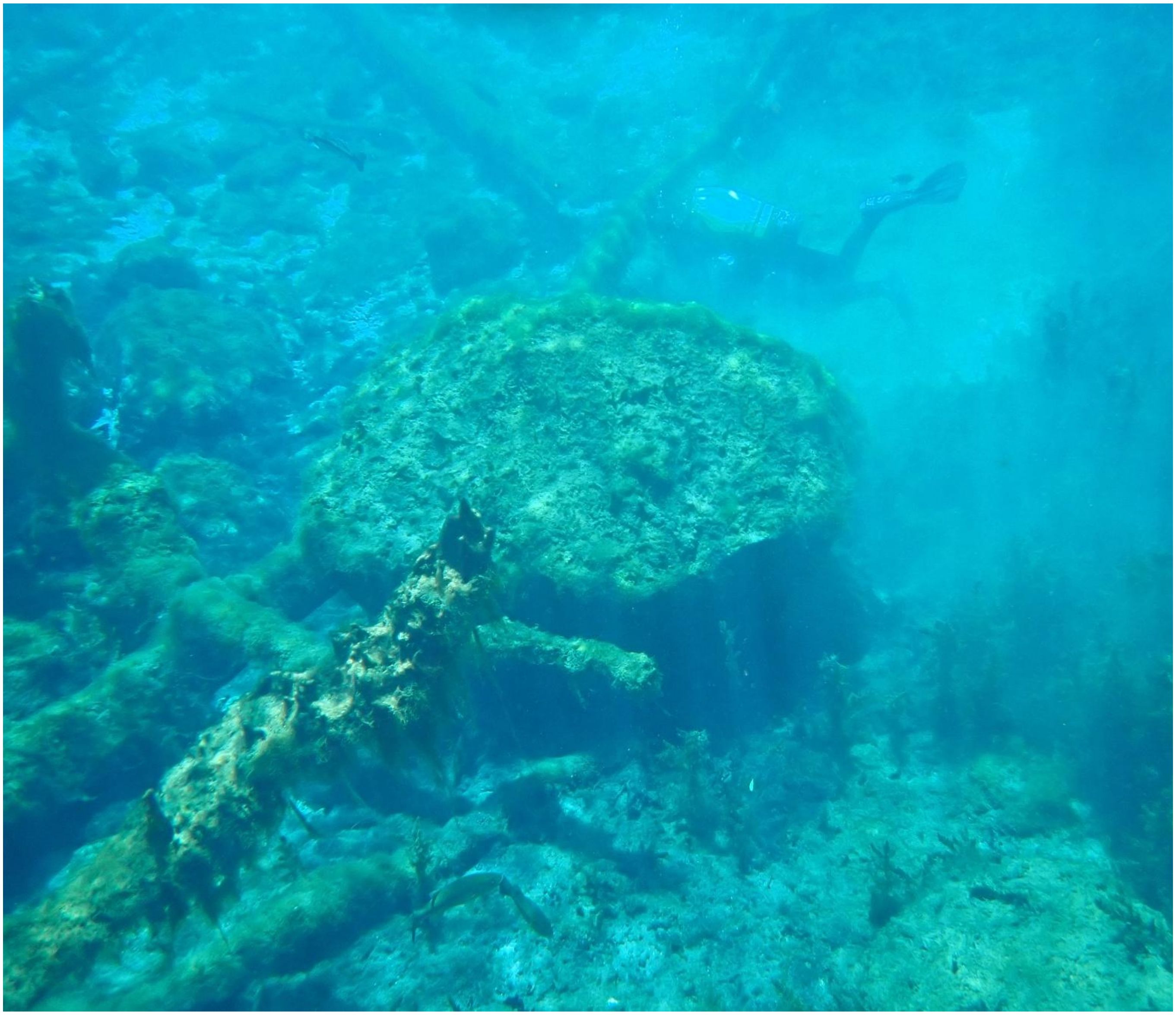
Spring Lake Sampling 2025 Results and 2026 Study Design and Recommendations



Contents

Peter Diaz
Texas Fish and Wildlife Conservation Office
San Marcos Texas, 78666

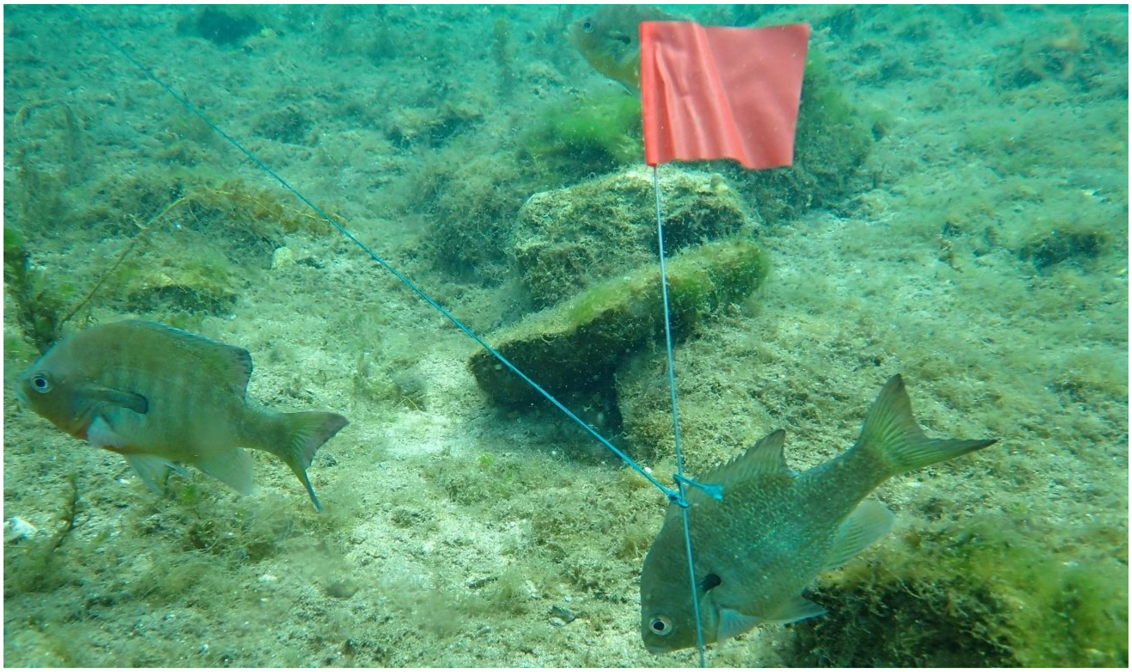
Sommerly Swarm, Justin Crow, and Randy Gibson
San Marcos Aquatic Resources Centroid
San Marcos Texas, 78666

Hannah Gilbreath and Matt Johnson
Austin Ecological Services
Austin Texas

Dr. Robin E. Russell
Science Applications

Tables

Figures



Introduction

The upper San Marcos River is a spring-fed ecosystem located along the margin of the Edwards Plateau in Central Texas. Spring Lake (SL) is the hydraulically controlled part of the headwaters of the San Marcos River. The remaining part of the headwaters is the area just on the other side of the Spring Lake Dam. There are two outlows over Spring Lake Dam, the eastern (ES) and the western shores. For this project we are sampling SL and the Eastern Shore (ES) where salamanders are known to be present. The artesian spring water from the Edwards Aquifer emerges into the lake from ~200 openings. The headwaters of the San Marcos River harbor a relatively large number of federally listed taxa, including the San Marcos salamander (*Eurycea nana*), Texas wild-rice (*Zizania texana*), the fountain darter (*Etheostoma fonticola*), and the Comal Springs riffle beetle (*Heterelmis comalensis*). The San Marcos salamander is a small-bodied (<60 mm total length), plethodontid salamander that has an exclusively aquatic life cycle and is limited to the San Marcos River headwaters (Chippindale et al. 1998).

The USFWS and the state of Texas currently lists the San Marcos salamander as threatened and the critical habitat is designated as the headwaters of the river (SL and ES) and the first 50 m of the river below the Spring Lake Dam at the end of Spring Lake (USDI 1980); this area is thought to encompass virtually all the geographic range of the San Marcos salamander. Although the San Marcos salamander can be abundant in portions of the headwaters (e.g., Tupa and Davis 1976), a limited geographic range and threats of declining ground water levels led to its listing (USDI 1980). Estimation of the population size of San Marcos salamanders in the San Marcos River headwaters region has been attempted multiple times (Tupa and Davis 1976; Nelson 1993; Lucas 2006). Habitat associations and distribution of the San Marcos salamander within the headwaters of the San Marcos River were documented by Diaz et al. (2015). Diaz found that only 7% of SL was occupied by *E. nana*. Laboratory studies subsequently found that San Marcos salamanders exhibit a relatively low upper thermal tolerance, indicating adaptation to thermally stable environments (Berkhouse and Fries 1995) and prefer relatively slow current velocities (1cm/s; Fries 2002).

Monitoring of the San Marcos salamander from 2002 – 2015 was analyzed by three different consultants in addition to the Nation Academy of Science (NAS 2018). Although, some correlations were exhibited in the data, in most cases the correlations were “likely coincidental and not cause and effect" (Beaver Creek Hydrology LLC). Suggestions by the NAS propose using occupancy estimates to gather additional data to inform the biological goals. Results from all the analysis indicate improvements could be made to the study design and easily implemented in order gather the data necessary to answer the questions asked by Biological Objectives.

The goal of this project is to is to develop future field sampling and analytical protocols for a San Marcos salamander monitoring program that can effectively monitor occupancy and abundance trends for the surface population of the San Marcos salamander in the headwaters of San Marcos River.

2025 Prescribed Sampling Methods

The design incorporated the distribution of the San Marcos salamander and divided SL into four zones, which reflect differences in percent area of available salamander habitat and salamander densities (Diaz et al. 2015). Within each zone, sampling occurred at areas previously established as salamander habitat by Nelson (1993) and Diaz (2015) (Figure 1). The nine areas (polygons) to be sampled were selected based on the following criteria: 1) classified as occupied habitat areas with associated EAHCP Biological Objectives and 2) distributing survey effort by the amount and number of salamander occupied areas in each zone.

Approximately two weeks prior to the first sampling event of each sampling period, FWS staff coordinated with Meadows Center staff to perform aquatic vegetation site maintenance at three areas in Zone 1 (i.e., Hotel, Diversion, and Riverbed). These three areas have proposed habitat-based coverage goals for Biological Objectives (BOs) for the San Marcos salamander. The purpose of the vegetation maintenance is to maintain targeted amounts of suitable salamander habitat at each of the three areas.

To maintain closure between events in a primary period the duration between sampling events for each primary period were kept close. Sampling occurred on April 21st, 25th, and 29th for the first primary period. The second primary period was on May 19th, 23rd, and 27th. The final primary period was on June 16th, 20th, and 24th. Monday, Friday and Tuesday were the preferred option for sampling days.

Study Sites

The number of transects per polygon depended on the approximate amount of salamander habitat available (Table 1). The random points generator tool in ArcGIS Pro's "Toolbox" was utilized to identify specific site locations within designated polygons. The generator was instructed to maintain 5 m buffer around all sites. These maps were transferred to a Trimble Geo X7, a submeter device, to establish sites for each primary period. Two days before the first sample date of each primary period divers and surface support determined the locations of sites from the map and marked the transect with flags and line between. Sites (targets) were selected on the GPS device map and navigated to by zooming in when necessary to pinpoint the location. If sites were overlapping by coming into a range of 5 m of another or perceived to be heading out of the polygon, the initial downstream placement of the following flag was changed to upstream, left, or right, if necessary and in that order. Once set, these transects were left in place for the duration of each primary period. Due to faster flows and activity below the dam, sites at that polygon were marked the day of the first survey for each primary period. These transects were removed at the end of sampling each event. There were only four transects in this ES polygon each primary period and the sites visited on the first event were easily recognizable from the first time given the short duration between events. Time and effort expenditure was considered when determining the total number of transects (i.e., 20-25) since surveying should be completed in a day. With only one sampling area currently identified in Zone 3, exploratory surveying to identify any other potential salamander habitat that could be used as an additional sampling area was done.

Prior to the search for salamanders the transect should be quantified by the diver for substrate categories (Table 2; Figure 2) and percentages and vegetation species. For vegetation divers calculate the percentages of each species summing to a total of 100% of the vegetation present, then a percentage of vegetation as a whole in the transect. For example, a quadrat could be 100 percent filamentous algae with 20% quadrat coverage.

After the transect has been described the search for salamanders can begin. The diver makes counts of adult and juvenile (<1 in) salamanders. A clicker was used to count adult salamanders. The search is active in that rocks, debris, and vegetation are disturbed and moved to visually observe the salamanders present. Duration of the search time is recorded for each transect. Divers noted the presence of springs along the transect. The spring presence score was binary (1 = Spring). The spring size score ranged from 1 – 3. A one was given if the spring was diffusive, no real spring orifice was observed. A two was given for small orifices about the size of a fist. A three was given for a larger orifice. In areas where the substrate may be stacked, like a rock pile, the searchers will disturb substrates up to ~3 inches deep. Following the search in the transect for salamanders an embeddedness score should be made. The categories for embeddedness are presented in Table 3. Depth and flow should be measured after the search. Depth was measured at around center of the transect. Flow was measured along the transect, at the middle of depth for the transect and at the bottom, at a distance about the height of the probe off the substrate and allowed to log a 20 second cycle.

