

Alternate composite models

Goal

To compare alternative models for estimating the composite species indicators used in the State of Canada's Birds, i.e., the "Summary Models" for the State of Canada's Birds. These indicators are intended to represent the mean status of species within a group through time; e.g., the mean overall change in grassland bird populations since 1970, for all years from 1970 through the present.

Conceptual steps

The modeling effort needs to start with estimated population trajectories for each species, smooth the population trajectories, calculate the index values to support a given summary model, fit the summary models, and plot the resulting indicators.

Load estimated trajectories

Here, just using output from the BBS as an example.

```
all_bbs_inds <- read_csv("c:/github/CWS_2022_BBS_Analyses/website/BBS_2022_annual_indices_
  filter(area_name == "Canada",
         period == "Long-term")

saveRDS(all_bbs_inds, "data/Canadian_BBS_indices.rds")
```

Selecting out the grassland bird annual indices of abundance from the BBS. Then log-transforming the index values and calculating a log-scale standard error of the indices by transforming the credible limits and approximating the standard error as the credible interval width divided by 3.9 (assuming an approximately normal 95 percent CI).

Constant number of species

The code below generates the input data for the smoothing model and the *Constant* scenario, where the number of species in the indicator is constant across all years because each species has data in all years.

```
grassland_inds <- bbs_canadian_indices %>%
  filter(species_name %in% grassland_sp,
         !is.na(index)) %>%
  mutate(ln_index = log(index),
         ln_lci = log(lower_ci),
         ln_uci = log(upper_ci),
         ln_index_sd = (ln_uci-ln_lci)/3.9) %>%
  group_by(species_name) %>%
  mutate(yearn = year-(min(year)-1)) %>% # sets a yearn value specific to each species
  arrange(species_name,year)

saveRDS(grassland_inds,"data/grassland_inds.rds")
```

Variable number of species

To demonstrate how each model handles missing data, we can create an identical dataset where five of the species do not have monitoring data across the entire time-series. The code below removes the first 20-years of data for five of the species in the grassland birds indicator. These data form the raw data for the *Variable* graph facet in Figure 1.

```
grassland_inds <- bbs_canadian_indices %>%
  filter(species_name %in% grassland_sp,
         !is.na(index),
         !((species_code == "AMKE" & year < 1990) |
            (species_code == "BOBO" & year < 1990) |
            (species_code == "CCLO" & year < 1990) |
            (species_code == "VESP" & year < 1990) |
            (species_code == "GRSP" & year < 1990))) %>%
  mutate(ln_index = log(index),
         ln_lci = log(lower_ci),
         ln_uci = log(upper_ci),
         ln_index_sd = (ln_uci-ln_lci)/3.9) %>%
  group_by(species_name) %>%
  mutate(yearn = year-(min(year)-1)) %>% # sets a yearn value specific to each species
  arrange(species_name,year)
```

```
saveRDS(grassland_inds,"data/grassland_inds_w_missing.rds")
```

Smooth indices accounting for uncertainty

We want the composite indicators to track long- and medium-term changes in populations that are shared across many species in a group. Sharp annual fluctuations may represent true variation in the species' population, but they can mask the longer term changes that are primary interest of this high-level summary. Following previous State of Canada's Birds reports, we use a generalized additive model (GAM) that accounts for the uncertainty of each estimated annual index of abundance with a measurement-error structure. Fitting this GAM for each species generates smooth population trajectories. Within the GAM, we can estimate the log-scale input data for the two composite summary models. The code is not shown here, but the process involves log-transforming the annual indices of abundance for each species and year, then fitting the smooth GAM to each species separately, and then calculating the input data representing the change since the species first year of data for the standard model and the series of annual differences to generate the input data

Apply the composite summary models

There are two conceptual models to compare here:

Standard model

The Standard model has been used in previous SoCB reports and estimates the annual values of the mean change since the first year of data across species. Data are the log-scaled indices of annual status relative to a base-year (usually 1970) for each species. The model estimates the mean change since the base-year across all species, and independently in each year.

