

SOP #5. Data Analysis, version 1.0

Revision History Log:

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #

5.1 Purpose

This Standard Operating Procedure (SOP) explains methods for data analysis of Shenandoah salamander monitoring data contained in the NEARMI local database.

5.2 Scope and Applicability

This SOP applies to the data analyst.

5.3 Procedures and General Requirements

- This type of analysis is complex, and care should be taken to understand the principles behind it prior to applying it. Preferably, a quantitative ecologist or biometrician should perform these analyses. There are many ways in which the multiple visit data collected can be combined for analyses, and these alternatives should be discussed with someone familiar with options within the analysis program.
- Carefully read and review Mackenzie et al. (2005). This text provides the foundation for understanding the application and interpretation of the PRESENCE model for detection/non detection data. Carefully read and review Burnham and Anderson (2002). This text provides a foundation for understanding multi-model inference.
- Use program PRESENCE (downloadable from <http://www.mbr-pwrc.usgs.gov/software.html> or <http://www.proteus.co.nz/home.html>; MacKenzie et al. 2002, MacKenzie et al. 2003) to conduct routine (annual) and long-term trend analyses. The program PRESENCE uses maximum likelihood techniques to estimate the probability of detection (p) and the probability of occupancy (ψ) at a set of habitats within a defined area of inference (e.g., “historic” range). This program allows the inclusion of site and sampling covariates to test the association between abiotic and biotic variables on detection and occupancy.
- Site-specific covariates do not change across sampling occasions within years (e.g., many habitat variables) and may influence either occupancy or detection probabilities. Sampling covariates (e.g., temperature, weather) often change for each sampling occasion and may

primarily influence detection probability. Covariates are included in a set of candidate models, and these competing models are tested using Akaike's Information Criterion (AIC) to determine the model (and therefore, covariates) which best explain variation in the observed data. Because of the potential for a very large number of candidate models with the inclusion of a variety of covariates hypothesized to influence detection or occupancy of a species, much attention should be devoted to choosing relevant covariates prior to constructing models. Ideally the choice of covariates should be informed by *a priori* hypotheses and determined, by definition, before the initiation of sampling. Hypotheses about occurrence of the Shenandoah salamander include the influence of elevation, aspect, and micro-habitat covariates that are included in this monitoring program and will be incorporated into the candidate model set. Other covariates may be added over time, as observations and results suggest. Additional site-level and remotely-sensed covariates (e.g. percent microhabitat cover categories, forest type) can be related to estimates of occupancy and dynamics, as robust relationships with environmental covariates will be valuable for developing effective management strategies for the species.

- Refer to SOPs #4 to determine data field definitions and which data fields are applicable for analyses. Covariates used in constructing candidate models should reflect the *a priori* hypotheses outlined prior to collection of field data (i.e., found in the protocol narrative).
- The final data file (species matrix table) should consist of detection (1) and nondetection (0) data at each visit to each site for species detected during surveys (Table 5.1). This species presence/absence (1/0) data should be assembled or downloaded from the database in the following matrix format (Table 5.1) for each of the species observed and from the following database data fields: Species (e.g., Shenandoah salamander), Presence (e.g., 0 for absence, 1 for presence), Site Identifier (e.g., name of site – site 1, site 2, etc. where site is the combination of the site identifier and the replicate transect), and Visit Number (e.g., Occasion 1, Occasion 2, etc.):

Table 5.1. An example of the Shenandoah salamander species matrix table.

Site	Occasion 1	Occasion 2	Occasion 3	Occasion 4	Occasion 5
Site 1	0	1	1	1	1
Site 2	1	0	0	0	0
Site 3	0	0	1	1	1
Site 4	1	1	1	0	0
Etc.	...	...	...	...	...

- For each species matrix table, prepare corresponding Sampling Covariate Tables to match the species presence/absence matrix data (Table 5.2). For sampling covariates (covariates that change with each sampling occasion), which are used for analyses in relation to species

detection probabilities and site occupancy rates, use the following table format (air temperature in °C is used as the example here):

Table 5.2. An example of the Shenandoah salamander sampling covariate table.

Site	Occasion 1	Occasion 2	Occasion 3	Occasion 4	Occasion 5
Site 1	14.5	15.6	17.1	18.2	20.2
Site 2	16.0	14.2	17.8	19.1	20.9
Site 3	13.2	13.9	15.4	17.0	18.3
Site 4	12.2	12.9	14.1	15.6	17.8
Etc.	...	...	...	...	...

- For each species matrix table, prepare corresponding Site Covariate Tables to match the species presence/absence matrix data (Table 5.3). For site covariates (covariates that do not change with each sampling occasion), which are used for analyses in relation to site occupancy rates, use the following table format (e.g. distance of site from historic range in meters is used as the example here):

Table 5.3. An example of the Shenandoah salamander site covariate table.

Site	Site Covariate
Site 1	55
Site 2	19
Site 3	233
Site 4	5
Etc.	...

- Once all the tables are proofed, copy these tables into Program PRESENCE and conduct analyses. A variety of models will be tested. The simplest model ($p(.)\psi(.)$) assumes that site occupancy and detection probabilities are constant across time and sites. Alternative models include variation across time or sites. More complicated models will be used to test whether site occupancy and detection probabilities are influenced by a variety of site and sampling covariates (see habitat covariates in SOP #4), which have been specified in a set of *a priori* hypotheses to be updated each year based on the current available science. A goodness-of-fit test is available in program PRESENCE to assess how well the global or most parameterized model represents the data, and whether the data are overdispersed, in which case a variance inflation factor can be estimated and used to adjust standard errors and model selection procedures (Burnham and Anderson 2002; MacKenzie and Bailey 2004). Further descriptions of occupancy estimation methods are available elsewhere (MacKenzie et al. 2002, MacKenzie et al. 2003, MacKenzie et al. 2004; MacKenzie and Bailey 2004,

MacKenzie et al. 2005). Analyses are best performed by someone familiar with model selection techniques and with program PRESENCE specifically.

5.4 Reference Documents

- Burnham, K. P. and D. R. Anderson. 2002. Model selection and multi-model inference: A practical information-theoretic approach. New York, Springer-Verlag.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- MacKenzie, D. I., L. L. Bailey, and J. D. Nichols. 2004. Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology* 73:546–555.
- MacKenzie, D. I. and L. L. Bailey. 2004. Assessing the fit of site occupancy models. *Journal of Agricultural, Biological and Environmental Statistics* 9:300–18.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and, J. E. Hines. 2005. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press, New York, USA.