**Breeding Bird Monitoring Protocol for the Heartland Inventory and Monitoring Network**

**Standard Operation Procedure 7: Data Summary and Analysis**

**Version 3.00 (06/03/2019)**

**Revision History Log:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Previous Version # | Revision Date | Author | Changes Made | Reason for Change | New Version # |
| 1.01 | May 2008 | D.G. Peitz | Entire document | Edited to reflect that bird monitoring has been expanded to ten additional network parks | 2.00 |
| 2.00 | May 2018 | D.G. Peitz | Updated throughout to NRR format. Updated analysis to reflect those more appropriate for long-term data sets. | Made SOP NRSS compliant. As the project transitions into one with long-term data records, more appropriate trend analysis is needed. Removed annual data summary and reporting requirement and supporting Appendix. | 3.00 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

This Standard Operating Procedure gives step-by-step instructions for analyzing variable circular plot bird count and habitat survey data collected at parks within the Heartland Inventory and Monitoring Network. This Standard Operating Procedure also outlines steps to preform to complete “Trend Analysis,” which is to be completed once every fourth year. Trend Analysis evaluates changes in bird population sizes over time, as well as how closely bird trends in a park reflect those seen in the larger bird conservation region(s) in which the park is located. In the Trend Analysis section of this Standard Operating Procedure guidance on summarizing habitat variables is provided as well as suggestion on approaches for exploring bird-habitat relationships.

**Annual Data Summary and Analysis**

Bird Community Data Summaries

Prior to annual summary analysis, the residency status (migrant, permanent resident, summer resident, winter resident, etc.) of each bird species recorded needs to be determined and reported ((see Supplemental File: GWCAData2(SpeciesList-5minObservations).xlsx for example)). Only permanent and summer residents should be included in analyses. Proportion of plots occupied by each bird species (total number of plots occupied by a species/plots surveyed) observed in a 5-min survey is calculated and reported. Park-wide abundance is calculated and reported for each species by first deriving a density from observations recorded within a radius around each plot center equaling half the distance between plots (Table 7.1), and then calculating abundance based on average plot densities (see Supplemental File: GWCAData4(Density-Abundance).xlsx for example analysis)).

|  |  |  |
| --- | --- | --- |
| **Table 7.1**. Sample grid sizes and analytical distance or area sampled around plots for parks in the Heartland Inventory and Monitoring Network. | | |
| Park | Grid size | Analytical distance (area sampled in ha) |
| Arkansas Post National Memorial, Arkansas | 200 m x 200 m | 100 m (3.14 ha) |
| Effigy Mounds National Monument, Iowa | 400 m x 400 m | 200 m (12.57 ha) |
| George Washington Carver National Monument, Missouri | 100 m x 100 m | 50 m (0.79 ha) |
| Herbert Hoover National Historic Site, Iowa | 100 m x 100 m | 50 m (0.79 ha) |
| Hopewell Culture National Historical Site, Ohio | 400 m x 400 m | 200 m (12.57 ha) |
| Homestead National Monument of America, Nebraska | 100 m x 100 m | 50 m (0.79 ha) |
| Lincoln Boyhood National Memorial, Indiana | 100 m x 100 m | 50 m (0.79 ha) |
| Pea Ridge National Military Park, Arkansas | 400 m x 400 m | 200 m (12.57 ha) |
| Pipestone National Monument, Minnesota | 100 m x 100 m | 50 m (0.79 ha) |
| Tallgrass Prairie National Preserve, Kansas | 400 m x 400 m | 200 m (12.57 ha) |
| Wilson’s Creek National Battlefield, Missouri | 400 m x 400 m | 200 m (12.57 ha) |

Annual analysis and reporting serve two functions: 1) a quick snapshot of the bird community observed on a park; and 2) a data quality check that allows the non-network birder an opportunity to review the data that will become a permanent record in the database. For example, the observer may be able to identify from the annual report whether species were correctly entered into the database in the numbers observed during surveys. Transcription errors can occur when data is entered into the database, so these reports serve as a vital step in the Quality Assurance / Quality Control of data.

**Trend Analysis**

Bird Analysis

The bird community variables and indices selected for trend analysis are descriptive, easily interpretable, and provide resource managers feedback to help assess management practices. Like annual reports, the proportion of plots occupied by each bird species (total number of plots occupied by a species/plots surveyed) is calculated and reported. For bird species with greater than 60 observations recorded, Distance R Package (R version 4.1.3 or later version, see Appendix A) is used to determine the park-wide abundance of each (Buckland et al. 2001, Thomas et al. 2010). A central part of the analysis in Distance is the modeling of a detection function to account for individuals present but not observed before calculating species abundance. Four candidate functions plus series expansion – Half-normal + Cosine, Hazard-rate + Cosine, Half-normal + Hermite polynomial, and Hazard-rate + Simple polynomial – should be considered in determining the detection function of each species, and the most robust models selected by Distance based on the lowest Akaike Information Criteria (AIC) values obtained. For species with fewer than 60 observations, park-wide abundance is calculated as in annual reports by first deriving a species density from observations recorded within a radius around each plot center equaling half the distance between plots (Table 7.1) and then calculating abundance based on average plot densities.

For species with adequate abundance (those with greater than 60 observations), trends are calculated by regressing abundance against survey years in the R application (Version 2.1.1) of the statistical software TRIM Version 3.54 (Pannekoek and Van Strien 2005, see Appendix B). TRIM is a program developed for the analysis of count data obtained from the monitoring of wildlife populations. It analyzes time series of counts using Poisson regression and produces estimates of yearly indices and trends. A linear trend model with changepoints is selected by a stepwise procedure. Serial correlation in count data among years and overdispersion are taken into account with this software. Although TRIM has the capacity to estimate missing data, regression analysis should be restricted to plots sampled most years. Weather, the ability to locate qualified birders each year, and time available to survey birds may result in some plots not being sampled every year. However, by restricting analysis to plots sampled in most years we analyze a consistent number of plots across years.

For reports, we obtain regional breeding bird trends for the bird conservation region that each park is located within from the Breeding Bird Survey website of the USGS Patuxent Wildlife Research Center. It is possible to determine trends for many bird species and many regions of interest for periods ranging from 1966 to two-years prior to current year by using the interactive calculator available at: <https://www.mbr-pwrc.usgs.gov/bbs/trend/tf15.html> (Sauer et al. 2017). However, a period of available data that closely matches the sampling period on each park should be chosen to maximize the accuracy of regional trend results without extrapolation. We can then compare regional trends with those calculated for the park using ggplot2 in the R statistical package.

**Example ggplot2 script in the R statistical package**. See Supplemental File: “GWCAData5(Grassland\_ParkvsRegionTrendComparison).csv” for example data set.

setwd("c:\\David\_files\\R\_practice\_folder\\2020\_GWCA")

library(ggplot2)

trendests <- read.csv("GWCAData5(GrasslandParkvsRegionTrendComparison).csv",

header = TRUE, as.is = TRUE)

head(trendests)

names(trendests) <- c("CommonName",

"GWCA.mul.trend",

"GWCA.mul.trend.SE",

"GWCA.mul.trend.LCI",

"GWCA.mul.trend.UCI",

"GWCA.AnnualPct",

"GWCA.AnnualPct.LCI",

"GWCA.AnnualPct.UCI",

"BBS.AnnualPct.Trend",

"BBS.AnnualPct.Trend.LCI",

"BBS.AnnualPct.Trend.UCI")

birds <- rev(trendests$CommonName)

trendests$CommonName <- factor(trendests$CommonName, levels = birds)

fig1 <- ggplot(trendests, aes(x = GWCA.AnnualPct, y = CommonName)) +

geom\_point(aes(color = "GWCA")) +

geom\_errorbarh(aes(xmin = GWCA.AnnualPct.LCI,

xmax = GWCA.AnnualPct.UCI,

height = 0.2,

color = "GWCA")) +

geom\_point(aes(x = BBS.AnnualPct.Trend, color = "BBS"),

position = position\_nudge(y = 0.1)) +

geom\_errorbarh(aes(xmin = BBS.AnnualPct.Trend.LCI,

xmax = BBS.AnnualPct.Trend.UCI,

height = 0.2,

color = "BBS"),

position = position\_nudge(y = 0.1)) +

scale\_color\_manual(name = '', values = c("GWCA" = "black", "BBS" = "blue")) +

geom\_vline(xintercept = 0) +

xlab("Annual Percent Change") +

ylab("") +

ggtitle("Trend Estimates for Common Grassland Species") +

theme\_bw() +

theme(plot.title = element\_text(hjust = 0.5)) +

theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank())

fig1

fig1 + theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

Regional trends with a confidence interval that straddles zero are classified as uncertain for comparison with results from the park. It should be noted that trends determined by the Breeding Bird Survey are calculated using a different methodology. Due to limitations in the Breeding Bird Survey field data collections, hierarchical modeling is used to produce an annual index of abundance, and trends are then estimated as constant annual rates based only on the first and last years of the intervals selected.