Statistical Analysis

Basic examination of site level covariates and count data were examined using Pearsons correlation coefficient. There were no strong correlations (±0.5) between the averaged count data per transect and any of the covariates measured. The average salamander/m^2 for each transect and each polygon were calculated for each primary period. To calculate the average salamanders per meter squared the count of salamanders per an individual transect was averaged then divided by five. These totals were then averaged for each primary period providing a salamander/m^2 estimate each time as a whole and per polygon. The final salamander/m^2 estimate was calculated as the average from the averages per primary period. Standard errors are reported for polygons averages.

Principal component analysis (PCA) was run using the "prcomp" function to identify spatial variation between measured abiotic parameters in the sections of SL surveyed with the intent of trimming down variable for abundance analysis. Parameters included in the analysis were: depth, substrates percentages, vegetation percentages, vegetation species percentages, embeddedness score, presence of a spring, spring size, and zone. Prior to PCA analysis the data was z-scored.

To examine occupancy (ψ, psi) and covariates that may influence the probability of detection (p), MacKenzie et al. (2002) single season models were run using the program Presence (Hines 2006). Count data was transferred to a binary data set (1's and 0's). Site covariates included substrate percentages, depth, vegetation types and percentages, transect overall vegetation percentage, embeddedness, spring size and spring presence, and zone. No covariate data was transformed for these models. Models examined the probability of detection and how the covariates may influence the detection of the salamanders.

Model selection was based upon the Akaike's Information Criterion (AIC). The AIC score is a way of ranking models using parsimony (Akaike 1973). Typically, a lower AIC score, the better fit of the model to the data. In addition to the AIC score, models were also evaluated for performance by significance of covariates in the model and ranges in confidence intervals of estimates. Once a model with the appropriate probability of detection covariates determined we began to analyze the data set for occupancy using the site covariates and the selected probability of detection covariates. Models were also created with the assessment of the goodness of fit (GOF) function in Presence.

To determine the sites level estimates of abundance within the sampled transects, closed and open N-mixture models were run (Royle 2004; Dail and Madsen 2011) in R using the package "unmarked" (Fiske and Chandler 2011) and the codes "pcount" and "pcountOpen". Site covariates were the same for the abundance analysis as with the occupancy analysis. Initial examination of the data set was conducted to determine which covariates influence the probability of detection. For N-mixture models the probability of detection is calculated differently than for occupancy models. The selected covariates were then used to create a suite of abundance models. Most models were run using either the Poisson or negative binomial distribution. For negative binomial models K was initially selected by the computer then allowed to range in other models to determine changes in abundance. Closure was examined using the community parameter (c) which ranges from 0 – 1. The scores quantifies proportion of individuals shared across visits, a score closer to 1 shows closure (Hyun et al. 2025). Model selection was based upon the AIC, significance of covariates in model and goodness of fit. All abundance models used the "AICcmodavg" package in R to run GOF tests. All R code will be submitted as an appendix to this document.

Other methods to determine site level and polygon abundance included running generalized linear mixed models (GLMM). We used generalized linear mixed models to examine how measured covariates might predict salamander abundance using count data as the response variable with random effects of site and transect. These models were run in the package "glmmTMB" and "AICcmodavg" all in R to calculate site and polygon level abundance estimates. The package "DHARMa" (residual diagnostics for hierarchical regression models) was used to look at model fit for the GLMMs. Covariates (fixed effects) for the GLMM's were determined by the highest positive and negative loadings from PCA axis one and two on the Lake data. For GLMM's the data was switched to a long format (135 rows) and using transects and the sampling events as random effects. To determine random effects REML was marked as

“TRUE” to determine the AIC score (Faraway 2006). The “nbinom2” family which handles overdispersal was used for all models with “ziformula = ~ 0” with a log link. For all other fixed effects it was left out of the equation, as FALSE is the default. To examine variance partitioned in the random effects the intraclass correlation coefficient (ICC) adjusted and unadjusted was calculated overall and for groups. The adjusted ICC measures how much variance is explained by the random effects and unadjusted accounts for both fixed and random effects. The ICC score will also be spilt out by groups, quantifying the portion of the adjusted value to each group.

2025 Results

Three primary periods with three sampling events for each were conducted in the months of April, May, and June of 2025 from SL and the ES. A total of 57 samples were collected, 45 from SL, and 12 from the ES. Each site with a total of three visits. During each primary period the conditions at the sites remained the same (e.g., no large rain events, etc.). There was an increase in flow between the second primary period and the third. The average flow of the San Marcos River was 87, 81 and 101 for each month respectively.

Dive time spent on each transect varied based on salamander observations (Figure 3). The longest transect took about 20 minutes, while the shortest took around 25 seconds. On average, a 5m x 1m transect took slightly over 8.5 minutes, or 1.56 minutes per meter. The maximum time per meter was 3.47 minutes and the minimum was 1 minute. Divers completed surveys at 19 sites using a single tank, averaging about 47 minutes of dive time.

In total over 2,100 E. nana (n = 2,127) were collected during sampling events in 2025, with the vast majority being collected in SL (n=2,121) and the remainder collected in the ES (n = 6). Of the total salamanders collected 1,542 were adult and 585 were considered juvenile (<1 in). In April, 77% of all transects were occupied with 763 salamanders observed, with 255 of those considered juveniles. In May, 75% of all transects were occupied with 960 salamanders observed, with 268 of those being juveniles. In June, only 57% of the transects were occupied with 404 salamanders observed, with 62 of those sampled being juvenile.

The Hotel polygon had the most salamanders observed followed by the River Bed polygon. A rough estimate of abundance in the polygons based on the sal/m^2 density of each polygon, not accounting for detection, a total of 3,438 salamanders would be expected in the sampled polygons (Table 4). Therefore, Zone 1 would have 1,449 salamanders estimated to be in the Hotel and the Diversion polygons. In Zone 2 there would be 1,989 salamanders in the sampled polygons. These estimates of abundance should be viewed as more of an exercise and not an actual estimate of the salamanders present within the Zones or polygons given the low sample size and inherent heterogeneity in each polygon.

Results from the PCA of SL and the ES showed major separation between the two different mesohabitats (Figure 4). Results from the PCA explained 22% of the variance by PC I and another 11% in PC II for a cumulative score of 33% of the variance explained by PC axis II. Zone, *Z. texana*, and *Hydrocotyle* sp. had the highest negative loadings on PC I while springs, spring size, and depth had the highest positive loadings. The gradient on PC II shifted from negative loadings of silt, embeddedness, and *Hydrocotyle* sp. to positive loadings for cobble, sand and *Sagittaria* sp. Based on the total PCA analysis another separate PCA was run on just SL polygons.

Results from the Lake PCA explained 16% of the variance by PC I and another 14% in PC II for a cumulative score of 30% of the variance explained by PC axis II. Filamentous algae, Zone, and depth had the highest negative loadings on PC I while *Ceratophyllum* sp., cobble, and *Sagittaria* sp. had the highest positive loadings (Figure 5A). The gradient on PC II shifted from negative loadings of gravel, filamentous algae, and depth to positive loadings for spring presence, spring size and *Sagittaria* sp. There is a small separation between the habitats in Zone 1 and Zone 2. For example, the upper right quadrat of the PCA space is dominated by the habitat in the Hotel polygon (Figure 5B). Many of the measured habitat variables were correlated as shown by the strength and direction of individual vectors.

Occupancy models were run using the pooled 45 site data set from April, May, and June. Many of the top models had issues with the output of the models. The issues included no estimates for covariates, large standard error of occupancy covariates, no individual estimates of psi, and no significance for other covariates. The fourth model, with depth in the probability of detection parameter (Table 5) was the only model with a significant covariate ($p < 0.05$) that presented reasonable estimates of psi and exhibited a model fit ($p = 0.15$). The estimate of psi was 0.93 (0.81 - 0.97) and averaged site detection was estimated at 0.91 (0.81 – 0.95) for the fourth model. The fifth model with the addition of embeddedness to the probability of detection parameter was very similar to the fifth model in terms of their AIC scores, however, embeddedness was not significant. No covariates were shown to describe occupancy within the transects sampled during these sampling periods.

The N-mixture abundance models used the pooled 45 site data set. Initial models were run with a Poisson distribution. Closure test of the community parameter was 0.91; P(>|z|) = <0.0001; suggesting the primary periods were closed (Appendix A1). All Poisson models produced estimates ranging from 1,200 to around 2,000 for the sum of the 45 site level estimates. All Poisson models produced large c-hats = 23- 39 and had a lack of fit. Due to the high ranges in c-hat the distribution of the models was shifted to negative binomial models to decrease overdispersion shown in the Poisson models. Negative binomial models significantly reduced c-hat estimates to around 1.3 – 1.8. Estimates of site level abundance increased and in turn the overall sum for each of the binomial models compared to their Poisson counterparts.

Negative binomial models were run with covariates to determine the significant ones that affect the probability of detection. The computer was allowed to select K (the upper limits) for the creation of the candidate models. Covariates influencing the probability of detection were Zone and Springs. C-hat in these two detection models ranged from 1.44 to 1.17 respectively. The two covariates were then examined in combination with each other, and this produced a model with significant covariates for detection and the lowest AIC score among the candidate models influencing detection with this data.

The model was then run with covariates hypothesized to influence abundance. Covariates run included cobble, gravel, rocks, silt, subcode (sum of rocks / embeddedness score), total vegetation, and other ratios. As shown in Table 6, the model with the lowest AIC was where cobble was in the abundance parameter and Springs and Zone were in the detection parameter. This model passed the GOF test with a c-hat of 1.32 and a *p*-value = 0.18. Residual plots seem to show a good scatter (Figure 6).

Once the selected model was established the examination of K within the model was tested at different levels (Figure 7). The computer selected a K of 248 during the initial model runs. K was changed from 1,000 to 150 to test relationships with N, GOF tests, and the probability of detection with the models (Table 7). The selected NB model was unidentifiable as the estimates of N increased when K was adjusted.

Other models were run using the “pcountOpen” function which allows a relaxation of the closure assumption to compare the estimates of abundance and detection with the selected NB model (Figure 8). The open NB model of the selected NB closed model produced similar estimates of abundance (open = 2,256 to closed = 2,744) and a better detection (open average = 37 to closed average = 27; Table 8). The open model still showed fit with a c-hat of 1.3 and GOF *p*-value = 0.16. Log likelihood scores for the open model were -578.77 and for the closed model -572.83. However, the open NB model also proved to be unidentifiable.

Due to these infinite estimates, potentially due to large count data, in the NB models the zero-inflated Poisson distribution was run to examine the effects on overdispersion. However, the models still had larger than expected c-hats ranging from 10 – 12 for the top models. Models never showed any fit through GOF tests.

Generalized linear mixed models were initially investigated with the global model to determine the random effects of transect and site individually, along with the combination of the two. Models with transect and site as random effects had the lowest AIC score of 949. The AIC score for transect was 953 and the AIC for site was 1008. Therefore, all proceeding models were run with the random effects of transect and site in the equation.

Sixteen models were run to examine the influence of the fixed variables and determine fit of models (Table 9). The third lowest AIC model with Springs, Cobble, Gravel, Depth and *Ceratophyllum* sp./*Cabomba* sp. was selected. The top two models showed degrees in their lack of fit based on quarterly residual analysis. The second model had an average quartile p-value of 0.14 however, the individual residuals showed a lack of a fit (Figure 9). The third model exhibited better fit based on the higher average quartile p-value of 0.34 and passed all individual quartile tests (Figure 10). Dispersion p-values were both insignificant.

Transect level estimates for the third model ranged from 0.19 to 103.6 salamanders per transect. Estimates by primary period were 231.70 salamanders for April, 288.01 salamanders for May, and 137.38 salamanders for June of 2025. The total predicted by the model was 657 salamanders (±225; Table 10). The Hotel polygon had the highest salamanders per m^2 followed by the River Bed polygon. Zone 1 had more salamanders per m^2 than Zone 2 (1.2 sal / m^2; 0.81 sal / m^2; respectively; Table 11). The predictions tended to track the count data (Figure 11).

Random effects standard deviation estimates were larger for transect (stdev = 1.09; 0.82 / 1.44); than site (stdev = 0.756; 0.305 / 1.87). This is expected as the transect group (45) is larger than the site group (7). The ICC adjusted score of 0.84 and an unadjusted score of 0.62 highlight the similarity among the groups used. Transect (0.567) as a random effect had a larger proportion of the variance than site (0.273) based on the group ICC scores.

