

Comal Springs Riffle Beetle Population Assessment Work Plan Contract 21-019-TES (May 2022)

Biological considerations

Comal Springs riffle beetle, (CSRB) *Heterelmis comalensis*, is found at surficial interfaces where springs are active at the Comal Springs. Our current life-history knowledge indicates that larvae take 9 – 11 months to reach maturity in captivity (BIO-WEST 2017). Wild caught adults may live over a year in captivity, but often do not live as long (Fries 2003) and captive reared adults rarely live to one year old (personal observations). Females produce eggs soon after becoming adults and are iteroparous (Kosnicki 2022). Therefore, new breeding cohorts can be expected within less than one year and interbreeding among cohorts is possible. These life-history aspects complicate mark and recapture methods, assumptions of N-mixture models, and other census methods. However, we can assume that each population census is a representation of the population size and distribution at the time of that census and if surveys are conducted far enough apart, sampling the same individuals is highly unlikely.

Field sampling design

Sample sites

Sampling will be conducted over three of the sub-populations as recognized by Lucas et al. (2016) plus the headwaters area of Comal Springs. The Spring Run 2 area will not be sampled since there are few springs along this reach and since it will be under recovery from restoration activities. Sampling will include the 30 designated sites that are used for the Edwards Aquifer Habitat Conservation Plan (EAHCP) bi-annual biological monitoring. In addition, 50 randomly selected springs from each sub-population area have been selected to represent roughly 20% of the mapped springs designated by TPWD (**Map 1**). The areas and number of sites has been selected as follows:

Spring Run 1: 10 sites

Spring Run 3: 20 (including 10 biomonitoring sites)

Western shoreline + Spring Island + Backwater: 42 (including 20 biomonitoring sites)

Spring Run 4 + Spring Run 5 + Comal headwaters + Blieder's Creek: 8 sites

Upwelling and margin habitats

Springs types will be divided into upwellings and margin spring habitats. After discussions with Marcus Gary (Edwards Aquifer Authority), it was clear that flow measures from these should be considered separately (see *Flow index measures* below). Upwellings are represented by spring flow that is vertical in direction, originating from “alluvial clusters” or karst orifices. A specific unit of area around this spring-type will be used to record flow as a means of standardizing flow measures. Margin spring habitats will be associated with more horizontal flows where the area of spring activity will be measured.

Flow index measures

A flow index will be based on velocities measured over the area around the lure. For upwelling habitats, a “bucket flow-measuring device” will be used to isolate flow from spring upwellings, incorporating a 660 cm² area. The bucket flow-measuring device (BFMD) will consist of an inverted 5-gal bucket with the bottom cut off. A ¾-in PVC-pipe will be positioned horizontally through the top of the bucket so that it can support a flow-meter probe in the center of the bucket and ca. 2.4 cm from the surface of the spring

source to be measured. Four measures will be taken around the spring, one measure directly above the position of the lure (before lure is placed and after it is removed), and three equal-distant spaced measures around the central measure at a radius of 10 cm.

Margin spring-type habitats will be delineated around a linear plane interpreted as perpendicular to the main concentration of spring-flow associated with the placement of the lure. A standardized device such as a bucket cannot be used in these types of habitats because they are based at or near the water surface and/or due to their 3-D structure. For these habitats, sample areas that are < 10 cm X 10 cm a single min and max flow will be measured, separately, by physically holding the probe in the spring flow until stable readings can be taken. Larger margin habitats will be subdivided into triangles so that Heron's formula can be used to find the area. Flow measures will be taken within each triangle.

A field survey will be conducted to examine the variation of this measuring strategy where at least six springs (three of each spring-type) will be measured with this protocol ca. five times each, over a single day as a means of assessing our precision.

Sampling level covariates

Julian day lure is retrieved (reflects the calendar day)

Cumulative river Q measured from USGS gauge station (average over lure placement period)

Cumulative precipitations measures (taken from closest gauge stations)

Sub-population (as delineated above)

Covariates for each lure (see datasheet)

Recorded during lure set and retrieval

Temperature °C

DO (mg/L)

Conductivity ($\mu\text{S}/\text{cm}$)

pH

Water depth (cm)

Lure depth within spring (cm) – The depth of the lure within the substrate

Percent substrate coverage

Recorded during lure retrieval only

Biofilm color and percentage coverage categorization (compare with unconditioned cotton sample)

Lure condition

Organic material present – Note the types of organic materials at the spring surface

Number of days deployed

Beetle counts and removal considerations

Upon retrieval, lures will be inspected with a stereoscope in the field. All individuals will be identified and counted according to maturity level. Larvae will be qualitatively identified as small, medium or large. All individuals will be returned back to the spring from which they were collected. The US Fish and Wildlife Service has requested four specimens per lure for an upcoming genomics assessment. It is unknown how many beetles will be attracted per lure. In the event that a large number of individuals (larvae and/or adult) are collected from a lure, four individuals can likely be removed without an anticipated influence on future sampling events. In these cases, a nuisance parameter will be created to account for percentage of individuals removed from the previous sampling event. For lures that only

attract two or three individuals, removing 100% of the individuals is not considered appropriate at this time.