Trends in the diversity, richness, and species distribution evenness of the breeding bird community are assessed by regressing each metric against survey years in the add-in statistical software of Microsoft Excel 2010, and then graphing the results ((see Supplemental File: GWCAData3(Diversity-Richness-Evenness).xlsx for example analysis and graphing)). Prior to trend analysis, bird community diversity values are calculated annually using the Shannon Diversity Index:

*H’ = -Σ(n1/N)ln(n1/N)*

where n1/N is the proportion of the total number of individuals in a population consisting of the ith species (Shannon 1949). Species richness values are determined as the total number of bird taxa recorded annually. Species distribution evenness values are calculated using Pielou (J’):

*J’ = H’ / Hmax*

where H’ is the Shannon Diversity Index and Hmax is the maximum possible diversity for a given number of species if all species are present in equal numbers ((ln(annual species richness)). J’ is a measure of how evenly individuals are distributed within a community when compared to the equal distribution and maximum diversity a community can have (Pielou 1969).

Because not all species occurring in an area may actually be observed in a survey (i.e., rare species may be missed), recorded species richness is often an underestimate. Statistical species richness estimators utilize the information in species distribution and abundance patterns to produce an estimate of true species richness. Species richness estimators are also useful in comparing surveys with unequal sampling effort (e.g., different numbers of plots) since more species are usually discovered with greater sampling effort. Different species richness estimators will produce varying estimates, however, and no single estimator is consistently superior to others. Nonparametric statistical estimators have generally performed better than parametric types (Walther and Moore 2005). Reese et al. (2014) recently reviewed nonparametric species richness estimators and found that two coverage-based estimators, the Abundance Coverage-based Estimator (ACE) and Incidence Coverage-based Estimator (ICE), provided less biased and more accurate estimates than many of the others. Thus, we employ these two species richness estimators and report estimated species richness along with observed species richness. The software application EstimateS (Colwell 2013) is used to calculate the ACE and ICE estimators (see Appendix C).

Habitat Summary and Analysis

Similar to the bird community data, habitat variables and indices selected for data summary purposes are descriptive and easily interpretable. Habitat data is collected on plots centered on each variable circular plot location. Once estimates for all parameters have been obtained for each plot, averages and standard deviations among plots can be obtained for individual study units (management units or reference frames) or for park-wide inferences. **Standard deviation is a descriptive statistic that** measures the amount of variation in the yearly observations of habitat parameters at the chosen scale. Cover data, collected by class intervals, should be converted to median cover values prior to data summary and analysis (i.e., class 1 = 0.5, class 2 = 3.0, class 3 = 15.0, class 4 = 37.5, class 5 = 62.5, class 6 = 85.0, class 7 = 97.5). Numbered subsections below reflect the order data summaries and analyses should be performed.

*1) 50-m Plot Data Analysis*

Summary reports for habitat attributes at the 50-m radius plot level should be provided for “permanent” and “semi-permanent” features. For permanent features, a report should include a listing of the values for slope of plot, aspect of slope, and topographic features for each plot. These values are measured only once and are assigned to a permanent locations table within the database. Semi-permanent plot features are not expected to change much, however they could, therefore these features including percent cover of habitat types, road, and water are recorded each time a survey is conducted and should be reported ((see Supplemental File: GWCAHabitatData1(VegType).xlsx for example analysis)). Permanent and semi-permanent plot features provide an overall description of the area available to birds and in which their habitat is evolving. For example, terrain was steep or flat with intervening streams or no water sources at all, a paved road traversed the plot if one did, and the overall vegetation was upland prairie, brome field or whatever it might be.

*2) 5-m Subplot Data Analysis*

Like permanent attributes measured for plot locations, permanent features for subplots (i.e., azimuth, direction from plot center if more than one subplot is sampled, slope across subplot, and aspect of slope) are measured only once and stored in a separate table within the database. A listing of these values should be given in the initial report for a park and not in subsequent reports. Permanent subplot features provide context in which the sampled bird habitat has evolved. For example, terrain was steep with a northerly exposure.

Tree Tally, a measure of stems per hectare (ha), is calculated as stems recorded per 0.00785 ha surveyed ((stems/0.00785 ha; see Supplemental File: GWCAHabitatData2(TreeTally).xlsx for example analysis)). Species are grouped to family and DBH size class (<1.0 cm, 1.1-2.5 cm, 2.6-8.0 cm, 8.1-15.0 cm, 15.1-23.0 cm, 23.1-38.0 cm, and >38.0 cm) prior to calculating stems per hectare. Tree tallies illustrate size distribution of trees by family across a study unit or park. Knowing types and sizes of trees present serves as one measure for assessing habitat availability for bird species that utilize trees for nesting and food.

Woodland measures represent the forest structure in which a plot is located, and generally (but not always) denote a forested or riparian area. Woodland measures help delineate the habitat available for forest bird species. Canopy height in meters should be reported for hardwoods and conifers each time they are encountered on or near (close enough to influence) a plot ((see Supplemental File: GWCAHabitatData3(CanopyHeight).xlsx for example analysis)). Canopy cover, recorded in the field as dots covered on a spherical densiometer, should be converted to percent coverage (dots recorded x 1.04) before being reported for analysis by tree type and collectively ((see Supplemental File: GWCAHabitatData4(CanopyCover).xlsx for example analysis)). The percent values for canopy cover are then averaged (n=4) to generate a plot level mean. Basal area measurements of plot tree density are reported as m2/ha ((see Supplemental File: GWCAHabitatData5(BasalArea).xlsx for example analysis)). Basal area for a plot is determined by multiplying the field stem counts by 2.5 for both hardwoods and conifers encountered.

The Horizontal Vegetation Profile represents the area obscured by vegetation, or screening cover for nesting birds, at various heights to 2 m. Average (+ std dev) values are reported for each height class (0 – 0.25, 0.25 – 0.5, 0.5 – 0.75, 0.75 – 1.0, 1.0 – 1.25, 1.25 – 1.5, 1.5 – 1.75, and 1.75 – 2.0 m) by individual study units (management units or reference frames) or for park-wide inferences ((see Supplemental File: GWCAHabitatData6(HorizontalVegetation).xlsx for example analysis)).

The Structural Diversity Index, a measure of vertical habitat available to birds to a height of 7.5 m, is calculated for each plot by summing the percentages of possible touches (8 for herbaceous and deciduous or 12 for herbaceous, deciduous, and coniferous if conifers are present in the park) from vegetation within each 1-m height increment that were actually touched; dividing this value by the number of height increments measured (8); adding the resulting value to the percent of increments occupied; multiplying this value by 100; and then dividing it by two. Average (+ std dev) annual vertical structure diversity is estimated and reported by habitat type or park-wide, calculated as:

|  |  |  |
| --- | --- | --- |
|  | *Structural Diversity Index =* | *((∑pi / 8) + a) \* 100* |
| *2* |

where pi is the observed frequency for vegetation in the ith interval touching a measuring rod out of 8 or 12 measuring events, and a is the percent of intervals with recorded vegetation in eight height increments. Vertical structure diversity values are weighted equally to represent both the vertical height of vegetation and how dense the vegetation is within each height increment ((see Supplemental File: GWCAHabitatData7(VegetationProfile).xlsx for example analysis)).

*3) 1.78-m Subplot Data Analysis*

Percent Ground Cover is a measure of non-vegetative subplot attributes and includes bare soil, conifer litter, deciduous litter, grass litter, rock, unvegetated surface, and woody debris (>2.5 cm diameter). Average (+ std dev) values for these variables are calculated and reported for individual study units (management units, reference frames, or habitat types) or for the park as a whole ((see Supplemental File: GWCAHabitatData8(GroundCover).xlsx for example analysis)).

Percent Foliar Cover is a measure of vegetative cover recorded for cool season grass, forbs, moss and lichen, shrubs and vines, total vegetation cover (< 1.50 m), tree seedlings, and warm season grass plant guilds. Average (+ std dev) values for these variables are calculated and reported for individual study units (management units, reference frames, or habitat types) or for the park as a whole ((see Supplemental File: GWCAHabitatData9(FoliarCover).xlsx for example analysis)).

Bird – Habitat Relationships

Descriptive statistics (e.g., means, standard deviations) summarizing habitat data will be provided with all reports. Inferential statistics such as hypothesis testing should be reserved for instances in which relevant ecological questions or hypotheses have been specified. This distinction is made because with so many bird species and habitat variables, a large number of statistical comparisons could potentially be made, and many spurious results are possible. Ideally, hypotheses regarding habitat effects or relationships will be generated by patterns observed in the bird species abundance data. Park managers may also have questions that can be stated as testable hypotheses.

