Human–Wildlife Interactions 6(2):261–272, Fall 2012

A model to predict the likelihood of cliff swallow nesting on highway structures in northern California

Robert W. Coates, University of California, Department of Biological and Agricultural Engineering, One Shields Avenue, Davis, CA 95616, USA

Michael J. Delwiche, University of California , Department of Biological and Agricultural Engineering, One Shields Avenue, Davis, CA 95616, USA *mjdelwiche@ucdavis.edu*

W. Paul Gorenzel, University of California, Department of Wildlife, Fish, and Conservation Biology, One Shields Avenue, Davis, CA 95616, USA

Terrell P. Salmon, University of California, Department of Wildlife, Fish, and Conservation Biology, One Shields Avenue, Davis, CA 95616, USA

Abstract: Cliff swallows (*Petrochelidon pyrrhonota*) are colonially breeding migratory birds that frequently nest on highway structures. Under the Migratory Bird Treaty Act of 1918, people cannot harm swallows or their active nests. This restriction causes problems and delays for construction and maintenance divisions of many departments of transportation. In planning future projects, it would be useful for these divisions to have a habitat selection model that can predict the likelihood of cliff swallow nesting on a particular highway structure. We used logistic regression on data collected from 206 highway structures and 2 different land cover data sets to develop habitat selection models for northern California. The models indicated that low urban development and structure undersurfaces with multiple junctures were the 2 most important predictors of cliff swallow occupancy. Both the presence of water under a structure and a large underpass opening were also factors included in the models. The models correctly predicted 59% of sites occupied by cliff swallows and 88% of sites not occupied. The occupancy classification rate may offer departments of transportation useful insight into the nesting behavior of cliff swallows.

Key words: bridge, classification, cliff swallow, habitat selection model, highway structure, human–wildlife conflicts, logistic regression, nest, occupancy, *Petrochelidon pyrrhonota*, prediction, transportation

Cliff swallows (*Petrochelidon pyrrhonota*) are protected by the Migratory Bird Treaty Act of 1918. Under the act, completed nests cannot be disturbed during the breeding season, which is defined by the California Department of Fish and Game to be February 15 to September 1. Cliff swallows nest in colonies that often contain 200 to 400 nests (Brown and Brown 1995). They build gourd-shaped nests composed of mud pellets carried to the nest site from the surrounding area (Figure 1). The original nesting habitat of cliff swallows was rocky cliffs (Emlen 1954), but their range has expanded in North America over the last half century due to the availability of suitable habitat from bridges, culverts, and buildings, which serve as surrogates for cliffs.

Most existing literature concerned with control of cliff swallow nesting focused on preventing nesting on buildings (Gorenzel and Salmon 1982, Salmon and Gorenzel 2005), but did not discuss highway structures. Emlen (1954) wrote that the 3 main factors for a cliff swallow nesting site are (1) an open area for foraging, (2) a vertical surface with overhang for nest attachment, and (3) a mud supply suitable for nest construction. Brown and Brown (1995) reported that cliff swallows typically use a mud source within 0.5 km of the nesting location. Brown et al. (2002) indicated that colony selection is a complex behavior and that flowing and standing water and land-use diversity (Simpson's index) were correlated with colony size and repeated site use between years. Cultivated cropland was correlated with reduced colony size. Brown and Rannala (1995) suggested that cliff swallows may not simply choose a site based on local resources, but may also judge site fitness based on the size of a colony that has already begun to form.

Cliff swallows nesting on man-made structures create challenges for construction, maintenance, and repair, which cannot be performed during the breeding season. Departments of transportation frequently struggle with this impediment and are actively seeking solutions. There have been demolition projects where unsuccessful swallow prevention has caused project delays, bird mortality, and cost increases. We have previously reported on nonlethal methods used to exclude cliff swallows from nesting on highway structures (Conklin et al. 2009, Delwiche et al. 2010).

In planning future work, it would be useful for state departments of transportation and other government agencies to have a model that predicts the likelihood of cliff swallow nesting on a particular highway structure. Our objectives in this study were to (1) conduct a field survey of highway structures (bridges) to count cliff swallow nests and record structural and surrounding habitat characteristics, (2) develop a habitat selection model to predict the likelihood of cliff swallow occupancy, and (3) identify characteristics most likely to result in nesting.

Materials and methods

We randomly selected 300 highway structures from the California Department of Transportation (Caltrans) state bridge log (Caltrans 2011). Caltrans also provided us with additional structure information not listed in the log. A table of the data used for this work is available from Caltrans (Coates et al. 2009). Highway structures were limited to those within a 161-km radius of the University of California–Davis (UC–Davis) and with a length <152 m. The 161-km radius allowed multiple site analyses in single-day trips but at the same time provided geographical diversity (e.g., Coast Range, Sacramento Valley, San Joaquin Valley, Sierra Nevada foothills, and mountains). As part of this requirement, we selected only highway structures within Caltrans districts 1, 3, 4, and 10. Distance to each highway structure was determined by converting latitude and longitude from bridge log entries to Universal Transverse Mercator (UTM) coordinates and calculating the vector length to UTM coordinates for UC–Davis. Highway structures >152 m in length were considered too long to be surveyed without the use of boats or in a reasonable amount of time. We obtained encroachment permits from districts 1, 3, 4, and 10 for the selected highway structures. Cliff swallows and their nests were not disturbed during our visits.