Sample schedule

The sampling schedule includes four sampling events, following the spring and fall EAHCP biomonitoring schedule which minimizes the frequency of habitat disturbance. . The field sampling for this schedule would be initiated in fall 2022 and would be concluded in spring 2024. Each sampling event will reflect similar protocols for the current EAHCP biological monitoring program. Lures will be set for ca. one month to allow for biofilms to develop and attract beetles before retrieval.

Statistical analysis

N-mixture models

The request for proposal for this study implicated that the survey data should consider analysis with N-mixture models. We have extensively explored the utility of using the N-mixture model developed by Royle (2004), speaking with the author of the model and other statisticians. However, as described in the sampling design, life-history aspects of CSRB complicate the assumptions, mainly that the population is not closed between sampling events. We have also considered the use of an open N-mixture model by Dail and Madsen (2011), which is a generalized form of the Royle (2004) model that assumes population status between repeated sampling events is open according to a Markov process, where abundance at site i at sampling event t only depends on sampling event $t-1$. However, true sample replication is probably not achievable to satisfy the underlying assumptions. Open N-mixture models are also likely an unsuitable approach for insect populations that are subject to high levels of over-dispersion (J.A. Royle, personal communication). This modeling approach was previously used by Diaz et al. (2020) to address similar questions for *Heterelmis cf. glabra* within spring systems of the Devils River basin. Even though their candidate model appeared to perform well (AICc $w = 0.84$), their population estimates did not seem realistic for an insect. In as much, we will experiment with the N-mixture models; however, we are also offering an alternative analysis which is detailed below.

General linear mixed model overview

The proposed experimental design anticipates that the data to be collected will be highly structured, containing non-independent observations at multiple hierachal levels (**Fig. 1**). Based on the presence of more complex data structure, a generalized linear mixed effects model (GLMM) framework will be used to quantify spatiotemporal patterns in population performance of CSRB. GLMMs are an extension of generalized linear models that includes a combination of fixed and random effects. Fixed effects are predictor variables that are hypothesized to be ecologically meaningful in relation to the response variable. In contrast, random effects represent a grouping variable, such as levels within a hierarchy that are repeatedly sampled from a larger level. GLMMs are flexible methods for modeling non-normal data and the incorporation of random effects helps control for non-independence within the data. Moreover, explicitly accounting for within/among group structure and variation provides more reliable inferences about the fixed effects that better generalizes to the entire population (i.e., partial pooling) (Bolker et al. 2009; Kéry & Royle 2015; Harrison et al. 2018). Provided below outlines the general protocol of statistical procedures that will be used to fit, validate, and evaluate GLMMs to quantify population trends of CSRB and assess the efficacy of this approach (Zuur & Ieno 2016; Harrison et al. 2018).