The relevant statistical analyses will depend on the question and underlying structure of the data set (e.g., whether the data are distributed normally enough for a parametric test, or a non-parametric test is required). It is not possible to anticipate and describe how to conduct all possible analyses that could be applied in this SOP. A qualified analyst should be consulted in the context of any bird-habitat interaction analyses.

Literature Cited

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press. 432 pp.

Colwell, R. K. 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at: http://purl.oclc.org/estimates.

Pannekoek, J. and A. J. Van Strien. 2005. TRIM 3.54 manual (TRends and Indices for Monitoring data). Statistics Netherlands, Voorburg, The Netherlands.

Pielou, E. C. 1969. An introduction to mathematical ecology. John Wiley and Sons, New York, New York. 286pp.

Reese, G. C., K. R. Wilson, and C. H. Flather. 2014. Performance of species richness estimators across assemblage types and survey parameters. Global Ecology and Biogeography, 23: 585-594.

Sauer, J. R., D. K. Niven, J. E. Hines, D. J. Ziolkowski, Jr, K. L. Pardieck, J. E. Fallon, and W. A. Link. 2017. The North American Breeding Bird Survey, Results and Analysis 1966 - 2015. Version 2.07.2017 USGS Patuxent Wildlife Research Center, Laurel, MD. <https://www.mbr-pwrc.usgs.gov/bbs/trend/tf15.html>

Shannon, C. E. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois. 177 pp.

Thomas, L., S. T. Buckland, E. A. Rexstad, J. L. Laake, S. Strindberg, S. L. Hedley, J. R. B. Bishop, T. A. Marques, and K. P. Burnham. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size.  Journal of Applied Ecology 47: 5-14.

Walther, B. A. and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography, 28, 815–829.

**Appendix A.** Step-by-step instructions for conducting Distance analysis bird population abundance using the R package Distance.

**Obtaining “Distance R Package”:** The Distance R Package can be downloaded free of charge at: [Distance R packages · distancesampling.org](http://distancesampling.org/R/)

A user’s manual is also available at this website. The following is a script with notations for performing an abundance analysis.

**Example Distance R Package script**. See Supplemental File: WoodCARW.csv in the GWCADistanceExample folder for example data set.

library(Distance)

setwd("C:\\Users\\DPeitz\\Desktop\\SupplementalFile\\GWCABirdDataAnalysis\\GWCADistanceExample")

WoodCARW <- read.csv("WoodCARW.csv",

header = TRUE, as.is = TRUE)

head(WoodCARW)

summary(WoodCARW)

hist(WoodCARW$distance, xlab="distance (m)",

main="Woodland Bird Data point transects")

conversion.factor <- convert\_units("meter", NULL, "hectare")

bird.hn <- ds(WoodCARW, key="hn", adjustment=NULL,

transect="point", convert\_units=conversion.factor, truncation="5%")

summary(bird.hn)

cutpoints <- c(0,5,10,15,20,30,40,max(WoodCARW$distance, na.rm=TRUE))

plot(bird.hn, breaks=cutpoints, pdf=TRUE, main="Woodland Bird Data point transects.")

bird.hn.cos <- ds(WoodCARW, key="hn", adjustment="cos",

transect="point", convert\_units=conversion.factor, truncation="5%")

summary(bird.hn.cos)

bird.hr.cos <- ds(WoodCARW, key="hr", adjustment="cos",

transect="point", convert\_units=conversion.factor, truncation="5%")

summary(bird.hr.cos)

bird.hn.herm <- ds(WoodCARW, key="hn", adjustment="herm",

transect="point", convert\_units=conversion.factor, truncation="5%")

summary(bird.hn.herm)

bird.hr.poly <- ds(WoodCARW, key="hr", adjustment="poly",

transect="point", convert\_units=conversion.factor, truncation="5%")

summary(bird.hr.poly)

AIC(bird.hn.cos, bird.hr.cos, bird.hn.herm, bird.hr.poly)

gof\_ds(bird.hn)

knitr::kable(summarize\_ds\_models(bird.hn.cos, bird.hr.cos, bird.hn.herm, bird.hr.poly),digits=3,

caption="Model selection summary of Woodland Bird Data point transect.")

**R Output**

R version 4.1.3 (2022-03-10) -- "One Push-Up"

Copyright (C) 2022 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

> library(Distance)

Loading required package: mrds

This is mrds 2.2.6

Built: R 4.1.3; ; 2022-03-21 09:17:27 UTC; windows

Attaching package: ‘Distance’

The following object is masked from ‘package:mrds’:

create.bins

> setwd("C:\\Users\\DPeitz\\Desktop\\SupplementalFile\\GWCABirdDataAnalysis\\GWCADistanceExample")

> WoodCARW <- read.csv("WoodCARW.csv",

+ header = TRUE, as.is = TRUE)

> head(WoodCARW)

Region.Label Area Sample.Label Effort object distance Study.Area

1 single\_stratum 22.59 GWCA1 1 1 30 CARW

2 single\_stratum 22.59 GWCA1 1 2 56 CARW

3 single\_stratum 22.59 GWCA18 1 3 34 CARW

4 single\_stratum 22.59 GWCA2 1 4 22 CARW

5 single\_stratum 22.59 GWCA2 1 5 38 CARW

6 single\_stratum 22.59 GWCA27 1 6 29 CARW

> summary(WoodCARW)

Region.Label Area Sample.Label Effort

Length:75 Min. :22.59 Length:75 Min. :1

Class :character 1st Qu.:22.59 Class :character 1st Qu.:1

Mode :character Median :22.59 Mode :character Median :1

Mean :22.59 Mean :1

3rd Qu.:22.59 3rd Qu.:1

Max. :22.59 Max. :1

object distance Study.Area

Min. : 1.0 Min. : 10.00 Length:75

1st Qu.:19.5 1st Qu.: 26.50 Class :character

Median :38.0 Median : 43.00 Mode :character

Mean :38.0 Mean : 54.44

3rd Qu.:56.5 3rd Qu.: 70.00

Max. :75.0 Max. :300.00

> hist(WoodCARW$distance, xlab="distance (m)",

+ main="Woodland Bird Data point transects")

> conversion.factor <- convert\_units("meter", NULL, "hectare")

> bird.hn <- ds(WoodCARW, key="hn", adjustment=NULL,

+ transect="point", convert\_units=conversion.factor, truncation="5%")

Fitting half-normal key function

Key only model: not constraining for monotonicity.

AIC= 649.087

> summary(bird.hn)

Summary for distance analysis

Number of observations : 71

Distance range : 0 - 102.6

Model : Half-normal key function

AIC : 649.0875

Detection function parameters

Scale coefficient(s):

estimate se

(Intercept) 3.710708 0.06593049

Estimate SE CV

Average p 0.3039319 0.0344215 0.1132540

N in covered region 233.6049552 35.1420151 0.1504335

Summary statistics:

Region Area CoveredArea Effort n k ER se.ER cv.ER

1 single\_stratum 22.59 46.29911 14 71 14 5.071429 0.8351106 0.1646697

Abundance:

Label Estimate se cv lcl ucl df

1 Total 113.9792 22.77947 0.1998564 75.9451 171.0612 27.08191

Density:

Label Estimate se cv lcl ucl df

1 Total 5.045561 1.008388 0.1998564 3.36189 7.572431 27.08191

> cutpoints <- c(0,5,10,15,20,30,40,max(WoodCARW$distance, na.rm=TRUE))

> plot(bird.hn, breaks=cutpoints, pdf=TRUE, main="Woodland Bird Data point transects.")

Specified endpoints > 102.6; values reset.

> bird.hn.cos <- ds(WoodCARW, key="hn", adjustment="cos",

+ transect="point", convert\_units=conversion.factor, truncation="5%")

Starting AIC adjustment term selection.

Fitting half-normal key function

Key only model: not constraining for monotonicity.

AIC= 649.087

Fitting half-normal key function with cosine(2) adjustments

AIC= 647.046

Fitting half-normal key function with cosine(2,3) adjustments

Warning in check.mono(result, n.pts = control$mono.points) :

Detection function is not strictly monotonic!

AIC= 643.753

Fitting half-normal key function with cosine(2,3,4) adjustments

Warning in check.mono(result, n.pts = control$mono.points) :

Detection function is not strictly monotonic!

AIC= 645.416

Half-normal key function with cosine(2,3) adjustments selected.

Warning in mrds::check.mono(model, n.pts = 20) :

Detection function is not strictly monotonic!

> summary(bird.hn.cos)

Summary for distance analysis

Number of observations : 71

Distance range : 0 - 102.6

Model : Half-normal key function with cosine adjustment terms of order 2,3

Strict monotonicity constraints were enforced.

