

Standard Operating Procedure #10:

Multi-Year Data Analysis

Version 1.0

In Bat monitoring protocol for the Great Lakes Inventory and Monitoring Network: Version 1.0

Katy Goodwin

National Park Service
Great Lakes Inventory and Monitoring Network
2800 Lakeshore Drive East
Ashland, Wisconsin 54806

Month 2020

Please cite this publication as:

Goodwin, K. 2020. Standard operating procedure #10: Multi-year data analysis (version 1.0). *In* K. R. Goodwin. 2020. Bat monitoring protocol for the Great Lakes Inventory and Monitoring Network: Version 1.0. Natural Resource Report NPS/GLKN/NRR—2020/XXXX. National Park Service, Fort Collins, Colorado.

Revision History Log

The following table lists all edits and amendments to this document since the original publication date. Information entered in the log must be complete and concise. Users of this standard operating procedure will promptly notify the GLKN protocol lead and/or project coordinator about recommended and required changes. The protocol lead and/or project coordinator must review and incorporate all changes, complete the revision history log, and change the date and version number on the title page and in the header of the document file.

Previous Version #	Revision Date	Author (with title and affiliation)	Location in Document and Concise Description of Revision	Reason for Change	New Version #
Add rows as needed for each change or set of changes tied to an updated version number					

Contents

	Page
1.0 Introduction.....	1
1.1 Parameters Measured	1
Activity Levels	1
Spatial and Temporal Distribution	2
Species Richness	2
1.2 Preparing Data	3
1.3 Analyzing Data	5
Manual Vetting.....	5
Activity Levels	5
Spatial and Temporal Distribution	6
Species Richness	6
1.4 Reporting.....	7
1.5 Potential Additional Analyses.....	9
Literature Cited	11
Appendix A. Error Rate Example	13

1.0 Introduction

The Great Lakes Inventory and Monitoring Network (GLKN) is implementing a monitoring program to document changes in bat populations in national park units in the upper Midwest. We are using acoustic monitoring as an index for assessing long-term bat population trends in response to ecosystem changes including climate, habitat, and disease such as white-nose syndrome (WNS). The goals of this monitoring program are to assess probable presence and absence of bat species, and to determine relative activity levels of those species for each national park unit over time. This Standard Operating Procedure (SOP) describes how data will be analyzed to compare results from multiple years and identify population trends.

1.1 Parameters Measured

Acoustic monitoring, although a very useful method of studying bat populations, cannot provide an estimate of abundance. When examining recordings of bat echolocation calls, individual animals are indistinguishable. For example, if we have ten recordings from the same location, all identified as little brown bat calls, we cannot determine if they were produced by ten different little brown bats, or by one little brown bat passing the microphone ten times. Therefore, rather than quantifying abundance we focus on several other parameters, as described below: bat activity levels, spatial and temporal distributions, and species richness.

Activity Levels

Bat activity levels are measured using the metric of *call files per deployment night*. As defined for this project, a “call file” refers to an audio file that has been determined to contain bat echolocation sounds and is identified to a specific species. A “call sequence” is defined as a series of vocalizations made by one individual bat as it passes the microphone (Loeb et al. 2015). Although it is possible for an audio file to contain multiple call sequences, for this study we are equating one call sequence to one call file.

A “deployment night” is defined as the period from the evening of one day to the morning of the following day, when one acoustic detector was deployed. Simultaneous deployment of multiple detectors increases the number of deployment nights. For example, if three different detectors are deployed at three different sample sites, all on 10 June, this would be equivalent to three deployment nights, not one deployment night.

Bat activity level can be assessed at the site-level, park-level, and for all parks combined. Activity level can be calculated for each species individually, for all species combined, or for smaller groups of species such as the *Myotis* group. The value is calculated by taking the total number of call files identified to that species or species group, divided by the total number of deployment nights over which those call files were collected. These calculations should use the results of the automated species classifications, with accompanying data specifying the software error rates derived from manual vetting.

Spatial and Temporal Distribution

The distribution of a particular species may include both spatial and temporal components. Spatial distribution is assessed by calculating the proportion of sample sites within a park where a given species was detected, for each year of sampling. Similarly, temporal distribution is assessed by calculating the proportion of deployment nights at a park during which a given species was detected, for each year of sampling. These calculations should use the results of manual vetting rather than automated classifications. A species is considered to be “detected” if at least one call file from that site or deployment night is identified to that specific species (identification to a species group is not counted).

Species Richness

Species richness is calculated by totaling the number of species detected per park or per sample site during each year of sampling. These calculations should use the results of manual vetting rather than automated classifications. As above, a species is considered to be “detected” if at least one call file from that park or site was identified to that specific species (identification to a species group is not counted). Mean, minimum, and maximum species richness per park or per site would also be useful to report.

