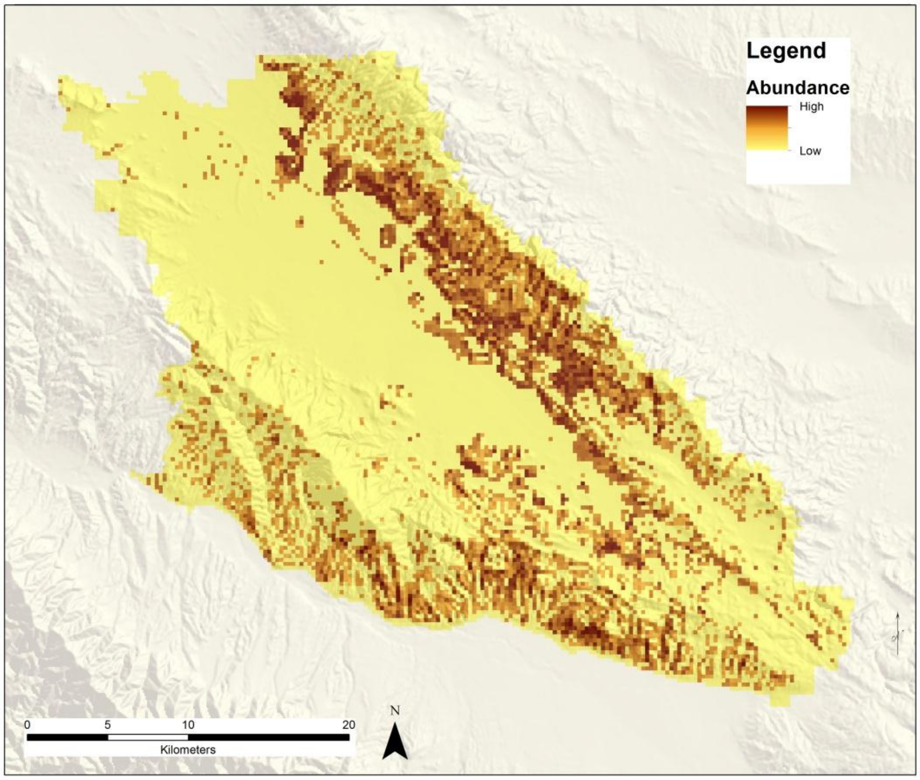
Le Conte's Thrasher Abundance Modeling

2014 Summary Report to the Bureau of Land Management



Dennis Jongsomjit, Leo Salas, Jim Tietz, Geoffrey R. Geupel

Point Blue Conservation Science

3820 Cypress Drive # 11

Petaluma, CA 94970

www.pointblue.org

Contents

Introduction

Le Conte's Thrasher (*Toxostoma lecontei;* LCTH) is an uncommon desert species associated with hot and dry climates throughout the southwestern United States. Its range is contiguous throughout the Mojave and Sonoran desert portions of Arizona, Nevada, Utah, and California except for a disjunct population found in California's San Joaquin Valley. This population, whose range partially overlaps with the Carrizo Plain National Monument (CPNM), has been described as a sub-species (*T. l. macmillanorum*), but this distinction has not been widely recognized (Sheppard 1996). This population has, however, been recognized as a California Bird Species of Special Concern (Shuford and Gardali 2008) due to greatly reduced range and population size, high endemism, and habitat loss and degradation (Fitton 2008).

This project was initiated in 2010. In 2010-2011, monitoring methods were tested and refined and 120 area search survey plots (250m x 250m, approximating the size of a LCTH territory) were established within the CPNM (Jongsomjit et al. 2013). We continued to survey these plots in 2012 and 2013.

Habitat characteristics can be useful for various spatially explicit management applications. In the case of the Carrizo Plain, we used a habitat suitability model to help guide the placement of new survey sites (Jongsomjit et al. 2012). This model was based on presence-only information and a GIS layer provided by the BLM delineating vegetation types at a relatively large scale. However, our surveys and previous research indicates that Le Conte's thrashers respond to vegetation types and variables at a small scale, including the percent cover of shrubs and the availability of open ground. Accordingly, in our report we recommended that improved vegetation maps be produced that included this type of information. Recently, the California Native Plant Society (CNPS) completed a comprehensive vegetation survey of the CPNM. The result of this work was a fine scale vegetation GIS layer that includes information on tree, shrub, and herb cover, vegetation alliance (56 types), as well as heterogeneity and disturbance measures (Buck-Diaz and Evens 2011). While some potentially important vegetation characteristics were not measured by this effort (such as shrub height), this layer is a considerable improvement over the previous layer.

The development of a geospatial population model, showing species distribution across the study area and potential abundance, may be possible using the improved vegetation layer developed by the CNPS. A geospatial abundance model would use Le Conte's thrashers count and detection data, providing a more informed map of greater conservation value of the potential distribution of LCTH within the CPNM and beyond. This type of data can be used to prioritize or identify management, conservation, or restoration actions. For example, species population targets can be calculated at different scales and used to guide, prioritize, and evaluate management actions.

In this second phase of our study of Le Conte's Thrasher, we analyzed all three years of data using imperfect detection models of occupancy and abundance (the "site-level models"), and then used these results to train a geospatial population model (the "landscape-level model") of presence and abundance for the species in the CPNM. We present the results of both these analyses, identify areas for improvement of our understanding of the species locally and at landscape levels, and provide recommendations for management action.

Site-Level Models

Our goal was to create a site-level abundance model that could then be used to inform a landscape-level abundance model across the CPNM. In order to create the site-level abundance model we first had to take into account that some individuals would go undetected during our surveys due to low detection probabilities (imperfect detection) which in turn could lead to biased models of abundance.