Executive Summary of 2026 Sampling Design

The main goal of the study is to develop future field sampling and analytical protocols for the San Marcos salamander monitoring program that can effectively monitor occupancy and abundance dynamics for the surface population of the San Marcos salamander in Spring Lake (SL) and the upper portion of the San Marcos River. Data collected during the prescribed sampling in Spring of 2025 (Task 1) was analyzed (Task 2) to prepare for the sampling season in Spring of 2026 (Task 3). This document provides the guidelines and recommendations for sampling in Spring of 2026 (Task 3) at SL in San Marcos Texas for *Eurycea nana* (San Marcos salamander).

Originally SL was divided into 4 zones to establish the distribution of *E. nana* across the lake systematically (Diaz et al. 2015). Diaz et al. (2015) determined that *E. nana* distribution within SL was isolated to areas where course type substrates (i.e., gravel and cobble) were present. This translated to about 7% (~6,086 m^2) of SL being occupied by *E. nana*. A longitudinal distribution pattern also appeared in SL where there were more salamanders closer to the headwaters and decreased going downstream. The exception was below the dam in the Eastern Spillway (ES) where numbers increased again. For these reasons the idea of zones is going to be dropped for this sampling in 2026. A measurement of distance to the headwaters will be used for each site to model the potential longitudinal distribution in SL. Modelling and analysis will be separate for SL and the ES.

One large problem of the study design is that the area covered in SL where the salamanders are known to occur was too small, around ~1,195 m^2, around 20% of the available habitat for *E. nana* in SL. In addition, most of the transects in the polygons were always occupied. Although it is important to monitor the areas where the salamanders occur it is also important to sample a gradient of salamander densities to populate the models and provide a range in count data over as many different habitats as possible. Sampling at similar habitats within different sections of SL will help remove the potential polygon or site effect on abundance and allow for the ranges in abundance to be gathered along different longitudinal points in SL.

Based on previous published literature (Diaz et al. 2015) we know the areas where salamanders are residing in SL and can sample those areas probabilistically. Doing this can only be done by covering more of the 7% (~6,086 m^2) available area in SL where the salamanders occur (Diaz et al. 2015). The total area available for sampling in the new polygons is around 2,500 m^2, around 42% of the available *E. nana* habitat in SL. Although sampling isn't probabilistic across the entire lake, the polygons to be sampled are representative of the available salamander habitat and cover around 42% of the distribution in SL, based on previous work.

Analysis of the 2025 sampling data has provided insights for developing recommendations for the Spring 2026 design. The 5m transects in 2025 created large heterogeneity, making site categorization difficult. In 2026 we will use ½ m^2 quadrats to sample the study sites. There will be 50 to 60 quadrats in SL and 23 in the ES per event. Therefore, from 150 to 180 quadrats from SL and 69 quadrats from the ES all sampled three times. The quadrat will create fixed area to sample and not arbitrary boundaries that may change due to diver or whims. Additionally, the categories for the embeddedness scores were too large in our 2025 data collection. The categories for the embeddedness metrics will be modified to smaller categories in 2026 (Table 3), and an additional category will be added to separate the large category of gravel (Table 2). One final check will be the group calibration of the dive team, in regard to individual diver estimates of substrates and vegetation categories and percentages.

N-mixture models have been shown to have issues with indefinability for negative binomial models. In the case of the 2025 data exhibiting overdispersion and in some cases with individual counts two orders of magnitude over other counts, the issue of "large" counts might be a problem and a reason for the issues with the N-mixture models. Switching to the ½ meter quadrats should handle the issue with larger count data. For example, 148 is the largest count for a transect, in the Hotel polygon, with reasonably homogenous habitat. That means, if the salamanders are distributed evenly, there around 9 salamanders per meter for this transect. So, a count for a ½ meter quadrat might be somewhere around 3 to 9 salamanders. There might be an area where we get lucky and land on a site with maybe 30, but that still is closer to the actual mean of the survey than the larger 148 count.

2026 Methods

As in the prescribed sampling, polygons will be sampled and were selected based on the fitting two of the following criteria: 1) classified as occupied habitat areas with associated EAHCP Biological Objectives; 2) distributing survey effort by the amount and number of salamander occupied areas in SL and 3) part of the old river channel. Sampling will occur within established polygons within the old river channel and other salamander occupied areas in SL and the ES (Figure 12).

Approximately two weeks prior to the first sampling event of each sampling period, coordination with Meadows Center staff is necessary to inform them of the sampling event and allow them to perform aquatic vegetation site maintenance at three areas in headwaters section (i.e., Hotel, Diversion, and Riverbed). Team members will be going with the Meadows team to map and observe the areas cleaned. The larger polygons in the Hotel, the River Bed, and the Diversion sites should allow for comparisons between maintained and non-maintained sites.

Study Sites

The number of quadrats or sites per polygon depended on the approximate amount of salamander habitat available (Table 1). The random points generator tool in ArcGIS Pro's "Toolbox" will be used to identify quadrats or site locations within designated polygons. The generator tool was instructed to maintain 5 m buffer around all sites. These maps will be transferred to a Trimble Geo X7, a submeter device, to locate sites in SL and the ES for each primary period. Before the first sample date of each primary period divers and surface support determined the locations of sites from the map and marked the location with a flag. Sites were set as targets on the GPS device map and navigated to by zooming in when necessary to pinpoint the location (Figure 12). Once set, these flags were left in place for the duration of each primary period and each flag was established as the upper left corner of the quadrat. Due to faster flows and activity below the dam, sites in the ES polygon will be marked before the sampling day and left in place over each primary period, however, the method for the marking is yet to be determined. There will be 50 to 60 quadrats in SL and 23 in the ES per event. Therefore, from 150 to 180 quadrats from SL and 69 quadrats from the ES all sampled three times. Time and effort were considered in 2026 when determining the total number of sites since surveying should be completed in a day.

Before each survey begins the dive team will use three quadrats to standardize each other at the starting polygon. In the shallow water of the Hotel polygon the dive team will gather. A quadrat will be placed on the substrate. Then each member of the dive team will determine their individual scores for the substrate and vegetation percentages and types. Then the team will share their scores and compare them with each other to get synced into how the work will proceed.

Each site is demarcated by a ½ m^2 quadrat. At the previously placed flag, the upper left corner of the quadrat should be placed with the flag inside that corner and the sides running parallel to the flow or in the upstream direction. Prior to the search for salamanders the quadrat should be quantified for substrates and vegetation types and percentages. After the quadrat has been described a 2.5 minute search for salamanders can begin. In areas where the substrate may be stacked, like a rock pile, the searchers will disturb substrates up to ~3 inches deep. Following the search in the quadrat for salamanders an embeddedness score should be made. Depth and flow should be measured after the search. Depth will be measured at the center of the quadrat. Flow will also be measured at the center of the quadrat, at a distance about the height of the probe off the substrate and allowed to log a 20 second cycle.

Statistical Analysis

Basic examination of site level covariates and count data will be examined using Pearsons correlation coefficient. Strong correlations (±0.5) between the averaged count data per quadrat and any of the covariates will be noted for univariate analysis. The average salamander/m^2 for each quadrat and each polygon will be calculated for each primary period. To calculate the average salamanders per meter squared the count of salamanders per an individual quadrat will be averaged then multiplied by 2. These totals will be averaged for each primary period providing a salamander/m^2 estimate each time as a whole and per polygon. The final salamander/m^2 estimate will be calculated as the average from the averages per primary period.

Principal component analysis (PCA) will be run using the “prcomp” function to identify spatial variation between measured abiotic parameters in the sections of SL surveyed, to reduce variables for models, and to examine correlations. Parameters will include in the analysis: depth, substrates percentages, vegetation percentages, vegetation species percentages, embeddedness score, presence of a spring, spring size, and zone. Prior to PCA analysis data will be z-scored.

To examine occupancy (ψ, psi) and covariates that may influence the probability of detection (p), MacKenzie et al. (2002) single season models will be run using the program

Presence (Hines 2006). Count data will be transferred to a binary data set (1's and 0's). Site covariates will include substrate percentages, depth, vegetation types and percentages, quadrat overall vegetation percentage, embeddedness, spring size and spring presence, and zone. There are no data transformations needed for these models. Models will examine the probability of detection and how the covariates may influence the detection of the salamanders.

Model selection will be based upon the Akaike's Information Criterion (AIC). The AIC score is a way of ranking models using parsimony (Akaike 1973). Typically, a lower AIC score, the better fit of the model to the data. In addition to the AIC score, models will be evaluated for performance by significance of covariates in the model and ranges in confidence intervals of estimates. Once a model with the appropriate probability of detection covariates is determined we will analyze the data set for occupancy using the site covariates and the selected probability of detection covariates. Model validation will use the assessment of the goodness of fit (GOF) function in Presence.

To determine the sites level estimates of abundance within the sampled quadrats, closed and open N-mixture models will be run (Royle 2004; Dail and Madsen 2011) in R using the package "unmarked" (Fiske and Chandler 2011) and the codes "pcount" and "pcountOpen". Site covariates will be the same for the abundance analysis as with the occupancy analysis. Initial examination of the data set will be conducted to determine which covariates influence the probability of detection. For N-mixture models the probability of detection is calculated differently than for occupancy models. The selected covariates will then be used to create a suite of abundance models. All models will be run using the Poisson and Negative Binomial distribution. For negative binomial models K will initially be selected by the computer then allowed to range in other models to determine changes in abundance and determine identifiability. Closure will be examined using the test code from Dail and Madsen (2010). Model selection will be based upon the AIC, significance of covariates in model and goodness of fit. All abundance models will use the "AICcmodavg" package in R to run GOF tests. All R code will be submitted as an appendix to this document.

Other methods to determine site level and polygon abundance will include running generalized linear mixed models (GLMM). Generalized linear mixed models with random effects will be run in the package "glmmTMB" and "AICcmodavg" all in R to calculate site and polygon level abundance estimates. The package "DHARMa" will be used to look at model fit for the GLMMs. Covariates for the GLMM's will be determined by the highest positive and negative loadings from PCA axis one and two on the Lake data. For GLMM's the data will be switched to a long format and using quadrats and the sampling events as random effects. To determine random effects REML will be marked as "TRUE" to determine the AIC score. For all other fixed effects it will be left out of the equation, as FALSE is the default. All other covariates will be considered fixed effects. To examine variance partitioned in the random effects the intraclass correlation coefficient (ICC) adjusted and unadjusted will be calculated overall and for groups. The adjusted ICC measures how much variance is explained by the random effects and unadjusted accounts for both fixed and random effects. The ICC score will also be spilt out by groups, quantifying the portion of the adjusted value to each group.

Part of the final analysis will examine how many sites are required at a minimum to create viable and powerful models. Sample size for N-mixture models has been shown to be around 45 at a minimum. Sample sizes for GLMM's should also be over 30 for each sampling event. This is a minimum and will only allow simple models to be run. For complex models a larger sample size will be required. In the end the sample size will be based on what type of question is asked going into the future.

Discussion

Large scale results from the 2025 data set were very similar to Diaz et al. 2015 regarding the available habitats, the habitat differences between SL and the ES, and the longitudinal densities of salamanders decreasing farther away from the Hotel polygon (headwaters) into Zone 3. There is a large difference in the count data from Diaz et al. 2015, Comal Springs data (Diaz et al. 2018; federal report unpublished), and the data collected for this round of sampling in 2025. The largest count for Diaz et al. 2015 in a $1m^2$ quadrat was 20 salamanders. For the Comal Springs data the largest count in a ½ m^2 quadrat was seven. The largest value for the count data using the 5m long transect with an approximate width (no real grid was created, the divers estimated shoulder width for the transect) of 1m was 148 salamanders. The large range in count data between transects created overdispersion within the data set. This created issues when attempting to fit Poisson distributions. In addition to the overdispersion issues, the N-mixture models have been shown not to perform well with larger count data as in our data set (Parker et al. 2022; Madsen and Royle 2023). The proper statistical analysis will be determined by the data collected in 2026. I predict lower count data overall given the fixed and smaller prescribed "site", and the ability to use N-mixutre models with the Poisson distribution. If the N-mixture models fail for some reason, the 2025 data set showed pliability with the GLMM approach.

The long-term monitoring average for the Hotel area from the 2024 Biological Monitoring Program is 15.5 sal/m^2 with the long-term average for the River Bed area at 14.5 sal/m^2. The data from the Spring of 2024 was 15.5 sal/m^2 for the Hotel and 12 sal/m^2 for the River Bed. The count data from the Spring 2024 sampling primary periods was calculated at 5.6 sal/m^2 for the Hotel and 3.96 sal/m^2 for the River Bed. The model estimates from the third GLMM were similar to the count data with 5.11 sal/m^2 for the Hotel and 3.73 sal/m^2 for the River Bed. In the methods for the Biological Monitoring Program a macrophyte area is haphazardly selected as a site with no mention of site size. Therefore, the sites they selected for each sample in any year might be hypothesized to have larger abundances of salamanders due to the site habitat type selected (i.e. rocks). In the sampling design for 2025 and 2026 the sites are randomly selected and will include suboptimal habitat (smaller gravel, silt, macrophytes), therefore the estimates for sal/m^2 might be expected to be smaller than the reported historical sal/m^2.