AIC : 643.7534

Detection function parameters

Scale coefficient(s):

estimate se

(Intercept) 3.708867 0.06905588

Adjustment term coefficient(s):

estimate se

cos, order 2 0.3713982 0.1638415

cos, order 3 -0.2453750 0.1623802

Estimate SE CV

Average p 0.2458388 0.08441689 0.3433830

N in covered region 288.8070803 103.54203319 0.3585163

Summary statistics:

Region Area CoveredArea Effort n k ER se.ER cv.ER

1 single\_stratum 22.59 46.29911 14 71 14 5.071429 0.8351106 0.1646697

Abundance:

Label Estimate se cv lcl ucl df

1 Total 140.9131 53.6633 0.3808255 67.75291 293.0724 80.58072

Density:

Label Estimate se cv lcl ucl df

1 Total 6.237854 2.375534 0.3808255 2.999243 12.97355 80.58072

> bird.hr.cos <- ds(WoodCARW, key="hr", adjustment="cos",

+ transect="point", convert\_units=conversion.factor, truncation="5%")

Starting AIC adjustment term selection.

Fitting hazard-rate key function

Key only model: not constraining for monotonicity.

AIC= 641.708

Fitting hazard-rate key function with cosine(2) adjustments

AIC= 643.163

Hazard-rate key function selected.

> summary(bird.hr.cos)

Summary for distance analysis

Number of observations : 71

Distance range : 0 - 102.6

Model : Hazard-rate key function

AIC : 641.7076

Detection function parameters

Scale coefficient(s):

estimate se

(Intercept) 3.511271 0.2411925

Shape coefficient(s):

estimate se

(Intercept) 0.9249545 0.2030985

Estimate SE CV

Average p 0.2453246 0.065121 0.2654483

N in covered region 289.4124309 82.414989 0.2847666

Summary statistics:

Region Area CoveredArea Effort n k ER se.ER cv.ER

1 single\_stratum 22.59 46.29911 14 71 14 5.071429 0.8351106 0.1646697

Abundance:

Label Estimate se cv lcl ucl df

1 Total 141.2085 44.11017 0.3123762 76.88085 259.3602 74.08866

Density:

Label Estimate se cv lcl ucl df

1 Total 6.250929 1.952641 0.3123762 3.403313 11.4812 74.08866

> bird.hn.herm <- ds(WoodCARW, key="hn", adjustment="herm",

+ transect="point", convert\_units=conversion.factor, truncation="5%")

Starting AIC adjustment term selection.

Fitting half-normal key function

Key only model: not constraining for monotonicity.

AIC= 649.087

Fitting half-normal key function with Hermite(4) adjustments

AIC= 651.01

Half-normal key function selected.

> summary(bird.hn.herm)

Summary for distance analysis

Number of observations : 71

Distance range : 0 - 102.6

Model : Half-normal key function

AIC : 649.0875

Detection function parameters

Scale coefficient(s):

estimate se

(Intercept) 3.710708 0.06593049

Estimate SE CV

Average p 0.3039319 0.0344215 0.1132540

N in covered region 233.6049552 35.1420151 0.1504335

Summary statistics:

Region Area CoveredArea Effort n k ER se.ER cv.ER

1 single\_stratum 22.59 46.29911 14 71 14 5.071429 0.8351106 0.1646697

Abundance:

Label Estimate se cv lcl ucl df

1 Total 113.9792 22.77947 0.1998564 75.9451 171.0612 27.08191

Density:

Label Estimate se cv lcl ucl df

1 Total 5.045561 1.008388 0.1998564 3.36189 7.572431 27.08191

> bird.hr.poly <- ds(WoodCARW, key="hr", adjustment="poly",

+ transect="point", convert\_units=conversion.factor, truncation="5%")

Starting AIC adjustment term selection.

Fitting hazard-rate key function

Key only model: not constraining for monotonicity.

AIC= 641.708

Fitting hazard-rate key function with simple polynomial(4) adjustments

AIC= 643.708

Hazard-rate key function selected.

> summary(bird.hr.poly)

Summary for distance analysis

Number of observations : 71

Distance range : 0 - 102.6

Model : Hazard-rate key function

AIC : 641.7076

Detection function parameters

Scale coefficient(s):

estimate se

(Intercept) 3.511271 0.2411925

Shape coefficient(s):

estimate se

(Intercept) 0.9249545 0.2030985

Estimate SE CV

Average p 0.2453246 0.065121 0.2654483

N in covered region 289.4124309 82.414989 0.2847666

Summary statistics:

Region Area CoveredArea Effort n k ER se.ER cv.ER

1 single\_stratum 22.59 46.29911 14 71 14 5.071429 0.8351106 0.1646697

Abundance:

Label Estimate se cv lcl ucl df

1 Total 141.2085 44.11017 0.3123762 76.88085 259.3602 74.08866

Density:

Label Estimate se cv lcl ucl df

1 Total 6.250929 1.952641 0.3123762 3.403313 11.4812 74.08866

> AIC(bird.hn.cos, bird.hr.cos, bird.hn.herm, bird.hr.poly)

df AIC

bird.hn.cos 3 643.7534

bird.hr.cos 2 641.7076

bird.hn.herm 1 649.0875

bird.hr.poly 2 641.7076

> gof\_ds(bird.hn)

Goodness of fit results for ddf object

Distance sampling Cramer-von Mises test (unweighted)

Test statistic = 0.332572 p-value = 0.10972

> knitr::kable(summarize\_ds\_models(bird.hn.cos, bird.hr.cos, bird.hn.herm, bird.hr.poly),digits=3,

+ caption="Model selection summary of Woodland Bird Data point transect.")

Table: Model selection summary of Woodland Bird Data point transect.

| |Model |Key function |Formula|C-vM p-alue|$\hat{P\_a}$|se($\hat{P\_a}$)|$\Delta$AIC|

|:-|:--------------------|:----------------|:------|-----------:|----------:|--------------:|----------:|

|2 |\texttt{bird.hr.cos} |Hazard-rate |~1 | 0.827| 0.245| 0.065| 0.000|

|4 |\texttt{bird.hr.poly}|Hazard-rate |~1 | 0.827| 0.245| 0.065| 0.000|

|1 |\texttt{bird.hn.cos} |Half-normal with |~1 | 0.882| 0.246| 0.084| 2.046|

cosine adjustment terms of order 2,3

|3 |\texttt{bird.hn.herm}|Half-normal |~1 | 0.110| 0.304| 0.034| 7.380|

>

**Literature Cited**

**Appendix B.** Step-by-step instructions for conducting Trim analysis of bird population trends.

**Obtaining ‘rtrim’:** ‘rtrim’ can be downloaded free of charge at: <https://rdrr.io/cran/rtrim/> “[rtrim: Trends and Indices for Monitoring Data version 2.1.1 from CRAN (rdrr.io)](https://rdrr.io/cran/rtrim/)”

The latest version is 2.1.1 (accessed 30 November 2020). User’s guides and analysis examples are available at this website. The following is a condensed guide for performing trend analysis on Heartland Inventory and Monitoring Program bird data.

**1) Data preparation**

Each species must have its own data file. The data file must be a comma delimited (.csv) file containing one line (a record) for each combination of site and time. So, for I sites and J timepoints, the number of records is I x J. Each record contains the following variables (the order is important!), separated by one or more spaces. Sites that were not sampled at particular times must be included as missing records and count is left blank.

|  |  |  |
| --- | --- | --- |
| Variable | Values | Required/Optional |
| Site | identifier integer not exceeding 9 digits | Required |
| Time-point identifier | integer not exceeding 5 digits | Required |
| Count | integer in range (0...2.0 x 109) or missing code (left blank – unlike the TRIM software, ‘rtrim’ will not filter out negative number. This can be accomplished by leaving “Count” blank) | Required |

The records must be sorted by I site and J timepoint so that the first record should correspond with the first J time-point for the first site and the next record should correspond with the second J time-point for the first site, and so on. The order of the sites is unimportant. It is only required that the records for the same site are kept together and the time-points for a site must be sorted in increasing order.

Each site has to contain at least one positive count (> 0). ‘rtrim’ (v. 2.1.1) will automatically delete any sites with no positive counts, so this does not have to be done as part of data preparation.

The three variables in the table above reflect the minimum information necessary to run the analysis. Other variables (covariates and weights) may also be input; see the TRIM manual or website listed above for information on the use of these variables.

See Supplemental File / Folder GWCATRIMdataandanalysis for example .csv data sets, .TFC run files, and .out output files.

**2) Modeling**

To analyze HTLN data with ‘rtrim, we start with a model with time effects (model 3), ignoring covariates if there are any. Model 3 is chosen because it makes no assumption about how population changes over time. Year effects are strictly independent of each other.

z1 <- trim(count ~ site + year, data=GWoodCARW, model=3, serialcor=TRUE, overdisp=TRUE)

Output from a model with time effects includes:

* The call to ‘trim’ used to estimate the model.
* Details on the model, estimation method, and the numerical solution.
* The estimated model parameters (from both the additive and the multiplicative perspective), and the associated standard errors.
* Estimations of overdispersion and serial correlation parameters, if applicable.
* Several model goodness-of-fit measures.

