**Biodiversity and Biocultural Conservation Laboratory**
**University of Texas at Austin**

**Technical Note** No. 56

**Date** 10 February 2009

# Habitat Suitability for the Black-capped Vireo in Comal and Hays County, Texas: Final Report to Smith|Robertson

Trevon Fuller, Tiffany Hollon, and Sahotra Sarkar

We evaluated the suitability of habitat for the Black-capped Vireo (BCVI), *Vireo atricapilla* (Woodhouse, 1852), in Comal and Hays Counties of Texas. Suitable, good, and high-quality habitat for the species are reported on the basis of predictions from an ecological niche model. This report supersedes Fuller *et al*. (2008b) which was a preliminary report based on less data. The additional data used in this report came largely from southern Travis County, especially from Shield Ranch, a known location of BCVI nesting sites. These data are particularly important because southern Travis County is adjacent to Hays County and the machine-learning techniques used in this analysis are sensitive to the geographical location of data points.

The main new results are the following: (i) in Hays County, particularly in the north, the analysis now predicts a much larger area of high and good quality habitat than before; and (ii) there is very little such habitat in Comal County. However, the second result should be interpreted with caution because there is little data on BCVI presence records for Comal County (we did not have access to a single reliable occurrence record). This could reflect a genuine absence of nesting individuals in Comal County or it could simply reflect inadequate monitoring and data collection. More data from Comal County are necessary before there can be much confidence about the quality of habitat there. We suggest that the collection of such data and adequate monitoring be part of any Habitat Conservation Plan (HCP) for Comal County.

The analysis is based on an ecological niche model (ENM) constructed using a maximum entropy algorithm implemented in the Maxent 3.2.19 software package (Phillips and Dudík 2008). This method has been shown to be one of the most reliable techniques for predicting species' distributions from presence-only (occurrence) data (Elith *et al*. 2006). As input, these models take occurrence data and environmental layers. The output consists of the fundamental niche of a species which must then be modified to obtain its realized niche and projected onto geographical space to obtain its predicted distribution or suitable habitat map. In January 2009 Maxent was used to predict wintering habitat for the BCVI in México and independently collected data were used to test the predictions of these models (S. Sarkar and T. Hollon, unpublished results). All presence points for the BCVI in the independent data set were found to occur in areas identified as high-quality habitat by the niche models. Given that winter habitat of the BCVI is believed to be even more difficult to predict than nesting habitat, this result

boosts confidence in this technique.

## 1. **Study Region**.

The analysis was performed for a region that consisted of a rectangle on a sphere (bounded by longitude and latitude lines) so as to include Comal, Hays, and Travis Counties and the part of the Balcones Canyonlands National Wildlife Refuge that falls in Burnet and Williamson Counties (see Figure 1). This extended study region was used for the analysis because of the paucity of records from Comal and Hayes Counties alone while a large number of records were available from Travis County and the Balcones Canyonlands National Wildlife Refuge. The study region was bounded by 97°21´ and 98°39´ of longitude and 29°35´ and 30°42´ of latitude. The analysis was performed at a 30 m $\times$ 30 m resolution which corresponds to the resolution of Landsat images used for some of the environmental layers. There were 1 304 304 data points or "sites."

## 2. **Ecological Niche Modeling Methodology**.

Maxent uses a machine-learning algorithm to identify the fundamental niche of a species, that is the areas in which the environmental conditions are suitable for the species in question. This is treated as a pattern recognition problem similar to standard statistical procedures to find correlations but better able to tease out signals from noise in the data. In an ENM two types of input data are needed: (i) occurrence data for the species; and (ii) environmental layers (for instance, layers incorporating climatic, topographical, and land use land cover [LULC] data) which are treated as the explanatory variables. The critical advantage of the maximum entropy method used in Maxent (and a suite of other machine-learning methods for constructing ENMs) is that the occurrence data may be presence-only (using all reports of the occurrence of a species) rather than presence-absence (in which case a data-deficient site must be identified as such through a systematic survey). The data used for this analysis are described in the next two sections.

Results are internally evaluated by Maxent. A subset of the data are not used to construct the model but to test its accuracy. The user can specify how much of the data should be retained for testing. If a large number of data points are available, this fraction can be high. The test data are then chosen randomly so as not to introduce bias in the analysis. Model performance is judged using several criteria including the AUC or area under the receiver operating characteristic (ROC) curve, which plots the true positive rate versus the false positive rate using the test data. If the predictions of the model are no better than random, then the AUC would be equal to 0.5. If the predictions are perfect the AUC would be 1. There are circumstances in which the AUC can be misleading;

consequently, we supplement the use of the AUC by a set of 11 binomial tests that are performed by Maxent during the process of ENM construction.

