pacific Ocean	4	Raw Data
Eastern Waterfowl Breeding Ground Survey (helicopter only data)	2	Published Indices
Great Lakes Marsh Monitoring Program	2	Published Indices
American Woodcock Singing-Ground Survey	1	Published Indices
North American Breeding Bird Survey (Continental Estimates)	1	Raw Data
Seabird Colony Monitoring Program - Arctic Ocean	1	Raw Data
Eastern Waterfowl Breeding Ground Survey (Integrated Estimates)	1	Published Indices
Piping Plover Survey	1	Published Indices
Southern Ontario Ground-based Plots Survey	1	Published Indices
North American Trumpeter Swan Survey	1	Published Indices
Wild Population of Whooping Crane Survey	1	Published Indices
Mid-winter Tundra Swan Counts (sum of eastern and western populations)	1	Published Indices
No acceptable data available	107	NA
Total	449	NA

We excluded data for 9 species that have expanded their ranges into Canada since 1970, because of their undue influence on our indicator values. For these range-expansion species, estimates of the percent-change in population size since 1970 are extreme (theoretically infinite) and difficult to estimate accurately, because our best estimate of their Canadian population in 1970 is approximately 0. Because our indicators represent averages of percent change values across species, these extremely large values overwhelm the influence of the other bird species in the group. These 9 range-expansion species include: Wild Turkey, Anna's Hummingbird, Black-necked Stilt, Great Egret, Red-bellied Woodpecker, Bushtit, Carolina Wren, Blue-gray Gnatcatcher, and Blue-winged Warbler. On average according to our best estimate of their trajectories, the species in this group have increased by > 3000% since the first year each species occurs in the available data. Although each one of these species represents a real and potentially important change in Canada's bird populations, their estimates of percent change in population size do not represent the same biological or conservation-relevant processes as all other species (e.g., most of these range expansions reflect responses to human land use change, climate change, or purposeful re-introductions). In addition, if included in a group with other species, their extreme rates of increase would mask the other important patterns of change for the rest of the species in the group. We have included the raw data for these species in the supplementary data files, so interested users can explore the patterns of population change.

SMOOTH POPULATION TRAJECTORIES

To reduce the influence of varying statistical models used to produce published estimates of each species' population trajectory (i.e., the series of annual estimates of relative abundance) and to focus our analyses on long- and medium-term patterns in population change (i.e., our primary interests are not related to annual fluctuations), we applied a similar non-parametric smoothing model to all species trajectories. There are important differences among the underlying statistical models used in many of the regularly published estimates of species-level monitoring data. For example, the North American Breeding Bird Survey (BBS) and the Christmas Bird Count (CBC) use models that incorporate a log-linear regression approach smoothing the time-series of annual estimates. As a result, estimates from these models are particularly well designed to estimate long-term trends (i.e., rates of change) but not as well designed to estimate more recent changes in trends (e.g., change points in trends, see Smith et al. 2014 for an example). Similarly, estimates from the annual waterfowl breeding-ground surveys use models that do not smooth the time-series at all; instead, they estimate each year's estimate independently of all other years. This structure is particularly well suited for modeling the annual fluctuations in waterfowl populations, but can allow estimates of annual abundance to vary much more than many other models.

We used Generalized Additive Models (GAM) to either smooth the published estimates of a species annual abundance (51 species from 12 data sources, Table 1), or to re-analyse the underlying raw data from a given survey (292 species from 7 data sources, Table 1). The Living Planet Index (Collen et al. 2009), a comparable group-level indicator of mean population status across species, also uses GAM smooth of published annual indices in a similar way. The GAM smooth allowed us to accommodate the wide variation in the underlying population trajectory estimates across the various datasets. The flexible smoother retains the most important medium- and long-term patterns in the species' population, interpolates estimates for years with missing data for a given species, and reconciles the level of annual variation among species. One of the major benefits of this approach to modeling species population trajectories is that the GAM structure is capable of modeling almost any medium- and long-term pattern. Estimates can be a flat-line if the species data support such a pattern or if the species data are extremely noisy and uncertain. Alternatively, estimated trajectories can follow as many complex, non-linear or cyclical patterns as the underlying data support.

For most species, we reanalysed the raw monitoring data with a hierarchical Bayesian status and trend model that used a GAM smoothing function to estimate the temporal parameters (i.e., the change over years). The GAM reanalysis was specific to each of the three primary survey types within the seven data sources (i.e., 2 sources using the CBC, 2 sources using the BBS, and 3 sources using Seabird Colony Counts, Table 1) and each survey's model accounted the same survey-specific sources of variation that have been used in previous analyses (NABCI 2012, Smith et al. 2014, Meehan et al. 2018).

