COPPERHEAD
ENVIRONMENTAL CONSULTING

North Carolina Department of Transportation ATLAS Bats in Bridges Generalized Linear Model

Testing the efficacy of engineered data for predicting bat roosts in North Carolina DOT bridges: A summary report of which features influence use

Prepared for:

Morgan Weatherford
Modeling Team Lead
North Carolina Department of Transportation
Transportation Building
1 S. Wilmington St.
Raleigh, NC 27601

Prepared by:
Richard Borthwick
Senior Biologist

16 July 2021

COPPERHEAD ENVIRONMENTAL CONSULTING, INC.
P.O. BOX 73 ■ 471 MAIN STREET ■ PAINT LICK, KENTUCKY 40461
(859) 925-9012 OFFICE (859) 925-9816 FAX

www.copperheadconsulting.com

EXECUTIVE SUMMARY

Severe population declines in numerous North American bat species due to white-nose syndrome have prompted increased concern for monitoring bat populations at various life-history stages. There are also increased legislative protections for species at the state and federal levels to help limit impacts. As part of its broader effort to avoid and minimize impacts on federally listed bat species, the North Carolina Department of Transportation (NCDOT) is improving their internal processes regarding species surveys and monitoring efforts. This effort is a key component of NCDOT's Advancing Transportation through Linkages, Automation, and Screening (ATLAS) program. In particular, NCDOT is using machine-learning models to refine species maps for federally listed species. Bats were originally excluded from NCDOT's machine learning models because the federally listed species had too few records to develop an adequate model; bat habitat usage is highly variable through seasons, life-history stages, and across species; and approved programmatic habitat conservation plans (HCPs) are already in place for much of the state and for most priority bat species. To provide linkage, automated, and screening model tools relevant to bat research needs and NCDOT efforts to streamline bridge repair and maintenance, we developed a probabilistic model (i.e., generalized linear model [GLM]) for bat-use in bridges. The GLM was originally developed from a dataset of 709 NCDOT bridges that had undergone bat habitat assessments and surveys looking for roosting bats. It was then modified based on expert review and field testing. The results of the GLM can be applied to a larger, statewide dataset of NCDOT bridges, and the outcome of that application has been applied as final shapefiles.

Because environmental data available for many of the 709 bridges were not available for other bridges throughout the state, we considered engineered specifications that may serve as proxies for bat selection parameters in bridges. We developed these GLM inputs with bat experts from non-government organizations (NGOs), federal agencies, NCDOT, and other state agencies. Our refined list of structural variables was used to assess which, if any, were relevant in predicting bat roosting in bridges, how they contributed to bat roosting, and which model structure was the most effective in predicting potential bat roost use.

We shared our preliminary results with NCDOT and performed field testing to quality control the GLM's efficacy. Field testing feedback was incorporated into the GLM and the model was re-scored and re-evaluated. The results of these processes are presented in this report.

TABLE OF CONTENTS

LIST OF TABLES

APPENDICES

INTRODUCTION

The North Carolina Department of Transportation (NCDOT) is improving systems processes across their range of operations by implementing the Advancing Transportation through Linkages, Automation, and Screening (ATLAS) program. A targeted element of project ATLAS is modeling threatened and endangered species throughout the state to better understand how they interact with NCDOT infrastructure including bridges. Copperhead Environmental Consulting, Inc. (Copperhead) was originally retained to develop a machine-learning model for bat species including maximum entropy (MaxEnt) and random forest (RF) models. However, during initial project meetings, Copperhead and project bat experts at NCDOT (Melissa Miller) and CALYX (Heather Wallace) expressed concerns about the ability to adequately develop a MaxEnt or RF model. Concerns included limited element occurrences for the target bat species within the state's Natural Heritage Program (NHP) database, complex phenology of bat-habitat use, locational sampling bias for a landscape-wide model, and diverse and species-specific landscape use. To overcome these limitations, Copperhead proposed developing a generalized linear model (GLM) to model bat roost-use in bridges in North Carolina. This proposal, approved by NCDOT, benefits the ATLAS program by increasing the currently limited information on bat roost-use in bridges and by allowing NCDOT to better screen and prioritize bridge survey efforts. In addition, the GLM could leverage existing bat data, existing bridge-survey data, and the existing NCDOT engineering data for bridges (WIGINS data).

Increasingly bats are being found using bridges and culverts throughout various seasons (Allen et al. 2010, Stepanian and Wainwright 2018). Current research is focusing on micro-habitat quantification to understand why specific roosts are selected within a bridge (Ferrara and Leberg 2005, Bektas et al. 2018), and this has fundamental ecological relevance and importance, particularly if transportation departments can use this data to incorporate bat roosting spaces as an onsite mitigation tool. However, models with such targeted data collection cannot easily be expanded to statewide probabilistic models which can have wide ranging utility for project screening and risk assessment for bridge repair and maintenance.