**Model simplification**

The advantage of Model 3 is, as argued above, the absence of any assumptions regarding the temporal trend. This, however, comes at a price: Positive counts are required for all individual years to allow estimation of the model parameters. So, this model cannot be used for cases where one or more years are missing, or a monitoring site has zero positive observations. Furthermore, the model is far from being parsimonious.

For both reasons it is preferable to replace model 3 with model 2 (piecewise loglinear), especially because in one extreme case these models are equivalent. This is the case when all years are treated as changepoints, and each year the trend changes into a different one.

z2 <- trim(count ~ site + year, data=GWoodCARW, model=2, serialcor=TRUE, overdisp=TRUE)

The breaks of the piecewise loglinear model can be specified with the changepoints option. The trim function will give an error when too little observations are present in a time segment, except when the autodelete option is set to TRUE. In that case time segments are combined until enough observations are present for a model to be estimated.

z2a <- trim(count ~ site + year, data=GWoodCARW, model=2, changepoints=”all, autodelete=TRUE, serialcor=TRUE, overdisp=TRUE)

If it is important to know (generally when more than 20 years of data are available), and if the tests for the significance of changes in slopes show that the only significant changes are for a subset of years, which mean that the slope between one or more subsets of years are different from other subsets of years, it may be possible to describe data with a model with /less than the full set of changepoints. To investigate this possibility, the stepwise procedure for selection of changepoints can be used by including stepwise=TRUE in the call to trim():

z3 <- trim(count ~ site + year, data=GWoodCARW, model=2, changepoints=”all”, stepwise=TRUE, serialcor=TRUE, overdisp=TRUE)

The difference between the models of this run (z3) and the previous (z2) can be tested by comparing their Likelihood Ratio's.

gof(z2)

LR2 <- gof(z2)$LR$LR # get raw LR info for run 2

Df2 <- gof(z2)$LR$df

gof(z3)

LR3 <- gof(z3)$LR$LR # get raw LR info for run 3

Df3 <- gof(z3)$LR$df # test the differece by using the fact that the difference of two LR measures is asymptotically Chi^2 distributed

LR <- abs(LR3 – LR2)

df <- abs(df3 – df2)

p <- 1 - pchisq(LR, df=df) # use Chi-squared distribution

p

The explicit setting of initial changepoints can be replaced by the more convenient changepoints=”auto”, which combines changepoints=”all” with stepwise=TRUE

z3a <- trim(count ~ site + year, data=GWoodCARW, model=2, changepoints=”auto”, serialcor=TRUE, overdisp=TRUE)

**3) Data input and analysis by example using WoodCARW.csv data.**

**R Code for HTLN breeding bird analysis**:

rm(list=ls())

library(rtrim)

setwd("C:\\Users\\DPeitz\\Desktop\\rTRIM\\2020GWCATRIM")

WoodCARW <- read.csv("WoodCARW.csv",

header = TRUE, as.is = TRUE)

head(WoodCARW)

summary(WoodCARW)

idx <- which(names(WoodCARW)=="year") # rename year->season

names(WoodCARW)[idx] <- "season"

count\_summary(WoodCARW, year\_col="season") # show that it works

names(WoodCARW)[idx] <- "year" # revert to original name

z1 <- trim(count ~ site + year, data=WoodCARW, model=2, overdisp = TRUE, serialcor = TRUE)

summary(z1, which="both")

index(z1, which="both")

coefficients(z1, which="both")

z1$slope

overall(z1) #displays the overall imputed values, model significance value, and analysis meaning.

plot(index(z1)) #plots index values and std err.

plot(overall(z1)) #plots overall imputed values, std err and model slope.

**R Output**:

> rm(list=ls())

> library(rtrim)

Welcome to rtrim 2.0.6 Type ?`rtrim-package` to get started.

Attaching package: ‘rtrim’

The following object is masked from ‘package:stats’:

heatmap

Warning message:

package ‘rtrim’ was built under R version 3.6.3

> setwd("C:\\Users\\DPeitz\\Desktop\\rTRIM\\2020GWCATRIM")

> WoodCARW <- read.csv("WoodCARW.csv",

+ header = TRUE, as.is = TRUE)

> head(WoodCARW)

site year count

1 91 2008 2

2 91 2009 0

3 91 2010 0

4 91 2011 0

5 91 2012 0

6 91 2013 0

> summary(WoodCARW)

site year count

Min. : 91.0 Min. :2008 Min. :0.0000

1st Qu.:923.8 1st Qu.:2011 1st Qu.:0.0000

Median :932.0 Median :2014 Median :0.0000

Mean :828.6 Mean :2014 Mean :0.6146

3rd Qu.:937.8 3rd Qu.:2017 3rd Qu.:1.0000

Max. :958.0 Max. :2020 Max. :4.0000

NA's :8

>

> idx <- which(names(WoodCARW)=="year") # rename year->season

> names(WoodCARW)[idx] <- "season"

> count\_summary(WoodCARW, year\_col="season") # show that it works

Total number of sites 8

Sites without positive counts (0):

Number of observed zero counts 51

Number of observed positive counts 45

Total number of observed counts 96

Number of missing counts 8

Total number of counts 104

> names(WoodCARW)[idx] <- "year" # revert to original name

>

> z1 <- trim(count ~ site + year, data=WoodCARW, model=2, overdisp = TRUE, serialcor = TRUE)

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Overdispersion 0.9 <1; consider setting overdisp=FALSE

Warning: Serial correlation is very low (rho=0.104); consider disabling it.

> summary(z1, which="both")

Call:

trim(count ~ site + year, data = WoodCARW, model = 2, overdisp = TRUE,

serialcor = TRUE)

Model : 2

Method : GEE (Convergence reached after 10 iterations)

Coefficients:

from upto add se\_add mul se\_mul

1 2008 2020 0.1060695 0.03678252 1.111899 0.04089846

Overdispersion : 0.9040

Serial Correlation : 0.1040

Goodness of fit:

Chi-square = 78.65, df=87, p=0.7269

Likelihood Ratio = 86.06, df=87, p=0.5082

AIC (up to a constant) = -87.94

> index(z1, which="both")

time fitted se\_fit imputed se\_imp

1 2008 1.000000 0.00000000 1.00000000 0.0000000

2 2009 1.111899 0.04089846 1.07781576 0.5209683

3 2010 1.236320 0.09094993 1.00000000 0.5586899

4 2011 1.374663 0.15169073 0.09620517 0.4225463

5 2012 1.528487 0.22488640 0.25000000 0.4737805

6 2013 1.699523 0.31256376 1.00000000 0.6171557

7 2014 1.889698 0.41704728 1.21588808 0.6497147

8 2015 2.101154 0.54100029 1.25000000 0.7189703

9 2016 2.336272 0.68747177 1.00000000 0.6860953

10 2017 2.597699 0.85994922 2.18179876 1.0012439

11 2018 2.888379 1.06241873 2.50000000 1.1355031

12 2019 3.211586 1.29943279 1.90458700 0.9276526

13 2020 3.570960 1.57618722 2.00000000 1.0335446

> coefficients(z1, which="both")

from upto add se\_add mul se\_mul

1 2008 2020 0.1060695 0.03678252 1.111899 0.04089846

> z1$slope

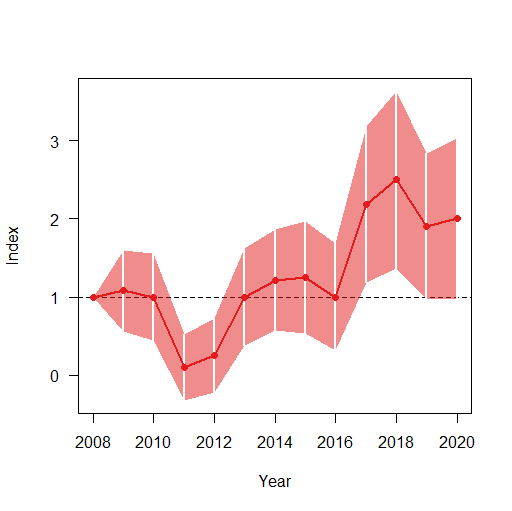
NULL

> overall(z1)

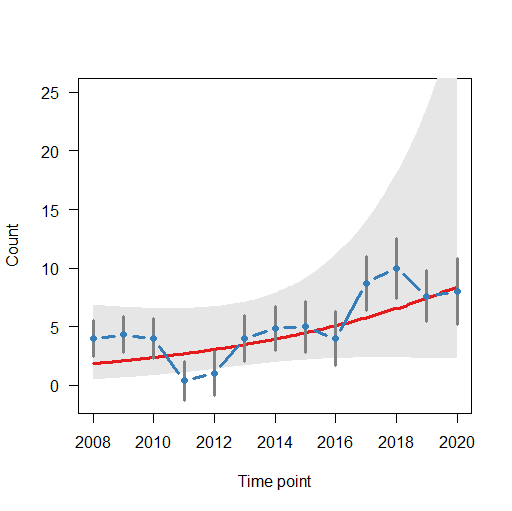
from upto add se\_add mul se\_mul p meaning

2008 2020 0.1265422 0.08455837 1.134897 0.09596508 0.162652 Uncertain

> plot(index(z1))



> plot(overall(z1))



**4) Results**

Check for “Goodness of Fit” of the model. The fit of the model is evaluated by two tests: the Chi-square test and the Likelihood Ratio test. Usually, the results of these tests are similar. If the P-value of (one of) these tests is below 0.05, the model is rejected. In case the model is rejected, one can try to find a better model by incorporating a covariate, if possible. In case of lack-of-fit, the model can still be used; TRIM converts any lack-of-fit into higher standard errors, which will make it less likely that any trend will be significant.