For the BBS and the CBC, the GAM reanalysis included exactly the same parameters and prior distributions as the standard models used for the BBS by the CWS (Smith et al. 2014) and USGS (Sauer and Link 2011), and the Audubon Society for the CBC (Meehan et al. 2018), except the random year-effects and log-linear slope parameters were replaced with the GAM function. For the seabird colony count analyses, the GAM function was used to model changes over time at each monitored colony, while accounting for the size of the colony and the years in which each colony was monitored (following the analyses conducted for NABCI 2012).

All changes in abundance through time were modeled by the GAM smooth function ($f_i(t)$). In these models, the i subscript indicates that the GAM function modeling changes in time (i.e. the shape of the species population trajectory) is specific to a geo-political stratum (the BBS and CBC use the intersection of Bird Conservation Regions and provinces, states, and territories) or a particular monitored colony in the seabird analyses. Because of the hierarchical structure of the model, the smoothing functions for each colony or geo-political stratum were shrunk towards a species-level mean smoothing function. This allows the model to share information among strata or colonies on the shape of the overall trajectory of a species population. Like all hierarchical models, this shrinkage is only relevant to the extent that it is supported by the data: in colonies or strata with more data and a clearer signal of population change, the hierarchical shrinkage is minimal.

To estimate the GAM parameters, we used the parameterization of Crainiceanu et al (2005).

$$f_i(t) = \sum_{k=1}^K b_{i,k} \chi_{t,k}$$

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 $b_{i,k}$ is the K -length vector of parameters that control the shape of the trajectory in stratum- i . Each $b_{i,k}$ is estimated as a random effect, centered on a hyperparameter: a mean across all strata (β_k).

$$b_{i,k} = \text{Normal}(\beta_k, \sigma_b^2)$$

The mean smoothing parameters (β_k) are themselves estimated as normally distributed random effects with a mean of zero, and variance σ_β^2 controlling the shrinkage towards a first-degree polynomial (i.e., a line). The variance parameters ($\sigma_\beta^2, \sigma_b^2$) control the complexity penalty of the smooths.

The design matrix (X) has a row for each year, and a column for each of K knots. The matrix was calculated following Crainiceanu et al (2005) so that the GAM function represented a 3rd-degree polynomial spline: $\chi_{t,k} = |t - t_k|$. For the CBC and BBS models, with 47-year time-series, we set the number of knots to 12, which means the trajectories are sufficiently flexible to capture all but the shortest-term variation (i.e., long- and medium-term variation but not yearly fluctuations). We explored models with more and fewer knots (9-14), but the estimated trajectories from the models with differing numbers of knots were almost indistinguishable from those presented here. We set the knots at even intervals along the time-series. For the Seabird analyses, most species are not monitored on an annual basis so the GAM functions included fewer knots. The number of knots varied among species (range 2 – 8), depending on the monitoring schedules. Species with colonies that were monitored more often (e.g., Northern Gannet) had more knots, those monitored less often (e.g., Leach’s Storm Petrel) had fewer.

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

$$\sigma_b^2 = 1 / \text{gamma}(10^{-2}, 10^{-4})$$

$$\sigma_\beta^2 = 1 / \text{gamma}(10^{-2}, 10^{-4})$$

The above description of the hierarchical Bayesian GAM reanalyses provides the basic structure of the GAM components. Full descriptions of the reanalysis of the BBS, CBC and seabird monitoring data are in preparation for publication in scientific journals. The full hierarchical Bayesian GAM for the BBS is available online through an R-package “bbsBayes” available on GitHub (<https://github.com/BrandonEdwards/bbsBayes>). In addition, the population trajectories for all species used in these State of Canada’s Birds analyses are available for download from the report website.

For some species and data sources it was not practical to re-analyse the raw data and so we smoothed published trajectories of annual indices using the R-package mgcv (Wood 2017). With this R-package, we used a hierarchical Bayesian GAM that accounted for the uncertainty of each annual index in the trajectory to model most species, and for the few species where published estimates of uncertainty were not available we used a simpler non-Bayesian GAM function from the same package.