Table 1. Polygons to sample for *E. nana* in Spring Lake and the Eastern Spillway in 2025 and 2026.

2026	**Area (m^2)**	**Percent**	**Samples to Take**	**Samples**	
Hotel	1034.6	40.29	19	19.34	23.37
Diversion	171.75	6.69	3	3.21	3.88
River Bed	904.75	35.23	16	16.91	20.44
Deep Hole	129.27	5.03	3	2.42	2.92
Arc	97.57	3.80	2	1.82	2.20
Spunk	39.07	1.52	2	0.73	0.88
Rio Grande	190.75	7.43	3	3.57	4.31
Lower Lake	1	0.04	2	2	2
Sums	2567.76	100	50	50	60
2025	**Area (m^2)**	**Percent**	**Samples Taken**	**Samples**	
Old Hotel	239	20.00	4	9.60	11.60
Diversion	171.45	14.34	2	6.89	8.32
River Bed	328.14	27.45	4	13.18	15.92
Deep Hole	129.27	10.82	2	5.19	6.27
Arc	97.57	8.16	1	3.92	4.73
Spunk	39.07	3.27	1	1.57	1.90
Rio Grande	190.75	15.96	1	7.66	9.26
Lower Lake				2	2
Sums	1195.25	100	15	50	60

Table 2. Substrate categories for sampling Spring Lake in 2026.

Substrates	Size (mm)
Silt	<1
Sand	1 - 2
Small Gravel (Dime and smaller)	2 - 16
Gravel	16 - 64
Cobble	64 - 258
Boulder	258+

Table 3. Embeddedness categories available for each sampling design from 2025 and into 2026.

Embeddedness Categories 2025	Percent	Embeddedness Categories 2026	Percent
None	0	None	0-1
Low to Moderate	<50	Low	2 - 25
Moderate to High	>50	Moderate	26 - 50
Completely	100	High	51 - 75
		Saturated	75+

Table 4. Counts of salamanders observed and the estimated metrics associated with each polygon.

Polygon	**Salamanders observed**	**Transects Occupied (%)**	**Meters (m^2)**	**Salamanders (m^2)**	**Estimated salamanders in polygon**
Hotel	1,008	97	239	5.60	1,338
River Bed	714	97	328	3.97	1,302
Diversion	58	39	171	0.64	110
Deep Hole	89	78	129	0.99	128
ARC Site	73	89	97	1.62	158
Spunk	106	89	39	2.36	92
Rio Grande	73	89	190	1.62	309
Eastern Spillway	6	14	615	0.03	40

Table 5. Occupancy models run for the 2025 prescribed data set. NS = not significant; NA's = estimates not available.

	Model	AIC	Δ AIC	no Par	Sig Psi Cov	Sig Det Cov	Issues
1	psi(Springs), p(Depth)	90.8	0	4	NS	Depth	Large st err, psi estimates conf int ranged 0 -1
2	psi(Zone), p(Depth)	92.46	1.66	4	NA	Depth	Na's for covariates and estimates
3	psi(Sag), p(Depth)	95.52	4.72	4	NS	Depth	Large st err, psi estimates conf int ranged 0 -1
4	psi(.), p(Depth)	96.24	5.44	3	NA	Depth	none
5	psi(.), p(Emb + Depth)	96.49	5.69	4	NA	Depth	Emb not sig
6	psi(Cobble), p(Depth)	97.55	6.75	4	NA	Depth	Na's for covariates and estimates

Table 6. Abundance models run. All models shown except for the last model were run with the negative binomial distribution.

Model	AIC	Δ AIC	no Par
λ (Cobble), r(Springs +Zone)	1157	0	6
λ (.), r(Springs +Zone)	1161	4	5
λ (SubCode), r(Springs +Zone)	1161	4	6
λ (Rocks), r(Springs +Zone)	1162	5	6
λ (Silt), r(Springs +Zone)	1163	6	6
λ (VegetationTotal), r(Springs +Zone)	1163	6	6
λ (.), r(Zone)	1198	41	4
λ (.), r(Springs)	1204	47	4
λ (.), r(*Ceratophyllum* sp.)	1219	62	4
λ (.), r(Silt)	1228	71	4
λ (.), r(Embeddedness)	1230	73	4
λ (.), r(Sand)	1233	76	4
λ (.), r(Sag)	1248	91	4
λ (.), r(FilamentousAlg)	1256	99	4
λ (.), r(Depth)	1257	100	4
λ (.), r(Null)	1261	104	4
λ (.), r(Rocks)	1262	105	4
λ (.), r(VegTot)	1262	105	4
λ (.), r(Cobble)	1263	106	4
λ (.), r(SubCode)	1263	106	4
(Poi) λ (Cobble), r(Springs +Emb)	2167	1010	5

Table 7. Null models and the selected negative binomial model; λ (Cobble), r(Springs +Zone), closed, open, and with variations with K.

K	n	GOF	C-hat	AIC	Prob of det
100	too	small	to	estimate	na
150	1815	0.06	1.49	1350	0.41
175	2047	0.08	1.42	1250	0.37
185	2142	0.08	1.36	1229	0.35
200	2284	0.14	1.32	1204	0.33
248	2744	0.21	1.31	1157	0.27
300	3249	0.23	1.22	1131	0.23
500	5147	0.21	1.25	1091	0.14
750	7473	0.19	1.31	1075	0.09
1000	9741	0.23	1.20	1067	0.07
200_Open	2016	0.37	1.30	1173	0.16
Open Null	1427	0.02	1.70	1316	0.49
Closed Null	1826	0.05	1.64	1261	0.38

Table 8. Estimates from the selected N-mixture models, λ (Cobble), r(Springs +Zone), under closed and open scenarios. Also presented are the sums for each abundance model and the average probability of detection for each model.

	Abundance Estimate			Probability of Detection	
Site	Closed	Open		Closed	Open
Hot1	18.67	13.36		0.47	0.62
Hot2	173.73	126.21		0.47	0.62
Hot3	65.21	68.18		0.47	0.62
Hot4	88.09	35.65		0.19	0.26
RB1	87.70	44.16		0.26	0.37
RB2	57.11	36.01		0.26	0.37
RB3	49.62	47.12		0.26	0.37
RB4	28.99	14.04		0.26	0.37
DIV1	30.12	9.38		0.19	0.26
DIV2	2.94	2.06		0.19	0.26
DH1	10.95	3.78		0.26	0.37
DH2	3.03	3.94		0.26	0.37
ARC1	174.77	125.42		0.09	0.11
SP1	226.42	157.69		0.09	0.11
RG1	65.37	64.98		0.26	0.37
Hot1	4.17	3.68		0.47	0.62

Hot2	244.82	178.32	[redacted]	0.47	0.62
Hot3	62.64	43.98	[redacted]	0.47	0.62
Hot4	11.36	14.99	[redacted]	0.47	0.62
RB1	62.16	79.53	[redacted]	0.09	0.11
RB2	239.16	182.98	[redacted]	0.26	0.37
RB3	143.34	120.71	[redacted]	0.26	0.37
RB4	4.08	3.95	[redacted]	0.26	0.37
DIV1	0.81	1.91	[redacted]	0.19	0.26
DIV2	30.43	32.14	[redacted]	0.47	0.62
DH1	18.16	19.16	[redacted]	0.26	0.37
DH2	47.62	59.67	[redacted]	0.26	0.37
ARC1	22.49	19.16	[redacted]	0.26	0.37
SP1	113.98	118.33	[redacted]	0.09	0.11
RG1	12.82	17.49	[redacted]	0.26	0.37
Hot1	4.54	5.27	[redacted]	0.47	0.62
Hot2	51.08	40.64	[redacted]	0.47	0.62
Hot3	102.05	50.89	[redacted]	0.19	0.26
Hot4	1.58	2.07	[redacted]	0.47	0.62
RB1	172.91	178.22	[redacted]	0.26	0.37
RB2	43.72	12.02	[redacted]	0.26	0.37
RB3	119.96	154.65	[redacted]	0.09	0.11
RB4	5.74	6.59	[redacted]	0.26	0.37
DIV1	0.81	1.91	[redacted]	0.19	0.26
DIV2	0.83	2.03	[redacted]	0.19	0.26
DH1	14.97	16.60	[redacted]	0.26	0.37
DH2	31.58	49.05	[redacted]	0.26	0.37
ARC1	32.91	47.49	[redacted]	0.09	0.11
SP1	40.70	12.70	[redacted]	0.09	0.11
RG1	20.33	28.26	[redacted]	0.26	0.37
Sums	2744.46	2256.37	Averages	0.27	0.37

Table 9. Results from GLMM analysis and AIC scores. CC = *Certopylum* sp and *Cabomba* sp.; Sag = *Sagitarria* sp.; FA = filamentous algae; quartiles column designates if a models residuals conforms to all expected quartiles; quantile average and dispersion columns represent p-values from observed vs expected tests in the "Dharma" package in R.

Model	Fixed Effects	AIC	logLik	2*log(L)	df(resid)	Quartiles	Quartile Average	Dispersion
1	Spring + Cobble + Gravel	903.62	-444	889	128	Fail	<0.0001	NA
2	Spring + Cobble + Gravel + Depth	904.26	-444	888	127	Fail	0.1472	0.51
3	Spring + Gravel + Cobble + Depth + CC	905.96	-443	887	126	Pass	0.35	0.52
4	Spring + Cobble + Gravel + Depth + FA	906.16	-444	888	126	Fail	0.0018	0.56
5	Spring + Cobble + Gravel + Depth + Sag	906.25	-444	888	126	Pass	0.34	0.54
6	Gravel	906.41	-448	896	130	Fail	<0.0001	NA
7	Spring + Gravel	907.38	-447	895	129	Fail	<0.0001	NA
8	Spring + Cobble + Gravel + Depth + Sag + CC	907.95	-443	887	125	Fail	0.08	0.62
9	Null	908.65	-450	900	131	NA	NA	NA
10	Spring	908.8	-449	898	130	NA	NA	NA
11	Cobble	909.7	-449	899	130	Fail	<0.0001	NA
12	Global	909.83	-443	887	124	Pass	0.28	0.57
13	Depth	909.96	-449	899	130	Fail	<0.0001	NA
14	Sag	910.39	-450	900	130	Fail	<0.0001	NA
15	CC	910.55	-450	900	130	Fail	<0.0001	NA
16	FA	910.62	-450	900	130	Fail	<0.0001	NA

Table 10. Estimates from the third GLMM for each site. Sites are in primary period order and set up in a longitudinal order from the headwaters to the farthest downstream site.

Transect	N	SE	Zone	m2	Average Count
Hot1	8.38	3	1	1.68	8.33
Hot2	69.59	22	1	13.92	80.33
Hot3	27.44	9	1	5.49	27.67
Hot4	13.64	5	1	2.73	15.67
Hot1_2	2.89	1	1	0.58	2.00
Hot2_2	103.60	32	1	20.72	123.67
Hot3_2	29.11	10	1	5.82	28.33
Hot4_2	4.58	2	1	0.92	4.00
Hot1_3	2.47	1	1	0.49	2.00
Hot2_3	25.65	9	1	5.13	24.00
Hot3_3	18.39	6	1	3.68	19.33
Hot4_3	1.12	1	1	0.22	0.67
DIV1	4.79	2	1	0.96	5.33

DIV2	0.64	0	1	0.13	0.33
DIV1_2	0.19	0	1	0.04	0.00
DIV2_2	12.17	4	1	2.43	13.67
DIV1_3	0.31	0	1	0.06	0.00
DIV2_3	0.50	0	1	0.10	0.00
RB1	20.87	7	2	4.17	22.67
RB2	13.72	5	2	2.74	15.00
RB3	13.35	5	2	2.67	12.67
RB4	7.12	3	2	1.42	7.67
RB1_2	5.32	2	2	1.06	5.33
RB2_2	59.28	19	2	11.86	68.33
RB3_2	35.06	11	2	7.01	38.00
RB4_2	1.49	1	2	0.30	1.00
RB1_3	44.04	15	2	8.81	44.67
RB2_3	10.88	4	2	2.18	11.00
RB3_3	10.56	4	2	2.11	10.33
RB4_3	1.94	1	2	0.39	1.33
DH1	2.69	1	2	0.54	2.33
DH2	1.00	1	2	0.20	0.67
DH1_2	4.89	2	2	0.98	4.67
DH2_2	9.85	3	2	1.97	11.67
DH1_3	3.79	2	2	0.76	3.67
DH2_3	5.95	2	2	1.19	6.67
ARC	14.32	5	2	2.86	16.00
ARC1_2	6.22	3	2	1.24	5.67
ARC1_3	2.90	1	2	0.58	2.67
SP1	18.78	6	2	3.76	21.67
SP1_2	9.78	4	2	1.96	10.33
SP1_3	3.66	2	2	0.73	3.33
RG1	15.36	5	2	3.07	16.67
RG1_2	3.58	2	2	0.72	3.00
RG1_3	5.24	2	2	1.05	4.67
Zone 1	325.46		1	1.20	
Zone 2	331.63		2	0.81	
Total	657.09		All	2.92	

Table 11. Estimated salamanders per meter squared from the third GLMM for each polygon and primary period. PP1 = primary period 1 in April, PP2 = May, PP3 = June.