The site-level models are the imperfect detection models. We fit a model to estimate the occupancy of Le Conte's Thrasher in our data, and a model to estimate the abundance. Both include correction for imperfect detection, which is possible to estimate if there are repeated surveys at the same location. For example, in repeated visits to a site LCTH detection may be 0 in some visits and 1 in others. If we assume the population is closed (i.e., no immigrations or emigrations, births or deaths), we must assume the site is clearly occupied and the survey events when a bird was not recorded must be solely due to imperfect detection. That is, the variation in detection events throughout the repeated visits to the site informs a probability of detection. The same logic applies to the abundance models, except that now probability of detection is inferred from the variation in the total number of birds detected throughout the visits. Notably, a site with 0 detections may be occupied, because birds there may be difficult to detect. Further, a site that has a detection of one individual may indeed host more than one individual, such that the count of one bird is the result of the true abundance with imperfect detection.

The imperfect detection models thus include two functions. One function fits the best estimate of the probability of detection, and the other fits the abundance or occupancy. These are fit simultaneously and each has its own set of covariates, which we discuss below. The statistical procedure seeks to find the set of covariate coefficients such that the predicted abundance/occupancy, corrected for imperfect detection, fits the observed data as best as possible. Because the functions are fit simultaneously, there may be several competing top models, all fitting the data approximately equally.

Models were fit using package "unmarked" (Fiske et al. 2014) in the programming language "R" (R Core Team 2014).

Site-Level Covariates

Area searches were conducted within 250m x 250 plots, an approximation of a LCTH territory size. During each area search, surveyors conducted vegetation releves recording various measurements characterizing the plot. Some of these measurements were chosen *a priori* as variables hypothesized to be important to the occurrence of LCTH and were used to develop the site-level imperfect detection and occupancy models. Variables included linear and quadratic forms of: percent *Atriplex polycarpa* cover (common saltbush), percent *Ephedra californica* cover, percent bare ground, percent grass cover, and percent total shrub cover. These were considered for both the detection and the abundance/occupancy functions at the site level.

Site-Level Imperfect Detection and Abundance Model

We evaluated all possible combinations of the covariates, resulting in evaluation of 1,048,576 models. We considered all models within 2 Akaike Information Criterion units of the top model and reviewed these to determine the one that made the most biological sense. For example, the top abundance model included common saltbush for detection (linear and quadratic), and only amount of bare ground for the abundance or occupancy function. However, we know that the presence of shrubs is important for the species. Competing models included *Ephedra californica* cover, or just the percent total shrubs, so we opted to include the effect of total shrub cover, as it increase the likelihood of the model fit, even if not enough to outweigh the (Akaike) penalty for including an additional variable in the model.

Regarding the abundance model, we also realized that abundance was the most important predictor of occupancy. So, we opted to take advantage of this relationship and conservatively estimate abundance using the approach suggested by Royle and Nichols (2003).

Site-Level Results

Our top model estimated counts of 0.09 to as many as 5.54 birds per plot, though the median (see Figure 1) is 0.72 and the mean is 1.07. Coefficient estimates are shown in Table 1. Note how percent shrub cover does not significantly add to the explanation of the variance in the data. The difference in the log-likelihood of the model with and without that variable was 0.2 (-177.4 vs. -177.2), which resulted in a difference in AIC value of 1.6, thus making it a top competing model (Burnham and Anderson 2002).

Table 1. Top model coefficient estimates.

Parameter	Estimate	SE	z	P(>|z|)
(Intercept abundance)	-1.790	0.740	-2.417	0.016
% bare grounds	0.040	0.009	4.495	0.000
% shrub cover	-0.009	0.013	-0.657	0.511
(Intercept detection)	-3.128	0.452	-6.919	0.000
% cover Saltbush	0.192	0.049	3.916	0.000
% cover Saltbush^2	-0.005	0.002	-3.165	0.002

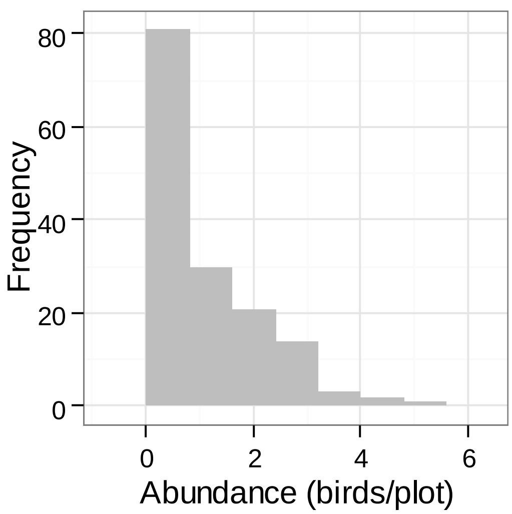


Figure 1. Frequency of abundance of LCTH per plot, as estimated by the site-level model.

We also evaluated the goodness of fit of the model. Unlike conventional generalized linear models, hierarchical mixed-effect models such as ours cannot be readily evaluated through metrics pertaining to amount of variance explained. This is because abundance and detection coefficients are co-dependent on each other. But it is possible to use a procedure that samples from the posterior distribution of the model coefficients and uses the observed data to predict values and calculate a metric (e.g., the squared sum of error, SSE), and compare this metric against the same metric estimated directly from the data. A good model fit should result with most predicted values for the metric surrounding the observed value. Figure 2 shows the results for our choice of top model, evidencing an excellent model fit.