CALCULATING THE INDICATORS

Standardizing across species and data sources

The data for the final models used to calculate the composite species indicators in the State of Canada’s Birds were the annual indices for each species ($i_{s,y}, \sigma_{s,y}^2$), either from the GAM reanalyses or the GAM smooths of published trajectories. Before they were entered into the final model, each species’ trajectory was re-scaled to a common base-year (1970 for most species) and log-transformed. This was necessary because the original annual indices from each data source were in different units, e.g., BBS estimates represent the average number of birds seen on a single route and CBC estimates represent the average number observed in an average count-circle. In the final model, each species trajectory must be on a comparable scale, to allow for the hierarchical structure of the model that pools information across groups of species (e.g., grassland birds). For some species (15%) the available data did not span the full 1970 –

2016 time-series (Figure 2). For these species, the first year data were available was used as the base-year, and all years previously were given a value equal to the base year, but with an arbitrarily large variance so that these missing data had very little influence on the final result (more detail on this below). The re-scaled annual indices ($\hat{\theta}_{s,y}$) were the data for the final model.

$$\hat{\theta}_{s,y} = \ln\left(\frac{i_{s,y}}{i_{s,1970}}\right)$$

Where, $\hat{\theta}_{s,y}$ is the log-transformed standardized annual estimate for year y and species s ($i_{s,y}$) and represents the status of the species in year- y , as a log-transformed proportion of the original estimate in the base-year, 1970 ($i_{s,1970}$).

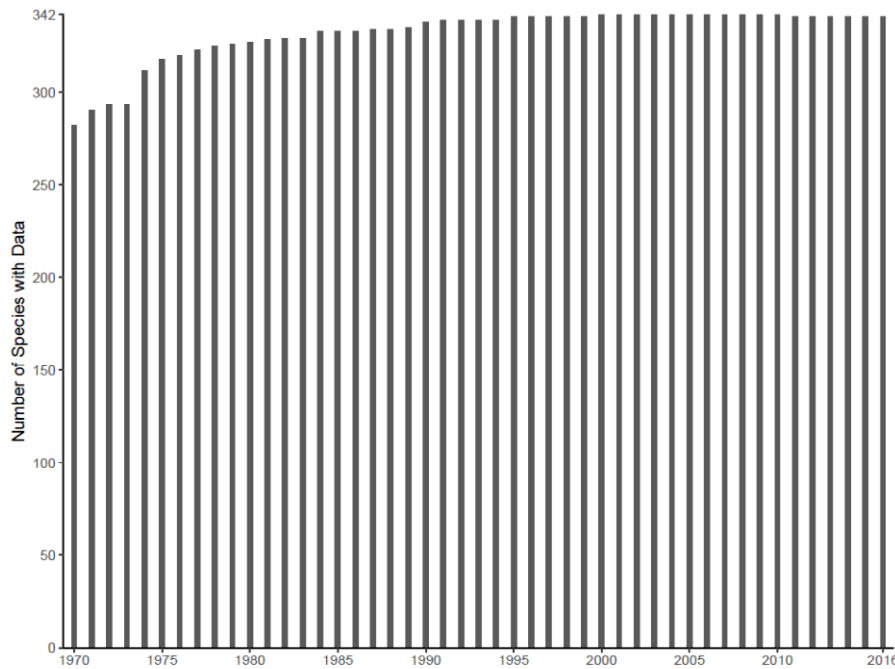


Figure 1. The number of species in each year, with annual estimates of relative abundance suitable for inclusion in the analyses for the State of Canada’s Birds indicators. Approximately 85% of species have trajectories that span the full 1970 – 2016 time-series, and by 1980, 95% of species have data.

Accounting for relative precision among species and years

Species-specific estimates of population status vary in their precision among species and among years. To calculate the indicators in the State of Canada’s Birds, we used a model that explicitly considers the relative precision of each species’ estimate of population status. Accounting for the relative precision ensures that extreme values for species with relatively imprecise estimates of status do not dominate our indicators. Precision is defined as the inverse of variance (high precision = low variance and *vice versa*) and in most statistical literature, all of the equations below and the data sources used in our analyses variance is the quantity used. Hereafter, both terms are used, but the reader should be aware that conceptually they are measures of the same thing—uncertainty.

We estimated the variance around each re-scaled species-specific annual estimate in a way that accounts for the variance in both a given year (year- y) and in the base-year (i.e., 1970 for most species). We calculated the variance of $\hat{\theta}_{s,y}$ using an approximation of the log transformation of the variance of a ratio of two random variables (Cochran 1977, pg. 183).

$$\sigma_{\hat{\theta}_{s,y}}^2 = \ln\left(1 + \frac{\sigma_{i_{s,y}}^2}{i_{s,y}^2} + \frac{\sigma_{i_{s,1970}}^2}{i_{s,1970}^2}\right)$$

It is an approximation because we made the simplifying assumption that the annual estimates are independent in time, and therefore that the covariance term equals 0 and can be ignored. We acknowledge that this assumption of independent estimates in time is certainly invalid for adjacent years, but becomes more plausible as the length of the time-series increases.