Primary Period	Polygon	Salamanders / m^2
PP1	Hotel	5.95
	Diversion	0.54
	River Bed	2.75
	Deep Hole	0.37
	Arc Site	2.86
	Rio Grande	3.07
	Spunk	3.76
PP2	Hotel	7.01
	Diversion	1.24
	River Bed	5.06
	Deep Hole	1.47
	Arc Site	1.24
	Rio Grande	0.72
	Spunk	1.96
PP3	Hotel	2.38
	Diversion	0.08
	River Bed	3.37
	Deep Hole	0.97
	Arc Site	0.58
	Rio Grande	1.05
	Spunk	0.73
Average	Hotel	5.11
	Diversion	0.62
	River Bed	3.73
	Deep Hole	0.94
	Arc Site	1.56
	Rio Grande	1.61
	Spunk	2.15

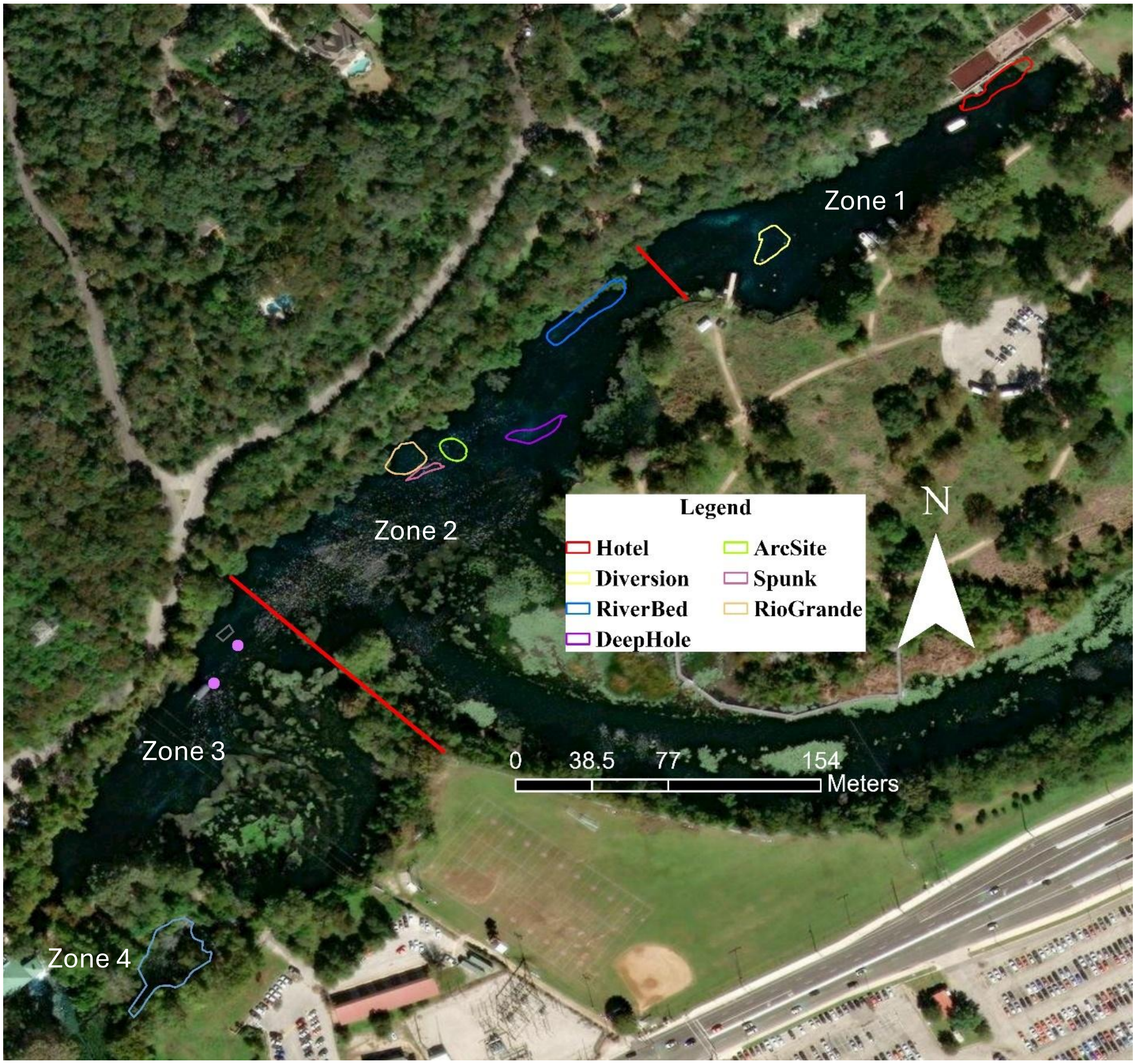
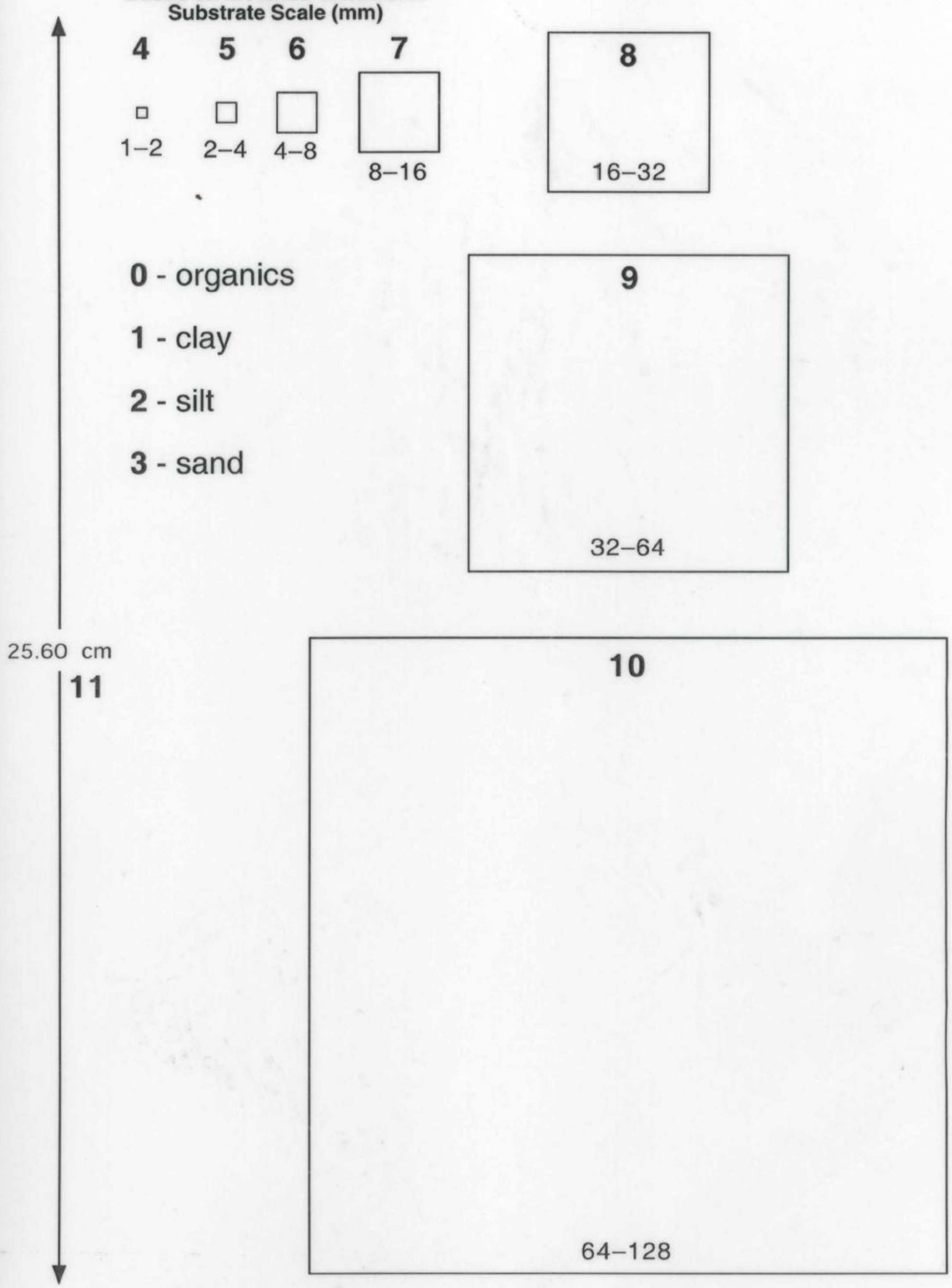


Figure 1. Map of Spring Lake Zones and the polygons available for sampling in 2025. Pink dots are locations in Zone 3 where salamanders were observed.

Based on Modified Wentworth Substrate Scale (mm)



0 - organics

1 - clay

2 - silt

3 - sand

Figure 2. Substrate standards for sampling Spring Lake in 2025 and 2026. In 2025, 5-9 were considered Gravel. In 2026, sizes 5, 6, and 7 (dimes and smaller) will be considered small gravel, and sizes 8 and 9 and will be gravel.

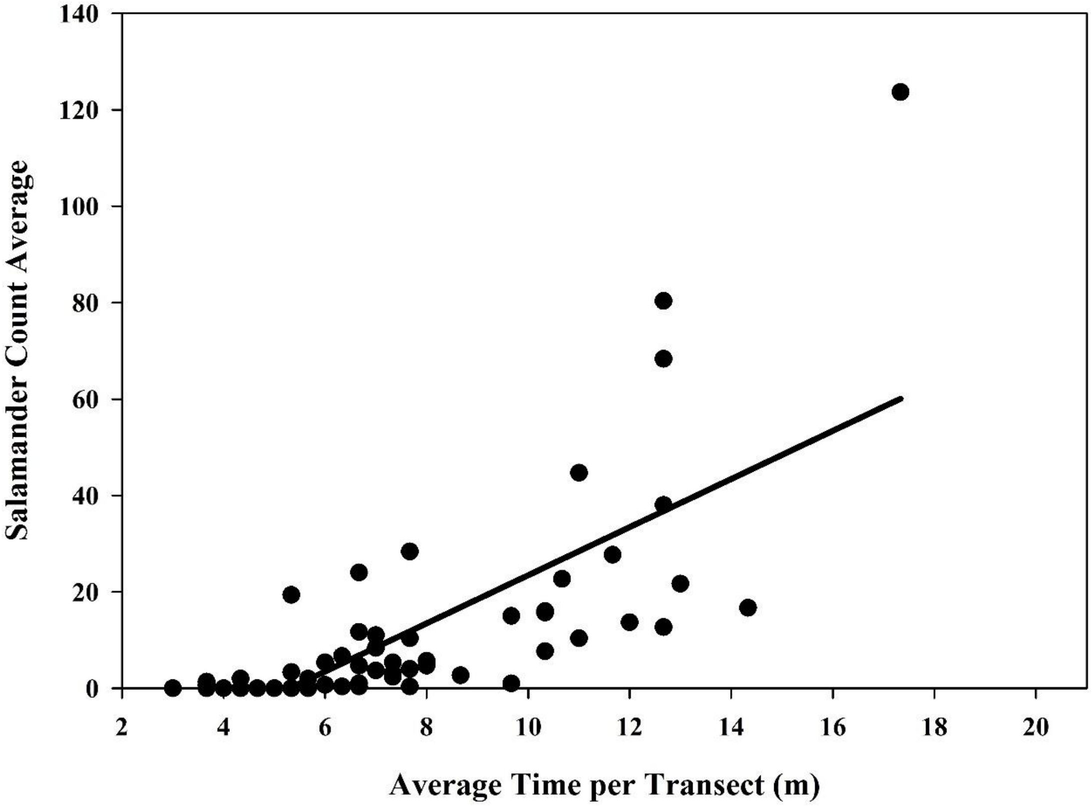


Figure 3. Diver time per 5 m transect in the Spring 2025 and the relationship with the salamander count data.