Although it is widely accepted that bats use bridges as roosting sites (Keeley and Tuttle 1999), little attention has been given to understanding the combined characteristics of bridge and habitat associated with their use of bridges as roosting sites in North Carolina (Keeley and Tuttle 1999, Bektas et al. 2018). We investigated how engineered bridge features corresponded to bridge roost selection by bats. Because our surveyed bridges included any sign of bats as a "used" bridge, this model considers any roost including short-term night roosts. A major goal of this research was to better understand when bridge replacement, repair, or rehabilitation projects have the potential for "taking" (i.e., harassing, injuring, or killing) bat species listed as federally threatened or endangered under the US Endangered Species Act (ESA 1973). The federally endangered gray bat (*Myotis grisescens*), federally threatened northern long-eared bat (*M. septentrionalis*), and the little brown bat (*M. lucifugus*) that is currently under review as a candidate species for listing, are all found in bridge structures in North Carolina.

METHODS

Initial Model Development

Original model development was completed using two primary data sources: bat habitat assessments (ecological data) in North Carolina and bridge engineering specifications (WIGINS). The ecological data-list was developed from bridges surveyed for bat use by trained wildlife biologists and literature review. The WIGINS database contains the engineered specifications for each bridge, including measurements and materials. Ecological data forms were considered to help develop corollaries between necessary ecological needs for bats and engineered WIGINS information to develop a robust GLM framework that could be applied statewide.

Bridge habitat assessments are carried out according to the *NCDOT Preliminary Bat Habitat Assessments (Structures, Caves, & Mines)* (NCDOT 2019) and include quantifying land cover, identification of bridge features likely important to bat species (e.g., structure material, crevices present), and presence of bats (Keeley and Tuttle 1999). Originally, we hoped to use the ecological data almost exclusively, but both datasets (ecological and WIGINS) included data gaps, superfluous information that would not inform the model, and differing structure identification numbers. Consequently, we collected remotely sensed data to augment ecological data. However, we still needed to link surveyed bridges from the existing ecological data to engineered data and remotely sensed data. Because these datasets were not designed to be integrated with each other and integration was complex, the project team (Jessica Tisdale [HDR], Richard Borthwick [Copperhead], and Eric Wilson [KCI at the time]) met with NCDOT staff (Walt Tallman) to understand the definitions of the engineered features and which features (as defined by an engineer) correspond to bat use features (as defined by a biologist). These were supported through expert opinion and literature review (Bektas et al. 2018, Cleveland and Jackson 2013, Gore and Studenroth 2005, Keeley and Tuttle 1999, Lance et al. 2001, Ferrara and Leberg 2005). Two data gaps posed challenges in coordinating the two datasets: incomplete bridge structure numbers and inaccurate or missing GPS locations. The following are the specific sources of these issues:

- **Missing county code:** All WIGINS structure numbers (i.e., unique bridge identification number) consist of six-digit codes with the leading two digits corresponding to a county code. In many of the habitat assessment structure numbers, the county code had been omitted and the structure number (consequently less than 6 digits) did not correspond with the WIGINS database. Both datasets also include the county name, so we used this data (where complete) to augment the structure number from the habitat assessment data with the appropriate code.

- **Missing structure number:** Without a structure number, the structure was untraceable through the WIGINS database. Where this occurred, GPS data were used to find "nearest neighbor" bridges in the WIGINS database.
- **Missing GPS data**: The lack of GPS data resulted in no geo-referenced information for the assessed bridge. If the structure number was complete, the bridge was directly linked to WIGINS through the structure number. Two bridges were missing GPS data and a structure number. These sites were deleted from the GLM training source data. Although they could have been manually identified through aerial photos on a map based on location descriptions, neither site had bats roosting in them and were not likely to contribute significantly to the GLM.

In addition to these issues, the coordinates provided from both datasets were not being displayed correctly when imported into GIS software (i.e., ArcMap or QGIS) (Ryan Dugger, HDR, Pers. Comm.). Therefore, bridge GPS locations from both datasets were added to Google Earth Pro, each dataset was exported as a *.KML file, and each *.KML was imported into ArcMap to spatially join the two *.KML files into a single shapefile (Appendix 1).

Because the habitat assessment data is dependent on a field visit and our intention was to test models that could be sourced from existing data and applied to bridges that have not yet been surveyed, we used ArcMap and remotely available data to help determine proxies for ecological variables. The remotely sensed data included the latest iteration of National Land Cover Database (NLCD) (Yang et al. 2018) to determine presence of water at each site and surrounding land cover, and we used the mineral resources data systems (MRDS) embedded in the ATLAS layer[1] to determine cave or mine presence within 0.5-miles of the crossing structure. Due to missingness, we were forced to remove sun exposure, an important consideration for bridge-roosting bats (Ferrara and Leberg 2005, Keeley and Tuttle 1999, Lance et al. 2001).

Our team interpreted the WIGINS engineering data from 13,910 bridges to provide proxies for ecologically relevant variables. For example, joins and seals are entered as total lengths in unique columns depending on the join or seal type. However, because we considered these data a proxy for crevice availability in a bridge, we combined varied seal and join data into one column used as total crevice potential. Crevices have been shown to be important bat roost selection (Keeley and Tuttle 1999), but identifying remote parameters to define crevices is challenging. In addition to seals and joins, we considered the number of spans and age of the bridge to further refine proxies for crevices.