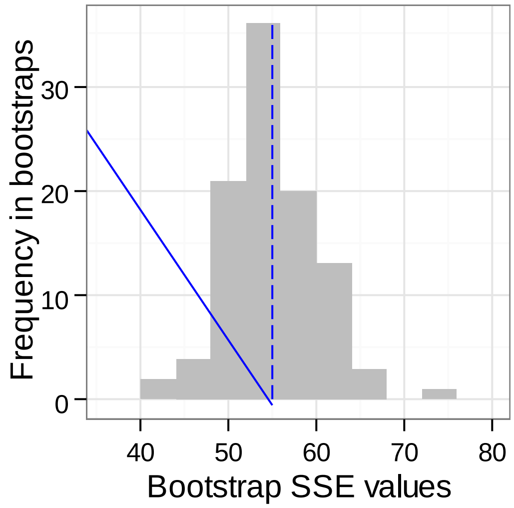


Figure 2. Test of goodness of fit for the site-level imperfect detection model. The blue dotted line indicates the observed sum of squared errors in the dataset; the gray bars reflect the frequency of SSE values in 100 bootstrap simulations.

Figure 3 shows how abundance (A & B) and probability of detection (C) relate to the predictor covariates. The curves span the value of the covariates in the dataset. LCTH is more abundant in areas with high percentage of bare grounds, yet not entirely denuded of vegetation. The species seems to like areas clear of brush but only to some extent. It likes low but not entirely missing, shrub cover. Note that our data does not include samples with values of 90% or higher bare ground cover, or > 70% shrub cover. These limits are of great consequence for the landscape-level model, because there will be areas in the landscape whose % bare ground cover and % shrub cover exceed these limits in the data, and the model will likely over-predict the abundance of LCTH in these locations.

The relationship of detectability with % saltbush cover is depicted in Figure 3 C. Note that detection seems to increase as % saltbush cover approximates 30% and remains high thereafter. However, though saltbush helps detect the species, at high levels of % cover the species becomes less and less abundant.

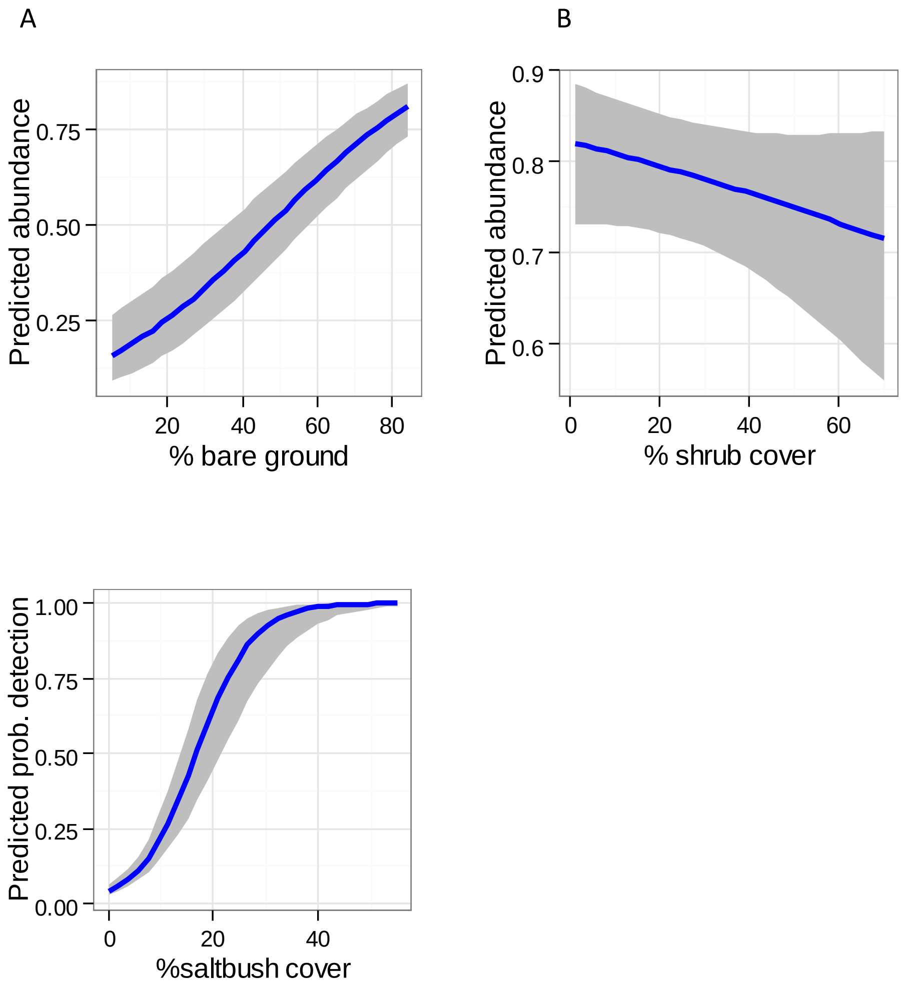


Figure 3. Partial dependence plots of abundance (A & B) and probability of detection (C) covariates for the top model.

Landscape Level Models

Landscape-Level Covariates

We developed several landscape variables that we considered could help determine LCTH distribution and abundance within the study area (Table 2). These included both geophysical and vegetation based variables, with the vegetation variables based off of the CNPS Vegetation Map layer (Stout et al. 2013) (Figure 4).

The CNPS layer was developed using a combination of extensive ground surveys and aerial images to produce a fine-scale vegetation map which included tree, shrub and herb cover classes, and the identification of over 50 vegetation types. We selected potentially important vegetation types by identifying which types occurred within our survey plots. We then developed landscape level metrics with this reduced set of vegetation types using the programs ArcMap v10.1 (ESRI 2012) and Fragstats v4.2 (McGarical et al. 2012). Because vegetation cover classes were defined by a set range of values, we first converted the ranges to equal the median value of the range. For example, a homogeneously-defined range of 20-30% cover was converted to equal 25% cover.