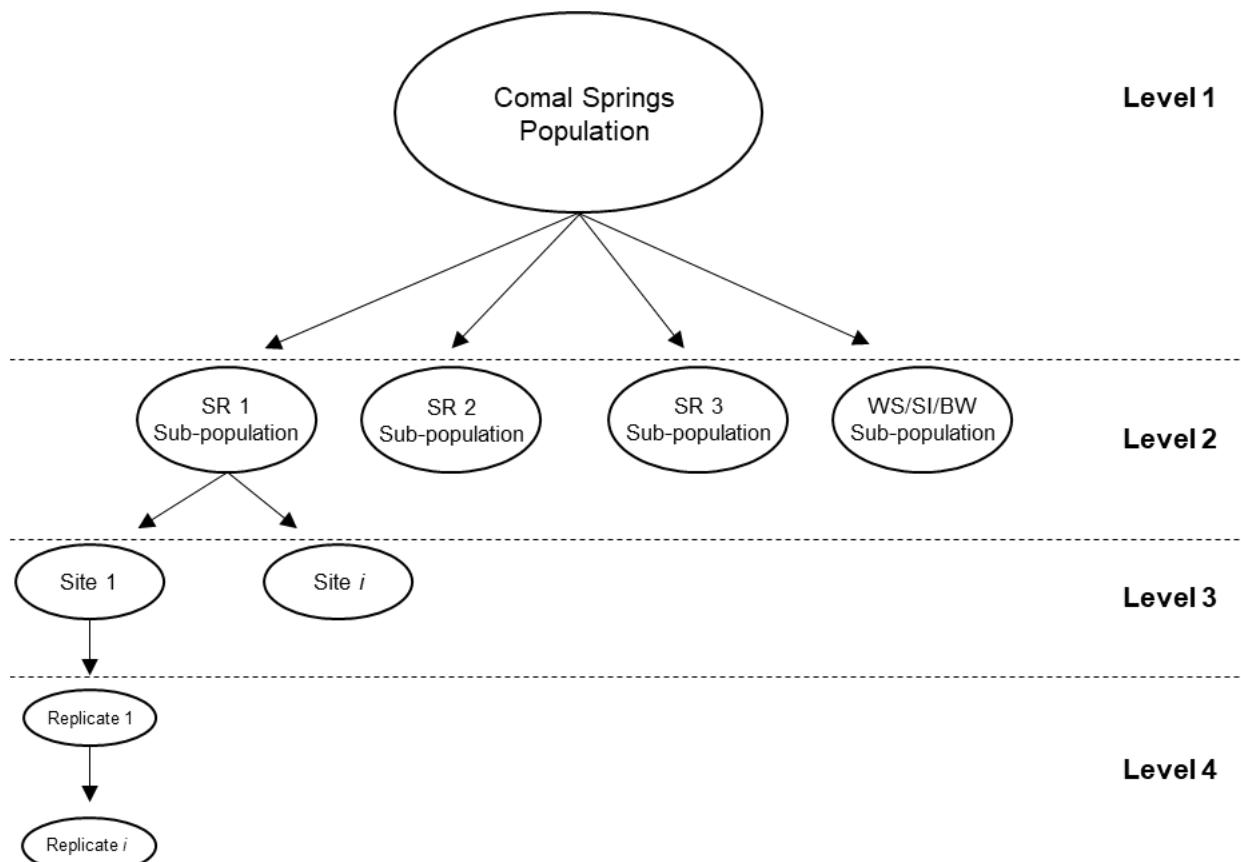


Fig. 1. Example of the hierachal structure for the CSRB data to be collected. Levels one and two characterizes the metapopulation structure of CSRB as described in Lucas et al. (2016), which include four sub-populations. Levels three and four represents the nested survey process (example provided via one sub-population), where fixed sites are repeatedly sampled.

General Linear Mixed Model Statistical Procedures

Step 1: Select appropriate population metrics and conduct exploratory data analysis

Interpretations of count data from GLMMs differ from N-mixture models, mainly because they don't explicitly model the underlying detection process that generated the observed counts (Royle 2004; O'Brien 2011). This integration of detection probability distinguishes estimates of population size provided by N-mixture models compared to hierarchical models like GLMMs, which instead estimate a population index (i.e., metric assumed to be correlated with the true population size) (O'Brien 2011). Therefore, 'relative abundance' will be used as a population index via lure counts, under the assumption that lure counts are expected to vary with population size, meaning that the direction of change in relative abundance will be used to infer trends in the population. Moreover, due to the cryptic nature of CSRB and potential observation variance (e.g., measurement error, random variability) associated with lure sampling, count data collected at a given time may be zero-inflated or highly skewed, which could make it difficult to provide reliable estimates of relative abundance via statistical inference (MacKenzie et al. 2006). Therefore, presence/absence will be used as an additional metric to quantify population trends.

Patterns in the dataset will be explored to describe/understand response and predictor variables within the data with summary statistics and data visualization. Data exploration may also help identify potentially meaningful trends and group structures. Predictor-variable relationships with the response

variable will be explored using summary statistics (e.g., central tendency, variation, zero counts, skewness, kurtosis) and visual tools as appropriate for continuous (e.g., scatterplots with loess trend lines, histograms) and categorical (e.g., boxplots) predictors. Variation within predictors and covariation among predictors will also be explored (Bolker 2008; Wickham & Grolemund 2016).

Step 2: Present the statistical models

Estimates of CSRB presence/absence will be fit with a binomial error distribution (link function = logit): $Y \sim \text{Binomial}(\eta, \Phi)$, where η and Φ denote the number of trials and probability of presence, respectively. Relative abundance will be estimated using count data and may be fit with a Poisson error distribution (link function = log): $Y \sim \text{Poisson}(\lambda)$, where λ represents mean counts and assumes λ equals the variance σ^2 . If the assumption for the Poisson distribution does not adequately represent the count data (i.e., overdispersion; $\sigma^2 > \lambda$), a negative binomial distribution (link function = log) will be used: $Y \sim \text{NegBinom}(\lambda, \kappa)$, where the second parameter κ controls the dispersion of the distribution by allowing σ^2 to exceed λ (Bolker 2008; Zuur et al. 2009). Models will be encoded to account for underlying structure within the data via nested and crossed random effects. Specifically, repeated measures are nested in sites and sites are nested in each sub-population, which are crossed with spring-type. Random intercept models (Eq. 1) will be fit and compared with random intercept and slope models (Eq. 2) for each population metric. Using notation similar to the R package ‘lme4’, both GLMMs can be described as:

$$\text{Eq. 1} \quad Y_{hijk} \sim X_{hijk} + (1 | \text{sub-population}_i) + (1 | \text{sub-population}_{i:\text{site}}_j) + (1 | \text{spring-type}_h)$$

$$\text{Eq. 2} \quad Y_{ihjk} \sim X_{ihjk} + (1 + X_{hijk} | \text{sub-population}_i) + (1 + X_{hijk} | \text{sub-population}_{i:\text{site}}_j) + (1 + X_{hijk} | \text{spring-type}_h)$$

where Y_{hijk} is the k th repeated measure in site j within sub-population i and spring-type h , and X_{ijk} is the chosen fixed effects that may include, but are not limited to, the covariates listed previously. Interactions between fixed effects that are identified as important may also be included. For the random effects component, sub-population_i (Eq 3.) is a random intercept that allows for variation between the four sub-populations, $\text{sub-population}_{i:\text{site}}_j$ (Eq. 4) is a second random intercept, allowing for variation between sites j of the same sub-population i , and spring-type_h (Eq. 5) is a third random intercept that allows variation between upwelling and margin spring habitats. Random intercepts are assumed to be normally distributed and defined as:

$$\text{Eq. 3} \quad \text{sub-population}_i \sim \text{Normal}(0, \sigma^2_{\text{sub-population}_i})$$

$$\text{Eq. 4} \quad \text{sub-population}_{i:\text{site}}_j \sim \text{Normal}(0, \sigma^2_{\text{sub-population}_{i:\text{site}}_j})$$

$$\text{Eq. 5} \quad \text{spring-type}_h \sim \text{Normal}(0, \sigma^2_{\text{spring-type}_h})$$

with a mean of zero and variance σ^2 , which determines the level of variation between these groupings. The random slopes component shown by Eq. 2 also allows for the effect size of fixed effects X (i.e., regression coefficients) to vary among groups (Zurr et al. 2009).

If the count data contains more zeros than expected from a Poisson or negative binomial distribution, a zero-inflated GLMM may be used instead. Zero-inflated models are fit using a mixture distribution with two parts that are modeled from the same data, which includes the probability of presence and mean counts when present (Zurr et al. 2009; Harrison 2014; Brooks et al. 2017). Since zero-inflated models

estimate probability of presence and mean counts in tandem, separate analyses of presence/absence and relative abundance would not be required.

Step 3: Pre-process data for model fitting

Insights from data exploration will be used to facilitate data pre-processing prior to fitting each model (Kuhn & Johnson 2013). Predictors that exhibit near-zero variance or are highly correlated ($r > 0.7$) will be removed from the dataset. Data transformation may be required if models do not converge or if assumptions are violated, which may include centering/scaling or other techniques (e.g., square root, log).

Step 4: Fit and validate the model

Each model will be fit using the R package ‘lme4’, or ‘glmmTMB’ if using a zero-inflated model is warranted. Model fit will be validated by assessing the diagnostics of the model to check whether basic distributional and structural assumptions of the model have been violated. Model diagnostics that may be checked include:

1. Overdispersion
2. Inspection of residuals
 - a. Pearson residuals vs. fitted values
 - b. Pearson residuals vs. predictor variables (fixed effects)
 - c. Pearson residuals vs. fitted values per grouping level of the random intercept
 - d. Spatiotemporal independence of Pearson residuals
3. Stability of variance components and significance of random effects
4. Goodness-of-fit (e.g., R^2)

Model fit will be assessed by checking diagnostics from the global model directly. Simulation procedures (e.g., Monte Carlo, parametric bootstrapping) will also be used to check model diagnostics, as well as examine sampling error (i.e., natural variability) of parameter estimates and uncertainty (e.g., bias, variance) of estimates for the response variable. To do this, a large number of datasets are randomly generated from a fitted model. Each simulated dataset is then used to refit the model, all of which are used to produce sampling distributions for model parameters and chosen fit statistics. Lastly, simulation results are compared to the global model to identify whether assumptions are met and if the chosen statistical model is a reasonable representation of the system (Harrison 2014; Kéry & Royle 2015).

Step 5: Model selection and evaluation

A two-step procedure will be used to select the most parsimonious model and evaluate its predictive performance. Model selection will first be used to identify which covariates best explain CSRB presence/absence and relative abundance and choose the best model for data inference. All candidate models will be ranked using Akaike Information Criteria corrected for small sample size (AICc). Differences in AICc scores will be used to calculate each candidate models weight (w) and the model with the lowest AICc score and highest w will be considered the best supported (Burnham & Anderson 2002). Models within two AICc scores will be considered equally supported, unless variables in the top model are a subset of the competing models (i.e., uninformative parameters; Arnold 2010). Model averaging may be used if selecting a single final model is not warranted.