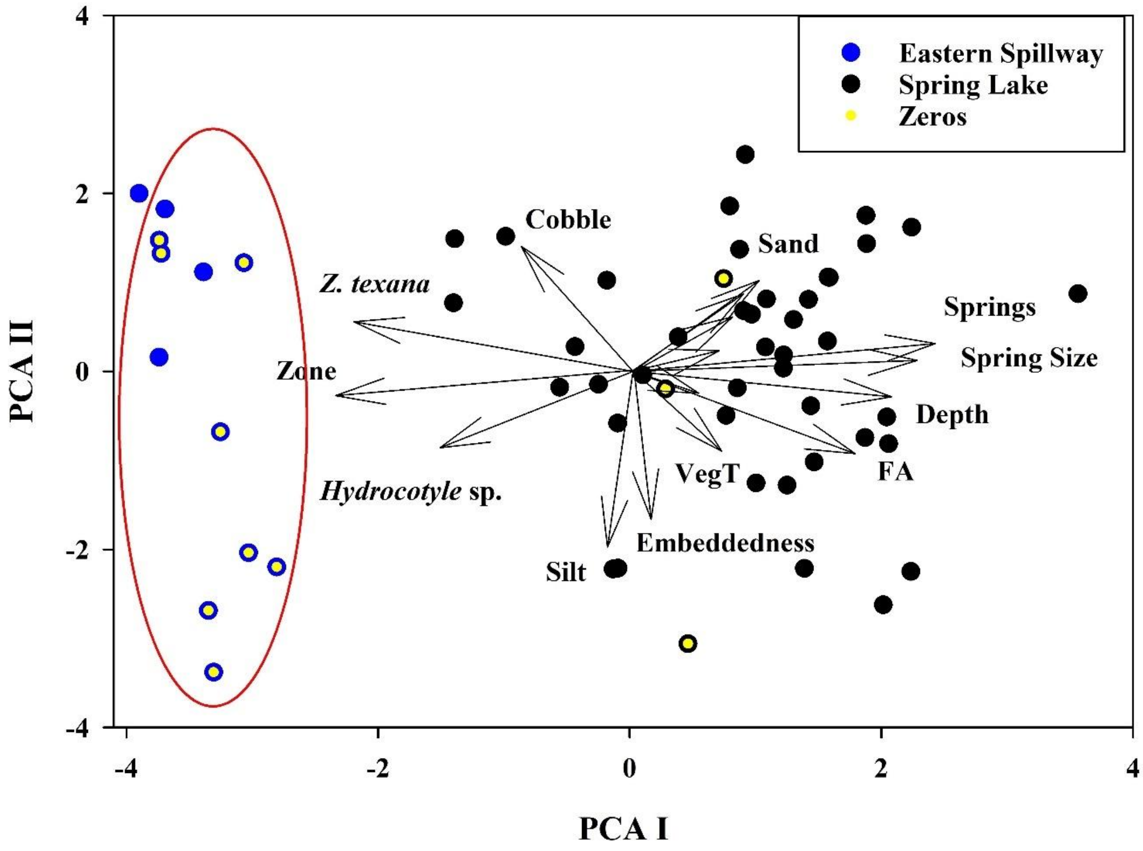


Figure 4. Principal component analysis from 2025 of the entire study area. Veg T = total vegetation along transect; FA = filamenotous algae.

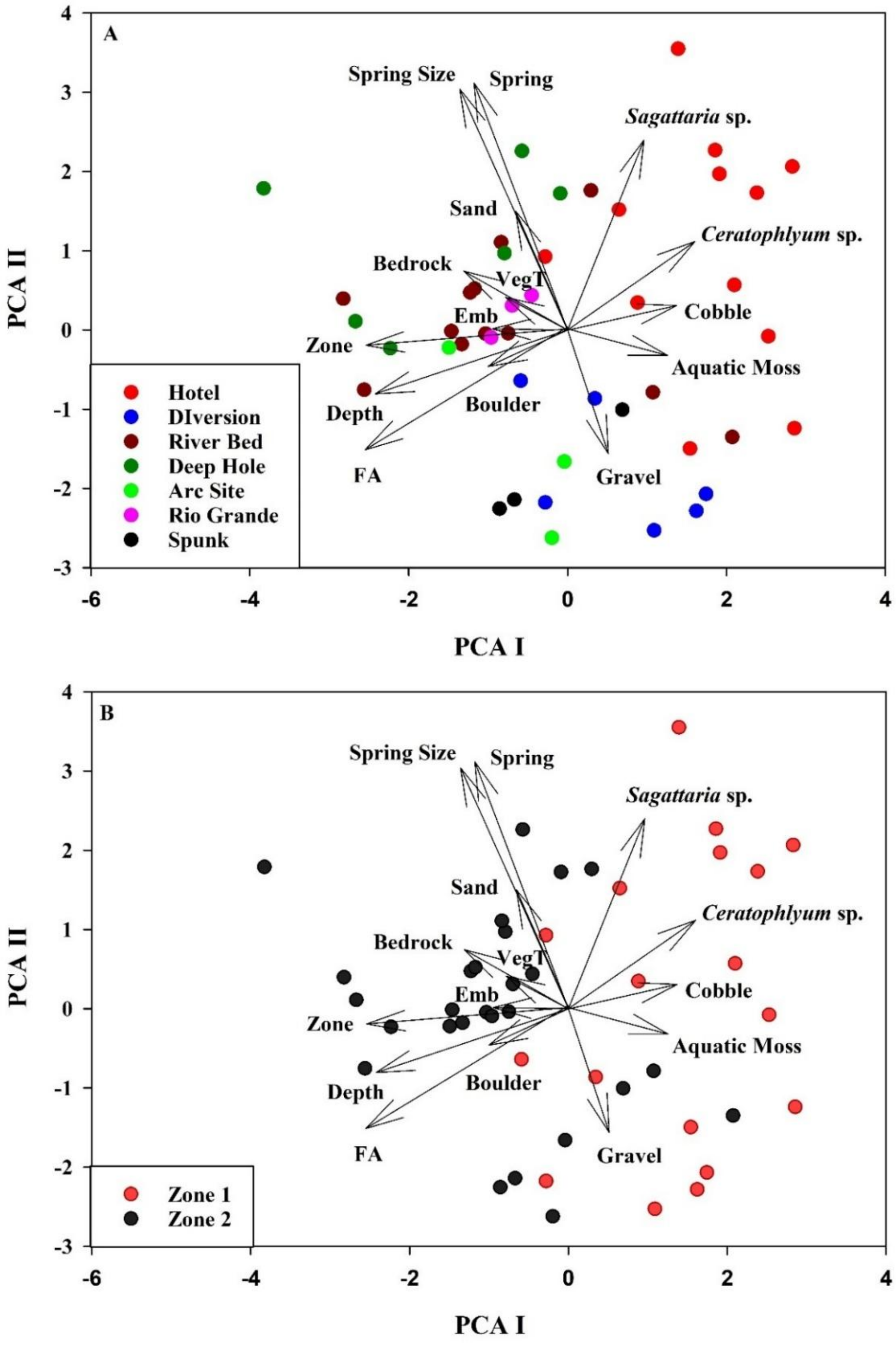


Figure 5. Principal component analysis from 2025 of the Lake transects showing habitat structure between zones and polygons. Veg T = total vegetation along the transect; FA = filamentous algae, Emb = embededdness.

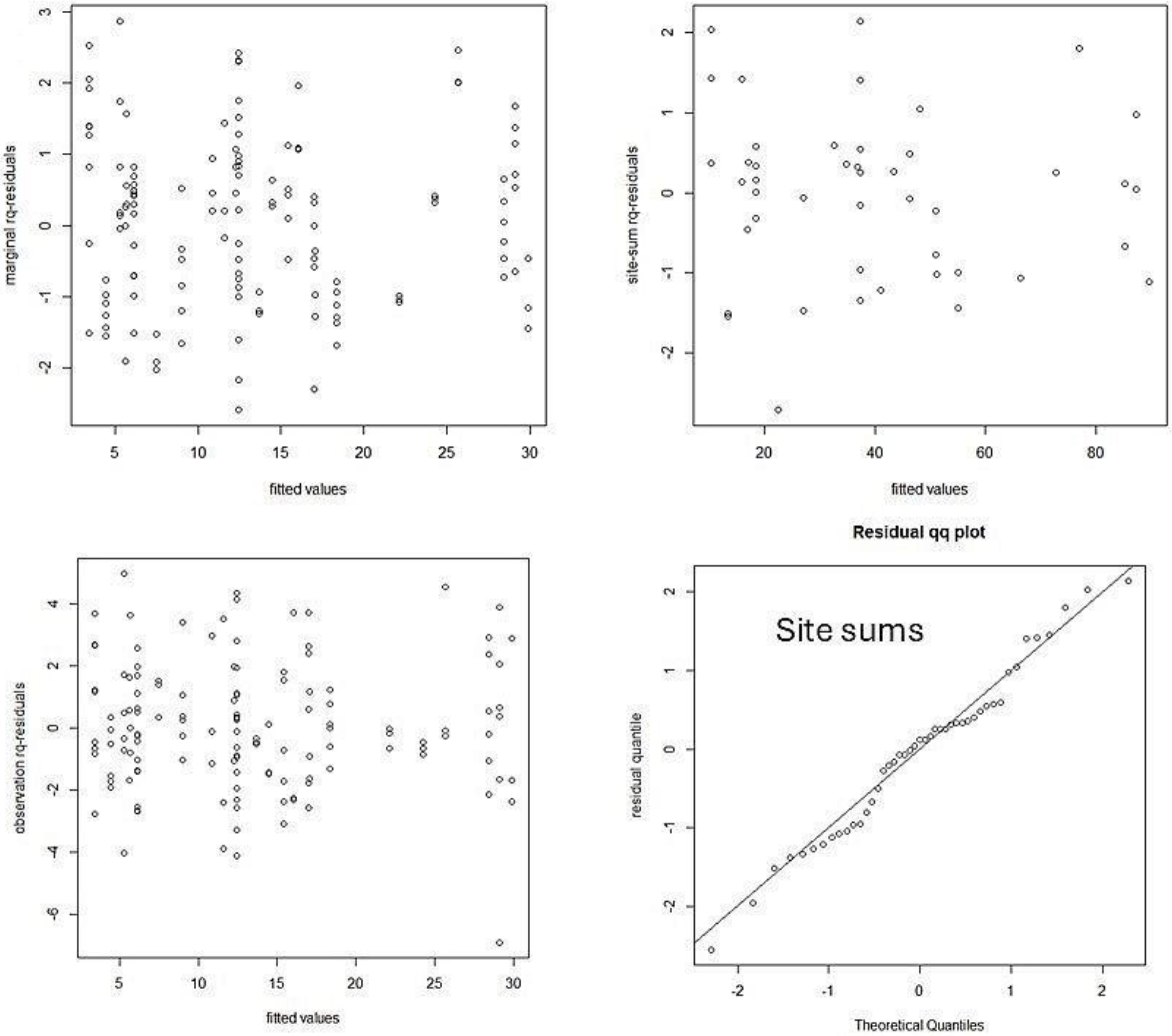


Figure 6. Residual plots from the selected N-mixture abundance negative binomial model.

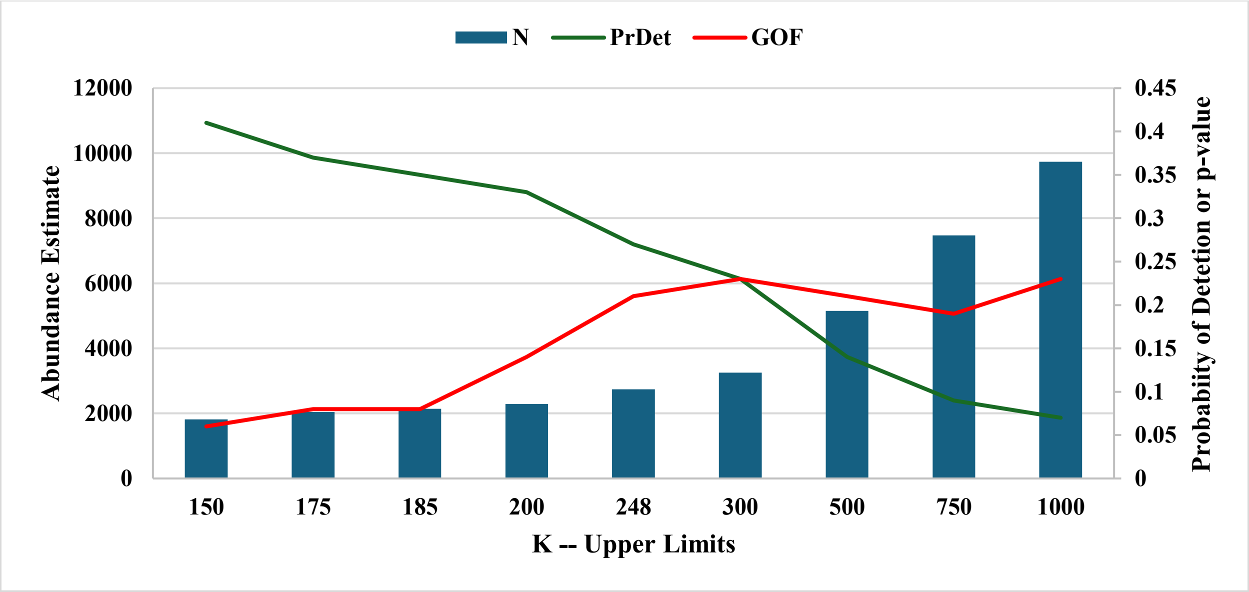


Figure 7. The selected negative binomial model and the relationships with K, N, GOF tests, and the probability of detection.