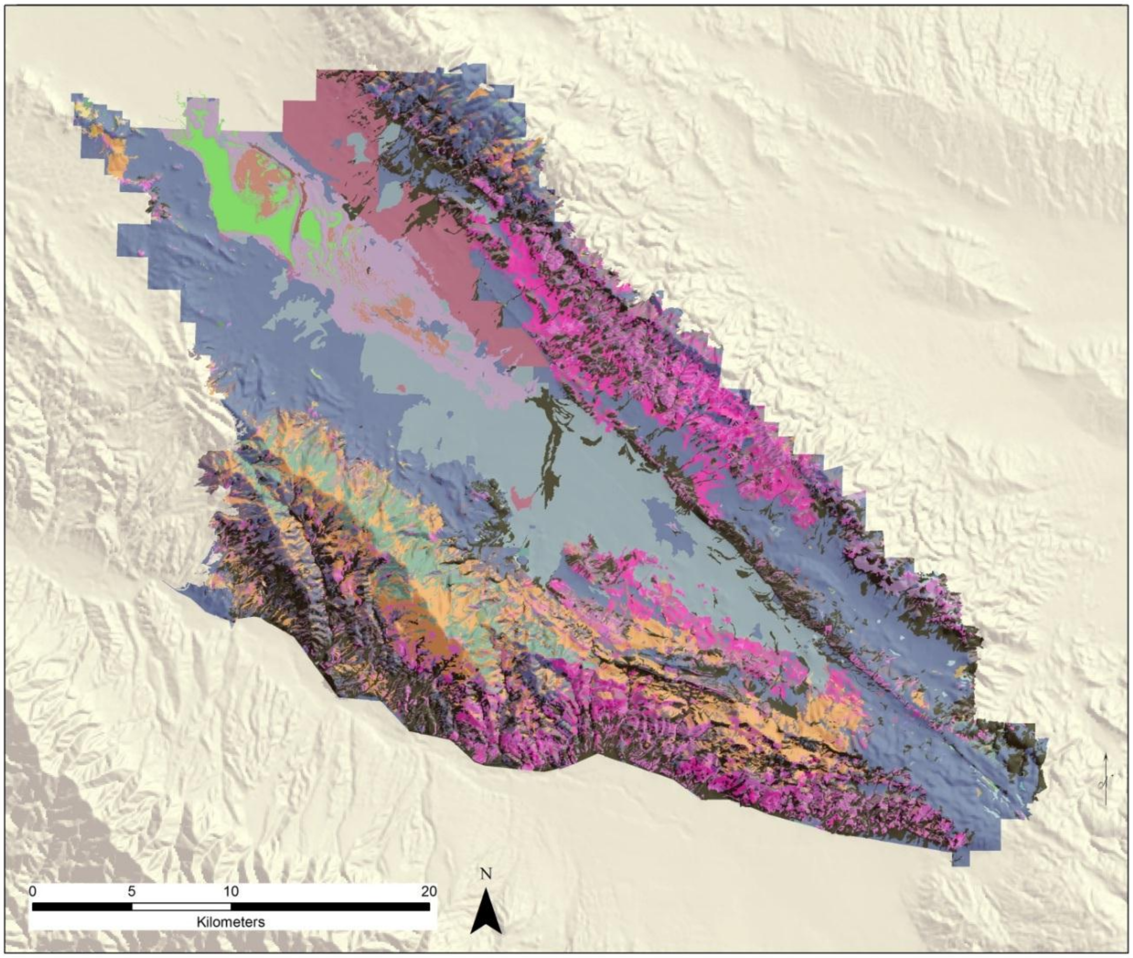


Figure 4. Map of vegetation types identified by the California Native Plant Society within the Carrizo Plain National Monument boundary.

Mean values of each covariate were defined by taking the average value of all raster cells within a given area search plot. We also ran a moving window analysis that returned the proportion of each vegetation type within 375 m of a given raster cell. These results were then summarized by taking the average value within a given area search plot.

Table 2. Variables considered in the landscape boosted regression tree models. Variables were summarized within each 250m x 250m cell across the landscape

Variable	Definition	Source
Percent Atriplex polycarpa	Mean proportion of vegetation type within a 375m moving window.	CNPS Vegetation Layer
Percent Ephedra californica	Mean proportion of vegetation type within a 375m moving window.	CNPS Vegetation Layer
Percent California Annual and Perennial Grassland Macrogroup	Mean proportion of vegetation type within a 375m moving window.	CNPS Vegetation Layer
Percent Lasthenia californica - Plantago erecta - Vulpia microstachys	Mean proportion of vegetation type within a 375m moving window.	CNPS Vegetation Layer
Percent Ericameria linearifolia - Isomeris arborea	Mean proportion of vegetation type within a 375m moving window.	CNPS Vegetation Layer
Aspect	Mean aspect in degrees	USGS National Elevation Dataset
Slope	Mean slope in degrees	USGS National Elevation Dataset
Stream distance	Total sum of the distance to nearest stream or creek	National Hydrography Dataset
Mean shrub cover	Mean shrub cover	CNPS Vegetation Layer
Mean shrub cover landscape	Mean shrub cover within a 375m moving window	
Mean herb cover	Mean herb cover within a 375m moving window	CNPS Vegetation Layer

Landscape-Level Abundance Models

Using the values from the site level imperfect detection and occupancy model described above, we first fit boosted regression tree (BRT; Elith et al. 2008) models using the landscape covariates as potential predictors to estimate species abundance for each 250m cell in the landscape. In basic terms, BRTs work by fitting an ensemble of models (trees) in succession with each successive model built for the variation in the response that is not yet explained by all previous fits. The final model combines the results of all the trees. BRTs have been shown to have better predictive performance than other statistical model algorithms (Caruana and Niculescu 2006, Elith and Graham 2009, Hastie et al. 2009) and they can incorporate non-linear responses and interactions between covariates.