Using the final model selected, predictive performance of each model will be further evaluated to examine how they generalize to new data. Resampling procedures (e.g., k-fold cross-validation,

bootstrapping) will be used to simulate new data and estimate out-of-sample predictive error. For each resampling iteration, a subset of the data is used to train the model, and the remaining data is used to independently examine model accuracy, which in total estimates mean generalization error (Hastie et al. 2009). Predictive performance for each model could be assessed with any of the following metrics:

1. Presence/absence model
 - a. Area under the receiving operating curve (AUC)
 - b. Sensitivity
 - c. Specificity
 - d. True skill statistic
2. Relative abundance model
 - a. Correlation
 - b. Root mean squared error
 - c. Mean absolute error
 - d. R²

Step 6: Model interpretation

Summary statistics will be presented for each model fit with the full dataset, which will include estimated variance for random effects and estimated coefficients for the fixed effects included in the final model selected. Generalization error will also be summarized based on mean (\pm error) out-of-sample predictive performance. Relative importance (0-1) of each fixed effect will be calculated based on AIC_c. Partial dependence plots will also be built to compare the strength of response-fixed effect relationships and display spatiotemporal population trends throughout the study duration.

These results will help facilitate a critical post-study review and recommendations for future research. For example, identifying specific data points with the largest predictive error may elucidate what components of the model failed to distinguish signal from noise and suggest how future work can improve predictive accuracy. Partial dependence plots may also show whether the environmental covariates used are ecologically rationale and identify covariates that have strong functional relationships with CSRB occurrence or abundance. Ultimately, we will compare the results of the GLMM to those of the N-mixture model.

Other considerations

Lure efficacy considerations

Although the GLMM does not require lure-efficacy data, understanding lure efficacy could be useful for estimating the population size at Comal Springs. *Efficacy (E)* can be interpreted from a recent luring study (BIO-WEST 2021); even though there were adverse conditions for each of the trials, results indicated that 0 – 80 % of the beetles would reside on cotton, with an average of ca. 20 %. This information could be used as a means of estimating the number of beetles in the vicinity of a lure and therefore, one adult observed on a lure during a check could be interpreted as four others in the vicinity (total of five adults). However, this cannot be done for larvae as it is expected that their movements are less than that of adults.

Additionally, we plan to place 5 lures (separated by the length of a lure) at select locations where large numbers of *H. comalensis* are expected to be found. After ca. 30 days of conditioning, the number of beetles will be counted among those lures. The beetles will be replaced and a single lure will be placed in the middle of where the set of five was and the lure will be inspected ca. one week later as a means of

developing field-based E that may also be useful for larvae. We plan to do this for at least 3 separate spring locations that are not part of the study sample sites. However, this is not a major focus of the study since the results may not be useful (i.e., in the event that no beetles are retrieved from the single lure).

Beetles per unit area

A simplistic approach to estimating the population size of *H. comalensis* is to take the fraction of the overall area sampled and extrapolate the number of beetles sampled to the total area of spring activity (corrected by lure efficacy) where a is the proportion of spring area sampled by the entire set of lures per sampling event with a given E , the surficial population N can be estimated based on the total number of sampled individuals n :

$$N = \frac{n}{a \times E}$$

If for instance a survey in a single sampling event finds 1,000 adults among 80 lures that represent 20% of the surficial spring area (a) and E is considered 20% effective, the total estimate of adults at the near surface would be 25,000. The coefficients a and E can be adjusted at later times as better information becomes available.

Biofilm considerations

The quality of the biofilms that form on the poly-cotton lures are thought to be an important factor with regard to the attraction of the beetles to the lures. Preliminary work by Dr. Camila Carlos-Shanley indicated that the biofilms found on these lures can be highly variable in terms of bacteria taxa and relative abundances of those taxa (personal communication). Having a diverse community of bacteria per lure elicits many metric measures that can be delineated as covariates of riffle beetle presence and abundance. The extraction of such data would require the expense of genomic sequencing for each lure and the time for a technician/student to perform the bioinformatics. However, the lures can be kept in 95% EtOH and stored for a few years. We would like to offer the service of saving these biofilms if there is an interest in pursuing this type of data acquisition in the future. It is also noted that a better understanding of the microbial community affiliated with *H. comalensis* could also help focus habitat restoration efforts.

Literature Cited

BIO-WEST. 2021. Comal Springs riffle beetle luring analysis. Final report. Prepared for the Edwards Aquifer Authority.

BIO-WEST. 2017. Comal Springs riffle beetle (*Heterelmis comalensis*): life history and captive propagation techniques. Final report. Prepared for the Edwards Aquifer Authority.

Bolker, B.M. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.

Bolker B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens M.H.H., and J.S.S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127–135.

Brooks, N.E., K. Kristensen, K.J. van Benthem, A. Margnusson, C.W. Berg, A. Nielsen, H.J. Skaug, M. Mächler, and B.M. Bolker. Modeling zero-inflated count data with glmmTMB. bioRxiv
DOI:10.1101/132753.