1.2 Preparing Data

Data processing procedures are described in SOP #8. These procedures will result in the exclusion of files that were recorded during deployments that did not meet protocol specifications. This includes the following:

- Deployments are excluded if they do not have at least four valid nights of recording, with a valid night defined as having a minimum of six hours of recording. The duration of recording is determined by examining the summary file automatically generated by the Song Meter, which indicates the times at which the device turned on at night and off in the morning. If the summary file is not available, duration is determined from the date/time of the first and last audio files.
- Deployment length is cut off at a maximum of sixteen nights, with any data recorded after the sixteenth night excluded from analysis.
- Deployments are also excluded if less than four nights fell within the protocol sampling period (1 June–15 August). If at least four nights fell within these dates, and additional nights extended beyond, all nights in that deployment will be included.

Additional filtering steps must be performed prior to multi-year analyses. The project database can be used to identify files or sites that fall into these categories. First, all files classified as non-bat noise or unknown bats by Kaleidoscope Pro must be excluded from analyses, since our focus is on status and trends of each particular species. Secondly, for some analyses it may be appropriate to exclude any sample sites where a successful deployment was not completed in the same exact location every year. This includes situations where a year's deployment was unsuccessful due to equipment failure or logistical challenges. It also includes situations where the location of the sample site was moved significantly (i.e., if the site name changed from "A" to "B").

1.3 Analyzing Data

Analytical methods are constantly changing and improving. The methods described below may eventually need to be adjusted if better methods become available. Updates may be needed both in terms of the selected statistical tests and the software used to run the tests/models. As of this publication, it is recommended that statistical analyses be performed using the software program R (R Core Team 2019).

Manual Vetting

Manual vetting on a minimum of 1% of call files per park/species combination per year will be performed as described in SOP #8. The results of this manual vetting will be used to evaluate the error rate associated with Kaleidoscope Pro's species classifications. We will calculate the proportion of call files that underwent vetting that resulted in an identical species-level classification by both the software and the manual reviewer, versus the proportion that did not have identical classifications. Classifications will not match in cases where the software identified a call file to a particular species, but the manual reviewer determined it to be a different species, or non-bat noise. The resulting error rate will account for the software's false positives but not for false negatives. An example of error rate calculations is provided in Appendix A.

Software error rates should be calculated in this manner for each species and each year separately. Then, the error rates can be applied to the raw counts of call files, to obtain "corrected" counts that remove the influence of false positives. For example, if we calculated an error rate of 20%, we would adjust the raw count downward by 20% to remove the 20% of call files that are estimated to be false positives. The corrected count will be a conservative number representing the estimated *minimum* number of call files attributed to that species. The corrected counts can be used for activity level trend analysis, as described below.

In addition to calculating error rates associated with the Kaleidoscope Pro classifications, we will also use the results of manual vetting to create a confusion matrix to illustrate the relationship between the software and manual classifications. Counts in each cell of the confusion matrix will indicate the number of call files classified to each species by the software and by the manual reviewer. The confusion matrix will show where agreements and disagreements most commonly occur, and will incorporate classification categories beyond a single species, for example if the manual reviewer assigns a call file to the *Myotis* group.

Activity Levels

Changes in bat activity levels can be analyzed using linear mixed effects modeling. This type of modeling examines the relationship between a response variable and a set of fixed and random effects. Briefly, fixed effects are variables being tested for, while random effects are variables being controlled for. For this project, we will use the bat activity level (call files per deployment night) as the response variable. The fixed and random effects could include variables such as park, sample site, species, sample date or week, year, and the interactions among them. Including sample date or week is particularly important to account for the presumed increase in activity due to young bats becoming volant in mid-season.

The first step will be performing preliminary model selection to evaluate different possible combinations of fixed effects, with the goal of identifying the best-fitting model. During model selection, all candidate models will have the exact same response variable and random effects, to allow valid comparisons. Models can be compared to each other using Akaike's Information Criterion corrected for small sample size (AICc). Each model is given an AICc "score" with lower scores suggesting better fits.

Once the best-fitting model has been selected and run, calculate 95% confidence intervals around each of the model coefficients. If a confidence interval overlaps zero, this suggests that the variable does not have a significant effect on bat activity level. Multiple comparisons hypothesis testing can also be performed to complement the modeling results, for example to identify differences among parks, species, or years. The hypothesis testing results will include a p-value to help assess the significance level of the differences.