We applied a threshold value to the resulting BRT model to define where a LCTH could be present (above the threshold value) or absent (below the threshold value). This threshold value could be decided in several ways and can be guided by user objectives (Liu et al. 2005, Freemen and Moisen 2008). For example, a threshold can be based on minimizing errors that identify a species to be present when it is actually absent. Or a threshold can be based on trying to minimize both commission (predict presence when in reality it is not present) and omission errors (predict not present when it is). Since we were interested in identifying suitable habitat and potentially setting population targets within the Monument we decided to base our threshold on a target population density of 2 pairs per hectare. This is the minimum historic density identified in the Birds of North America species account (Sheppard 1996). Thus, our algorithm could be described as follows: sort all cells in descending order of LCTH abundance, add one cell at the time, starting with the highest count cell, until the sum of counts divided by the study area results in 2 pairs/hectare. We opted to use the area of the monument (1,050 km2) to estimate the density: we determined the count of birds per cell below which the species was not included as present so that the sum of counts in cells above this value results in the reported density. However, the density estimate in Sheppard (1996) is without description to the study area to which it applies.

We tested several combinations of BRT models with different learning rates of 0.01, 0.005, and 0.001 (determines the contribution of each tree added to the model) and different tree complexities of 1 to 5 (determines the interaction order in the response) with the aim of minimizing the predictive deviance while still achieving at least 1000 trees (Elith et al. 2008). The final model was built with a learning rate of 0.001 and a tree complexity of 2 using a Gaussian link function and was evaluated with a 10-fold cross-validation

Landscape-Level Results

Model predictive performance results are shown in Table 2. We projected our model results onto environmental variable grid surfaces to produce a map across the Monument (Fig 5). However, we caution that the point-level model evaluation indicates poor understanding of areas where the species is not present, because of the biased selection of survey locations. This causes the landscape-level to over-predict. We therefore strongly suggest that the results reported here should be seen as an index of relative abundance rather than actual abundance. In general, LCTH were predicted to occur within the distribution map polygons developed in the Bird Species of Special Concern (BSSC) account with some differences. For example, our model predicted LCTH to occur north and southwest of the largest primary distribution polygon, areas where we have indeed detected LCTH. The model also predicted LCTH to occur northeast of the primary polygon along the steeper slopes of the Temblor range. This is an area that has not been well surveyed (see discussion). Our model predicts thrashers will occur along the southeastern edge of the Caliente range. A smaller secondary distribution polygon covers part of this area but we have not surveyed it or the areas surrounding it. However, personal communication with the BSSC account author confirms that LCTH have been detected by him in this area before. Lastly, our model predicts LCTH to occur on the western slopes of the Caliente range. Part of this area is covered by a historic range polygon, but most of it is not identified by the BSSC range map.

Table 3. Model predictive performance cross-validation statistics. Values are calculated at each fold (10x) then averaged across all 10 folds to produce a mean and standard error value.

Mean deviance (Std. Error)	0.698 (0.05)
Mean correlation (Std. Error)	0.437 (0.05)

Table 3 shows the results of 10-fold cross-validation of the landscape-level model. These are intended to show high heterogeneity in the data (mean correlation), and yet relatively high deviance (estimated with the sample left out for testing, averaged across all bootstraps). Results suggest that the model may require better training by collecting more estimates from a wider range of habitat conditions within the CPNM. This result also evidences the lack of proper training in areas where the LCTH is not found, and this limitation may lead to over-predicting LCTH abundance in the landscape.