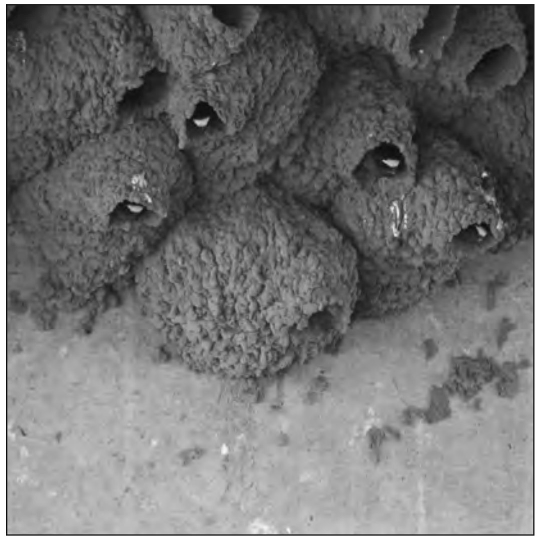

Figure 1. Cliff swallows nesting at the juncture of a highway structure abutment and deck.

Because the physical characteristics that we recorded at each site did not change much in the short term, the timing of the surveys was not restricted to the breeding season when birds were present. Between January and November 2007, we visited bridges to record physical characteristics of the structure, cliff swallow nesting evidence, and surrounding habitat. Several sites were not surveyed due to the time constraints of our daily trips or if they were considered to be unsafe or difficult to reach by car or foot. We ultimately surveyed 206 highway structures, which were well interspersed within the 81,000-km^2 region of study (Figure 2). Prior to site visits, we printed aerial photographs of each site (Google Earth, Google Inc., Mountain View, Calif.) that showed the surrounding habitat within a 4-km^2 area centered on each highway structure. During site visits, we annotated these maps with more detailed and current information on the habitat. The habitat classes we used were (1) fresh water, (2) salt water, (3) row crops, (4) orchards and vineyards, (5) trees and chaparral, (6) grass, fields, and bare ground, and (7) roads and buildings. The total land area of each habitat type was measured using a dot grid. A 900-cell transparent grid corresponding to 0.44 ha/cell was overlaid on each aerial photograph. The total land area covered was 393 ha. Each cell was assigned the habitat class that filled the greatest proportion of the cell. The number of

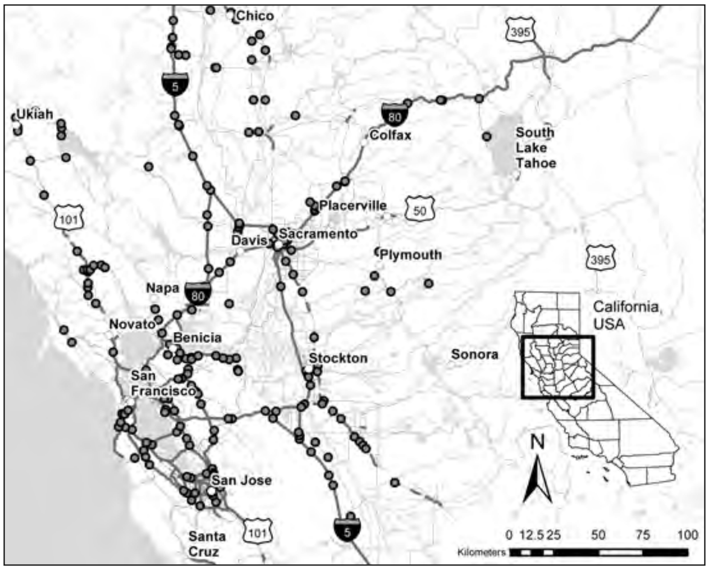

Figure 2. Map of 206 highway structure locations used in development of cliff swallow habitat selection model.

cells for each habitat class was divided by 900 to obtain a classification percentage. We also calculated Simpson's diversity index (Simpson 1949) for each site.

As a simpler alternative to our dot grid method of habitat classification, we obtained a separate set of data from the 2001 National Land Cover Database (NLCD) from the U.S. Geological Survey (Homer et al. 2007). The NLCD contained 16 noncoastal classes outside of Alaska, fifteen of which appeared in our data (the Perennial-ice-and-snow class was not present). Cover classes did not always have 1-to-1 correspondence with categories used in our dot grid classification. For example, the dot grid class Roads-and-buildings included 3 NLCD classes—Developed-high-intensity, Medium-intensity, and Low-intensity—and may have included Developed-open-space, as well. Conversely, the NLCD class Cultivated crops included 2 dot grid classes—Row crops and Orchards-and-vineyards. At the time of this work, we used the most current NLCD data available (2001), with a mapped resolution of approximately 0.4 ha. We obtained the total area of land for each classification within 0.5 km and 1.0 km (radii) of each highway structure, corresponding to 79 ha and 314 ha, respectively. NLCD data were provided by the GIS Lab at University of California Hopland Research and Extension Center. The advantage of using the NLCD is that classification data can be easily obtained for any highway structure.

In addition to habitat classification, we collected data related to the highway structure characteristics and cliff swallow nesting (Table 1). Some of these data could have been obtained from the Caltrans bridge log, but all data used in our analysis were recorded based on our field observations to ensure accuracy.