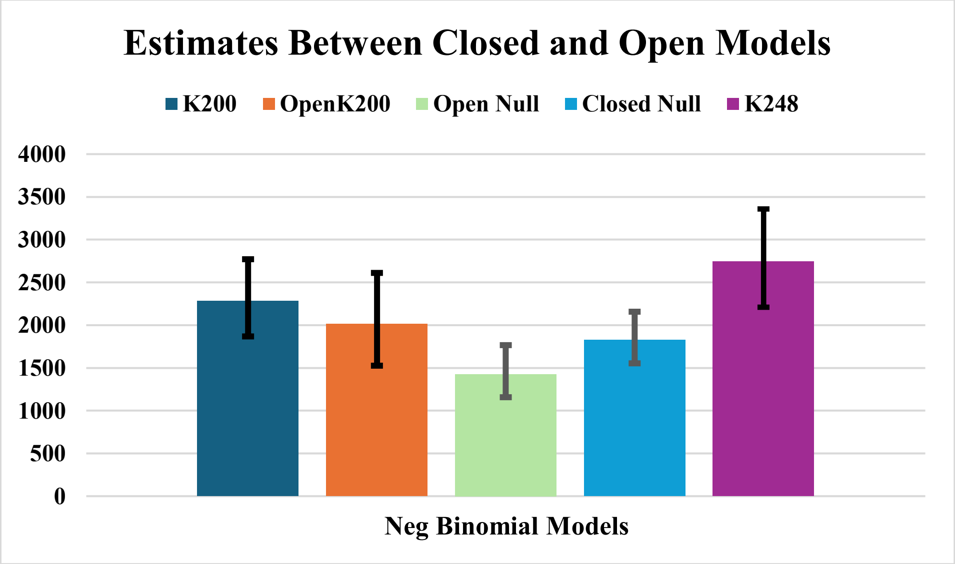


Figure 8. The selected closed negative binomial model and other estimates of the model with different levels of K. Also shown are a closed and open null models.

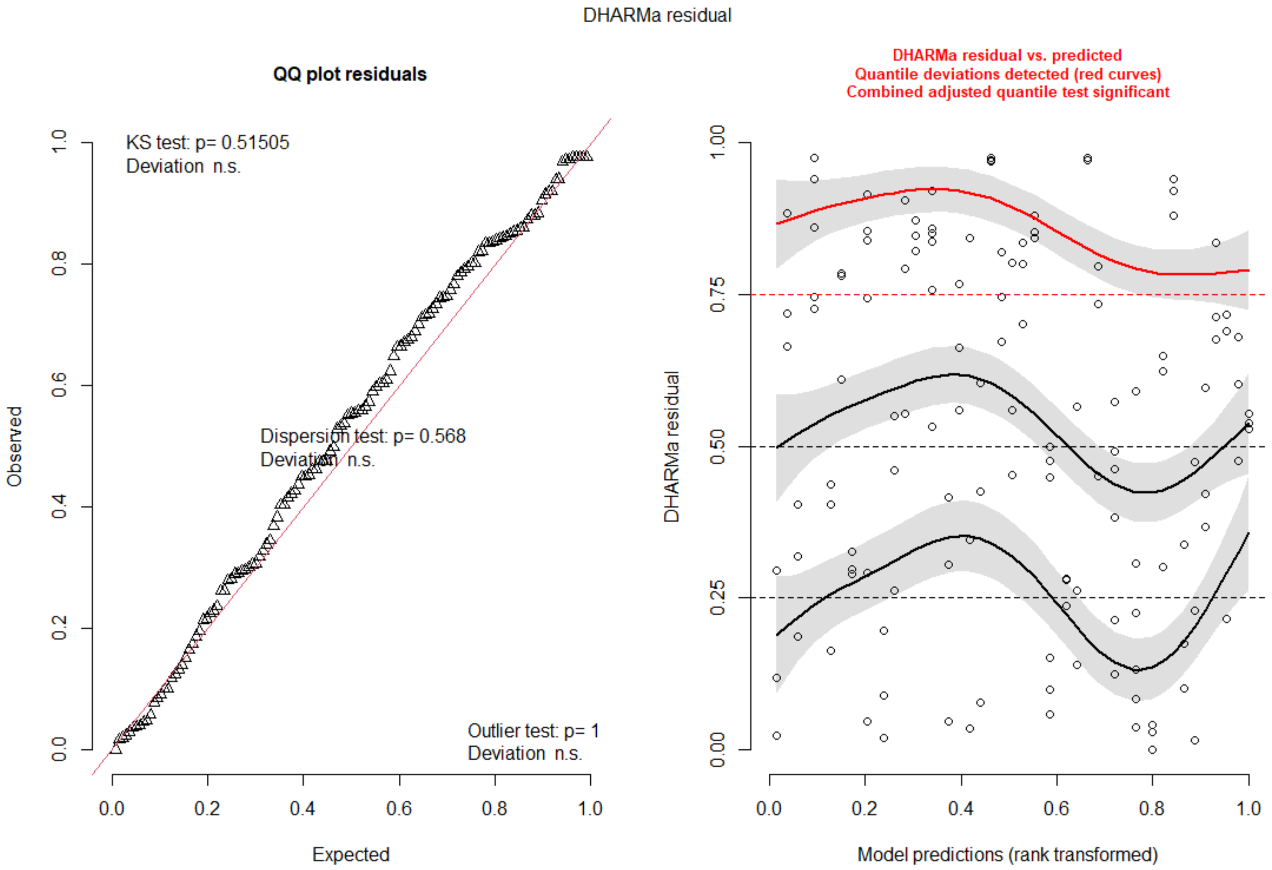


Figure 9. Results from the “Dharma” package in R for the second GLMM. The left side shows a QQ plot and the dipsersal test results. The right shows the results of the residual quartile test. Red line indicates significant deviation within the third quartile for the model demonstrating a lack of fit.

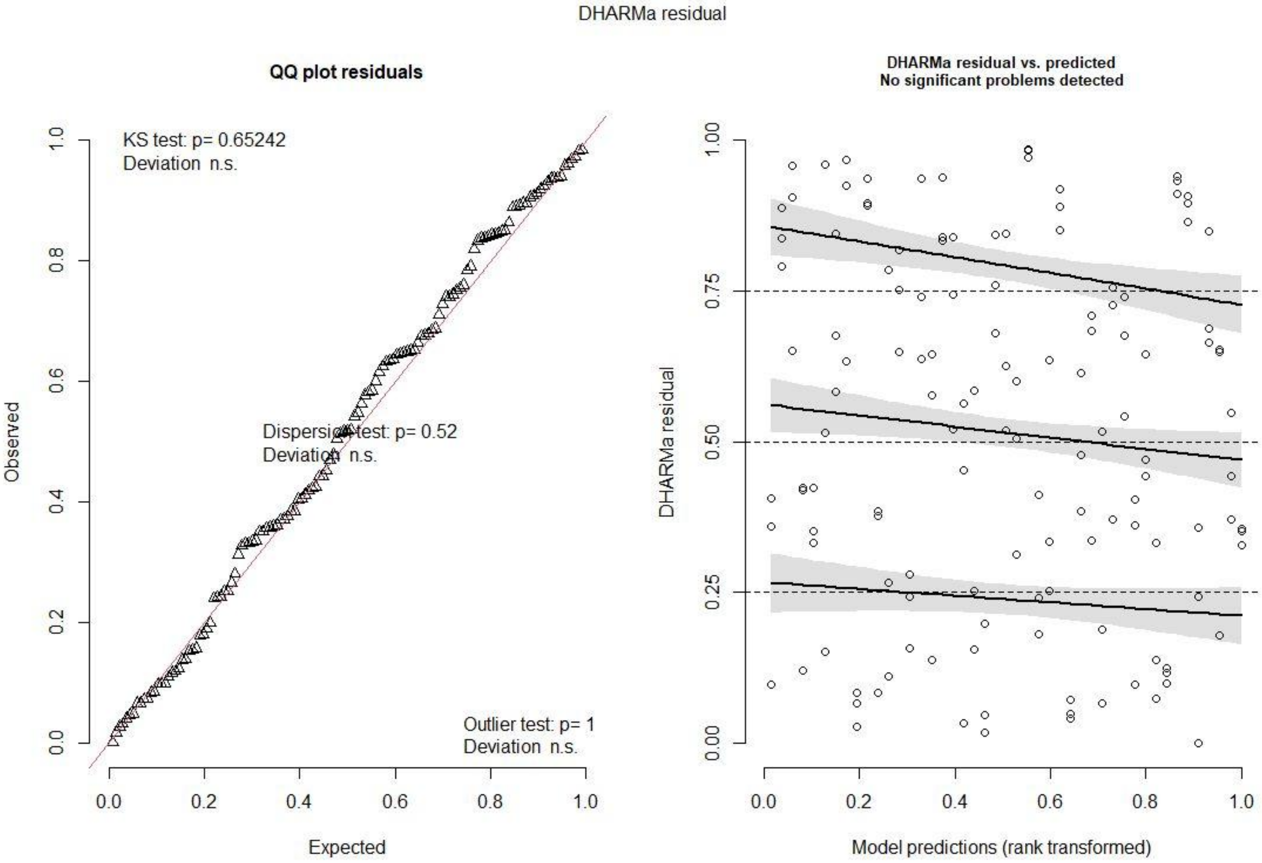


Figure 10. Residual quartile plots from "Dharma" package in R for the third GLMM. The left side shows a QQ Plot and the dipsersal test results. The right shows the results of the residual quartile test.

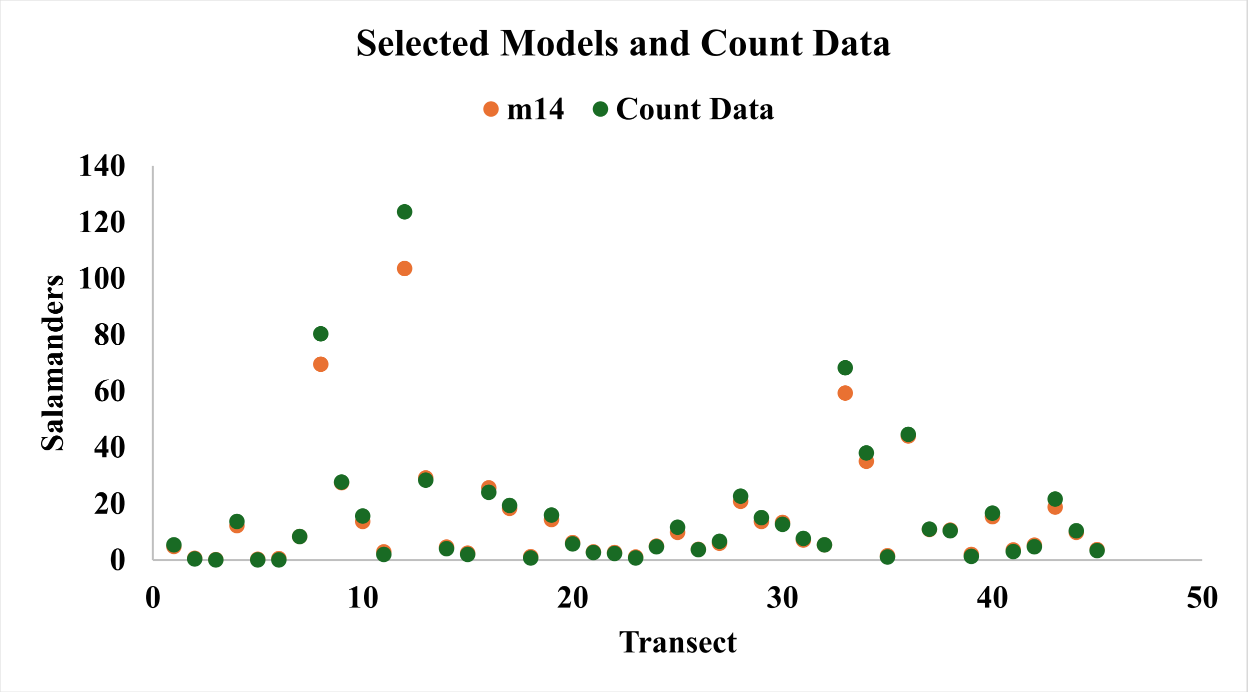


Figure 11. Scatterplot of count data and predicted abundances from the selected GLMM.