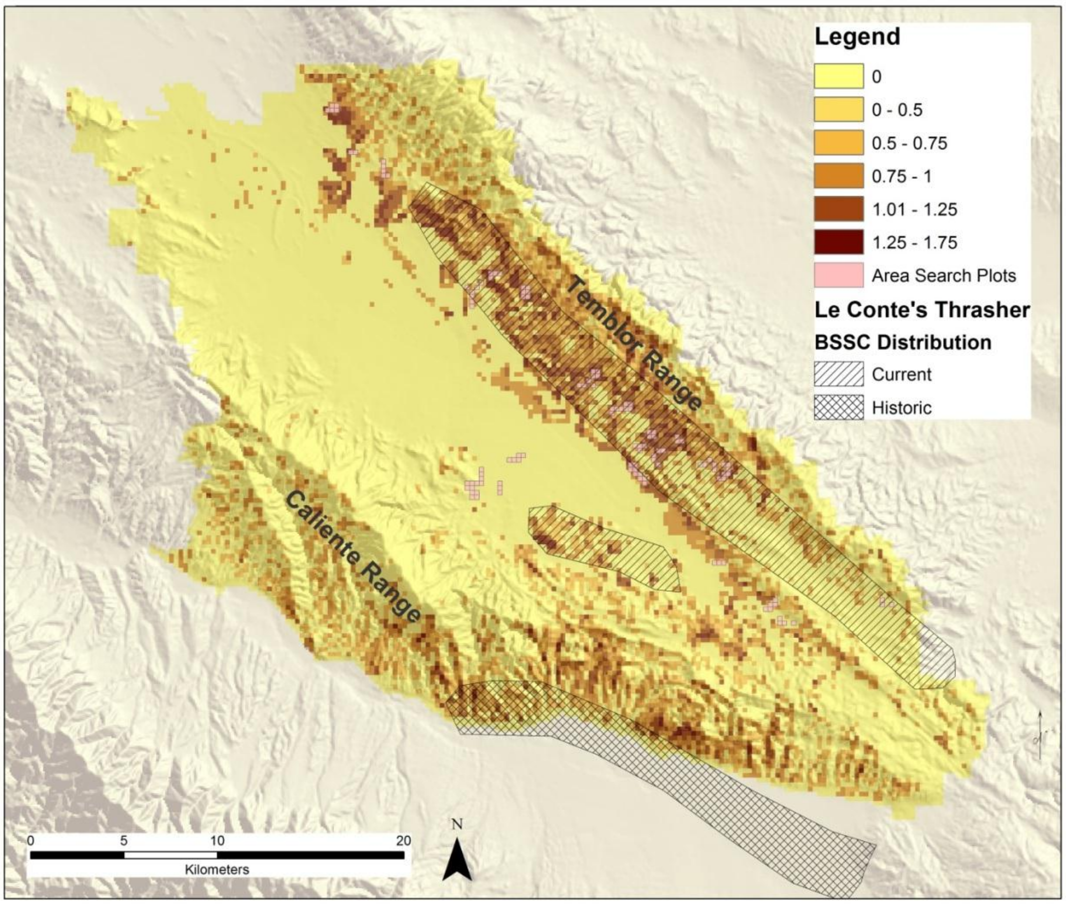
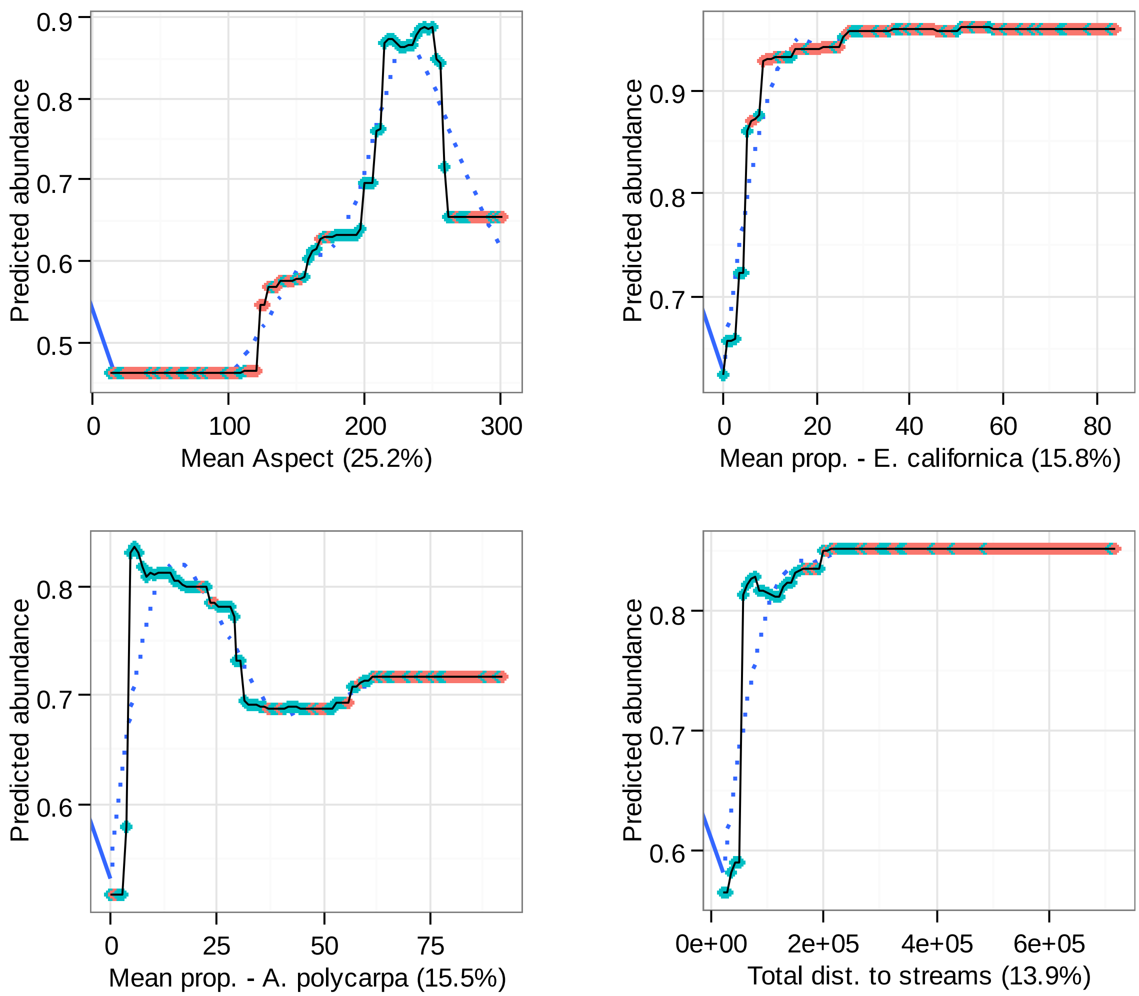


Figure 5. The landscape level abundance model for Le Conte's Thrasher projected onto the mapped covariate grids. Darker areas represent higher projected abundance values. Yellow areas indicate a predicted abundance of zero after application of the density threshold.