After consolidation was complete, the data had an imbalance in bat presence ("1") and absence ("0"). Consequently, we ran a series of GLMs with various techniques to address the imbalance. This included zero-inflated GLMs through the "pscl" package in the R software environment (R

[1]https://gis23.services.ncdot.gov/arcgis/rest/services/AtlasMapServicesStatic/NCDOT_NCMines/MapServer/0

Core Team 2019) and a random over-sampling example (ROSE) GLM using the "ROSE" package in the R software environment. Because model generation requires knowledge of use, inflated zeroes in a dataset (i.e., one case is rare) can result in a poor model fit or an over-fit predicting negative or absent data. Because a model tends to focus on prevalent data (Japkowicz & Stephen 2002), a data imbalance can compromise model training and (due to scarcity of a class) model accuracy testing (Menardi & Torelli 2010).

Several methods exist to address data imbalances (Barandela et al. 2004), but there are some conflicting opinions about how to best address the issue (e.g., Allison 2012, Greene 2007). We tested three model variations to explore which improved model predictive efficacy. First, we ran a conventional model with a reduced negative sample using a modified Wilson's editing. This method generally requires a reduction of both positive and negative occurrences weighted more heavily on negative samples. However, due to the large imbalance in this dataset, meaningfully reducing zero occurrences was only possible if extraction included only negative data. This method resulted in an unquantifiable loss of data and a relatively low-quality model. Second, we ran a zero-inflated negative binomial (ZINB) model which attempts to account for the imbalance within the model framework. This output focused on the prevalent (i.e., negative) data, a well-documented shortcoming (e.g., Menardi & Torelli 2010). Third, we considered a smoothed bootstrap form of re-sampling from the data to develop a balanced dataset for model training (ROSE). This was the only oversampling effort used and we deemed it appropriate as Barandela et al. (2004) found that when the data imbalance was severe, oversampling was effective.

A ROSE model consists of a random oversampling of sites with replacement for the rare class, and without replacement for the common class. These methods help avoid the issue of missingness that arises from undersampling (i.e., removing zeroes), but increase the risk of over-fitting as sample units of the minority class are duplicated (Menardi & Torelli 2010). To offset the issue of overfitting, Chawla et al. (2002) proposed randomly generating new sample unit observations based on nearest-neighbor values from existing sample points. The generation of new artificial data has the benefit of reducing over-fitting risks. However, there are inherent concerns in randomly generated new data. This has been subsequently studied for applicability and efficacy (Guo & Viktor 2004, Mease et al. 2007, Menardi & Torelli 2010). Due to the unknowns in bat-roost selection, we opted to use the ROSE model as opposed to generating new data. This resulted in better predictive power than either negative data removal or ZINB models.

The model was used to estimate the probability of bats using a bridge for a roost. To create a confusion matrix of successful predictions, a probability threshold was set to allow for the assignment of a binary "yes" (i.e., "batty") or "no" (i.e., "non-batty") based on the estimated probabilities. Thresholds were adjusted when running each GLM so that the probability value was optimized. To complete this assessment, we tabled various probability thresholds and their respective confusion matrices to maximize both specificity and sensitivity. We completed this trial-and-error contrasting table for each model iteration as it was possible that the optimal threshold would change.

We used threshold assignments to calculate model specificity and sensitivity. Specificity is the true negative rate (number of predicted negatives divided by the total number of actual negatives) while sensitivity is the true positive rate (number of predicted positives divided by the total number of actual positives) (Table 1). Specificity is our ability to accurately predict bat absence from bridges while sensitivity is our ability to accurately predict bat presence. For our purposes, sensitivity is more important as we do not want to miss bat bridges. As sensitivity increases, our rate of false positives also increases, resulting in over-prediction of bat presence. The optimal model was conservative in nature and low specificity meant that we were routinely predicting "non-batty" bridges as "batty" when using a 32% threshold. For both zero-inflated and ROSE model structures, specificity and sensitivity were considered to evaluate which method of addressing a zero-inflated dataset predicted most effectively. Akaike's Information Criteria (AIC) or Bayesian Information Criteria (BIC) are generally used to compare model structures because, when the response variable remains the same, these criteria indicate improvements in model fit as explanatory variables are altered. However, our model iterations altered the response variable distribution, so an alternative model scoring framework was required. Predicted outputs and probability scores were calculated for the statewide data, saved as a shapefile, and uploaded to the HDR and NCDOT Sharepoint hub for file storage.

Table 1. Confusion matrix format for calculating specificity and sensitivity for each model type.

	Actual Positive	**Actual Negative**
Predicted Positive	True Positives – Model predicted bats, and bats or bat sign were present during survey.	False Positives – Model predicted bats, but no bats or sign were present during survey.
Predicted Negative	False Negatives – Model predicted no bats, but bats or bat sign were observed during surveys.	True Negatives – Model predicted no bats, and no bats or bat sign were observed during surveys.

First Model Revisions

The statewide GLM shapefile was reviewed on ArcGIS Online by species experts at the US Fish and Wildlife Service (USFWS), NGOs, and NCDOT. Their comments (Appendix 2) indicated that overall the GLM was performing well. Reviewers recommended changes making the GLM less conservative (increasing specificity) and incorporating bridge height and deck materials which were being poorly reflected in the GLM. The model was comprised of several variables queried from the WIGINS database (Appendix 3), but bridge height was not available during this initial model revision.