After an ENM is used to identify potential habitat in a region (by projecting the environmental explanatory variables back to geographical space), further refinement is necessary to identify the realized niche (or actual distribution) of a species. This must take into account contingent factors (such as lack of suitable connecting habitat) that prevents a species from occupying all habitat that is ecologically suitable for it. This stage requires biological expertise. Section 7 provides detail of how this was done in this analysis (see Table 3).

3. **Occurrence Data**.

Occurrence data were collected from a variety of sources (see Table 1). Maxent requires data in the form of co-ordinates of presence for records of species' occurrence. When data were only available as accurate polygons for the home range of an individual, these were converted into point data by calculating centroids of the polygons in ArcGIS. Only data from 1990 or later were used to construct the ENM for the BCVI. This choice reflects the fact that it is only after the time at which the BCVI was listed as endangered under the Endangered Species Act (ESA)—which was 1987—that its habitat was protected. Earlier records may be from habitats that have been altered to become unsuitable for the species after the time at which the record was collected. A total of 955 points were used in the model. Figure 1 shows the location of data points.

4. **Environmental Layers**.

Topographic layers were obtained from public sources (see Table 2). A preliminary study of Travis County and the Balcones Canyonlands National Wildlife Refuge showed the importance of May temperature for determining habitat suitability for the BCVI (T. Fuller, unpublished results). Consequently, meteorological layers were constructed for the entire study area through interpolation of the raw data provided by weather stations. For a list of the weather stations used for the interpolations, see Fuller *et al*. (2008a).

5. **Model Parameters**.

Model settings in Maxent followed the recently published recommendations of Phillips and Dudík (2008) who designed the software package. The settings were: 60/40 training/test split, linear, quadratic, product, threshold, hinge, and categorical features, regularization multiplier: 1, and logistic output. Models were generated using 10 000 background points (as pseudo-absences). The regularization value was 0.05 for the linear, quadratic, and product features, 0.25

for the categorial features, 1 for the threshold features, and 0.5 for the hinge features.

6. **ENM for the Fundamental Niche**.

Figure 2 shows the fundamental niche of the species as projected into geographical space. The performance of such models is principally judged by the AUC which was explained in Section 2. The AUC for the model was 0.99, which is highly satisfactory. In other words, there were almost no false positives. However, the AUC should not be used as the only criterion for judging preformance. We also used all internal tests in Maxent. All 11 binomial tests on the model returned significant values (Table 3). The *p*-values are uniformly indistinguiable from 0.

The most important variables for determining habitat suitability for the BCVI were elevation and April precipitation. The former resulted in the highest AUC when it was the only variable included in the Maxent model. The latter resulted in the lowest AUC when omitted from the model.

7. **Model Refinement**.

Niche models predict the fundamental niches of species. To obtain the realized niche, these fundamental niches must be modified to take into account factors such as habitat modfication.

The logistic output from Maxent reports a suitability value (to be interpreted as the probability that the site is habitat) for each site in the study area. In the present analysis, sites predicted by Maxent to be BCVI habitat with probability 0.1 or greater are referred to as “suitable” habitat. Figure 3 shows the suitable habitat in the study region.

Because typical BCVI habitat is comprised of early successional shrub vegetation, the suitable sites were restricted to exclude sites without shrub/scrub vegetation as defined by the National Land-Cover Database (Homer *et al*., 2004) in Hays County (Figure 4). However, because there was so little predicted good or high-quality habitat in Comal County, this criterion was not used there.

Figures 5(a) and 5(b) show detailed habitat maps for Comal County. In both figures, good habitat was interpreted to be restricted to the top 50 % of the suitable sites; high-quality habitat was restricted to the top 10 % of the suitable sites.

Figures 6(a) and 6(b) depict the same BCVI habitat maps for Hays County.

Table 4 summarizes the Comal and Hays habitat models. Each row lists the acres of high and good quality BCVI habitat predicted to occur in the county according to the model, restricted to shrub/scrub habitat in the case of Hays County. Perhaps the most important result is that there is more than 8500 acres of high-quality BCVI habitat in Hays County.