In our landscape abundance model, aspect was the most influential covariate (Figure 6). This was followed by the mean proportion of *Ephedra californica* and *Atriplex polycarpa*. Total stream distance and mean proportion of annual/perennial grassland were the next two most influential variables, each accounting for over 10% relative contribution to the model. In general, abundance was positively correlated with aspect values above approximately 150 degrees (west facing slopes) then began to decline again after about 240 degrees. Abundance was positively correlated with increasing proportion of Ephedra. Abundance was also positively correlated with lower proportions of *Atriplex californica* and began to decline above approximately 15% cover.

Figure 6. Fitted function graphs of the top eight variables used the the landscape level abundance models for Le Conte's Thrasher. Numeric values in parenthesis represent the relative importance of the given variable for this model (out of a total of 100%). Dotted lines represent the smoothed version of the fitted function. Blue dots represent values in the training set; orange dots represent values for which the model was not trained.



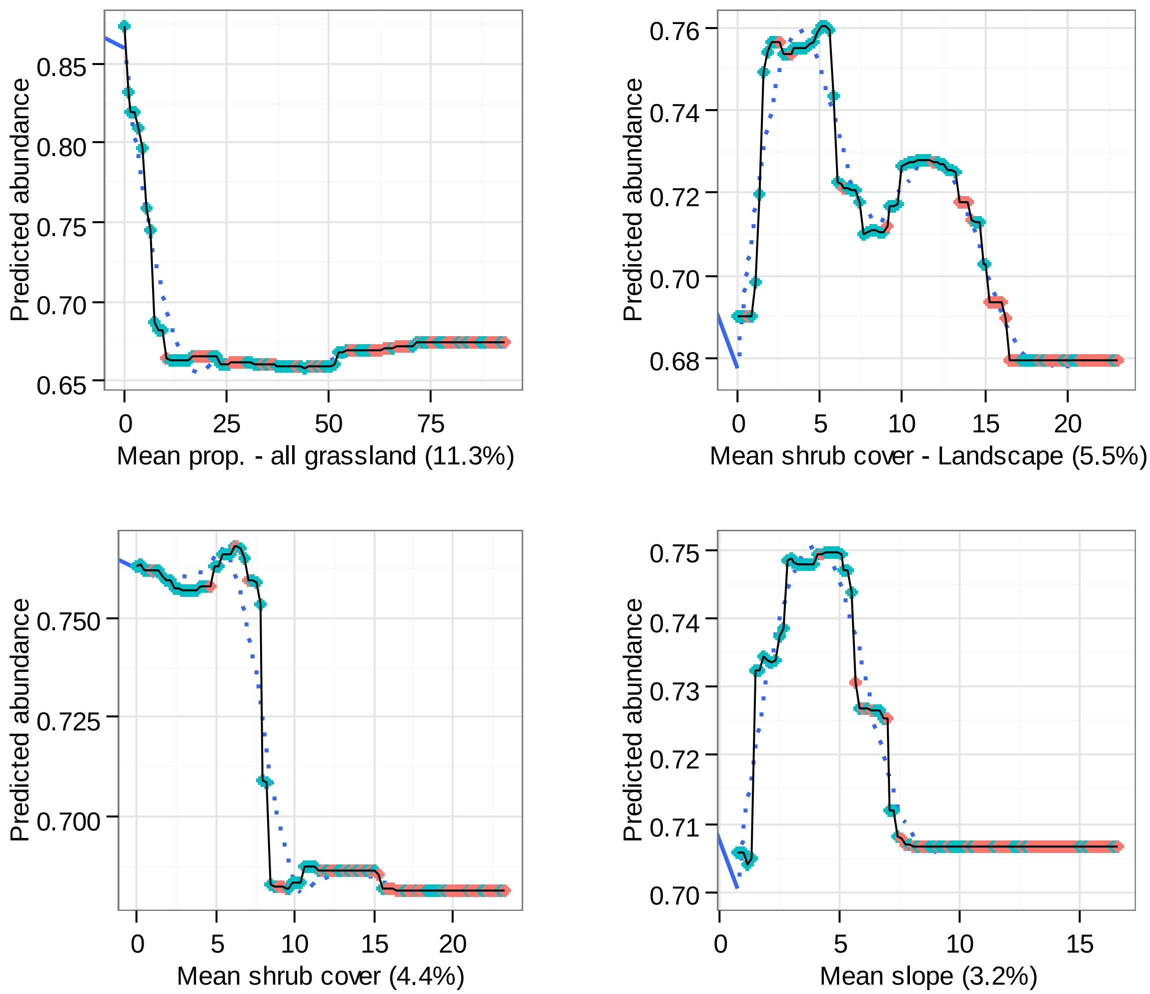


Figure 6 also shows that the observational data were collected in such manner that they sampled well the environmental conditions of the CPNM. Exceptions may include high values of total distance to streams, or mean slope. These results further suggest that over-prediction may be solely due to incomplete characterization of the conditions where the species is not present.

Discussion

Our landscape level abundance models provide an improved map of the potential distribution of Le Conte's Thrashers compared to the habitat suitability model produced in 2010.

The habitat suitability model used presence locations only and was informed by a coarse vegetation layer of habitat types. In contrast, the abundance model uses LCTH count data, corrected for imperfect detection, and is informed by a richer vegetation data layer (including shrub and herb cover) at a relatively high spatial resolution. However, there are several caveats that should be taken into account when interpreting the model results. First, our area search surveys have always targeted potential LCTH habitat and the model therefore lacks information on where LCTH are not in the landscape (e.g. unsuitable habitat). Therefore, the model may over-predict onto areas where thrashers may not persist. On the other hand, our model does not take into account factors besides suitable habitat that may be impacting the LCTH population. For example, historic and continuing habitat fragmentation may lead to artificially reduced numbers due to isolation of sub-populations within the San Joaquin valley (Fitton 2008). Thus, while our model may correctly be identifying suitable habitat, LCTH may be restricted by other factors unaccounted for in our models.