Burnham K.P., and D.R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Second Edition. Springer-Verlag, New York, NY.

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

Dail, D., and L. Madsen. 2012. Estimating open population site occupancy from presence-absence data lacking the robust design. *Biometrics* 69:146-156.

Diaz, P.H., J.R. Gibson, C. W. Norris, and C. L. Thompson. Abundance and distribution of *Heterelmis cf. glabra* (Coleoptera: Elmidae) within Dolan Falls Preserve and the Devils River State Natural Area, Texas, USA. doi: <https://doi.org/10.1101/2020.02.10.941641>

Fries, J.N. 2003. Possible reproduction of the Comal Springs riffle beetle, *Heterelmis comalensis* (Coleoptera: Elmidae), in captivity. *Entomological News* 114: 7–9.

Harrison, X.A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2: DOI: 10.7717/peerj.616

Harrison, X.A., L. Donalson, M.E. Correa-Cano, et al. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6: DOI: 10.7717/peerj.4794

Hastie, H., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. Springer, New York, NY.

Kéry, M., and J.A. Royle. 2015. Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance, and species richness in R and BUGS. Volume 1, Academic Press, Cambridge, MA.

Kosnicki, E. 2022. Fecundity of first generation captive reared *Heterelmis comalensis* Bosse, Tuff, and Brown 1988 (Coleoptera: Elmidae). *Journal of Insect Science* (accepted for publication).

Kuhn, M., and K. Johnson. 2013. Applied Predictive Modeling. Springer, New York, NY.

Lucas, L.K., Z. Gompert, J.R. Gibson, K.L. Bell, C.A. Buerkle, and C.C. Nice. 2016. Pervasive gene flow across critical habitat for four narrowly endemic, sympatric taxa. *Freshwater Biology* 61:993-946.

MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.L. Bailey, and J.E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier-Academic Press, San Diego, California.

O'Brien, T.G. 2011. Abundance, Density and Relative Abundance: A Conceptual Framework. Pages 71-96 in A.F. O'Connell, J.D. Nichols, and K.U. Karanth, editors. *Camera Traps in Animal Ecology: Methods and Analyses*. Springer, New York, NY.

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

Wickham, H., and G. Grolemund. 2016. R for Data Science. O'Reilly, Sebastopol, CA.