The differences model

The Difference model estimates the cumulative change over time by estimating the annual mean difference from the previous year across species, and then summing those annual mean differences through time. Data are the time series of log-scale annual differences for each species; i.e., the difference between the species' index in year(t) minus the index in year($t-1$). The model estimates the mean across all species of those annual differences for each year, and then estimates the time-series of the mean composite indicator by fixing the value of the first year at 0, then summing the mean annual differences through time.

Compare model outputs

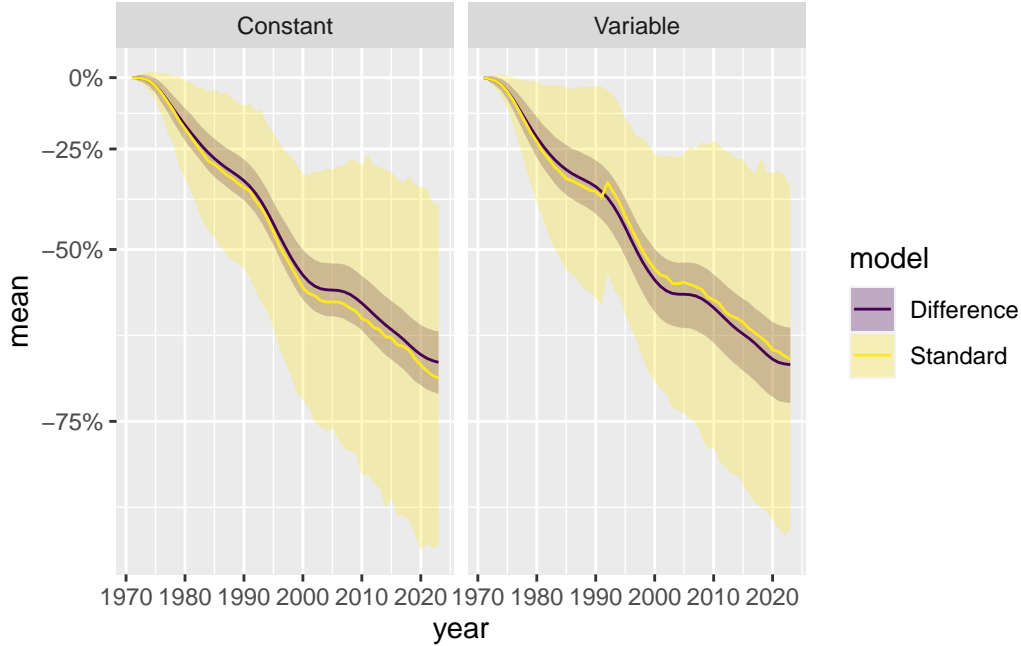


Figure 1: Comparison of two model for calculating the composite group indicator trajectories for State of Canada's Birds (SoCB). The facet on the left demonstrates that the models generate very similar mean trajectories when all species have data in all years. The facet on the right demonstrates the effect having a variable number of species through time. In the right facet, the data for five species from the first 20 years have been removed, and so their data enter the model in 1991

The standard model and the cumulative difference model generate very similar mean trajectories when the number of species is constant across all years (left panel, Figure 1). By contrast, the step-change in year 21 in the standard model results when the number of species varies through time (right panel, Figure 1) reflects the addition of the five species with missing data until 1990. Because these species have no data from before 1990, it is not possible to estimate a change since 1970. As a result, data for these five species represent the change since 1990 (i.e., in 1991, these species' input data reflect change over only 1 year, and so pull the mean indicator value back towards 0). In the cumulative difference model, there is no abrupt step-change because these species' input data in 1991 reflect the same metric as all other species: the change since 1990.

The treatment of uncertainty is worth revisiting for these models, as they clearly estimate different aspects of the uncertainty. On a related note, the entire assessment could be combined into a single hierarchical GAM, where the input data are the species annual indices and their SE. A model with species-level intercepts, hierarchical smooths for each species, and a global

mean smooth, should generate the same overall mean trajectory. This treatment could more easily propagate the uncertainty, and could generate smoothed species-level trajectories as well as explicit predictions for a hypothetical “new” species (or a species with missing data).