Reviewers were encouraged to comment on agreement or disagreement with the GLM, the reasons why, and whether the model should be changed or retained at specific bridges. The GLM was adjusted from these comments by including main structure material and by including beam type and material, then it was rerun. We reapplied the new probability scoring to bridges across the state and saved the updated shapefile for implementation into the digital data collection framework designed by HDR and for ease of reference and review. The new data collection framework was an automated data collection protocol that calculated surrounding habitat for each bridge structure and auto-populated digital data forms for ecological data collection. This initiative was intended to streamline field data collection, ensure completed data forms, and facilitate future data modeling efforts during bat habitat assessments at bridges. The automated system was beta tested in the fall of 2019 and included a field test for the probabilistic GLM's accuracy and applicability. Field testing was completed by teams from Copperhead and HDR and informed the second model revisions.

Second Model Revisions

Second model revisions incorporated a field test of the model and expert investigation of modeled bridge scoring. Increasing the GLM's specificity continued to be particularly relevant to wooden decked bridges where the GLM predicted high probability of bat use, but the field circumstances and past research (Ferrara and Leberg 2005, Keeley and Tuttle 1999) have ruled out that likelihood. Further, field testing documented that source data being used for deck structures was incorrect, and a new query was initiated from the WIGINs database.

The WIGINS database queries are highly detail specific. We encountered two primary issues with the queries: incomplete data and incorrect data columns. Incomplete data were an issue when regional or county data were separated during a query, i.e., a formatting error, or when data were still being entered and updated. Through the process improvements within NCDOT, staff are continuing to collect data to fill known information gaps and the lag between these data updates can result in unpredictable gaps in the data. This required careful screening prior to a variable's use in the model resulting in not all variables being used. For example, data gaps on bridge height prevented use of this variable during our initial model runs, but they were entered in fall 2019 and were subsequently incorporated. The later inclusion of bridge height data helped improve our model during the second round of revisions. Another example of missing data was guardrail materials which were inconsistently provided so were omitted from the model.

One primary data column – Structure Type – was uploaded incorrectly in the initial model. Because the WIGINS database is organized with coded data labels and because the labels are not ecologically related, we initially requested structure material type instead of deck type to model the bridge deck materials. Structure Type consists of a 3-digit code that includes the span materials and span design (FHWA 1995); however, this differs from the WIGINS coding system and definition for deck type. Deck type is a variable that is developed and coded within NCDOT and is not part of the FHWA (Federal Highway Administration) guidance document. It is, however, included in the WIGINS reports. After review and field verification, deck type better reflected the surface that the bats contacted under the bridges. By altering our query request, we improved our model, particularly pertaining to predictions of false positives on smaller wooden bridges. Accuracy increased on the field verified models recommended for change by 7%.

The final GLM incorporates repaired deck source data, newly available bridge height data, past reviewer comments, and comments from the field test. The results below include model specificity and sensitivity and some summary statistics for model improvements associated with revisions. The final GLM was a global model (i.e., includes all parameters and available for analysis) that was dredged using the "MuMIn" package in the R software environment. The top two reduced models were compared for the final model and are presented as "Reduced 1" and "Reduced 2" below. The GLM with the highest sensitivity was considered the most appropriate and was applied to all bridges within the state to produce a GIS shapefile for bat expert review. During GLM reduction, AIC was considered to identify the most parsimonious and powerful model within the ROSE model framework, not across model frameworks. Because parameters differed in their units, scales, and values, we scaled data prior to model development.

RESULTS

To address our zero-inflated dataset, we tested a "zero-inflated" model and a ROSE model. The zero-inflated models under-predicted bridges that housed bats (43% sensitivity) but more effectively predicted absences (96% specificity) (results not tabled). This model structure was ineffective at the scope and scale of the intended application, as bridges being used by bats were routinely missed. Consequently, the ROSE model was selected to develop the GLM. Under the **initial model development** (using all available i.e., 'global' model parameters), the ROSE model sensitivity was 87% and specificity was 44% at a 32% threshold.

Table 2. Confusion matrix for the initial model against test data

	Actual Positive	**Actual Negative**
Predicted Positive	86	420
Predicted Negative	13	330

The **initial** GLM was reviewed by five bat ecology experts from NGOs, state agencies, NCDOT, and USFWS. Each reviewer investigated a minimum of 12 sites independently and a total of 60 structure probabilities were adjudicated. Of these reviewed points, 44 were considered appropriate while 16 were recommended for reversal (Appendix 1). Of the 14 sites recommended for reversal, the model predicted no bat presence in 12, but expert opinion identified potential for roosting. Our first model adjustments addressed only 5 of these bridges because beam type and bridge height – the variables consistently driving recommendations to reverse model predictions – were unavailable at that time and were not included until second model revisions. The primary concern from reviewers was that the GLM would be used inappropriately to classify bridges.

With adjustments from expert feedback, we integrated **first model revisions**, which were applied to the predictive model and used for field testing. Field tests included the inspection of an additional 280 bridge structures. Of these, 19% were recommended by field teams to reverse from a model predicted "batty" to a real-world "non-batty" score and 4% were recommended to reverse from predicted "non-batty" to a real-world "batty" score. This corresponded to expectations in the model and resulted in the conservative estimates anticipated. Beam type and bridge heights were again the driving features that were poorly reflected. To date, an average of 9% of surveyed bridges have contained bats or bat sign, and models predict ~62% of all bridges are likely to contain bats.

