**Steps for basic analysis of daily migration counts, using multiple regression.**

Version 2009

Erica Dunn, Environment Canada, Ottawa Ontario (Erica.Dunn@ec.gc.ca)

David Hussell, Ontario Ministry of Natural Resources, (David.Hussell@ontario.ca)

Introduction

Here we present step by step instructions for basic analysis of migration counts. All that is required is a spread sheet program, and access to a statistical package that can handle multiple regression or ANOVA. At least two inexperienced analysts have successfully analyzed their data using these instructions, and the authors will be happy to work with others who are interested.

This document is mainly a cook-book for data analysis, and it contains very little explanation of why the different steps are necessary. The instructions may be more understandable if it is understood that the main problem with daily migration counts is their skewed distribution (many small counts and relatively few large ones). This distribution is caused by weather, and by seasonal variation in migration volume. As is true of many data that have a skewed distribution, the precision of the large counts is less than for the small counts. This comes partly from errors in counting the large flights, but also from unexplained variation in the large counts (due to unusual combinations of weather factors or other variables). It may seem appropriate to estimate annual indices by simply adding up all birds detected over the season; however, the errors in the big counts tend to swamp all other effects. Instead, what we need to do is try to detect overall annual effects on *all* of the daily counts, not just on the big ones. Log-transformation of the daily counts before calculating means is the single most important thing that can be done, because it tends to reduce the skew in the data and reduce the impact of the big counts and their errors. After doing that, additional variation can be explained by taking into account date and weather variables. The analysis steps given below include log transformation and adjustment for date in the season, but no adjustment for weather effects.

For more information on migration monitoring, its purpose, strengths, limitations and analysis, readers are referred to the following papers:

(1) Dunn, E. H. 2005. Counting migrants to monitor bird populations: state of the art. Pp. 712-717 *in* Bird Conservation and Implementation in the Americas: Proceedings of the Third International Partners in Flight Conference, vol. 2, C. J. Ralph and T. D. Rich, eds. USDA For. Ser. Gen. Tech. Rep. PSW-GTR-191. Albany, CA. Available online at http://www.fs.fed.us/psw/publications/documents/psw_gtr191/psw_gtr191_0712-0717_dunn.pdf.

(2) Francis, C.M., and D.J.T. Hussell. 1998. Changes in numbers of land birds counted in migration at Long Point Bird Observatory, 1961-1997. Bird Populations 4: 37-66.

Analysis instructions

1) Set up your data file in a spread sheet. Each row will have the data for a single day, and there should be columns for species, year, julian date, daily effort (e.g. total net-hr for

netting data, or hours of observation for count data), daily count, and daily count corrected for effort (e.g. birds/100 net-hr, or birds/1000 net-hr. Pick an effort-adjustment factor that will give you numbers that are >1, rather than very small numbers.) Make sure you include all days of operation, putting "0" in the count column for days when no birds were captured.

2) Determine migration 'window' for each season and species separately (the dates within which 95% of the birds are likely to pass through your area).

Crude method: Plot all counts for a species on date (all years combined), and picking a start and end date for migration by eye. Try to define the period in which the middle 95% of birds pass through, cutting out the 'tails.'

Quantitative method: Use a statistical package to define the 5th and 95th percentiles of dates with positive counts. (If there are local birds present before or after the migration season, this method can only be used if you first cut out data that clearly represent local birds.)

Note: If the species breeds or winters in your area, local residents will overlap with migrants. The analysis approach assumes that birds counted each day are new arrivals. Of course this isn't always true, but you should exclude local birds to the extent possible in order to meet assumptions of the method as closely as possible. One way is to exclude the portion of the migration season from analysis which is likely to overlap with the season in which local birds are present (e.g. if there are locally breeding birds that leave the area by a known date, exclude from analysis the portion of the fall migration season prior to that date). If local birds are present throughout migration season, take the average number of birds present before and after the migration window, and subtract that number from each daily count within the migration season. (However, such species are better left unanalyzed.)

3) Select species for analysis. You should only analyze species for which, within the migration window, you have an average annual total of 25 birds detected, and which were detected on at least 5 days/year. It's a good idea to exclude species for which you conducted counts on fewer than 75% of the days in the species' migration window. (It doesn't matter if the species was not detected on many of those coverage days.) If you conducted counts on fewer than 10 days within the migration window in any individual year, that year should be excluded from the final trend analysis (Hussell et al. 1992). If more than a few years would be eliminated using this rule of thumb, you should exclude the species from analysis entirely.

Species that are present in the count vicinity as breeders are poorly suited for analysis of migration counts, and preferably should be omitted (but see note above on selecting migration windows).

4) Calculate the middle date of the migration window. Construct a new "Date" variable (new column of data) which is calculated as follows: (julian date)-(julian date for middle day of the migration window). This data transformation sets the middle day to 0, with earlier dates having negative values and later dates having positive values. Similarly, calculate a new "Year" variable by taking a reference year within your series (e.g. middle year) and subtracting this from all other years. (That is, if reference year is 1989, new Year variable

for 1989=1989-1989, new Year variable for 1990=1990-1989, etc.) These steps improve statistical tolerance in the regressions to come.

5) In new columns (one for each year), create new variables: $Date^2$, $Date^3$, $Year^2$, and $Year^3$. (That is, take the Date variable constructed in step 4 and square it to make variable $Date^2$, etc.)

6) In new columns, add data for dummy variables for year (but see note to step 10). E.g. for the year 1989, the variable YR89 will have a value of 1, and in other years the variable YR89 will have a value of 0. Make a dummy variable for each year except one (which will be the reference year in the analysis). It doesn't matter which year is the reference, but it's best to choose one with lots of data.