For the species with missing data in years at the start or end of the time series, we assumed that the population did not change during the missing years. Years with missing annual indices (e.g., no data before 1974 for many shorebird species) were given values equal to the first year with data (i.e. a conservative assumption of no overall change) but we increased the estimated variance ($\sigma_{\theta_{s,y}}^2$) by the square of the number of years since non-missing data, so that these imputed data would have little overall effect on the final results. Because of the extremely high variance and the hierarchical structure of the model (see below), the population trajectories for these species have very little influence on the group-indicator during the years for which data are missing.

Statistical model for the indicators

For each year, the standardized individual species estimates ($\hat{\theta}_{s,y}$) were combined into a composite indicator (P_y) using a Bayesian hierarchical model, first described in Sauer and Link (2011). The model generates a Bayesian estimate of the geometric average population status across all species in the composite species group, which is consistent with other approaches to composite species indicators (e.g., Collen et al. 2009 and Gregory et al. 2005). It also accounts for the varying precision of each species' estimate, so that imprecise estimates have less influence on the indicator. The composite indicators represent our best estimate of the group's status, given the distributional assumptions of the model, the species for which we have data and the precision of those data.

As an example, the grassland birds indicator is a compilation of the annual estimates for 25 grassland bird species (Figure 2). The indicator line (thick black line with grey 95% credible interval) can be roughly interpreted as the average of the 25 species-specific lines (coloured lines), where each of the species lines influences the indicator to a degree dependent on its precision. Hence, the grassland bird indicator runs approximately through the centre of the species lines, but is not simply the mean of all species lines (e.g., the extreme fluctuations in the Lark Bunting line have relatively little influence compared to a simple average because its annual estimates are generally less precisely estimated than those for most other species in this group).

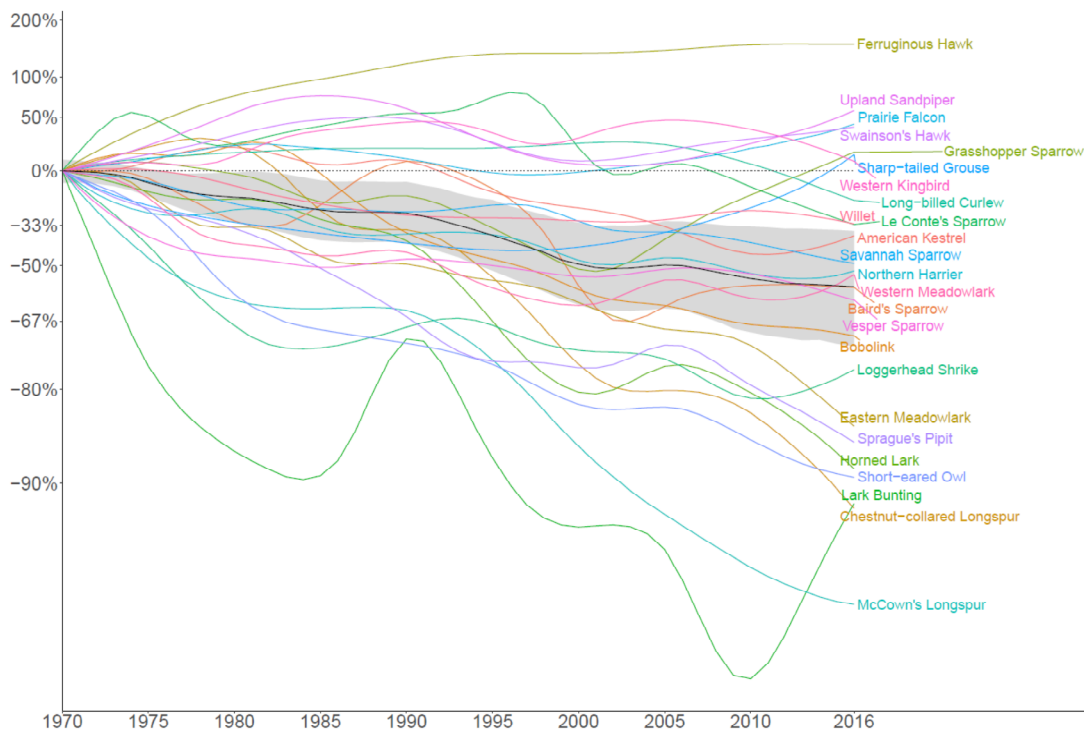


Figure 2. National grassland birds sub-indicator (thick line), 95% credible interval (grey uncertainty band) and lines showing the species-specific annual estimates of population status (percent change since 1970) for the 25 species included in the grassland birds group (coloured lines).