Table 1. Location, physical characteristics, and cliff swallow nest information recorded for highway structures.

Characteristic	Possible values (units) [details]
Latitude, longitude	(degrees)
Elevation	(m)
Material	concrete, concrete-steel, steel
Undersurface	steel I-beams, concrete girders with transverse diaphragms, concrete drop caps, none
Vertical support	concrete pile or column wall, steel column, abutments only
Deck-abutment angle	<90, 90, >90 (degrees)
Column-deck angle	<90, 90, >90, none (degrees)
Deck edge angle	≥180, <90, 90, 90–180 (degrees)
Road or water underneath	dry ground, waterway
Area of opening	(m^2)
Obstruction of opening	0–100 (percent) [4 quadrants, both openings of highway structure. Overall mean used for regression model.]
Nests	[Number of complete or partial nests]
Nest scars	[Number of scars on highway structure surface]

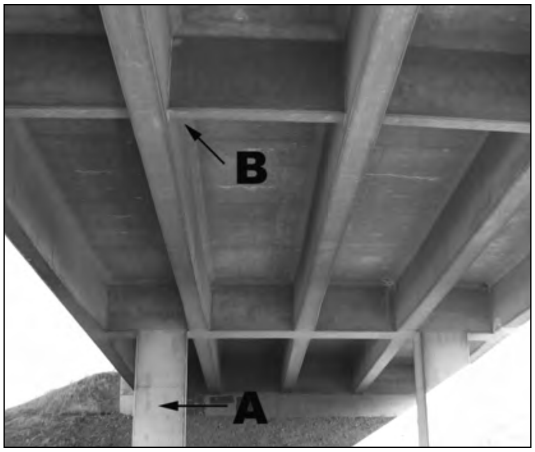

Figure 3. Highway structure showing (A) vertical support with concrete piles and (B) undersurface with concrete girders with transverse diaphragms.

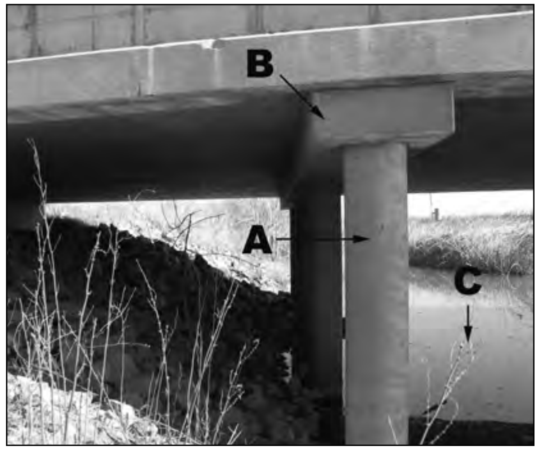

Figure 4. Highway structure showing (A) vertical support with concrete piles, (B) undersurface with concrete drop caps, and (C) water underneath.

Latitude, longitude, and elevation provided a broad geographic classification. Sites were also later grouped into 3 geographic regions: central valley, coastal range, and Sierra or Sierra foothills. The Material category was the primary material composing the highway structure surfaces; most commonly, this was concrete. The Undersurface category indicated the presence of steel I-beams, concrete drop caps, concrete girders with intermediate transverse diaphragms, or None (a smooth surface) on the underside of the deck. The Vertical-support category specified the presence of mid-span supports, including concrete piles, walls, steel columns, or the absence of mid-span supports (abutments only). Undersurface features and mid-span vertical supports could increase the possible nesting locations for cliff swallows. The deck-abutment angle was the angle at the juncture of the deck and abutment, classified as <90°, 90°, or >90°. The column-deck angle, the angle at the juncture of the deck and column, was likewise classified, but it included None as an option if mid-span supports were absent. The category Deck-edge-angle classified the angle on the outer edge (overhang) of the deck as <90°, 90°, 90° to 180, or ≥180° (no juncture). Angles of <180° indicated that the outer edges of the deck had an interior-angle overhang of likely interest to cliff swallows. The Road-or-water-underneath category specified the land feature over which the highway structure crossed, either dry ground (including roads and railroads) or a waterway. The Area-of-opening category was based on the opening height and width and provided a measure of the maximum flight-path area allowing access by cliff swallows to the underside of a highway structure. The Obstruction-of-opening category was visually estimated at 4 quadrants on each side of a highway structure to indicate how much of the openings were obstructed by trees, plants, or adjacent structures within 6 m. Values were between 0 and 100%, estimated to the nearest 5%, where a value of 0% indicated no obstructions within 6 m of an opening quadrant and 100% indicated complete obstruction. The overall obstruction was calculated as the mean of the 8 quadrant obstruction percentages. Finally, partial and completed cliff swallow nests and nest scars were counted, and each value was recorded. Nest scars were dark outlines caused by ectoparasite excrement on the structure surface that remains after a nest has fallen. Figures 3, 4, and 5 show a few examples of highway structures we visited.