The overall trend is a descriptive parameter for the trend slope over the period studied. ‘rtrim’ calculates the slope of the regression line based upon imputed indices. Use the multiplicative (mul), rather than the additive (add) slope, and associated SE. The multiplicative slope will be close to 1, the additive slope will almost always be <<1.

**Interpretation**

The multiplicative trend can be interpreted as the average percentage change per year. If the slope is equal to 1, there is no trend. A slope >1 indicates an increase, and a slope < 1 indicates a decline. For example, a slope of 1.08 indicates an increase of 8% per year; a slope of 0.93 indicates a decrease of 7% per year.

The TRIM manual (Pannekoek and Van Strien 2005) describes how to reach an overall conclusion based on both the statistical significance and biological importance of the calculated trend. Possible qualitative interpretations include five different categories (see also van Strien et al. 2001; Table 5, Appendices 1 and 2).

Trend classification types depending on statistical significance and magnitude (Pannekoek and Van Strien 2005, Van Strien et al. 2001), and follow rules outlined in Gregory et. al. (2007). The multiplicative overall slope estimate in ‘rtrim’ is converted into one of the following categories depending on the overall slope as well as its 95% confidence interval (= slope ± 1.96 times the standard error of the slope):

Strong increase – increase significantly more than 5% per year. Criterion: lower limit of confidence interval > 1.05.

Moderate increase – significant increase, but not significantly more than 5% per year. Criterion: 1.00 < lower limit of confidence interval < 1.05.

Stable – no significant increase or decline, and it is certain that trends are less than 5% per year. Criterion: confidence interval encloses 1.00 but lower limit > 0.95 and upper limit < 1.05.

Uncertain – no significant increase or decline, but not certain if trends are less than 5% per year. Criterion: confidence interval encloses 1.00 but lower limit < 0.95 or upper limit > 1.05.

Moderate decline – significant decline, but not significantly more than 5% per year. Criterion: 0.95 < upper limit of confidence interval < 1.00.

Steep decline – decline significantly more than 5% per year. Criterion: upper limit of confidence interval < 0.95.

Thus, basic reporting would include, for each species, the multiplicative trend and associated standard deviation, and the qualitative interpretation as indicated above (e.g., see Appendix 2 in Gregory et al. 2007 for an example).

**5) References**

Gregory, R.D., J. Reif, L. Fornasari, I.J. Burfield, P. Chylarecki, A.J. Van Strien, P. Vorisek, F.D.R. Jiguet, and A.W. Gmelig-Meyling. 2007. Population trends of widespread woodland birds in Europe. Ibis 149: 78-97.

Pannekoek J. and A.J. Van Strien. 2005. TRIM 3 manual (TRends and Indices for Monitoring data). Statistics Netherlands. Technical report.

rTRIM 2.1.1. <https://rdrr.io/cran/rtrim/> “[rtrim: Trends and Indices for Monitoring Data version 2.1.1 from CRAN (rdrr.io)](https://rdrr.io/cran/rtrim/)”. Accessed 30 November 2020.

Van Strien, A.J., J. Pannekoek, and D.W. Gibbons. 2001. Indexing European bird population trends using results of national monitoring schemes: a trial of a new method. Bird Study 48: 200-213.

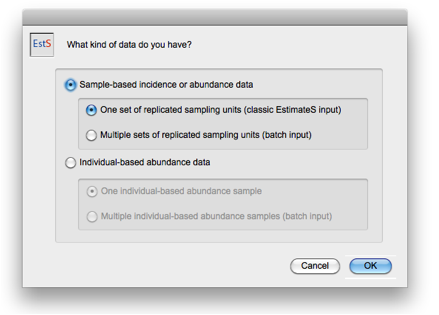
**Appendix C.** Step-by-step instructions for conducting EstimateS analysis of bird species richness.

**Obtaining EstimateS:** EstimateS can be downloaded free of charge at: [Robert K. Colwell Academic Website - EstimateS (robertkcolwell.org)](https://www.robertkcolwell.org/pages/estimates)

The latest version is 9.1.0 (assessed August 17, 2022). A user’s manual is also available at this website. The following is a condensed version of the user’s manual for obtaining the diversity indices.

**Preparing a Data Input File for EstimateS**

The Load Data Input File command (from the File menu) presents a set of four filetype options. This SOP describes the use of the sample-based incidence or abundance data option for one set of replicated sampling units. The EstimateS user’s manual should be consulted if it is desirable to use any of the other filetypes.



All Input Files in EstimateS must be in tab-delimited plain text (sometimes called tab-separated-values, or TSV). Excel files cannot be read. Save them first as tab-delimited text.

The Input File may have any name and may be located in any folder (directory).

The two required header records (rows) for this filetype are:

Record #1 (Title Record): Datafile Title

The first record (line) of the Input File must contain a title in the first field (column); any text will do.

Record #2 (Parameter Record): Number of Species<tab>Number of Sampling Units in the Sample Set

The second record (line) of the Input File must contain two obligatory control parameters: the number of species and the number of sampling units, separated by a <tab> character.

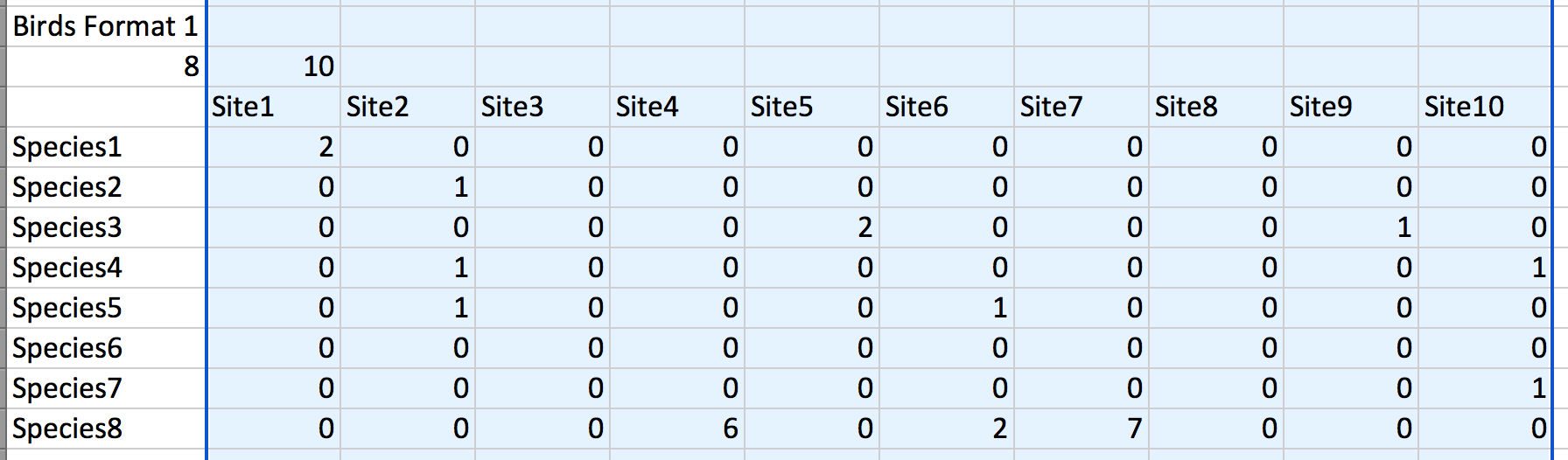
Record #3 etc.: The rest of the Input File contains the input data. (See examples files below.)

**Data Input Formats**

**Format 1.** Species (rows) by sites (columns). There will be one row for each species and one column for each sampling unit. The input file should contain an initial row of site labels and an initial column of species labels. There is no specific format for these labels; they are useful for us to understand the structure of the data but will be ignored by EstimateS.