Because our plots did not cover entire territories and because they were focused only on potentially suitable habitat, the partial dependency plots do not reflect the actual needs of an individual or pair. Rather, they reflect the best fit of the data to the covariates. Thus, they can provide useful information on the relationship between LCTH abundance and the environmental covariates but should not be used to provide strict guidelines for habitat management.

To improve our picture of LCTH distributions and abundance within the Monument we would recommend surveying plots outside of what is identified as suitable habitat. This would confirm that LCTH do not actually occur here and would provide our models with a more accurate picture of what may be defining suitable habitat. Areas to focus on may include more western sections of the Carrizo plain and the eastern slopes of the Caliente range, although the lack of roads and access to these areas may restrict our ability to survey them. The steeper upper slopes of the Temblor range are identified as potentially suitable habitat with some higher index of abundance values. We conducted a limited amount of surveys in this area in 2010 and did not find any LCTH. Given the model results, there is ample justification to return to these areas and confirm that LCTH are present or absent.

Recommendations

- Target priority areas for conservation based on high density areas identified by models (improvement over presence-only based model)
- Manage for LCTH-friendly habitat features using general guidance as identified in this project:
    - % ground cover
    - % shrub cover in general
    - % Atriplex polycarpa and Ephedra californica cover
- Manage areas with lower density projections to match areas with higher density projections
- Use model results to conserve areas outside the CPNM for the San Joaquin population
    - Habitat features
    - Target areas for conservation and connectivity

Potential next steps

- Validate the model: Survey high density areas not previously included in our sampling
- Improve area search surveys to estimate trends and set management targets based on trends
- Survey fully randomized plots with respect to relevant density and detection covariates
- Survey fully randomized plots with respect to relevant landscape-level covariates
- Evaluate models at different plot sizes (e.g. LCTH territory sizes)

Acknowledgements

This project was funded by the BLM through a NLCS Science Grant and the members and donors of Point Blue Conservation Science. We are grateful for the volunteers who came out to survey with us over the years: Matt Brady, Ryan DiGaudio, Michelle Gilbert, Oliver James, Adam Searcy, Brent Campos, Tom Edell, Geoff Geupel, Nora Livingston, Alex Metea, Kristin Sesser, Kathy Sharum, Maggie Smith, Megan Elrod, Mark Dettling, and Khara Strum. K. Sharum also helped initiate this project and provided valuable logistical support. We thank ESRI for providing ArcGIS software to us through their conservation grant program. We also thank Sherie Michaile and Doug Moody for providing assistance with data entry and management.

Literature Cited

Buck-Diaz, J., and Evens, J. 2011. Carrizo Plain National Monument vegetation classification and mapping project. http://www.cnps.org/cnps/vegetation/pdf/carrizo-vegetation_rpt2011.pdf

Burnham, K. P. and Anderson, D.R. 2002. Model Selection and multimodel inference: a practical information-theoretic approach. New York, Springer-Verlag.

Caruana, R., and Niculescu-Mizil, A. 2006. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168). ACM.

Elith, J., Leathwick, J.R., Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology. 77:802–13

Elith, J., Graham, C. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography. 32:66-77.

Fiske, I., Chandler, R., Miller, D., Royle, A., and Kery, M. 2014. Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software. 43:1-23. http://www.jstatsoft.org/v43/i10/

Fitton, S. D. 2008. Le Conte's Thrasher (*Toxostoma lecontei*) (San Joaquin population). Pages 322-326 *in* W. D. Shuford and T. Gardali, editors. California Bird Species of Special Concern: A ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California. Studies of Western Birds 1. Western Field Ornithologists, Camarillo, CA and California Department of Fish and Game, Sacramento.

Freeman, E. A., and Moisen, G.G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modelling 217:48–58

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The elements of statistical learning* (Vol. 2, No. 1). New York: Springer.

Jongsomjit, D., Tietz, J.R., Michaile, S., Fonseca, T., and Geupel, G.R. 2012. Le Conte's thrasher monitoring in the Carrizo Plain National Monument. Report to the Bureau of Land Management. PRBO contribution #1886.

Jongsomjit, D., Tietz, J.R., and Geupel, G.R. 2013. Carrizo Plain National Monument Monitorning. Report to the Bureau of Land Management.

Legendre, P., and Legendre, L. 1998. Numerical Ecology, 2nd Ed. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands.

Liu, C., Berry, P.M., Dawson, T.P., and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 3:385–393.

McGarigal, K., Cushman, S.A., and Ene, E. 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org

Royle, J.A. and Nichols, J.D. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777-790.

Salas, L., Fitzgibbon, M., Fonseca, T., Herzog, M., Ballard, G., Moody, D., and Nur, N. 2010. The Analyst: Data analysis tool for the California Avian Data Center. Petaluma, California. http://data.prbo.org/apps/analysts/

Sheppard, J.M. 1996. Le Conte's Thrasher (*Toxostoma lecontei*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved 7/15/2013 from the Birds of North America. http://bna.birds.cornell.edu.bnaproxy.birds.cornell.edu/bna/species/230 doi:10.2173/bna.230

Southwood, T. R. E., and Henderson, P.A. 2000. Ecological Methods, 3rd Ed. Blackwell Science Ltd. Oxford, England.

Unitt, P. 2008. Grasshopper Sparrow (Ammodramus savannarum). Pages 393-399 *in* W. D. Shuford and Gardali, T. editors. California Bird Species of Special Concern: A ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California. Studies of Western Birds 1. Western Field Ornithologists, Camarillo, CA and California Department of Fish and Game, Sacramento.