The habitat and highway structure classifications were analyzed using logistic regression in Number Cruncher Statistical System (NCSS) software (Hintze 2007). Logistic regression is a multivariate technique appropriate for habitat use-nonuse studies employing random sampling and can be used to model the conditional probability of occupancy (Keating and Cherry 2004). A binary-dependent variable was used to indicate presence (1) or absence (0) of nests and scars for each site. Habitat and highway structure

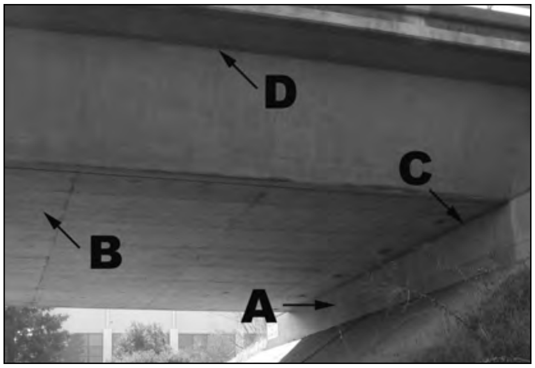

Figure 5. Highway structure showing (A) vertical support with abutments only, (B) undersurface with none, (C) 90° deck-abutment angle, and (D) 90–180º deck edge angle.

NCSS). Each nonreference, coded variable received a coefficient in the regression model, but this was used only when an observation had the corresponding value. For example, the Undersurface category had 4 possible values, and the Road-or-water-underneath category had 2 possible values. If the Undersurface category was classified as concrete drop caps, concrete girders with transverse diaphragms, or steel I-beams, the corresponding term in the regression model was assigned a 1 and 0, otherwise. Similarly, if the Road-or-water-underneath category classified as water, then the term was assigned a 1 and 0, otherwise. The value not appearing in the model (None for Undersurface and dry ground for Road-or-water-underneath) was the reference value for that categorical variable. Dummy coding is not considered appropriate when interactions with categoricals are present, so effect coding (called contrast with reference by NCSS) was used during the model selection process. The final models with dummy coding gave the same result as those with effect coding, but were easier to interpret.

Table 2. Numerical or categorical variables used in logistic regression, including dot grid data, National Land Cover Database (NLCD) 2001 data with 0.5-km radius, NLCD 2001 data with 1-km radius, and field survey data.

Dot grid	NLCD 0.5 km and 1 km	Field surveys
Numerical	Numerical	Numerical
Grass, fields, bare ground (ha)	Open water (ha)	Latitude (°)
Row crops (ha)	Developed–open space (ha)	Longitude (°)
Roads and buildings (ha)	Developed–low intensity (ha)	Elevation (m)
Fresh water (ha)	Developed–medium intensity (ha)	Year built
Orchards and vineyards (ha)	Developed–high intensity(ha)	Obstruction (%)
Trees and chaparral (ha)	Barren land (ha)	Area of opening (m^2)
Salt water (ha)	Deciduous forest (ha)	
Simpson’s diversity index	Evergreen forest (ha)	Categorical
	Mixed forest (ha)	Material
	Shrub, scrub (ha)	Undersurface
	Grassland, herbaceous (ha)	Vertical support
	Pasture, hay (ha)	Deck-abutment angle
	Cultivated crops (ha)	Column-deck angle
	Woody wetlands (ha)	Deck-edge angle
	Emergent herbaceous wetlands (ha)	Road or water underneath overpass, underpass region

classifications were entered into the model as independent numerical or categorical variables (Table 2). Numerical variables were continuous values, such as land area in hectares or bridge opening in square meters. Categorical variables with *n* possible values were converted by NCSS into n-coded variables, with 1 value used as a reference for the others. The final model used dummy coding (called binary by

Table 3. First 10 terms selected, in order of increasing log-likelihood, using hierarchical forward selection for dot grid data, along with number of free parameters, k, the Akaike information criterion, AIC, and pseudo-R^2, R_L^2, at each step.

Term	k	AIC	R_L^2
Intercept	1	263.31	0.000
RB	2	217.11	0.184
US	5	199.85	0.273
RW	6	196.71	0.293
AO	7	194.38	0.310
AO*AO	8	191.92	0.327
AO*US[1]	11	192.83	0.346
AO*RW[1]	12	189.82	0.365

[1]Term not included in final model.

Table 4. First 10 terms selected, in order of increasing log-likelihood, using hierarchical forward selection for 0.5 km NLCD data, along with number of free parameters, k, the Akaike information criterion, AIC, and pseudo-R^2, R_L^2, at each step.

Term	k	AIC	R_L^2
Intercept	1	263.31	0.000
DM	2	212.91	0.201
US	5	193.24	0.299
DO	6	190.72	0.316
AO	7	189.40	0.329
RW	8	187.07	0.345
H[1]	9	186.06	0.357
AO*AO[1]	10	184.74	0.370
AO*US[1]	13	186.10	0.387
AO*RW[1]	14	183.84	0.404

[1]Term not included in final model.

