Survival of Spring Salamanders in Hubbard Brook Experimental Forest

Carter Adamson

FISH 558 – Modeling Project

[intro]

Salamanders are of great interest to stream management and conservation groups due to their status as biological indicators (Lowe et al. 2018). This is especially true in small streams, such as many rivers’ headwaters, where a variety of factors such as stream size and barriers may prevent fish from establishing as competitors or predators (Lowe et al. 2018). As such, salamander monitoring is a useful tool to gauge the health of these headwater streams. This may have implications for the health of the headwater ecosystem, including associated riparian flora and fauna, as well as for river health further downstream in a given system. Trends in the success of these salamanders could therefore be incorporated into indices of overall riparian and biotic health. Due to this, my goal for the modeling project was to find data that would allow me to model the survival of salamanders in small streams. I investigate potential covariates such as salamander sex, life stage, natal stream, exposure to predation, and physical health, hypothesizing that each could impact rates of survival.

[methods]

The data I used to answer this research question comes from the Environmental Data Initiative (EDI) online repository, which is a curated platform containing environmental and ecological data from a wide array of scientific and governmental sources (Lowe 2022). This is a mark-recapture dataset which includes an observation for every time that a salamander was marked or recaptured.

Table 1: Table summarizing the spring salamander mark-recapture dataset from Hubbard Brook Experimental Forest (Lowe 2022). Only variables relevant to this assignment are listed.

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Possible Values** |
| **ElastomerID** | VIE tag code | Unique string |
| **PITTagID** | PIT tag number | Unique number |
| **Date** | Date the salamander was captured | Dates from 6/20/2012 to 8/16/2021 |
| **SurNum** | Survey number | 1-72 |
| **Sex** | Sex of salamander | M, F, or U (unknown) |
| **Stage** | Life stage of salamander | A (adult), L (larva), or M (metamorph) |
| **Stream** | Stream where salamander was captured | Bear, Canyon, Paradize, or Zigzag |
| **Reach** | Reach where salamander was captured | Upper or lower |
| **TailRemoved** | % of salamander’s tail missing | Number from 0 - 100 |

This dataset comes from the Hubbard Brook Ecosystem Study, a collaborative research effort located within the 8,000-acre Hubbard Brook Experimental Forest in New Hampshire (Hubbard Brook Ecosystem Study 2024). For more than 60 years, ecologists in Hubbard Brook have been intensively studying the system and generating long-term datasets (Hubbard Brook Ecosystem Study 2024).

The study that my dataset is sourced from was located in five headwater streams (Bear Brook, Cascade Brook, Canyon Brook, Paradise Brook, and Zigzag Brook), all of which are tributaries to Hubbard Brook itself (Lowe et al. 2018). The study used mark-recapture methodology to gather data on the population of spring salamanders (*Gyrinophilus porphyriticus*), a species which is widespread in headwater streams of eastern North America (Hebron 2022). The study was conducted in June through September from 2012-2014. Each stream was surveyed nine times each field season over three years. Within each year, streams were sampled in a random order within two-week survey periods (Lowe 2022). Surveys were conducted visually, maintaining a consistent effort by checking one random salamander “cover object” per meter of stream (Lowe 2022). When caught, salamanders were marked with Visual Implant Elastomer (VIE) tags, using colors and body locations to generate a unique code for every individual. Partway through the study, the VIE tags were switched out for Passive Integrated Transponder (PIT) tags, which also contain a uniquely identifiable code. The reason for this switch is unclear but it should not impact modeling results.

The spring salamander itself is a lungless species found in small, cool east coast mountain streams (Hebron 2022). They have aquatic larvae and adults that have some ability to forage on land but are also mainly aquatic (Hebron 2022). The spring salamander is listed as threatened in many states and in Canada, but not federally in the United States or locally in New Hampshire (Hebron 2022). Like many salamanders, they are sensitive to pollution and habitat modification. Introduced brook trout, which the Hubbard Brook Ecosystem Study has found in the study system, can also pose risk to spring salamanders through predation on larvae (Lowe et al. 2018).