Formally, following Sauer and Link (2011), the composite indicator of the percent change in species populations in a given year (P_y) is the median of the posterior distribution of the hyperparameter ρ_y , transformed to a percent-change scale. The hyperparameter ρ_y is the average across species of the normally distributed species-specific change parameters ($\theta_{s,y}$). The averaging accounts for the variance of each species annual change parameters ($\sigma_{\theta_{s,y}}^2$) and the variance among species in their annual change values ($\sigma_{\rho_y}^2$).

$$\hat{\theta}_{s,y} = N[\theta_{s,y}, \sigma_{\theta_{s,y}}^2]$$

$$\theta_{s,y} = N[\rho_y, \sigma_{\rho_y}^2]$$

$$P_y = ((e^{\rho_y}) - 1) * 100$$

Scaling of the indicator graphs

The vertical axes were scaled to be symmetrical on the log-scale (i.e., symmetrical on the scale of ρ_y) and accurately reflect the change in population required to balance out the opposite decrease or increase in each corresponding category. Values of percent change are not symmetrical around zero. For example, a population that has declined by 50% (i.e., halved) must then increase by 100% (i.e., double) to return to its original level. This asymmetry increases in a non-linear way for greater values of percent-change, because positive values of percent-change are theoretically unbounded (i.e., it is possible for a population to increase by values much larger than 100%) but negative values are limited at -100% (100% decrease is equal to extinction). For example, to recover from a 90% decrease, a population would then have to increase by 900%.

Variation among species

To depict some of the variation among individual species' trends within each indicator line, we presented info-graphics composed of groupings of coloured bird icons showing the number of species in each of 3 categories of long term (~47 year) changes in population (Table 3). The categories were defined to be symmetrical in the log-scale and respect the same asymmetry as the scaling of the vertical axis. The variation depicted in these info-graphics highlights the appropriate interpretation of these indicators—as indicators of the average or overall status of the group, not as indicators of the status of each species within the group. Our indicators give the best overall estimate of the group's status, but do not reflect the trends for all species in a group equally well; a stable indicator may reflect a group in which most or all species have stable trends, or it may reflect a group with an equal number of species with large increases and large decreases. Almost all of the indicators in the report, regardless of their overall pattern of change, include both species that are increasing and species that are decreasing. For example, populations of grassland birds as a group have decreased, but not all grassland bird populations have decreased (Figure 2).

Table 3. Categories of long-term population change used to group species for the coloured bird icons in the State of Canada's Birds.

Long-term population change category	Range of estimates of total population change over a 47-year period	Colour of bird icons indicating species trends
Increase	> 33% increase	Green
Little to Moderate Change	33% increase – 25% decrease	Blue
Decrease	> 25% decrease	Red

REFERENCES

- Cochran, W.G. 1977. *Sampling Techniques*. 3rd Edition. John Wiley & Sons, New York, USA. 448 pp.
- Collen, B., J. Loh, S. Whitmee, L. Mcrae, R. Amin, and J.E.M. Baillie. 2009. Monitoring Change in Vertebrate Abundance: the Living Planet Index. *Conservation Biology* 23:317–327.
- Crainiceanu, C.M., D. Ruppert, and M.P. Wand. 2005. Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, 10.18637/jss.v014.i14
- Environment and Climate Change Canada (ECCC). 2019. [The Status of Birds in Canada Website](#), Data-version 2017. Environment and Climate Change Canada, Gatineau, Quebec, K1A 0H3.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:515-533.
- Gregory, R. D., A. van Strien, P. Vorisek, A. W. G. Meyling, D. G. Noble, R. P. B. Foppen, and D. W. Gibbons. 2005. Developing indicators for European birds. *Philosophical Transactions of the Royal Society of London B* 360:269–288.
- Meehan, T.D., G.S. LeBaron, K. Dale, N.L. Michel, G.M. Verutes, and G.M. Langham. 2018. [Abundance trends of birds wintering in the USA and Canada, from Audubon Christmas Bird Counts, 1966-2017](#), version 2.1. National Audubon Society, New York, New York, USA.
- North American Bird Conservation Initiative Canada (NABCI). 2012. *The State of Canada's Birds, 2012*. Environment Canada, Ottawa, Canada. 36 pages.
- Sauer, J.R. and W.A. Link. 2011. Analysis of the North American Breeding Bird Survey Using Hierarchical Models. *The Auk* 128:87-98.
- Smith A.C., M.-A.R. Hudson, C.M. 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.
- Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R*, 2nd edition. CRC Press, Taylor & Francis Group. London UK. 476 pp.