All variables were analyzed using a hierarchical, forward-selection logistic-regression algorithm. The algorithm used prior probabilities of 0.67 for unoccupied sites and 0.33 for occupied sites, which were the actual proportions of occupancy at the 206 sites. Prior probabilities affect the intercept of the regression model. In the forward selection procedure, the model began with no variables. The algorithm tested the model with each variable and interaction term, one at a time, to determine which one produced the largest log-likelihood value. Once found, this term was added permanently to the model. The procedure was repeated to add additional terms to the model equation until the relative change in the log-likelihood from 1 step to the next was less than 10^{-6}. The procedure was completed twice for each data set. The first model included no interaction terms and generated a reduced set of 9 variables for the second model that did include interactions. Because the model was hierarchical, interaction terms were added only if both individual variables were already in the model. Because NCSS used log-likelihood for variable selection, and not *P*-values, we also calculated the Akaike information criterion (AIC) for each step of the second model, using the equation

$$AIC = -2ln(L) + 2k \, , \quad (1)$$

where L was the log-likelihood and k was the number of free parameters in the model (Akaike 1974). Typically, the model with the lowest AIC is selected. A pseudo-R^2 value, r_L^2, was calculated by NCSS as a comparative measure of the log-likelihood accounted for by the model in each step (Hintze 2007). The significance of each term in the full model was checked using chi-square tests on the model deviance. Deviance is -2 times the difference between the log likelihoods of the model with all possible terms and the model with a selected subset. The chi-square test determines whether removal of a single term results in a significant increase in the deviance, compared to the full model. *P*-values <0.05 indicated that the term was significant. The model with the lowest AIC was first selected and the deviance *P*-values were checked. If one or more terms had a *P*-value >0.05, the term added in the latest selection step was removed from the model. The remaining terms were used in the full model. Inclusion of too many variables in the final model would risk fitting idiosyncrasies in the data instead of the general patterns likely responsible for the differences in site occupancy. We also determined the Pearson's correlation coefficient for each pair of variables included in the final model. A pair with a correlation coefficient >0.6 was assumed to provide redundant information, and removal of 1 term from the model was considered. We used receiver operating characteristic (ROC) analysis to explore different cutoffs for classification of each

Table 5. Regression model terms, X_n coefficients, b_n, and standard error, SE, of the coefficients for equation 3 for dot grid data.

n	X_n	b_n	SE
0	Intercept	-0.90641	0.51685
1	RB	-9.1945E-03	1.7564E-03
2	US = dropcaps	1.1281	0.56547
3	US = diaphragms	2.1788	0.52876
4	US = steel	-1.1398	0.74428
5	RW = water	1.3168	0.44188
6	AO	3.3405E-03	1.5287E-03
7	AO*AO	-1.0780E-06	8.5200E-07

Table 6. Regression model terms, X_n, coefficients, b_n, and standard error, SE, of the coefficients for equation 3 for 0.5 km NLCD data.

n	X_n	b_n	SE
0	Intercept	0.21856	0.53667
1	DM	-8.6817E-02	1.6219E-02
2	US = dropcaps	1.28685	0.58918
3	US = diaphragms	2.40487	0.57144
4	US = steel	-0.93089	0.73486
5	DO	-5.3121E-02	2.5647E-02
6	AO	1.0393E-03	4.5560E-04
7	RW = water	0.89334	0.42782

site, and we used area under the ROC curve (AUC) to compare different models (Metz 1978, Hanley and McNeil 1982). We used histograms to illustrate trends in site occupancy and to aid in discussion of the results.

To validate the full model, we randomly selected 90% of the data (185 sites) to create a validation model using the same parameters selected for the full model. The remaining 10% of the data (21 sites) were evaluated using the validation model to determine whether site occupancy was correctly predicted. The procedure was repeated 10 times, each time using a different set of randomly selected data. The idea was to evaluate model stability through consistency in occupancy prediction.

Results

All 3 data sets yielded similar regression models. The Roads-and-buildings category, Developed-medium-intensity (NLCD data) category, and Undersurface category (both) were always selected first and second, regardless of the software settings or other analysis techniques. The variables selected for the final model using dot grid data, in order of contribution to the final log-likelihood, were Roads-and-buildings (RB), Undersurface (US), Road-or-water-underneath (RW), and Area-of-opening (AO; Table 3). The variables selected for the final model using NLCD data within a 0.5-km radius were Developed-medium-intensity (DM), US, Developed-open-space (DO), AO, RW, and Herbaceous (H; Table 4). The variables for the final model using NLCD data within a 1-km radius were the same, except that RW was selected before AO, and H was not selected. The order of selection did not matter in this case because both variables were included in the model. For all data sets, interaction terms were selected in subsequent steps. For dot-grid and 0.5-km NLCD data, addition of AO*AO, AO*US, and AO*RW to the model yielded the smallest AIC. For 1-km NLCD data, addition of AO*AO and AO*RW gave the smallest AIC. However, the presence of some interaction terms resulted in deviance P-values much <0.05 for one or more terms in each model. Only after eliminating these interaction terms were the deviance P-values of all terms >0.05. No interaction terms remained in the NLCD models and only AO*AO remained in the dot-grid model. The AIC of each final model was not appreciably different from the smallest AIC, so rejection of the interaction terms was considered a reasonable simplification.

In logistic regression, the logit transformation is defined as

$$logit(P) = ln\left(\frac{P}{1-P}\right), \quad (2)$$

where $logit(P)$ is the logit of the proportion, P, of observations with a response of 1, meaning that nests are present. The regression models were of the form

$$logit(P) = b_0 + b_1X_1 + b_2X_2 \cdots b_nX_n\,, \quad (3)$$

where b_n was the regression coefficient for each term, X_n. The resulting occupancy predictions for each model were similar. Tables 5 and 6 show the coefficients and terms selected for the final models using dot grid and NLCD data with 0.5-km radius. Detailed results for the 1-km model are not shown due to its similarity to the 0.5-km model. The intercept term, b_0, was dependent on the terms included in the model, their coefficients, and the prior probabilities. Analysis of deviance for each term yielded no rejections from the 3 final models because the largest chi-square P-value among all terms was 0.037. Correlation tests between selected variables in each of the 3 models resulted in no correlation coefficients >0.6, except for a correlation between *AO* and *AO*AO* in the model with dot grid data. *AO*AO* was kept in the model because it provided more information about the effect of Area-of-opening and was not simply redundant. R_L^2 for the dot grid, 0.5-km NLCD, and 1-km NLCD models were 0.327, 0.345, and 0.325, respectively. AUC for the same models were 0.855, 0.866, and 0.856.