We adjusted the global model based on these recommendations, when bridge height data became available and a final global model was defined. These adjustments **improved** the global model which was then dredged to determine the best model iterations: "Reduced 1" and "Reduced 2." The optimal model iteration had 93% sensitivity and 54% specificity at a 35% threshold (Table 2). The top two model iterations are provided in Table 1 with AIC and ΔAIC values.

Table 3. Confusion matrix after first model revisions.

	Actual Positive	**Actual Negative**
Predicted Positive	92	344
Predicted Negative	7	406

Model revisions from the initial model development increased true positive classifications from 89% to 95%, improved balanced model accuracy from 66.5% to 73%, and addressed shortfalls in data availability as beam type and bridge height became available for incorporation into the model.

Final GLM selection was based on the most parsimonious model with the highest specificity and sensitivity and lowest AIC values. Although the Reduced 2 model had the lowest AIC score, it had slightly lower than optimal sensitivity and specificity.

According to the GLM analysis using ROSE, the most important scaled predictors of bat use in bridges were wood decks (-), wood beams (-), steel decks (-), concrete beams (+), surrounding forest (+), concrete decks (+), steel beams (-), urban developments (-), and bridge heights (+) (Table 2). Importance is based on model coefficient weights from scaled variables (Table 3). Variables that had significant p-values are bolded in Table 3, and these variables also generally had the largest scaled coefficients. This indicates changes in these variables are the most likely, in our model framework, to influence bat roosting potential.

Table 4. Model iterations for random oversampling example (ROSE) model structures after expert review and field test adjustments. Explanatory variables are presented in Appendix 1. Specificity = "Spec", Sensitivity = "Sens"

Iteration	Model	AIC	∆AIC	Spec	Sens
Global	mainspans + approachspans + year + lanes.on + lanes.under + ADT + trucks + sub.grade + btwnrails + outtoout + beam.height + leftrail + rightrail + length + water_vici + RD_ANGLE + seal + seal.count + forest +dev + ag + open + wet +brg.ht + mat_cont2 + concrete_d + steel_d + wood_d + concrete_b + metal_b + wood_b	608.25	0	0.49	0.92
Reduced 1	mainspans + approachspans + year + outtoout + beam.height + length + water_vici + RD_ANGLE + seal + seal.count + forest +dev + open+ wet +brg.ht + mat_cont2 + concrete_d + steel_d + wood_d + concrete_b + metal_b+wood_b	596.13	10.19	0.54	0.93
Reduced 2	mainspans + approachspans + year +beam.height + length + seal + forest + dev + wet + brg.ht +wood_d + concrete_b + metal_b	587.35	20.9	0.47	0.90

Table 5. Variables in the optimal random oversampling estimates (ROSE) model, their predictive coefficient, and level of significance. Bolded variables were significant at *P* < 0.05.

Variable	Coefficient	Std. Error	z-value	*P*-value
Number of Main Spans	**0.23335**	**0.08255**	**2.827**	**0.004704**
Number of Approach Spans	**-0.25646**	**0.1241**	**-2.067**	**0.03878**
Year	-0.16696	0.08928	-1.87	0.061472
Out to Out	0.05758	0.11202	0.514	0.607242
Beam Height	**-0.24948**	**0.09138**	**-2.73**	**0.006333**
Length	**0.21166**	**0.10083**	**2.099**	**0.035794**
Water (0.5 miles)	-0.24502	0.26713	-0.917	0.359032
Mine (0.5 miles)	-0.1135	0.41276	-0.275	0.783331
Road Angle	-0.08666	0.08682	-0.998	0.318197
Seal Length	**0.19728**	**0.08611**	**2.291**	**0.021954**
Seal Count	-0.07997	0.06122	-1.306	0.191454
Forest Cover	**0.46071**	**0.10686**	**4.311**	**1.62E-05**
Developed Land	**-0.36054**	**0.10771**	**-3.347**	**0.000816**
Open Habitat	-0.02738	0.08404	-0.326	0.744572
Wet Habitat	**0.19341**	**0.06962**	**2.778**	**0.005467**
Bridge Height	**0.30877**	**0.10086**	**3.061**	**0.002203**
Deck Materials - continuous	0.07876	0.12449	0.633	0.526941
Concrete Deck	0.42848	0.42391	1.011	0.312123
Metal Deck	-0.73977	0.86707	-0.853	0.393557
Wood Deck	**-1.68017**	**0.473**	**-3.552**	**0.000382**
Concrete Beams	**0.68362**	**0.23805**	**2.872**	**0.004082**
Metal Beams	-0.41613	0.22702	-1.833	0.066794
Wood Beams	-1.18146	0.95559	-1.236	0.216323

DISCUSSION

This was a broad-scale effort to apply engineered bridge specifications and remotely sensed ecological data to a statewide probabilistic GLM for bat use in bridges. The final GLM predictors corresponded to those recommended in an *a priori* discussion with bat experts. The predictors hypothesized to be important were as follows: crevices, gaps, sun-exposure, bridge height and size, deck and beam material, and surrounding habitat including the presence of snags (Bektas et al. 2018, Cleveland and Jackson 2013, Gore and Studenroth 2005, Keeley and Tuttle 1999). We were unable to develop a proxy for sun exposure but found engineering data or remotely sensed data that was indicative of the remaining variables. Our study differs from these past studies in that we surveyed more bridges for bat use within North Carolina, we used a wider array of engineered specifications, and we investigated all bat species.