The row of site labels must follow the required Title and Parameter records and precede the data. The required Title and Parameter records begin in the first column.

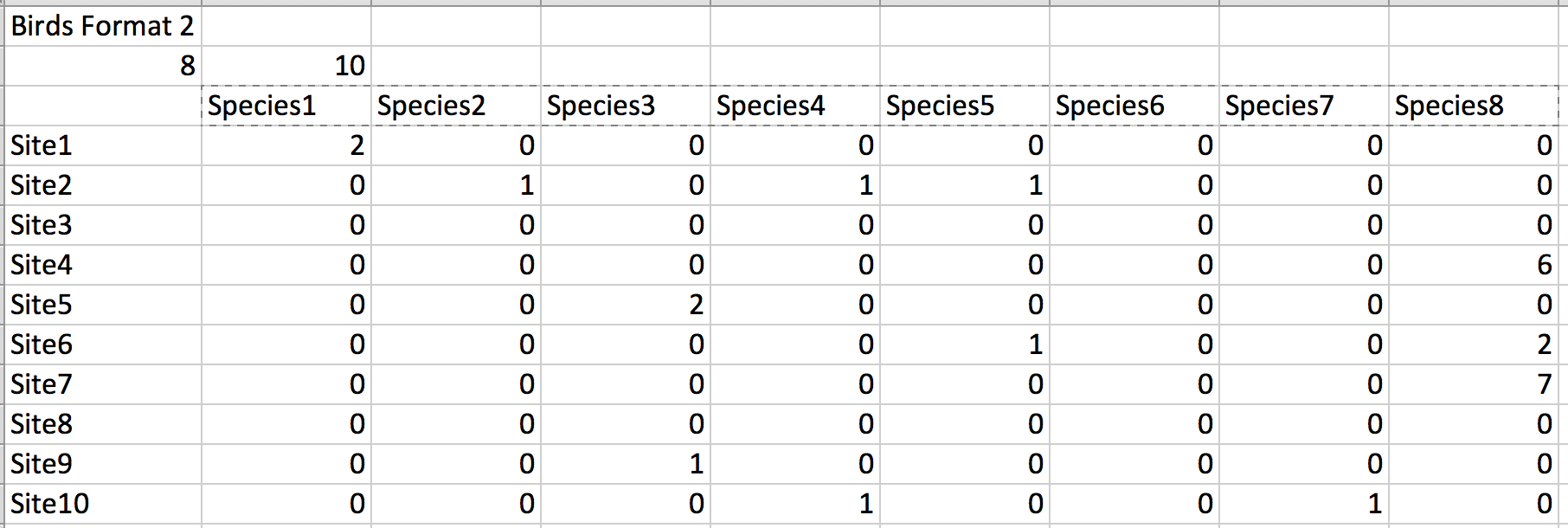
Format 1 Example: Below is a simple example of an EstimateS sample-based Input File in Format 1, for a dataset called "Birds Format 1" that includes data for 8 species (rows) and 10 sampling units (columns).



**Format 2.** Sites (rows) by species (columns). This is simply the transpose of format 1, with one row for each site and one column for each species. The input file should contain an initial row of species labels and an initial column of site labels.

The row of species labels must follow the required Title and Parameter records and precede the data. The required Title and Parameter records begin in the first column.

Format 2 Example: Below is a simple example of an EstimateS sample-based Input File in Format 2, for a dataset called "Birds Format 2" that includes data for 8 species (columns) in 10 samples (rows). The data are exactly the same as in the example above for Format 1.



**Running EstimateS**

Loading the Data Input File

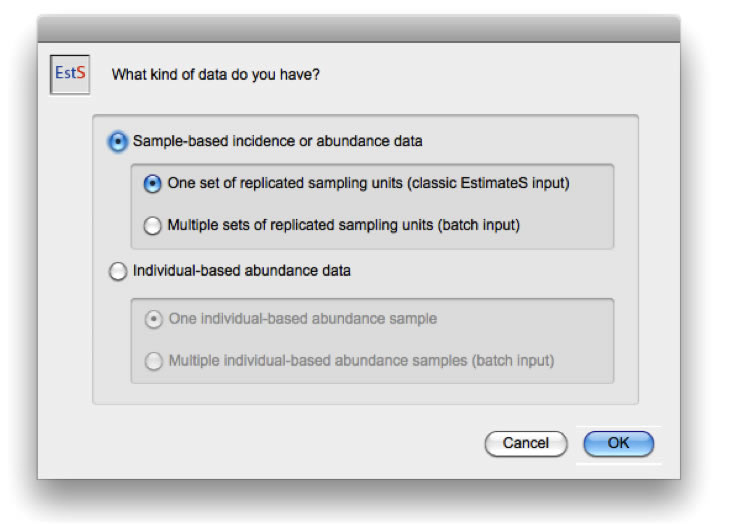
1. Launch EstimateS from the Programs section of the Start menu or double-clicking the EstimateS[version number].exe file.

2. IF a file navigation window appears requesting the selection of a "Data File," choose the file called Statistics.4DD. (This default file records the statistical output of Biota. This request may not appear; if not just skip the rest of this step. The Statistics file is of no practical use, but is required for EstimateS to function.)

Note 1: Do not try to load the input file at this point. If the Statistic Data file cannot be found, click the New button to create a new data output file. It can be given any name, using the extension .4DD (Windows).

Note 2: If one wants to create a new output data file or find a different existing one, the navigation window can be forced to appear as follows: Select the EstimateS icon or application name, then choose open from the Windows File menu, while holding down the Alt key.

3. From the File menu in EstimateS, choose Load Input File. The Filetype Selection dialog appears.



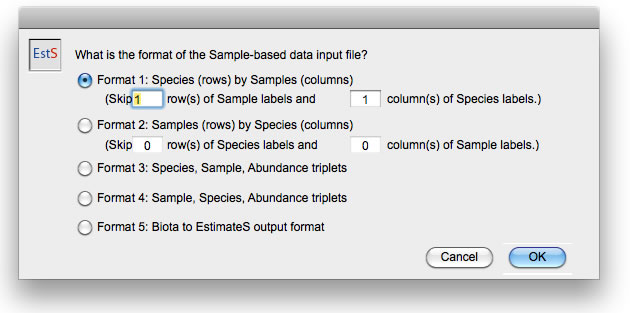
Choose the sample-based, single dataset option.

4. Click the OK button. The Open File window appears.

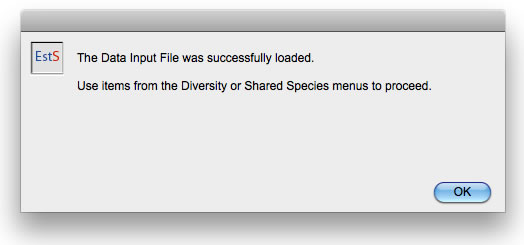
5. Find the Data Input File and open it.

For a single dataset, a confirmation screen appears, showing the parameter settings indicated by the Title Record and Parameter Record in the Data Input File (and default settings of several other parameters).

Click the OK button, and an input option dialog appears. Indicate either Input Format 1 (for species as rows and sites as columns) or 2 (for sites as rows and species as columns. Indicate how many row and column headers to skip (1 of each).



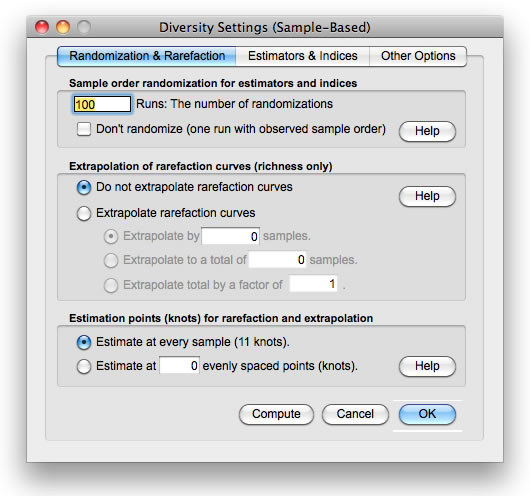
Click the OK button in the dialog. EstimateS completes the loading of the dataset and confirms that the file has been correctly loaded. (Input data errors will be flagged if they occur. Follow the onscreen instructions if this happens.)



**Setting and Running the Diversity Options**

Once the Data Input File has been loaded, the program is ready to set or check the Diversity options.

1. From the Diversity menu, choose Diversity Settings. The Diversity Settings screen appears. The defaults are generally what should be used, with the possible exception of selecting computation of Fisher’s alpha, Shannon and Simpson indices in the Estimators and Indices tab (see below).