Table 7. Receiver operating characteristic (ROC) table for 0.5 km NLCD model for several possible classification cutoffs.

Cutoff	Sensitivity	Specificity	Sensitivity + specificity	Proportion correct
0.1	0.97059	0.45652	1.42711	0.62621
0.2	0.89706	0.63043	1.52749	0.71845
0.3	0.82353	0.75362	1.57715	0.7767
0.4	0.70588	0.81884	1.52472	0.78155
0.5	0.58824	0.87681	1.46505	0.78155
0.6	0.52941	0.93478	1.46419	0.80097
0.7	0.41176	0.97826	1.39003	0.79126
0.8	0.36765	0.98551	1.35315	0.78155
0.9	0.14706	1.00000	1.14706	0.71845

Table 8. Prediction of occupied and unoccupied sites for 21 sites not used in creation of each validation model based on 185 sites. Predicted values shown as a fraction of the actual number of occupied or unoccupied sites in the validation set.

Validation model	Predicted/actual number of sites			
	Dot grid data		0.5-km NLCD data	
	Unoccupied	Occupied	Unoccupied	Occupied
1	12/15	3/6	12/15	3/6
2	14/16	1/5	14/16	2/5
3	11/13	5/8	13/13	5/8
4	16/17	1/4	16/17	1/4
5	8/11	7/10	10/11	7/10
6	9/11	5/10	11/11	7/10
7	11/11	6/10	11/11	6/10
8	13/16	3/5	13/16	3/5
9	12/14	5/7	12/14	5/7
10	12/14	3/7	12/14	3/7

Logit(P) can be converted to the proportion of sites with nests by solving equation 2 for *P*,

$$P = \frac{e^{logit(P)}}{1+e^{logit(P)}}. \quad (4)$$

Equation 4 can be combined with equation 3 to predict the likelihood of finding nests at a particular site. *Logit(P)* <0 corresponds to $P < 0.5$, indicating that nests have <50% chance of being present, and *logit(P)* > 0 corresponds to $P > 0.5$, indicating nests have >50% chance of being present. With this default classification cutoff, the dot grid model and 0.5 km NLCD model both correctly predicted 59% (40 of 68) of occupied sites and 88% (121 of 138) of unoccupied sites, giving overall correct prediction rates of 78%. The 1-km NLCD model correctly predicted the same number of occupied sites, but 2 fewer unoccupied sites. ROC analysis yielded similar sensitivities (true positive rate) and specificities (true negative rate) for all data sets, with differences of <0.05, but the AUC for the 0.5 km NLCD model was slightly higher, so its ROC results are shown in Table 7.

Ten validation models for each data set were created from a random selection of 90% of the sites (185), and the predicted occupancies of the remaining 10% of the sites (21) for the dot grid and 0.5-km NLCD models are in Table 8. The proportion of sites correctly predicted as occupied by the validation models was between 0.20 (1/5) and 0.71 (5/7). The proportion of sites correctly predicted as unoccupied was between 0.73 (8/11) and 1.00 (11/11).

Discussion

Overall, each term in the model can be interpreted in a fairly intuitive way. The first variable selected for inclusion in all models was always related to urban development (Roads-and-buildings for dot grid data and Developed-medium-intensity for NLCD data). Each had a negative coefficient, indicating developed land reduced the likelihood of cliff swallow nesting. This seems plausible because development would be likely to reduce the habitat available for food, water, and mud, as well as increase deterrence by people, pets, and vehicular traffic. A variable added to the model with NLCD data was urban development space, which is also an Urban-development category, but included mostly vegetation such as lawn grasses (Homer et al. 2007) that we might consider a source of food, water, and mud. Additional analysis helped validate the inclusion of Developed-open-space in the model. A logistic regression model with only DM and US variables was compared to one with only DM, US, and DO. Visual inspection of scatter plots of *P* versus DO, with points color-coded for occupancy and non-occupancy, showed that addition of DO to the model improved classification of sites that were close to $P = 0.5$ in the DM and US only model. Addition of DO to the model caused sites with greater Developed-open-space land area to be more likely classified as unoccupied. Based on its inclusion in the model as a negative predictor of nesting, we can assume that the developed land features present in this category serve more to deter cliff swallows than to attract them.