In order to assess the research question of what covariates impact the survival of this salamander population, I decided to run a Cormack-Jolly-Seber (CJS) model. This modeling approach is ideal for a question regarding survival rates, since the CJS has the built-in ability to explicitly account for the detection process. In ecology and fieldwork, it’s seldom the case that detection is perfect. With small juvenile salamanders, which are cryptic, aquatic, and often hide behind cover, detection rates may in fact be quite low. Without accounting for this, non-detections would be confounded with deaths, and the result would be a survival rate much lower than the truth (i.e. an individual not being seen does not mean that it is dead). In the project proposal, I had considered looking at abundance and/or growth in addition to or instead of survival. However, due to the nature of the data I was able to get my hands on, looking at survival with a CJS seemed most appropriate. CJS models require mark-recapture data, so it was a perfect fit.

However, that’s not to say that the data were already in the right format. I used the R software package RMark, which interfaces with Program MARK, in order to run the CJS models. RMark requires that its data be formatted in a very particular way, with capture histories condensed into a single string, and relevant covariates appended to the same row as the capture history they correspond to. This proved the sternest challenge of this entire assignment, since the data I obtained from EDI weren’t formatted in capture histories at all. Instead, each row corresponded to an observation of a salamander, regardless of whether it had been captured before. Salamanders that were recaptured therefore had their tag IDs listed more than once in the dataset. My challenge was to condense records of 7,021 observations into a unique capture history for each individual.

I did this by creating a new, empty data frame. I filled the first column with unique tag IDs, which itself was a slight challenge since there were two types of tag used, and thus two columns with tag IDs. Generally, PIT tagged salamanders had NA for VIE tags, and vice versa, though some salamanders were tagged both ways. I combined these into a single ID column, with the VIE taking precedent for salamanders who received both. I then created 72 blank columns on the new data frame, one for each sampling occasion, and filled them with zeros. Then, to generate the capture histories, I used a for loop that looked through each observation in EDI data, matched it to an ID and sampling occasion in the new data, and filled the appropriate box with a 1. In instances where sampling occasion was “NA,” I cross-referenced with the survey date to come up the proper occasion number. At the end of this process, each row of the new data frame contained a unique tag ID followed by a 72-column capture history. I then united those 72 columns into one string with no spaces or commas to create capture history in RMark’s preferred format.

With each individual now having its own capture history, I then had do assign covariates to them. Generally, this was a simple matter of adding covariate columns and looping through the original dataset to add the appropriate values. I did run into some issues, however. There were many, many NAs for the “Sex” variable, even though there was a “U” option for unknown sex built into the schema. I simply replaced these NAs with Us. I also changed the “Reach” variable to “Trout,” since the study design classified “upper” and “lower” reaches by trout presence. Those designations were determined a priori before the original study using long-term monitoring data from the Hubbard Brook Ecosystem Study. In this study design, upper reaches are by definition free of brook trout, and they are present in reaches labeled lower (Lowe et al. 2018).

An issue I ran into later was that RMark just wasn’t taking numeric covariates. After several unsuccessful attempts to troubleshoot this, I decided to work around the issue by converting my numerical covariate to a categorical one. To do this, I replaced the “% Tail Missing” covariate with a new one best described as “Tail Damage”, where any salamander with % Tail Missing > 0 was assigned to the group “Damaged”, and any salamander with 0% Tail Missing was assigned “Full.” There is some information lost when changing from continuous to categorical here, but I am still able to ask the question of whether visibly injured salamanders have different survival rates.

Based on the data available and a priori hypotheses, I generated 15 candidate models. I included life stage, stream, sex, tail damage, and trout presence as potential covariates affecting the probability of survival (Φ). It’s easy to imagine different life stages surviving at different rates, with larvae being more vulnerable to mortality. Different streams may also have different survival rates – for example, a stream nearer the edge of the Hubbard Brook Experimental Forest might experience more anthropogenic impacts impacting survival such as pollution. Physical damage and exposure to predators are other factors that I hypothesize may impact survival. As for sex effects, there is some sexual dimorphism in spring salamanders, with females being slightly bigger (Hebron 2022). Bigger individuals may be better survivors, since salmonids trout are gape-limited predators. I also allowed survival rates to vary over time, across the months and years of the study.

As for covariates affecting probability of detection (p), I whittled down the list of covariates to just stream. Different streams may have different morphology or turbidities that could impact detection. I also allowed detection rates to vary over time. Below are examples of equations for survival and detection that include all of these potential covariate effects:

I conducted model selection after the analysis, collecting 15 models that I tested and comparing them with AICc. To determine candidates for the top model, I chose an arbitrary cutoff of AICc = 4.