## 8. **Discussion**.

In general, because of the methodology used, predictions of high or good quality habitats are likely to be reliable; however, predictions of the absence of such habitat should be treated with more caution. The high value for the AUC indicates that the model makes very few false positive predictions. The implication is that predicted high-quality habitats should be treated as priority areas for conservation after further testing of the models through field work. Note that these tests should record whether sites have the type of habitat associated with BCVI nesting (which, in these counties, consists of mid-successional deciduous shrubland with sparse to moderate canopy). These models do not directly predict occupancy. Evidence against the model consists of finding occupied nesting sites that are predicted to have low suitability, not high or good quality sites with no occupancy.

Because of issues related to reliable data availability, the results for Hays County are much more reliable than those for Comal County. There was no reliable BCVI presence record available for Comal County. There were three clusters of such data points for Hays County; perhaps even more importantly, Hays is adjacent to data-rich Travis County. If presence records become available for Comal County, it is almost certain that the predicted configuration of habitats there will change considerably. Luckily, Maxent is computationally efficient enough for the analysis to be repeated with new data, if necessary, in a matter of hours. Results for Hays County, particularly northern Hays County, are unlikely to be very sensitive to moderate quantities of new data. We plan to produce new models for both these counties every year after updating our data set with whatever new data that is collected, either systematically or opportunistically, every breeding season. A similar exercise is currently being carried out for Travis County and the Balcones Canyonlands National Wildlife Refuge.

The predicted high-quality habitat regions of Hays County are of obvious conservation concern. We suggest that surveys of these areas during the 2009 breeding season be prioritized. The other good quality regions of Hays County also deserve some attention. Carrying out such surveys during the breeding season will allow the determination of both the availability of high or good quality habitat as well as occupancy. While occupancy only tests the models insofar as there is occupied habitat that is predicted as being marginal, information on occupancy is always highly pertinent for conservation planning. Consequently, survey efforts are best concentrated during the breeding season.

In the case of Comal County, the analysis predicted very little high-quality habitat and little good quality habitat. It found better habitat south of Comal in Bexar County. However, this result could simply be a result of the availability of some reliable data on the presence of BCVI individuals in Bexar County but no such data for Comal County. We suggest that systematic surveys for BCVI be carried out in Comal County during the 2009 breeding season. Moreover, any HCP for Comal County will best serve the interest of BCVI conservation if it includes surveys and monitoring as a regular component.

## References


Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberon, J., Williams, S. E., Wisz, M. S., and Zimmermann, N. E. 2006. "Novel Methods Improve Prediction of Species' Distributions from Occurrence Data." *Ecography* **29**: 129 -151.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D. 2007. "The Shuttle Radar Topography Mission." *Review of Geophysics* **45**: RG2004 DOI:10.1029/2005RG000183.

Fuller, T., Hollon, T., and Sarkar, S. 2008a. "Preliminary Report to Smith|Robertson: Habitat Suitability for the Golden-cheeked Warbler in Comal and Hays County, Texas." Technical Note No. 49. Biodiversity and Biocultural Conservation Laboratory. University of Texas at Austin. Austin, Texas.

Fuller, T., Hollon, T., and Sarkar, S. 2008b. "Preliminary Report to Smith|Robertson: Habitat Suitability for the Black-capped Vireo in Comal and Hays County, Texas." Technical Note No. 51. Biodiversity and Biocultural Conservation Laboratory. University of Texas at Austin. Austin, Texas.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M. 2004. "Development of a 2001 National Land-cover Database for the United States." *Photogrammetric Engineering and Remote Sensing* **70**: 829 -840.

Phillips, S. J. and Dudík, M. 2008. "Modeling of Species Distribution with Maxent: New Extensions and a Comprehensive Evaluation." *Ecography* **31**: 161 -175.

Woodhouse, S.W. 1852. "Descriptions of New Species of Birds of the Genera

Vireo, Vieill., and Zonotrichia, Swains." *Proceedings of the Academy of Natural Sciences of Philadelphia* **6**:60-61.

Yang, L., Huang, C., Homer, C. G., Wylie, B. K., Coan, M. J. 2003. "An Approach for Mapping Large-area Impervious Surfaces: Synergistic Use of Landsat-7 ETM+ and High Spatial Resolution Imagery." *Canadian Journal of Remote Sensing* **29**: 230 -240.