The GLM quantifies the contributions of engineered variables to bat roost selection in bridges in North Carolina. Although a bat may, in theory, use any bridge crevice as a roost, the model indicates that optimal bat roosting bridges are those with concrete beams that are relatively long and high, have multiple spans, and are near wet habitat (i.e., NLCD classified woody wetlands or emergent herbaceous vegetation) or forest cover. Bridges that have wooden decks, are relatively low, and lack gaps or crevices associated with multiple spans and high seal lengths are likely not to be used. Ecologically, these findings suggest that bats are looking for gaps and crevices, materials that exhibit thermal inertia, and offer protection from the elements, usually from beams.

Although we did not complete species-specific models, our findings align with past research in Louisiana that documented *Corynorhinus rafinesquii* rarely used wooden bridges but did use 48% of concrete bridges surveyed and that flat slab bridges were not highly selected (Lance et al. 2001). Ferrara and Leberg (2005) found that bats tended to roost in the darkest portion of the bridge and beams help provide darker environments. Our model shows that the level of traffic on, below, or around the structure does not seem particularly relevant for selection, which differs from the findings of Keeley and Tuttle (1999).

Corresponding to existing literature, we found that concrete substructure and beam type were positively correlated with bat roosting, as was surrounding forest habitat (Ferrara and Leberg 2005, Keeley and Tuttle 1999). Gaps and joins were generally important as well, as indicated by the seal length and number of spans.

Low bridges with wood decks were generally unsuitable for bat roosts potentially due to convective heat loss or creosote treatments. Wooden decks had a highly significant negative relationship and the highest predictive coefficient with bat-roost use and were more than four times as likely to deter bat roosting than concrete decks are to encourage bat roosting. Concrete decks did not have a significant relationship with roost-use, but concrete beams did and were the most important positive predictor of bat roost use in bridges based on the scaled model correlation coefficient.

Forest cover and bridge height were the second and third most important predictors of bat roost use in bridges, respectively. Urban developments, where forest cover is still present near bridges, are net positive for bats despite urban development in general having a negative influence on bat roost selection.

Guardrail material was not significant in final probability calculations, nor were the presence of mines or major water sources within 0.5 miles of a structure. These results were surprising as guardrail material was identified by reviewers and existing literature (Taylor 2006) as important for bat roost selection, as was the presence of water. Proximity to water may have been limited in this context as a consequence of the bridge type. Bridges that discourage bat roost use were primarily over water during our sampling. For example, small, low, wooden deck bridges are used almost exclusively to cross watercourses. These structures provide limited roosting potential. Consequently, the issue may be one of sample bias. This may also be an issue of thermal cover (Adam and Hayes 2000) as bridges over water may be subjected to higher convective heat loss. The low number of mines found within proximity to surveyed bridges may be a factor in its limited contributions to the model. Year the bridge was constructed was a significant factor, and this is likely due to the increase in gap or crevice size as a bridges age and erode which corresponded with past research (Gore and Studenroth 2005).

Model sensitivity performed well across the statewide GLM. Overall the model can be used to screen bridges for the likelihood of bat use, and this can help prioritize inventory and survey efforts. While the GLM should not be considered definitive in assigning potential bat use to a bridge, it helps in understanding variables driving roost selection and allows for a quantification of probability of use. The bats-in-bridges GLM achieved three major objectives identified by Copperhead, NCDOT, HDR, and CALYX/NV5: 1) quantify bridge features used by bats, 2) provide a probability scoring for bat use in un-surveyed bridges, and 3) incorporate these data into a field collection format that may advise a subsequent machine-learning model (not reported on here). As a result, the GLM will improve survey targeting and field methods for NCDOT crews assessing bridge repair, replacement, or routine maintenance. The GLM incorporated multiple bat species so that the dataset was large enough to develop a robust model, but individual species models were not tested.

As more bat surveys are completed on bridge or culvert structures, this model framework may be applied and refined to further our understanding of bat use of bridges. With additional data, GLMs could be completed for target species allowing a better understanding of species-specific effect. Ultimately, as the data becomes available, a similar model should be constructed for culverts as federally listed bat species are increasingly being found in these environments.

LITERATURE CITED

Adam, M.D., and J.P. Hayes. 2000. Use of bridges as night roosts by bats in the Oregon Coast range. *Journal of Mammalogy,* 81(2), 402-407.

Allison, P.D. 2012. *Logistic Regression Using SAS®: Theory and Application, Second Edition.* Cary, NC: SAS Institute Inc.

Allen, L.C., C.S. Richardson, G.F. McCracken, and T.H. Kunz. 2010. Birth size and postnatal growth in cave- and bridge-roosting Brazilian free-tailed bats. *Journal of Zoology, 280,* 8-16.

Barandela, R., R.M. Valdovinos, J.S. Sanchez, and F.J. Ferri. 2004. The imbalanced training sample problem: Under or over sampling? in *Proc. Joint IAPR Int. Workshops SSPR/SPR*, vol. 3138, *Lecture Notes in Computer Science*, 806–814

Bektas, B., Z. Hans, and B. Phares. 2018. Assessing Bridge Characteristics for Use and Importance as Roosting Habitats for Bats. Iowa State University, Final Report, pp. 40. Ames, IA. Available online at the following URL:

http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP20-07(397)_FR.pdf [Accessed November 22, 2019].