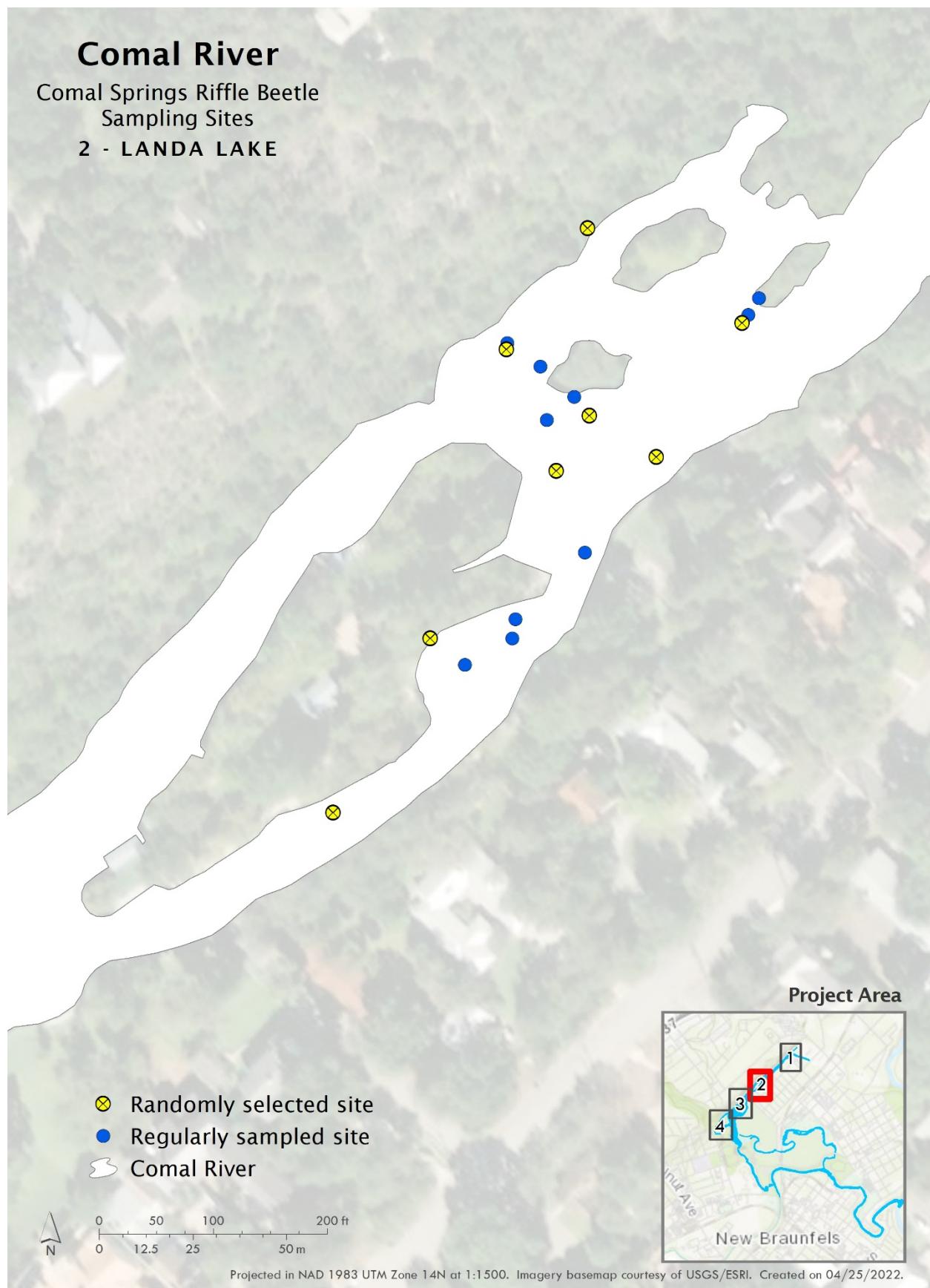


Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York, NY.

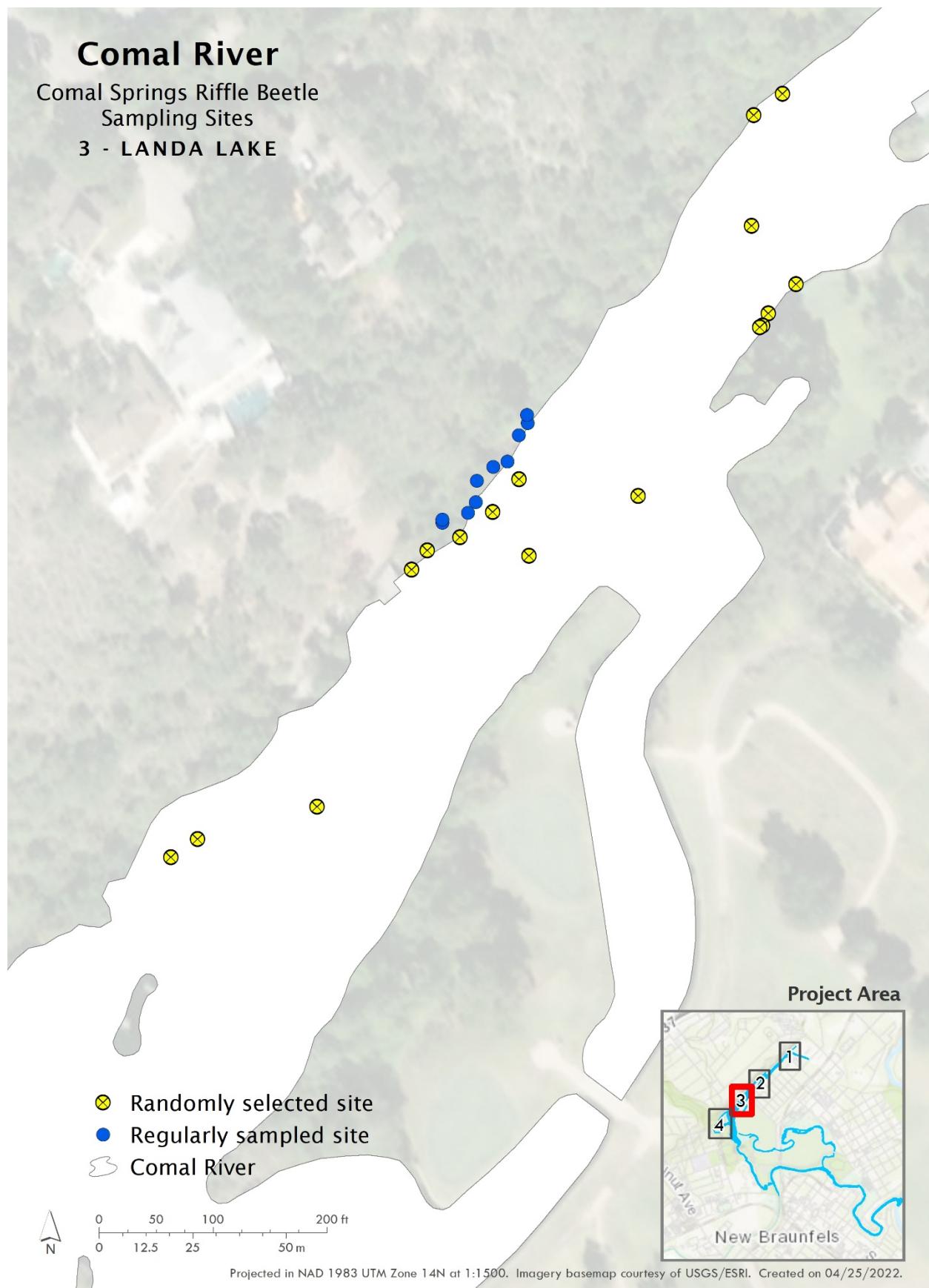
Zuur, A.F., and E.N. Ieno. 2016. A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution* 7:636-645.