Further analysis also helped explain the absence of Developed-low-intensity (DL) and Developed-high-intensity (DH) in the model. Because these land-cover classes were different levels of the same urban development features, we wanted to explore their importance to occupancy prediction. During forward selection of variables in the first model, DL and DH were not selected until steps 21 and 26, respectively, which indicates they were of little predictive importance. Even with DM removed from the data set, DL and DH were not selected until steps 18 and 20, indicating they were not suitable replacements for DM. Histograms were used to compare the number of occupied and unoccupied sites over the range of land area values for DH, DM, DL, and DO classes. A clear trend was that sites were more likely to be unoccupied if Development-medium-intensity was >26 ha. No other trends were apparent, which is the likely reason that only DM was selected for the model. Last, we combined DH, DM, DL, and DO into a single Total-development class by summing the land area of each. Similar to our other models, the Total-development variable was selected first, followed by Undersurface. This confirmed that urban land development was the most important variable for the cliff swallow occupancy model, though the DM development category was of primary importance when using the NLCD data for the region in this study.

The second variable included in all models was always Undersurface. This seemed reasonable because we expected that a highway structure undersurface with more interior junctures would provide better nesting surfaces than an undersurface with few or no junctures. We found that the percentage of occupied sites for each Undersurface classification was 22% (n = 137) for None, 52% (n = 25) for concrete drop caps, 67% (n = 30) for Concrete girders with transverse diaphragms, and 36% (n = 14) for steel I-beams. Because None was the reference category, the coefficients for the other three classification terms were relative to None. It is clear that sites with either Concrete-drop-caps or Concrete-girders with transverse diaphragms had a greater percentage of nests than sites with an Undersurface classification of None and this is reflected in the model. Steel I-beams had a negative coefficient and, thus, reduced the likelihood of observing nests. This seemed peculiar because Steel I-beam sites had a greater proportion of nests than the reference category, but the presence of other variables in the overall model likely altered the coefficient of Steel I-beams. For example, the classifications of US could differ in their average urban development, such that the difference between None and Steel I-beams is overridden by these other effects. This seemed to be supported by a comparison test that showed the mean values of Roads-and-buildings ($P < 0.001$) and Developed-medium-intensity (P = 0.001) differed between sites classified as having an undersurface of None and Steel I-beam. US coefficients do not suggest that concrete increased the likelihood of nesting compared to steel. Material was listed as a separate variable and was not found to be significant. Also, the US classification of None included structures made predominately of concrete, yet, it was the concrete structures with junctures on the undersurface (i.e., drop caps and girders with transverse diaphragms) that were more likely to be occupied.

The last 2 terms added to both models were Road-or-water-underneath and Area-of-opening. RW = water was added before AO for the model using dot grid data and vice versa for the 0.5-km NLCD data. Both had positive coefficients and indicated increased likelihood of cliff swallow nesting. Because the presence of a waterway eliminated vehicular traffic and likely increased food and mud availability, inclusion of this variable seemed logical. Similarly, a larger opening provided more open space for flight to and from the underside of the structure and also implied that the undersurface was higher from the ground, reducing the risk of predators.

The quadratic term, AO*AO was included in the dot grid model because it yielded a lower AIC (191.9 versus 194.4) and higher AUC (0.855 versus 0.850) than the model without the term. However, the difference between them was small enough that a model without AO*AO had a better overall classification rate, with 2 fewer false positives and one more false negative. AO*AO had a very small, negative coefficient, indicating that the response to AO was not linear. This means that AO had a lessening impact on nesting likelihood as the area of the opening became very large.

Statistical tests to compare AUC between models were not used, but it is reasonable to assume that the small differences in AUC were not indicative of 1 model have greater predictive ability then another. However, with the highest AUC and clear advantage of using publicly available land-classification data, we recommend the 0.5-km NLCD model for practical applications.

Without a predictive model, a department of transportation might guess as to whether or not a site is occupied by cliff swallows before visiting it. One approach, a coin toss, predicts a 50% likelihood of occupancy at any site. Applied to the data in this study, the true positive and true negative rates using this method were 17 and 33% (0.5 times prior probabilities), respectively. A second approach might presume that nests are less likely (33% likely, based on the observed proportion of sites occupied in this study) and this yielded true positive and true negative rates of 11 and 45% (i.e., the square of prior probabilities), respectively. In both cases, the regression models using a cutoff of P = 0.5 correctly predicted the likelihood of occupancy at a much higher rate than either of these methods based on chance.

ROC analysis can be used to choose a cutoff different from 0.5, depending on the desired balance between false positives and false negatives. A false positive (1-Specificity) is an unoccupied site that is predicted by the model to have nests. A false negative (1-Sensitivity) is an occupied site predicted to not have nests. A lower cutoff yields more false positives and fewer false negatives, while a higher cutoff yields the opposite. A department of transportation interested in sites with nests might select a lower cutoff and accept a higher false positive rate in order to reduce the likelihood of not addressing an occupied site. The potential cost is that many sites without nests may be classified as occupied, which could waste resources in examining these sites. To instead focus only on sites with a high likelihood of occupancy, a higher cutoff would be selected. In the absence of other preferences, optimum cutoffs could be chosen based on the maximum sum of sensitivity and specificity or the largest proportion of sites correctly classified. For the 0.5 km NLCD model, these cutoffs were 0.26 and 0.60, respectively.

One caution to using the models is that they should be considered valid only for the northern California region in which data were collected and, more specifically, within 161 km of Davis. Sites outside this region might introduce regional variations in cliff swallow nesting behavior or require other variables not included in this study. As an example of different cliff swallow behavior, nests are sometimes found on residential homes in southern California, but much less often in northern California. The effect this has on cliff swallow nesting on highway structures in southern California is unknown.