Chawla, N.V., L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic minority oversampling technique. *Journal of Artificial Intelligence Resources, 16,* 321–357, 2002.

Cleveland, A.G. and J.G. Jackson. 2013. Environmental Factors Influencing the Status and Management of Bats under Georgia (USA) Bridges. *Proceedings of the 2013 International Conference on Ecology and Transportation,* June 23-27, Scottsdale, AZ.

Endangered Species Act (ESA) of 1973, 16 U.S.C. 1531 note as amended through P.L. 113-287, enacted December 19, 2014.

Ferrara, F.J., and P.L. Leberg. 2005. Characteristics of positions selected by day-roosting bats under bridges in Louisiana. *Journal of Mammalogy, 86(4),* 729-735.

FHWA (Federal Highway Administration). 1995. Recording and coding guide for the structure inventory and appraisal of the nation's bridges. Report No. FHWA-PD-96-001, Washington, D.C. 20590.

Gore, J.A. and K.R. Studenroth, Jr. 2005. *Status and Management of Bats Roosting in Bridges in Florida. Transportation Research Board*, Report No. BD433, F214703-05-F, Tallahassee, FL.

Greene, W. 2007. Functional form and heterogeneity in models for count data. *Foundations and Trends* in *Econometrics, 1(2),* 113-218.

Guo, H. and H.L. Viktor. 2004. Learning from imbalanced datasets with boosting and data generation: The databoost-IM approach. *ACM SIGKDD Explorations Newsletter, 6(1),* 30-39.

Japkowicz, N. and S. Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6(5),* 429-449.

Keeley, B.W., and M.D. Tuttle. 1999. *Bats in American bridges*. Resource Publication 4. Bat Conservation International, Inc., Austin, TX.

Lance, R.F., B.T. Hardcastle, A. Talley, and P.L. Leberg. 2001. Day-roost selection be Rafinesque's big-eared bats (*Corynorhinus rafinesquii*) in Louisiana forests. *Journal of Mammalogy, 82(1),* 166-172.

Mease D., A.J. Wyner, and A. Buja. 2007. Boosted classification trees and class probability/quanitl estimation. *Journal of Machine Learning Resources, 8,* 409-439.

Menardi, G., and Torelli, N. 2014. Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, *28*(1), 92-122.

NCDOT. 2019. *Standard Operating Procedures (SOP): NCDOT Preliminary Bat Habitat Assessments (Structures, Caves & Mines)*. NCDOT Environmental Analysis Unit – Biological Surveys Group, Raleigh, NC.

Stepanian, P.M. and C.E. Wainwright. 2018. Ongoing changes in migration phenology and winter residency at Bracken Bat Cave. *Global Change Biology, 24,* 3266-3275.

Taylor, D.A.R. 2006. *Forest Management & Bats*. Bat Conservation International, online report: https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs143_009962.pdf

Yang, L., S. Jin, P. Danielson, C.G. Homer, L. Gass, S.M. Bender, A. Case, C. Costello, J.A. Dewitz, J.A. Fry, M. Funk, B.J. Granneman, G.C. Liknes, M.B. Rigge, and G. Xian. 2018. A new generation of the United States National Land Cover Database—Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing, 146,* 108–123, at https://doi.org/10.1016/j.isprsjprs.2018.09.006.

Appendix 1 – ArcGIS Work Flow to Reconcile Varied Spatial Dataframes

1. Coordinates provided by EAU NCDOT in the "Mergeddata" excel sheet were not being imported in ArcMap or QGIS properly. The following work around to get the points in the correct position was employed:
    a. Imported the points to Google Earth Pro
    b. Exported as KML file and Imported into ArcMap
    c. Joined exported points to "Mergeddata" excel sheet via UID (unique ID)
        i. This provided spatially correct points along with existing attribute data from the "MergedData" excel sheet. File named Bats.shp
2. Mines/Caves and Water analysis
    a. Buffered the Mine/Cave points to 1-mile diameter:
        i. If Bats.shp fell within the buffered zone of the Caves/Mines data, then they were given a Value of 1. Points that did not fall within the buffered zone were given a Value of 0
    b. NC_Wetlands.shp was used to see which Bats.shp points were over water.
        i. Used Select By Location tool to find which points were over waterbodies
            1. Note: The tool looked for points that were within 100ft to allow for inaccuracy of the NC_Wetlands line drawings. (added a 100ft buffer)
        ii. Bats.shp points that were over waterbodies were given a value of 1 and points that were not over waterbodies were given a value of 0
3. Reclassified Landcover Raster
    a. Raster used was nlcd_NC_utm17.tif
    b. Reclassify tool ran on the nlcd_NC_utm17.tif
        i. Water: Value 1
        ii. Developed: Land Value 2
        iii. Natural Land: Value 3
        iv. Agriculture Land: Value 4
4. Land Cover analysis using Tabulate Area tool in ArcMap
    a. Buffered the Bat.shp to .5 miles giving the buffered area a 1-mile diameter from the center point of the Bat point. Named BatsBuff.shp
    b. Ran Tabulate Area tool
        i. Input feature zone data: BatsBuff.shp
        ii. Zone field: UID
        iii. Input Raster: nlcd_Reclass
        iv. Class Field: Landcover
    c. Tabulate Area tool gave the Landcover area in Square Meters
5. The Tabulate Area tool did not initially process all 699 buffered areas. Step 4b was re-ran 5 times to get all 699 buffered areas processed

   a. The Tabulate Area produces an ArcMap Table. All 5 tables were converted to an Excel file then merged into one table called LandCoverPercentage_Working
6. Excel file was converted to an ArcMap table and joined to the existing Bats.shp using the UID field.
7. Exported the Joined Bats.shp to Bats_Final.shp.