[results]

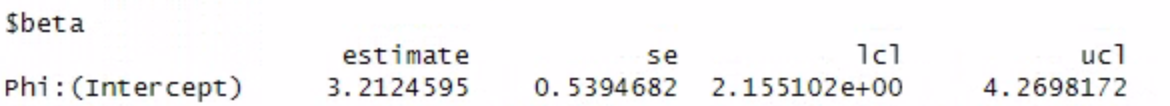
Table 2: Model selection table for the 15 a priori models considered.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model No.** | **Covariates** | **No. params** | **AICc** | **AICc** | **Weight** |
| 1 | Φ(t+sex+stage+tail+stream+trout), p(t+stream) | 156 | 16012.11 | 0.00 | 1.00 |
| 6 | Φ(t+stage+tail+stream+trout), p(t+stream) | 154 | 16179.40 | 167.29 | 0.00 |
| 9 | Φ(sex), p(t+stream) | 78 | 16346.64 | 334.53 | 0.00 |
| 7 | Φ(t+stage+stream+trout), p(t+stream) | 153 | 16413.32 | 401.21 | 0.00 |
| 13 | Φ(tail), p(t+stream) | 77 | 16424.83 | 412.72 | 0.00 |
| 8 | Φ(t+stage+stream), p(t+stream) | 152 | 16444.44 | 432.33 | 0.00 |
| 10 | Φ(stage), p(t+stream) | 78 | 16510.71 | 498.60 | 0.00 |
| 11 | Φ(stream), p(t+stream) | 80 | 16636.69 | 624.58 | 0.00 |
| 14 | Φ(t+stream), p(t+stream) | 150 | 16646.57 | 634.47 | 0.00 |
| 15 | Φ(trout), p(t+stream) | 77 | 16646.85 | 634.74 | 0.00 |
| 2 | Φ(.), p(t+stream) | 76 | 16687.93 | 675.83 | 0.00 |
| 12 | Φ(t), p(t+stream) | 146 | 16704.81 | 692.71 | 0.00 |
| 5 | Φ(.), p(t) | 72 | 16746.33 | 734.23 | 0.00 |
| 4 | Φ(.), p(stream) | 6 | 16856.45 | 844.34 | 0.00 |
| 3 | Φ(.), p(.) | 2 | 16890.91 | 878.80 | 0.00 |

Table 2, above, summarizes the findings of the model selection process. Of the 15 models I tested, only one had a AICc value beneath the cutoff of 4. As such, I have only generated one candidate for the best model. This model includes every covariate effect I described in the Methods section. Survival is affected by time, sex, life stage, tail damage, stream, and trout presence. Detection is affected by time and stream. The next best model includes all of the same covariates, but without the sex effect on survival, and performs significantly worse (AICc = 167).

Reporting the results of survival and detection for this top model is somewhat complex, due to the model calculating 156 separate parameters, where each combination of the five categorical variables also varies with time. This is very unwieldy to report with a table or figure, so I will report only a few values here. For example, survival for an unknown sex, tail-tamaged individual in a trout-free stretch of Bear Creek had a survival rate of Phi = 0.835 at t=3 (the third survey), but the same individual’s survival rate would have been Phi = 0.886 if it were known to be female.

Table 3: Model summary output table from RMark, curated from n = 156 estimates for ease of interpretation.



----- [additional timesteps trimmed]

A number on a white background

Description automatically generated

-----[timesteps trimmed]

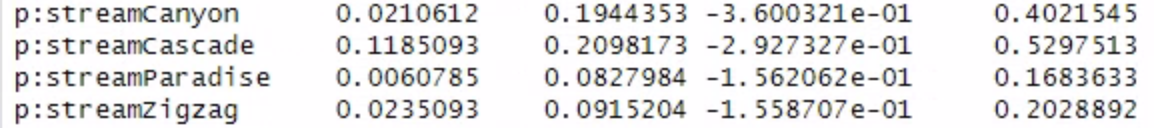


Table 3 consists of pieces of model output that I selected to help communicate the results. Again, since this model is so complex, it’s difficult to pick out direct patterns or produce simple, useful tables. That said, we can clean some useful information from this output. We can see that Male and Unknown sex individuals generally have lower survival than females (for example, the coefficient estimate for males is -0.099 less than that of females). The second column is a modifier for the intercept value that is associated with that particular level of a categorical variable. For example, Canyon and Cascade Brooks have a negative effect on survival when compared with Bear Brook, which is being used to generate the intercept estimate. Further, individuals with “full” tails (i.e. no tail damage) have higher survival, and trout presence is associated with lower survival.