7) In new column, add data for new variable, calculated as ln (daily count + 1), in which 'daily count' is the effort-corrected count (see #1). (You add 1 before taking the natural log, because you can't take the log of 0). Hereafter this log-transformed variable is referred to as the dependent variable "Count".

8) In a new column add data for new variable 'weight,' calculated as N x (daily effort) / (sum of all daily effort values), where N is the number of days of counts (across all years), daily effort is net-hr for that day (in the case of banding data), and the sum of daily effort is the sum across all years. This variable is used to weight regressions in the next steps.

9) For each species and season separately, run a weighted regression (read notes below first) in which Count is the dependent variable and the independent variables are Date, $Date^2$, Year, $Year^2$ and $Year^3$. Weight each case by the weigh variable (see step 8). Output will describe a curve that describes the seasonal pattern, adjusted for any long-term trend in the data. Look at the unstandardized predicted values that are produced by the output, and you may find that some dates early and late in the season of particular years have predicted counts of less than 0. Cut these specific cases out of your data file for the remainder of analyses, and recalculate the weights (step #8) for the reduced data set.

Note: If you have done a good job of defining the migration windows, and used a 95% cutoff to eliminate "tails" of the season, there may be no need to take this step (i.e. there will be no cases with predicted values less than zero).

This regression helps refine the migration window. Variables for year are added to the regression in order to adjust appropriately for any long-term trend in abundance. The reason for adding polynomial terms for year is to describe any non-linear trends in numbers.

Weighting is important if your daily counts vary widely in effort. Weighting by daily effort will cause the regression output to show more degrees of freedom than are appropriate, which is why the 'weight' variable is calculated in step 8. If daily effort is standardized, there is no need to weight the regression.

10) Run a weighted multiple regression with Count as the dependent variable and the independent variables are Date, $Date^2$, $Date^3$, YR89, YR90, etc.—adding in all your dummy

variables for year except the one for the reference year (see #6). The data you want to extract from the results are the intercept, and the coefficients for the independent variables. Calculate the initial index for the reference year as: intercept + (Date coefficient)(mean Date) + ($Date^2$ coefficient)(mean $Date^2$) + ($Date^3$ coefficient)(mean $Date^3$) + (½ the variance that was estimated by the regression, or "MS residual"). [The "means" here are the weighted means for each variable for all cases (days) used in the regression, using the same weights as in the regression. You add ½ the variance in order that the final indices (see below) represent the mean number of birds, rather than the median.] Calculate the initial index for other years as: initial index for reference year + (coefficient of dummy variable for the year in question).

Note: Analyses can be run in ANCOVA (which would give the same result as multiple regression), in which case you don't need to make dummy variables for year (step 6). GLM procedures make the dummy variables for you automatically. In ANCOVA output, the variance used in back-transformation of indices is the 'MS error.'

11) Take the initial indices from the step above (all years, including reference year), then back-transform them by exponentiating and subtracting 1. You have to back-transform in order to get back to real bird numbers, because the regression was done using log(daily count+1).

12) Set up new data file for trend analysis, with columns for year, annual index (from step 11) and annual weighting factor (calculated for netting data as (n years in analysis) x (annual total of net-hr) / (sum of annual net-hr totals across all years).

13) To calculate long-term trends in the data, first transform the annual indices by taking the natural log. (You don't have to add 1 before taking the log, as for the regressions above, because all annual indices are likely to be greater than 0.) Next, do a weighted regression (using annual weights from step 12) in which ln(index) is the dependent variable and Year is the independent variable. Take the coefficient of the Year variable in the output, and multiply by 100. The result represents the annual % change in population (linear change in the log scale).

Note: The trend produced in this analysis is in the same terms as those produced by the Breeding Bird Survey, so you can compare your trends directly to BBS trends in the region from which you think your birds originate. (See http://www.mbr-pwrc.usgs.gov/bbs/ if you want to get BBS trends for a specific region and set of years. Follow the links to "Regional Trend Analysis," on the home page listings under "Analytical Tools.") If you have >15 years of data, this linear analysis may not be appropriate. If your annual indices are not more or less linear, contact the author for additional instructions.

14) In interpreting your results, consider the following.

Weather is known to affect the number of birds present on a given day, as is moon phase. The regression which produces annual indices is based on log-transformed data, which helps to reduce the distorting effects of a few days in the season when there are unexpectedly large numbers of birds (usually caused by weather-induced fall-outs).

Weather effects may be especially important on oceanic coastal sites that are on the edges of migration routes, such that wind direction may have a particularly large effect on bird numbers. Any long-term trends in weather patterns could cause a change in bird numbers that is unrelated to true population change. However, frequency of different weather conditions is probably random over time, and the main effect of weather variation on annual indices may be that it will take more years of data collection before a change in population can be detected. Some of this “noise” in the annual indices can be removed statistically, if you are willing to go to the extra trouble of getting weather data and doing a more complicated analysis. If you are interested in doing that, contact the authors.

Vegetation change at the count site will also affect numbers of birds detected each day. If birds are being captured, vegetation growth will alter the capture probability of birds over time. If numbers banded are the daily counts, then growth of vegetation relative to net height may alter numbers captured, even though population has not changed. Similarly, vegetation growth can affect numbers of birds counted visually. Unlike weather effects on bird numbers, vegetation change is nearly always directional over time, and is likely to be the most important potential source of bias in migration monitoring. To the extent possible, vegetation should be controlled at long-term migration monitoring sites in order to avoid this kind of bias.