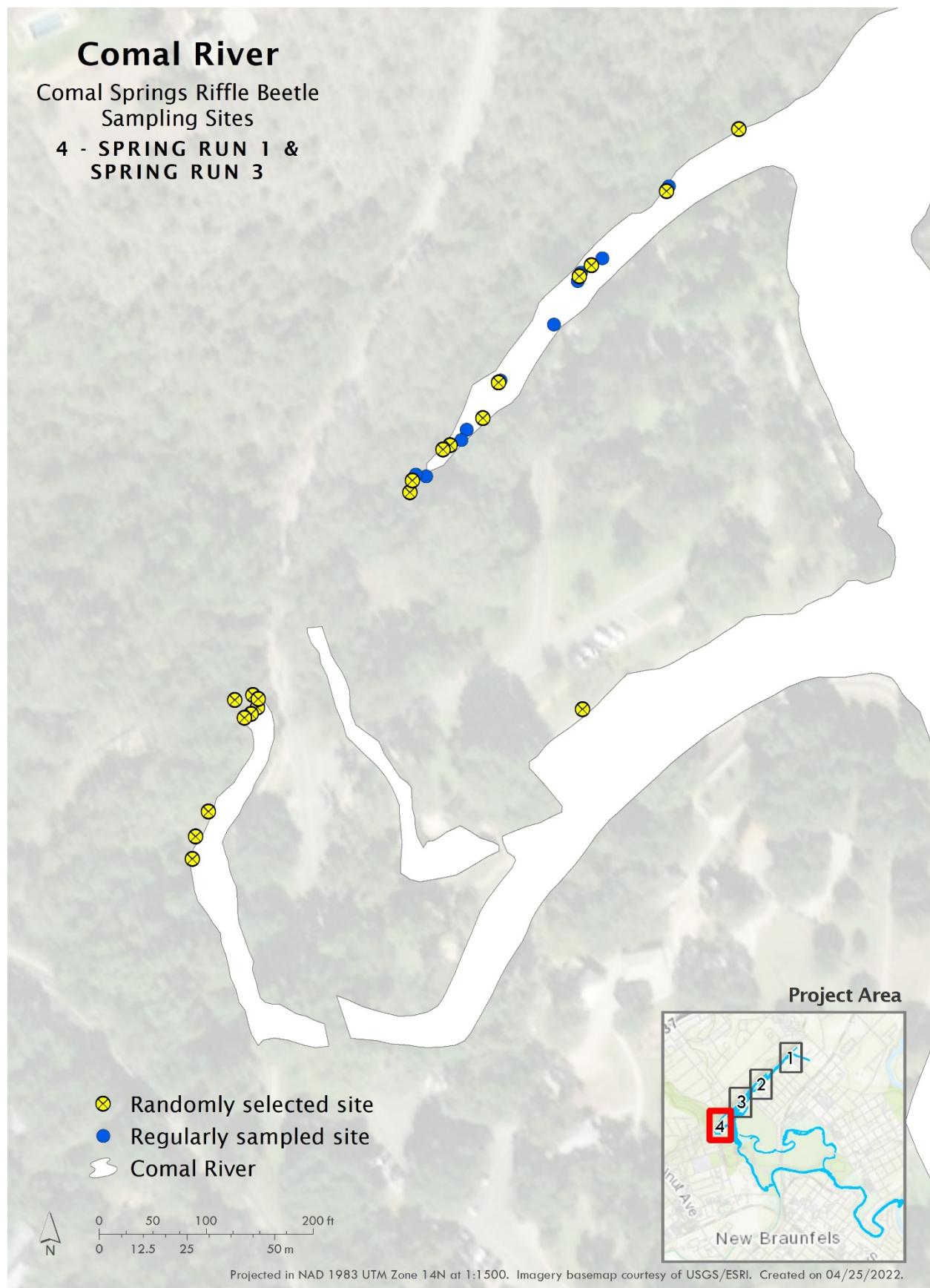
Map 1a. Map of Comal Springs and randomly selected sites for sampling *Heterelmis comalensis*.



Map 1b. Map of Comal Springs and randomly selected sites for sampling *Heterelmis comalensis*.



Map 1c. Map of Comal Springs and randomly selected sites for sampling *Heterelmis comalensis*.



Map 1d. Map of Comal Springs and randomly selected sites for sampling *Heterelmis comalensis*.