Due to my research question focusing on rates of survival, I am less interested in directly reporting and interpreting probability of detection. That said, I still have included some information about p in Table 3. Of the five streams, it’s apparent that Cascade Brook has the highest coefficient estimate for detection probability, and Bear Brook has the lowest.

Model assumptions for the CJS model include equal survival among individuals, no influence of tagging on capture probability or mortality, no lost tags, and independent fates of individuals. Individuals may have different survival rates if they are of different sexes or life stages, so building that variation into the model lets us account for this assumption. The impacts of tagging on the output parameters, and the possibility of tag loss, are both assumptions that would have to be addressed in the study design. Without having more information or getting in contact with the authors of the study, I cannot know whether there was tag loss or whether tags impacted detection probability. However, I believe it is safe to assume that there is negligible tag loss since VIE and PIT tagging are widely used procedures in ecology, and that these tags don’t impact detection probability because they are very small modifications to very small animals. With either type of tag, you would have to detect the animal first to detect the tag.

An assumption that I am less confident is being met is the assumption of instantaneous sampling, whereby there is no possibility for mortality or movement during a sampling occasion. I don’t have very much confidence in this assumption because each “sampling occasion” in this case was a two-week window. However, some sources suggest that mark-recapture studies are fairly robust to violations of this assumption (Lindberg 2012).

[discussion]

Though complex and difficult to show all at once, the results from this model did offer support to the hypothesis that sex, life stage, stream, damage, and trout presence can impact salamander survival. Specifically, undamaged female metamorphs in Zigzag Brook upstream of trout presence have high predicted rates of survival compared to individuals that have different levels for those variables. Contrastingly, damaged male adults in Canyon Brook in a trout-occupied reach have all the indicators for lower survival. Survival also varies with time, but without a clear pattern.

These results match up with my hypotheses well, in most cases. I expected that females would have higher survival rates due to their larger body size, and that body damage and trout presence would be associated with lower survival. I also predicted that survival would vary by stream and over time, which this model also supports. The results for life stage go against my modeling – specifically, the adult life stage is the one with the lowest survival rates. I would suggest that this is because of a limitation of the CJS model – the inability to tell between mortality and emigration. To be most accurate, what I have termed “survival” is actually “apparent survival,” which confounds those two processes. This is especially relevant in the case of adult salamanders, where juveniles have small home ranges and are limited [source], while adults are capable of moving out of the stream to forage. Though adult spring salamanders do not exhibit migratory behavior, even the possibility of adult salamanders moving out of the stream to forage coinciding with a survey suggests that survival estimates for these adults may be biased low.

A major limitation of my analysis here is that I was unable to analyze this data under the proper robust design framework of the study. This study was designed with a temporally nested survey, where surveys are nested within primary occasions, which correspond to the year in this case. I was unable to get a robust design mark-recapture model working in RMark, and so opted to ignore that aspect of the data. Instead, I simply treated each occasion as independent and sequential. The result of this is that the gaps between surveys can be wildly inconsistent, with two weeks elapsing after some surveys and nine months after others. At the very least, this certainly gives the salamanders time to move and undergo demographic processes in a way that may disrupt the assumption of instantaneous sampling, and make the stepwise time covariate calculated by RMark inaccurate. Incorporating the robust study design would be a critical step to improve the model.

I would also suggest that a better set of candidate models might be acquired to better tease apart the relative strength and importance of the covariates. By fitting a full model and comparing it to versions of the full model with only one covariate removed, we might be able to get a better idea of which covariates are the most important. However, my a priori list of candidate models did not incorporate this consideration and were chosen mostly on ecological and coding-related grounds. I wasn’t able to address this issue of a poorly chosen candidate model set myself, since the models were fairly time-intensive to run, and I was limited to a machine that was not my own since Program MARK (and therefore RMark) don’t function on Macs. However, if I were to do this project over again, I think that doing so could be informative. Determining the relative importance of covariates on salamander survival would be valuable information for a management team trying to determine how best to target restoration or protective actions.