Unless indicated otherwise (in the Other Options tab of the Diversity Settings screen or in the Shared Species Settings screen), EstimateS will remember whatever settings were last used, and display those as the default, although they may be overridden by Execution Control Parameters in the Data Input File.Col

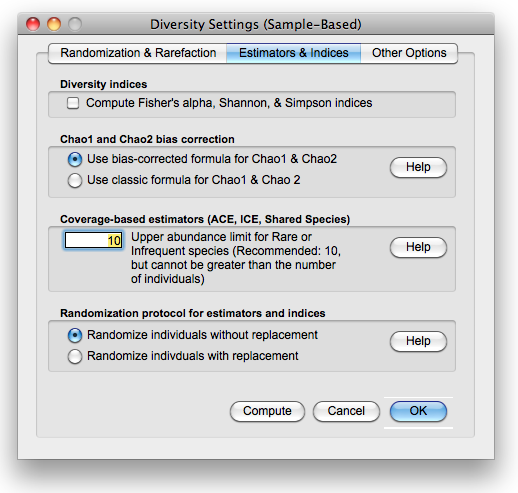
2. Set or check the options on the Randomization & Rarefaction Tab (illustrated above).

Sample order randomization for estimators and indices. Runs specifies the number of randomizations (resamples) to be carried out for rarefaction. Choose a reasonable number of randomizations (100 is usually enough) to get smooth curves for the estimators and indices. *Do not select the “Don't randomize” option.*

Extrapolation of rarefaction curves (richness only). If extrapolation is requested from the reference sample by selecting the "extrapolate rarefaction curves" option, EstimateS will estimate the expected number of species that would be found in an augmented sample. *Extrapolation should not be selected for routine analyses.*

Estimation points (knots) for rarefaction and extrapolation. EstimateS gives one a choice between computing, displaying, and exporting rarefied (and extrapolated) richness, asymptotic richness estimators, and diversity indices for every sample increment (the classic EstimateS approach); or, instead, computing, displaying, and exporting these statistics for a smaller number of sample increments, spaced at approximately even intervals along the rarefaction (and extrapolation) curve. The sampling points for the second approach are called "knots." *For routine analyses, just choose the traditional option and estimate for every sample increment.*

3. Set the options on the Estimators and Indices Tab.



Diversity Indices (Fisher's alpha, Shannon, Simpson). By default, the Compute Fisher's alpha, Shannon, and Simpson indices box is unchecked, so it must be manually selected if these indices of diversity are desired for rarefied subsamples of the reference sample. If this box is checked, be sure to indicate multiple runs (100 is suggested) on the Randomization and Rarefaction tab, so that the means among runs will produce a smooth rarefaction curve for the diversity indices.

EstimateS 9 computes Shannon exponential, as well as the Shannon information statistic. Simpson diversity is computed in its inverse form. Thus, EstimateS 9 computes the first three Hill numbers, for rarefied subsamples of the reference sample: q = 0 (richness), q = 1 (Shannon exponential diversity), and q = 2 (Simpson inverse diversity). Note that richness is computed analytically, whereas Shannon and Simpson diversities are computed by resampling.

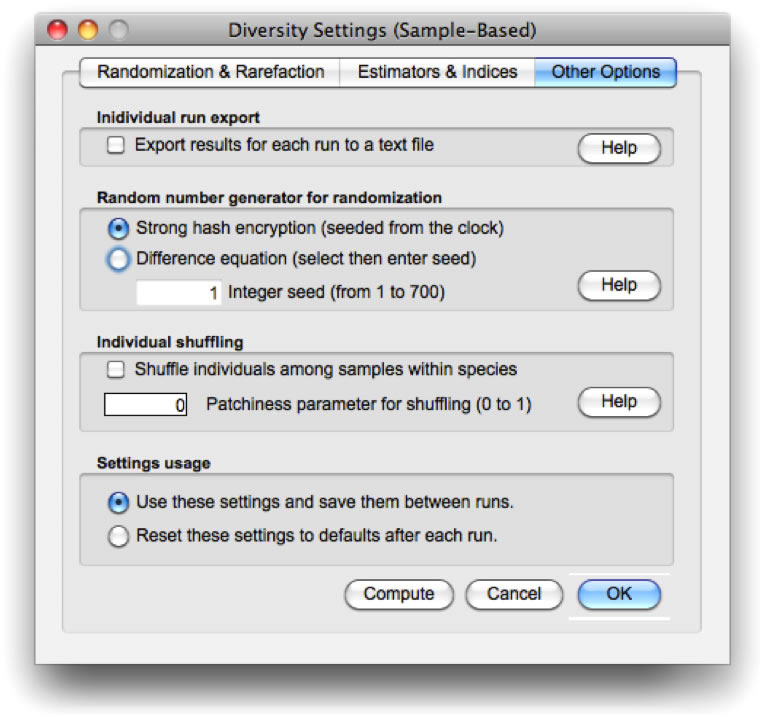
Chao1 and Chao2 bias correction. By default, EstimateS uses the bias-corrected form of the Chao1 and Chao2 richness estimators in all cases (the recommended default). If one chooses "Use classic formula for Chao1 and Chao2," instead, EstimateS uses the bias-corrected form only when either doubletons (Chao1) or duplicates (Chao2) are zero, and uses the approximate ("classic") formulas otherwise.

Coverage-based estimators (ACE, ICE, Shared Species). The recommended (and default) upper limit for Rare or Infrequent species is 10 individuals or 10 samples, respectively.

For cases in which all Rare species are Singletons, ACE is undefined. Likewise, for cases in which all Infrequent species are Uniques, ICE is undefined. EstimateS uses the bias-corrected form of the Chao1 and Chao2, respectively, for such cases. This setting also controls upper limit for Rare or Infrequent species for Shared Species estimation.

Randomization protocol for estimators and indices. Randomization without replacement is the default, which is highly recommended.

**4. Set the options on the Other Options tab.**



Individual run export. As an option, EstimateS records and exports results from individual randomizations to a text file, allowing computation of precision, accuracy, and other analyses, using Excel, R, or other applications. If one checks the "Export results for each run to a text file" checkbox, when the Compute button is clicked (or Compute Diversity from the Diversity menu is chosen), EstimateS displays an explanatory message, and requests the name and location of the text file that will contain the exported results when the randomizations are complete. The data for each randomization appear in the same format as the summary Diversity results that EstimateS creates by default. (The summary results appear onscreen as usual, and may be exported as usual.)

Random number generator for randomization. EstimateS offers two random number generator. The Strong hash encryption generator samples from a 160-bit strong hash (SHA) encryption function, seeded from the computer's clock. This procedure produces a non-repeating random number series that passes the most demanding tests.

The Difference equation alternative is based on a seed number that is supplied. Thus, it permits EstimateS to generate precisely the same results on repeated sets of resampling runs with the same dataset. Unless precise repeatability is required, *the strong hash encryption option is recommended.*

Individual shuffling This tool allows users to explore the effects of spatial patchiness on species richness estimators. If the "Shuffle individuals among samples within species" box is checked, EstimateS reassigns individuals at random to samples, within species, with a "tunable" degree of aggregation (patchiness). This is a research and simulation tool, not an estimator, and *this option should not be selected for routine analyses*.

Settings usage (saving settings). If one wants to save the current settings (the default) from one use of EstimateS to the next during a session, select "Use these settings and save them between runs." If one wants to start with default settings the next time the Diversity or Shared Species settings screens are opened, choose "Reset these settings to defaults after each run." Each time EstimateS is launched, all settings are returned to defaults.

5. Launch the Diversity computations.

To launch the Diversity computations directly, click the Compute button on the Diversity Settings screen, or click the OK button to save the settings, then choose Compute Diversity Stats from the Diversity menu. The results are displayed in the Diversity Statistics output screen.

6. Export the results of the Diversity computations.

To export the results of the Diversity computations to a tab-delimited text file, click the Export button at the bottom of the Diversity Statistics output screen or choose Export Diversity Stats from the Diversity menu. The exported file can be opened in Excel, R, or another application to analyze and plot the data.

7. (Optional) Export the input data and all current parameter settings to a tab-delimited text file.

**Interpretation**

EstimateS computes many diversity statistics, and the output contains estimates of all statistics for all sample sizes from 1 to the total number of samples. (It is not possible to compute only a subset of statistics or sample sizes; the program automatically calculates all diversity statistics for all sample sizes.)

To obtain species richness estimators based on all sites sampled (which should usually be the case), simply select the last row of the appropriate column. The Chao2, ICE and ACE species richness estimators all performed similarly in pilot tests of EstimateS with HTLN bird monitoring data. The use of one of these is thus suggested. After choosing an estimator, the same estimator should be used consistently for all years of data.

Note: The columns reporting Fisher’s alpha, Shannon and Simpson indices will all be zeros unless the “Compute Fisher's alpha, Shannon, and Simpson indices” box is checked, in the “Estimators and Indices” tab.

The user’s manual contains a detailed description of all the diversity statistics computed by EstimateS.