Appendix 2 - Reviewer Comments to Reverse Initial Model Predictions

General Notes	Model Prediction	Reviewer Agency
Too low, close to water	Bats	NCDOT
Concrete cross brace over bent often has bats	No Bats	NCDOT
Bridge is over water, receives sunlight and is in okay habitat. Bridge deck, concrete end walls and expansion joints may provide roosting habitat, if guardrails do not.	No Bats	NGO
Habitat is poor but bridge has concrete guardrails which could provide roosting habitat.	No Bats	NGO
Concrete guardrails and end walls may provide roosting habitat. This bridge should at least score high enough to warrant a survey.	No Bats	NGO
This bridge should score high enough to be surveyed. Habitat is poor but bridge has concrete guardrails and end walls which could provide crevices for roosting. Looks like bat bridge on US 1 south of Raleigh.	No Bats	NGO
Habitat is poor but bridge should at least rank high enough for a survey. Bridge has concrete deck; there could be suitable crevices in the expansion joints and in the base of the guardrail.	No Bats	NGO
This bridge does not look very batty to me. The only decent habitat is the end walls and the expansion joints, to a certain extent. I would expect that the expansion joints are completely seal	Bats	NGO
Although the surrounding landscape is forested and the bridge is in sun and high over water, the bridge itself offers little in the way of bat habitat (metal beams, expansion joints are closed, guard rails are open, etc.)	Bats	NGO
I've seen big brown in bridges just like this one.	No Bats	NGO
Metal beams not ideal, but overall bridge looks batty.	No Bats	NGO
Bat roosting habitat in gaps in concrete guardrails, bridge over stream in relatively wooded area (good habitat)	No Bats	NCWRC
Patchy habitat; over road, but stream close by; concrete guardrails and other features could provide some habitat; not ideal, but wouldn't rule it out	No bats	USFWS
While not great surrounding habitat, bridge is concrete and could provide habitat; we have had bats using these types of bridges	No Bats	USFWS
Habitat looks okay for bats; bridge construction doesn't look great for bats, but could have batty features underneath (crevices, bird/wasp nests)	No Bats	USFWS
I26 bridge over service road not ideal, but in good habitat, close to French Broad River and wetland; also very close to known bat roost; limited roost potential visible, but could be features underneath where bats could roost	No Bats	USFWS

Appendix 3 – Descriptions for Model Input Parameters.

Model Variable	Model Code	Definition
Number of Main Spans	Mainspans	The number of main spans
Number of Approach Spans	Approachspans	The number of approach spans
Year Built	Year	The year the bridge was built
Number of lanes carried	lanes.on	Number of lanes being carried by the structure
Number of lanes crossed	lanes.under	Number of lanes being crossed over by the structure
Average Daily Traffic	ADT	Average daily traffic volume for the structure
Percentage of Trucks in traffic	trucks	Percentage of trucks comprised in the ADT
Substructure Grade	sub.grade	Gradient of the substructure
Bridge width excluding guardrails	Btwnrails	Distance measured across the bridge surface between guardrails
Bridge width including guardrails	Outtoout	Distance measured from the outer edge of one guardrail to the other
Beam height	beam.height	Depth of the beam, not the height above ground/water
Material of left rail	leftrail	Material for the guardrail
Material of right rail	rightrail	Material for the guardrail
Bridge length	length	The overall length of the bridge
Water within 0.5 miles	water_vici	Binary response to the presence of water within 0.5 miles of bridge
Road Angle	RD_ANGLE	The Azimuth of the road
Seal length	Seal	The combined length of all seals in a bridge
Seal count	Seal.count	The total count of seal segments in a bridge
Acres of forest within 0.5 miles	Forest	NLCD forested habitat within 0.5 miles. Includes deciduous, coniferous and mixed forests
Acres of developed land within 0.5 miles	Dev	NLCD developed land cover within 0.5 miles; all NLCD levels
Acres of agricultural land within 0.5 miles	Ag	Acreage of hay/pasture and cultivated crops within 0.5 mi
Acres of barren land within 0.5 miles	open	Acreage of barren, herbaceous, or scrub landcover within 0.5 mi
Acres of water habitats within 0.5 miles	wet	Acreage of open water, woody wetlands, and emergent vegetation
Bridge height	Brg.ht	Height of the bridge above road or water surfaces
Deck materials scored on a continuum	Mat_cont2	Continuous ranking of deck materials: concrete=3,steel=2,wood=1
Concrete Deck	Concrete_d	Binary coding of deck material
Steel Deck	Steel_d	Binary coding of deck material
Wood Deck	Wood_d	Binary coding of deck material
Concrete Beams	Concrete_b	Binary coding of deck material
Metal Beams	Metal_b	Binary coding of deck material
Wood Beams	Wood_b	Binary coding of deck material