As with most animals, cliff swallow behavior is not easily predicted with a mathematical equation. Nonetheless, the models provided here can be used to predict the likelihood of cliff swallow nesting on hundreds of Caltrans highway structures and might provide insight into trends not yet evident in our data alone.

Management implications

Cliff swallows are a problem for state departments of transportation because they frequently colonize highway structures, and, according to federal law, their nests cannot be disturbed until the nesting season has passed. We used logistic regression to select significant terms from data that included bridge structural characteristics and surrounding habitat classifications from dot grid analysis and a land cover database. Based on our habitat selection model, the main factors increasing the likelihood of cliff swallow colonization are: (1) a lack of surrounding urban development, (2) an undersurface containing concrete drop caps or girders with transverse diaphragms, (3) the presence of water under the highway structure, and (4) a large opening. The models presented provide better predictions of cliff swallow nesting likelihood than estimates based on chance.

Acknowledgments

This research was partially supported by the California Department of Transportation. We thank N. Willits for guidance with the logistic regression analysis and S. Feirer for help with the NLCD data and map graphics.

Literature cited

Akaike, H. 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions 19:716–723.

Brown, C. R., and M. B. Brown. 1995. Cliff swallow (*Hirundo pyrrhonota*) *in* A. Poole and F. Gill, editors. The birds of North America, No. 149. Academy of Natural Sciences, Philadelphia, Pennsylvania, USA, and The American Ornithologists' Union, Washington, D.C., USA.

Brown, C. R., and B. Rannala. 1995. Colony choice in birds: models based on temporally invariant site quality. Behavioral Ecology and Sociobiology 36:221–228.

Brown, C. R., C. M. Sas, and M. B. Brown. 2002. Colony choice in cliff swallows: effects of heterogeneity in foraging habitat. Auk 119:446-460.

Caltrans. 2011. California log of bridges on state highways. California Department of Transportation, Sacramento, California, USA, <http://www.dot.ca.gov/hq/structur/strmaint/brlog2.htm>. Accessed April 2, 2012.

Coates, R. W., W. P. Gorenzel, M. J. Delwiche, and T. P. Salmon. 2009. Methods for excluding cliff swallows from nesting on highway structures. Report No. CA05-0926. California Department of Transportation, Sacramento, California, USA, <http://www.dot.ca.gov/research/researchreports/reports/2009/swallow_nesting.pdf>. Accessed April 2, 2012.

Conklin, J. S., M. J. Delwiche, W. P. Gorenzel, and R. W. Coates. 2009. Deterring cliff swallow nesting on highway structures using bioacoustics and surface modifications. Human–Wildlife Conflicts 3:93–102.

Delwiche, M. J., R. W. Coates, W. P. Gorenzel, and T. P. Salmon. 2010. Improved methods for deterring cliff swallow nesting on highway structures. Human–Wildlife Interactions 4:293–303.

Emlen, J. T., Jr. 1954. Territory, nest building, and pair formation in the cliff swallow. Auk 71:16–35.

Gorenzel, W. P., and T. P. Salmon. 1982. The cliff swallow—biology and control. Proceedings of the Vertebrate Pest Conference 10:179–185.

Hanley J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Hintze, J. 2007. Number Cruncher Statistical System. Number Cruncher Statistical System, LCC. Kaysville, Utah, USA.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham. 2007. Completion of the 2001 national land cover database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337–341.

Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat-selection studies. Journal of Wildlife Management 68:774–789.

Metz, C. E. 1978. Basic principles of ROC analysis. Seminars in Nuclear Medicine 8:283–298.

Salmon, T. P., and W. P. Gorenzel. 2005. Cliff swallows. University of California Agriculture and Natural Resources Publication 7482, Davis, California, USA, <http://www.ipm.ucdavis.edu/PMG/PESTNOTES/pn7482.html>. Accessed April 2, 2012.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

ROBERT W. COATES earned his B.S. and M.S. degrees from the University of California–Davis, with a major in biological systems engineering and minor in computer science. He currently works as an associate development engineer in the Department of Biological and Agricultural Engineering at the University of California–Davis. He has designed and implemented systems for wireless control and sensing in irrigation, GPS-guided soil fumigation, and cliff swallow nesting deterrence on highway structures.

MICHAEL J. DELWICHE is a professor in the Department of Biological and Agricultural Engineering at the University of California–Davis. He earned his undergraduate and graduate degrees at Cornell University and worked at Clemson University before joining the faculty at the University of California–Davis in 1986. His general area of research is bioinstrumentation and control, with current projects in biosensor development, sensor systems for precision field control, and biological control of plant pests.

W. PAUL GORENZEL received B.S. and M.S. degrees in wildlife biology, from the University of Massachusetts and Colorado State University, respectively. He worked at the University of California–Davis, both as a staff research associate for wildlife extension and the Department of Wildlife, Fish, and Conservation Biology since 1979. He recently retired. His professional interests include wildlife damage management and the development of a program for hazing birds at oil spills.

TERRELL P. SALMON earned his Ph.D. degree in ecology from the University of California–Davis. He has been a wildlife specialist for University of California Cooperative Extension for over 30 years, and has recently retired. His research focus is on human–wildlife conflicts, with emphasis on rodents and birds as they impact agricultural crops, structures and natural resources.