Spatial and Temporal Distribution

Spatial and temporal distributions can be compared among years, for each park/species combination, using the z-test for equal proportions. This test is appropriate because our occupancy data is in the form of proportions (e.g., proportion of sample sites where a species was detected). The null hypothesis will be that the proportions are equal. The alternative hypothesis will be that the proportions are not equal.

For spatial occupancy, the Yates' continuity correction must be applied when the expected number of sites with detections and/or sites without detections is less than five. Likewise, for temporal occupancy, the Yates' continuity correction must be applied when the expected number of deployment nights with detections and/or deployment nights without detections is less than five.

If a significant result is obtained, the z-test for equal proportions should be followed up with pairwise comparisons to identify which particular pair(s) of proportions were significantly different. Include a correction for multiple comparisons (e.g., Bonferroni).

Species Richness

Compare mean species richness between two years using the two-tailed paired samples t-test. A paired samples test is necessary because we are using repeated measures, sampling the same sites from year to year. This test will evaluate whether there is a difference between the two means (positive or negative).

1.4 Reporting

The multi-year report should include summary tables describing the following, for each park and each year:

- Survey effort (number of sites, number of deployment nights, average nights per site)
- Survey results (total number of files recorded that met protocol requirements, number of files that were identified to a specific species vs. noise vs. unknown bats)
- Species-level results (number of files identified to each species by Kaleidoscope Pro, indication of which species were/were not verified present through manual vetting, error rates per species obtained through manual vetting, total species richness per park, mean species richness per park)
- Species distribution results (number and percent of sites where each species was detected; number and percent of deployment nights on which each species was detected)

The report should also include results of modeling and statistical tests. For bat activity modeling, identify what candidate models were evaluated, provide AICc scores, and describe how the final model was selected. Also specify what variables were used in the final model. Provide a table showing model coefficients with upper and lower 95% confidence intervals, plus z-values and p-values from the hypothesis testing. For spatial and temporal distributions and species richness, indicate which tests had significant results and provide the associated values (chi-square, degrees of freedom, and p-value for occupancy; t-value, degrees of freedom, and p-value for species richness). Additional tables, figures, and explanatory text should be incorporated as needed.

The report should clearly distinguish between results of automated species classifications and results of manual vetting. It should also include the species-specific error rates derived from manual vetting and indicate if raw versus corrected counts of call files were used in the analysis. If automated classification detects a species outside of its known range, this should be manually verified prior to publication, or the report should include a qualification that this detection is unverified. The report should also include a statement indicating that automated species classification software is not perfect and results may include misclassifications (false positives and/or false negatives).

1.5 Potential Additional Analyses

Project personnel may decide that additional analyses should be included in the future. For example, the activity level, spatial and temporal distribution, and species richness measurements described above do not account for species that may have been present but went undetected during our sampling. They also do not account for possible misidentification of call files. Constructing false-positive occupancy models is one analytical method that would better incorporate these problematic factors. Methods for occupancy modeling are comprehensively described in MacKenzie et al. (2018). Application of occupancy modeling specifically to bat surveys is outlined in protocols developed for NABat (Loeb et al. 2015) and the Northwestern Bat Hub (Rodriguez et al. 2019) and demonstrated by Banner et al. (2019) and Clement et al. (2014). Note that occupancy modeling will require additional manual review of call files beyond the 1% minimum suggested in our protocol. Determination of which call files to review and how many files to review will depend on the modeling approach used; refer to Banner et al. (2018).

Another potential area of analysis would be community composition. It may be informative to examine whether species assemblages, relative activity levels, or activity patterns over the course of a season or night are changing, especially because some bat species are highly susceptible to white-nose syndrome and others are not affected. Previous researchers have examined this question (e.g., Ford et al. 2011, Pettit and O’Keefe 2017, O’Keefe et al. 2019, Nocera et al. 2019) and their work could provide a model for how to analyze GLKN bat data. Related to this and the above assessment of species richness, we could also construct species accumulation curves to determine how many nights of sampling are required to capture recordings of the full species assemblage at a site. This would help us evaluate the sufficiency of our sampling effort (Skalak et al. 2012, Richardson et al. 2019).

Future work could also include analysis of bat activity in relation to season, habitat, or weather. Evaluation of seasonal activity patterns could be accomplished by setting up acoustic detectors to record year-round. This would be logistically difficult in some parks but would potentially offer the opportunity to assess when bats emerge from hibernation, when young become volant, or when migration begins. Data collected for this project already include general information about the habitat type at and surrounding each sample site. More detailed habitat data could potentially be derived from existing land cover and vegetation datasets, aerial imagery, or lidar. Finally, although we are not collecting weather data for this project, it could be obtained from existing sources such as RAWS weather stations or state climatology offices.