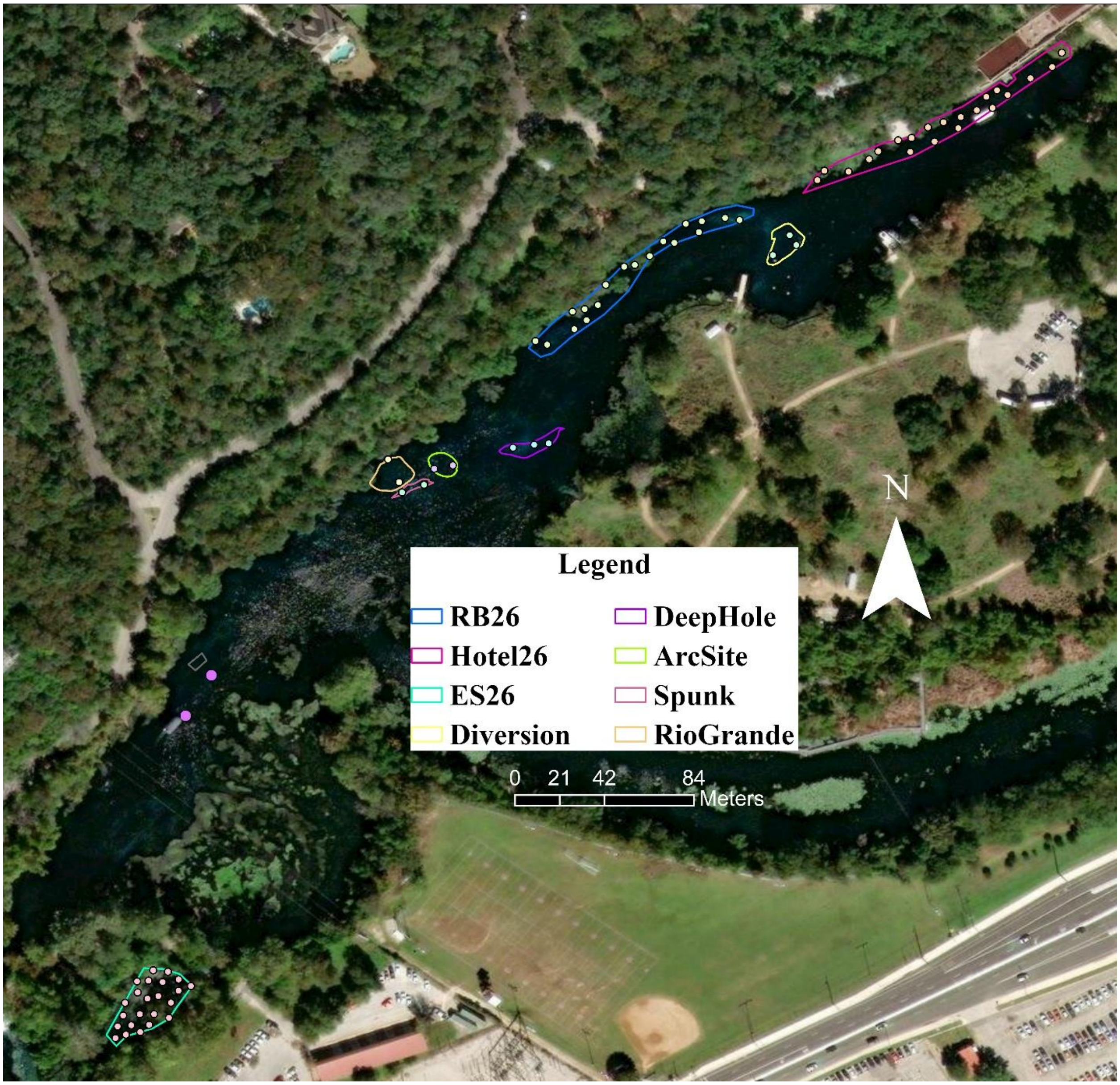


Figure 12. Map of Spring Lake showing updated polygons for sampling in the 2026 spring season.

Literature Cited

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* Second International Symposium on Information Theory (B. N. Petrov and F. Csaki, Eds.). Akademiai Kiado, Budapest.

Berkhouse, C. S., J. N. Fries. 1995. Critical thermal maxima of juvenile and adult *E. nana* (*Eurycea nana*). Southwestern Naturalist 40:430-434.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling." *The R Journal*, **9**(2), 378–400. doi:10.32614/RJ-2017-066.

Chippindale, Paul T., Andrew H. Price, and David M. Hillis. "Systematic Status of the San Marcos Salamander, Eurycea Nana (Caudata: Plethodontidae)." *Copeia* 1998, no. 4 (1998): 1046–49. https://doi.org/10.2307/1447356.

Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open metapopulation. Biometrics 67:577–587.

Faraway, J.J. 2006. Extending the Linear Model with R by J. J. Faraway, *Journal of the Royal Statistical Society Series A: Statistics in Society*, Volume 169, Issue 4, October 2006, Page 1008, https://doi.org/10.1111/j.1467-985X.2006.00446_8.x

Fiske, Ian & Chandler, Richard. (2011). Unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. Journal of Statistical Software. 43. 10.18637/jss.v043.i10.

Fries JN. 2002. Upwelling flow velocity preferences of captive adult San Marcos salamanders. *North American Journal of Aquaculture* **64**: 113-116.

Hartig F (2025). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.7, https://github.com/florianhartig/dharma.

Hines, J.E. 2006. PRESENCE, version 2: Software to estimate patch occupancy and related parameters. U.S. Geological Survey, Patuxent Wildlife Research Center. [Online.] Available at www.mbrpwrc.usgs.gov/software/presence.html.

Huynh HD, Royle JA, Hwang WH. 2025. A flexible framework for N-mixture occupancy models: applications to breeding bird surveys. Biometrics. 81(3):ujaf087. doi: 10.1093/biomtc/ujaf087. PMID: 40705489.

Lucas, L. K. 2006. Population and conservation of Texas Hill Country Eurycea. Master's thesis. Texas State University, San Marcos.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one.

Madsen, L and J. A. Royle. 2023. A review of N-mixture models. WIREs Comput. Stat. 15, 6 https://doi.org/10.1002/wics.1625

Mazerolle MJ (2023). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)*. R package version 2.3.3, https://cran.r-project.org/package=AICcmodavg.

National Academies of Sciences, Engineering, and Medicine. 2018. *Review of the Edwards Aquifer Habitat Conservation Plan: Report 3*. Washington, DC: The National Academies Press.

Nelson, J. M. 1993. Population size, distribution, and life history of *Eurycea nana* in the San Marcos River. Master's thesis. Southwest Texas State University, San Marcos.

Parker MRP, Cowen LLE, Cao J, Elliott LT. 2023. Computational Efficiency and Precision for Replicated-Count and Batch-Marked Hidden Population Models. J Agric Biol Environ Stat. 28:43-58.

Royle, J.A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60: 108–115.

Tupa, D. D., W. K. Davis. 1976. Population dynamics of the *E. nana*, *Eurycea nana* Bishop. The Texas Journal of Science 27:179-195.

USDI (U. S. Department of the Interior). 1980. Endangered and threatened wildlife and plants; listing of *E. nana* as threatened, the San Marcos gambusia as endangered, and the listing of critical habitat for Texas wild-rice, *E. nana*, San Marcos gambusia, and fountain darter. Federal Register 45(14 July 1980):47355-47364.

The views expressed in this paper are the authors' and do not necessarily reflect the views of the U.S. Fish and Wildlife Service or Texas Parks and Wildlife Department.

Author contributions: Pete Diaz: conceptualization, data curation, formal analysis, funding acquisition, writing. Randy Gibson: conceptualization, diving, writing. Somerly Swarm: diving, editing. Justin Crow: conceptualization, diving, writing. Hannah Gilbreath: data collection, editing. Matt Johnson: diving, data collection. Robin Russell: N-mixture occupancy models analysis (A1), editing.

Appendix

A1: San Marcos Salamander Experimental Modeling with Huynh et al. 2025 “A flexible framework for *N*-mixture occupancy models: applications to breeding bird surveys”

*Note N-mixture models were not designed for studies with wide fluctuations in the data or large numbers in a survey

The models do not produce accurate abundance estimates (by increasing the max N the estimated abundance continues to increase)

Models that “converged” with the caveat above

PN1-Poisson with detection <1(2755.8)

Parameter	Estimate	SE	z	P(>|z|)
Occupancy:psi	0.807	0.052	15.439	8.943258e-54
Effective Occupancy : adj-psi	0.807	0.052	15.439	8.943258e-54
Abundance: mu	28.191	1.061	26.566	1.678924e-155
p : detection	0.547	0.016	34.122	3.538433e-255

NB1-Negative Binomial with detection <1 (AIC : 1400.2)

Parameter	Estimate	SE	z	P(>|z|)
Occupancy:psi	0.99	0.146	6.801	1.035785e-11
Effective Occupancy : adj-psi	0.822	0.053	15.372	2.513244e-53
Abundance: mu	29.036	7.458	3.893	9.897357e-05
Detection:p	0.433	0.015	28.322	1.858773e-176
Aggregation Index : kappa	0.415	0.169	2.456	0.01405243

NBC-Negative Binomial with detection <1 and community parameter C (AIC : 503.3589)

Parameter	Estimate	SE	z	P(>|z|)
Occupancy:psi	0.939	0.063	14.955	1.441296e-50
Abundance: mu	186.084	13.917	13.371	8.957536e-41
Detection:p	0.994	0.011	88.316	0
Aggregation Index : kappa	3.778	0.702	5.382	7.347885e-08
Community Parameter :c	0.91	0.025	37.047	2.010565e-300

C: is the Community parameter that ranges from 0-1 with 0 indicating that different individuals are detected on each survey and 1 indicated that the same individuals are detected on each survey.

Pcount R Code from the unmarked pdf

```
> library(unmarked)
> library(AICcmodavg)
> sc45 <- read.csv("D:sc45.csv")..... site covariates
> y45 <- read.csv("D:y45.csv")..... count data
> fm1<- unmarkedFramePCount(y=y45, siteCovs=sc45)..... no sample covariates
> fm1 <- pcount(~1 ~ 1, umf1)..... first part of equation is the detection, the second part is
  the occupancy, Poisson distribution is the default
> fm1.nb <- pcount(~1 ~ 1, umf1, mixture = "NB")..... for negative binomial models
> fm1re <- ranef(fm1) .....Estimates for N per site
> fm1P <- getP(fm1)..... Estimates for Prob of Det
> obs <- Nmix.chisq(fm1)..... goodness of fit test
> cand.models <- list(mod1 = fm1, mod2 = fm2, mod3 = fm3, mod4 = fm4, .....)
> aictab(cand.models)..... to make table of models
obs <- Nmix.chisq(fm1)..... GOF Test
obs.boot <- Nmix.gof.test(fm1, nsim = 10)
residfit(fm1, type = "marginal")
residqq(lm17nb,type = "marginal",main = "Residual qq plot",plotLine = TRUE)
residqq(lm17nb,type = "observation",main = "Residual qq plot",plotLine = TRUE)
```

PcountOpen R Code from the unmarked pdf

```
> library(unmarked)
> library(AICcmodavg)
> primaryperiod<- matrix(as.integer(c(1,2,3,1,2,3,….)….. Setting up matrix for count data
> oc99<-list(flow=flow45,cond=cond45,temp=temp45)….. If there are sampling covariates
> umf1<-unmarkedFramePCO(y = y45, siteCovs = sc45,obsCovs = oc99, numPrimary = 3,
  primaryPeriod = primaryperiod)….. Default distribution is Poisson
> m1 <- pcountOpen(~lambdafor, ~gammaform, ~omegaform, ~pform, umf1, K=20))….
  Equation for open model; abundance is first, detection is last
> re <- ranef(m1)….. Site level estimates
> getP(m1)….. gets detection estimates
> obs.boot <- Nmix.gof.test(m1, nsim = 100)….. sort of a minimum of simulations.....GOF Test
```

GLMM R Code from the glmmTMB pdf

```
> library(glmmTMB)
> library(AICcmodavg)
> Model1.nb<-glmmTMB(Count ~ Coef1 + Coef2 + (1|Transect) + (1|Site), ziformula = ~ 0,
  family = nbinom2, data = nana.df)…..can have as many coefficients as necessary.
> summary(Model1.nb)
>  plotConventionalResiduals(Model1.nb)
> testQuantiles(Model1.nb, predictor = NULL, quantiles = c(0.25, 0.5,0.75),plot=TRUE)
> testDispersion(Model1.nb)
```

```
> coef(p_Spr_Gr_Cob_Dep_CC.nb, condVar = FALSE)$Transect
> rr <- ranef(Model1.nb)…. Random Effects
> confint(Model1.nb)
> predict(Model1.nb, type = "response", se.fit = TRUE)
> preds_link <- predict(Model1.nb,se.fit = TRUE, type = "response")
> alpha <- 0.05
> z_score <- qnorm(1 - alpha/2)
> lower_link <- preds_link$fit - z_score * preds_link$se.fit
> upper_link <- preds_link$fit + z_score * preds_link$se.fit
> conf_intervals_1 <- confint(Model1.nb)
> library(performance)
> icc(Model1.nb)
> icc(Model1.nb, by_group = TRUE)
```