Literature Cited

- Banner, K. M., K. M. Irvine, T. J. Rodhouse, W. J. Wright, R. M. Rodriguez, and A. R. Litt. 2018. Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecology and Evolution* 8:6144–6156.
- Banner, K. M., K. M. Irvine, T. J. Rodhouse, D. Donner, and A. R. Litt. 2019. Statistical power of dynamic occupancy models to identify temporal change: Informing the North American Bat Monitoring Program. *Ecological Indicators* 105:166–176.
- Clement, M. J., T. J. Rodhouse, P. C. Ormsbee, J. M. Szewczak, and J. D. Nichols. 2014. Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology* 51:1460–1467.
- Ford, W. M., E. R. Britzke, C. A. Dobony, J. L. Rodrigue, and J. B. Johnson. 2011. Patterns of acoustical activity of bats prior to and following white-nose syndrome occurrence. *Journal of Fish and Wildlife Management* 2:125–134.
- Loeb, S. C., T. J. Rodhouse, L. E. Ellison, C. L. Lausen, J. D. Reichard, K. M. Irvine, T. E. Ingersoll, J. T. H. Coleman, W. E. Thogmartin, J. R. Sauer, C. M. Francis, M. L. Bayless, T. R. Stanley, and D. H. Johnson. 2015. A Plan for the North American Bat Monitoring Program (NABat). General Technical Report SRS-208. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, North Carolina.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. Bailey, and J. E. Hines. 2018. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Second Edition. Academic Press.
- Nocera, T., W. M. Ford, A. Silvis, and C. A. Dobony. 2019. Patterns of acoustical activity of bats prior to and 10 years after WNS on Fort Drum Army Installation, New York. *Global Ecology and Conservation* 18:e00633. Available at: <https://doi.org/10.1016/j.gecco.2019.e00633> (accessed 15 April 2020).
- O’Keefe, J. M., J. L. Pettit, S. C. Loeb, and W. H. Stiver. 2019. White-nose syndrome dramatically altered the summer bat assemblage in a temperate Southern Appalachian forest. *Mammalian Biology* 98:146–153.
- Pettit, J. L., and J. M. O’Keefe. 2017. Impacts of white-nose syndrome observed during long-term monitoring of a Midwestern bat community. *Journal of Fish and Wildlife Management* 8:69–78.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org/.
- Richardson, S. M., P. R. Lintott, D. J. Hosken, and F. Mathews. 2019. An evidence-based approach to specifying survey effort in ecological assessments of bat activity. *Biological Conservation* 231:98–102.

- Rodriguez, R. M., T. J. Rodhouse, J. Barnett, K. M. Irvine, K. M. Banner, J. Lonneker, and P. C. Ormsbee. 2019. North American Bat Monitoring Program regional protocol for surveying with stationary deployments of echolocation recording devices: Version 1.0, Pacific Northwestern US. Natural Resource Report NPS/UCBN/NRR—2019/1975. National Park Service, Fort Collins, Colorado.
- Skalak, S. L., R. E. Sherwin, and R. M. Brigham. 2012. Sampling period, size and duration influence measures of bat species richness from acoustic surveys. *Methods in Ecology and Evolution* 3:490–502.

Appendix A. Error Rate Example

Manual vetting was performed on a minimum of 1% of call files per park/species combination for GLKN data from 2015–2018. An overview of the manual vetting results for two species are shown below as examples to illustrate how the error rate associated with automated species classifications is calculated. These two species were selected as examples because they represent the range of low to high average error rates. We calculated the proportion of call files that underwent vetting that resulted in an identical species-level classification by both the software and the manual reviewer (“agreement rate”), versus the proportion that did not have identical classifications (“error rate”).

Table A1. Manual vetting results for Little Brown Myotis (*Myotis lucifugus*) for 2015–2018 data. For each year of data, all parks have been combined. Software classification was done using Kaleidoscope Pro.

Year	Total Number of Files Manually Vetted	Number of Vetted Files with Identical Classifications by Software and Manual Vetter	Agreement Rate	Error Rate
2015	889	568	0.64	0.36
2016	1602	1310	0.82	0.18
2017	683	573	0.84	0.16
2018	462	383	0.83	0.17

Table A2. Manual vetting results for Northern Myotis (*Myotis septentrionalis*) for 2015–2018 data. For each year of data, all parks have been combined. Software classification was done using Kaleidoscope Pro.

Year	Total Number of Files Manually Vetted	Number of Vetted Files with Identical Classifications by Software and Manual Vetter	Agreement Rate	Error Rate
2015	54	3	0.06	0.94
2016	153	41	0.27	0.73
2017	108	18	0.17	0.83
2018	90	14	0.16	0.84