Table 1

**Point Data Sources**

| Source | Number of Points |
|---|---|
| Giri Athrey,<br>University of Louisiana Lafayette | 9 |
| Bob Gottfried,<br>Texas Natural Diversity Database,<br>Texas Parks and Wildlife Department | 33 |
| Paul Fushille,<br>Travis County ,<br>Department of Transportation and Natural Resources | 105 |
| Jayne Neal,<br>City of San Antonio,<br>Parks and Recreation Natural Areas | 12 |
| Chuck Sexton,<br>US Fish and Wildlife Service,<br>Balcones Canyonlands National Wildlife Refuge | 247 |
| Billy Simper,<br>Texas State University, San Marcos | 219 |
| Jenny Wilson,<br>US Fish and Wildlife Service,<br>Austin Ecological Services Office | 208 |
| Laura Zebehazy,<br>Travis County,<br>Department of Transportation and Natural Resources | 120 |

Table 2

**Layers Used to Construct Niche Models**

| Layer | Source |
|---|---|
| Aspect | Derived from elevation using the Spatial Analyst Extension of ArcGIS (DEM: Farr *et al*. [2007]; web-site: seamless.usgs.gov) |
| Slope | Derived from elevation using Spatial Analyst (DEM: Farr *et al*. [2007]; web-site: seamless.usgs.gov). |
| Elevation | DEM: Farr *et al*. (2007); web-site: seamless.usgs.gov. |
| Canopy | Homer *et al.* (2004), web-site: seamless.usgs.gov. |
| Distance to cropland: measured as the distance in meters from the current pixel to the closest pixel classified as cultivated crops. | Derived from the National Land-cover Database of Homer et al. (2001), web-site:seamless.usgs.gov. |
| Distance to impervious surface: measured as the distance in meters from the current pixel to the closest pixel classified as an impervious surface in the National Land-cover Database. In this context, an impervious surface is an urban area. | Yang *et al*. (2003); web-site: seamless.usgs.gov. |
| Distance to forest interior: distance in meters from the current pixel to the closest pixel in the interior of a forest, defined as a site 75 meters or more from the edge of a deciduous, evergreen, or mixed forest patch. | The methodology is based on DeBoer and Diamond (2006). The data were derived from the National Land-cover Database of Homer *et al*. (2001) (web-site: seamless.usgs.gov) using the Zonal Statistics Tool in ArcToolbox. |
| Site type: a categorical variable based on topography and exposure (examples: “high flats”, “protected | The methodology for the construction of the raster is in Diamond and True (2007). The data are from D. Diamond |

| | |
|---|---|
| slopes", "exposed slopes", "low flats and floodplains"). | and C. D. True (personal communication). |
| April temperature | The data were interpolated from weather stations using ordinary kriging in the Geostatistical Analyst Extension of ESRI Arc Map 9.2. Weather records were obtained from the National Weather Service (web-site: http://gis.ncdc.noaa.gov/website/ims-cdo/som/viewer.htm) and the Lower Colorado River Association (LCRA) (Bob Rose, LCRA, personal communication). |
| April precipitation | For details, see "April temperature". |
| May temperature | For details, see "April temperature". |
| May precipitation | For details, see "April temperature". |

Table 3

**Binomial Tests**

| Test | *p*-Value |
|---|---|
| Fixed cumulative value = 1 | 0 |
| Fixed cumulative value = 5 | 0 |
| Fixed cumulative value = 10 | 0 |
| Minimum training presence | $7.155 \times 10^{-25}$ |
| $10^{th}$ percentile training presence | 0 |
| Equal training sensitivity and specificity | 0 |
| Maximum training sensitivity plus specificity | 0 |
| Equal test sensitivity and specificity | 0 |
| Maximum test sensitivity plus specificity | 0 |
| Balance training omission, predicted area and threshold value | 0 |
| Equate entropy of thresholded and non-thresholded distributions | 0 |

Table 4

**Acreage of Predicted Habitat**

| Model | BCV Habitat (acres) |
|---|---|
| Comal | |
| Top 50% of Maxent predictions (Fig. 5(a)) | 92.48 |
| Top 10% of Maxent predictions (Fig. 5(b)) | 0.89 |
| Hays | |
| In shrub/scrub, top 50% of Maxent predictions (Fig. 6(a)) | 35067.45 |
| In shrub/scrub, top 10% of Maxent predictions (Fig. 6(b)) | 8673.66 |

Figure 1

**Occurrence Records for the Black-capped Vireo in the Study Area**

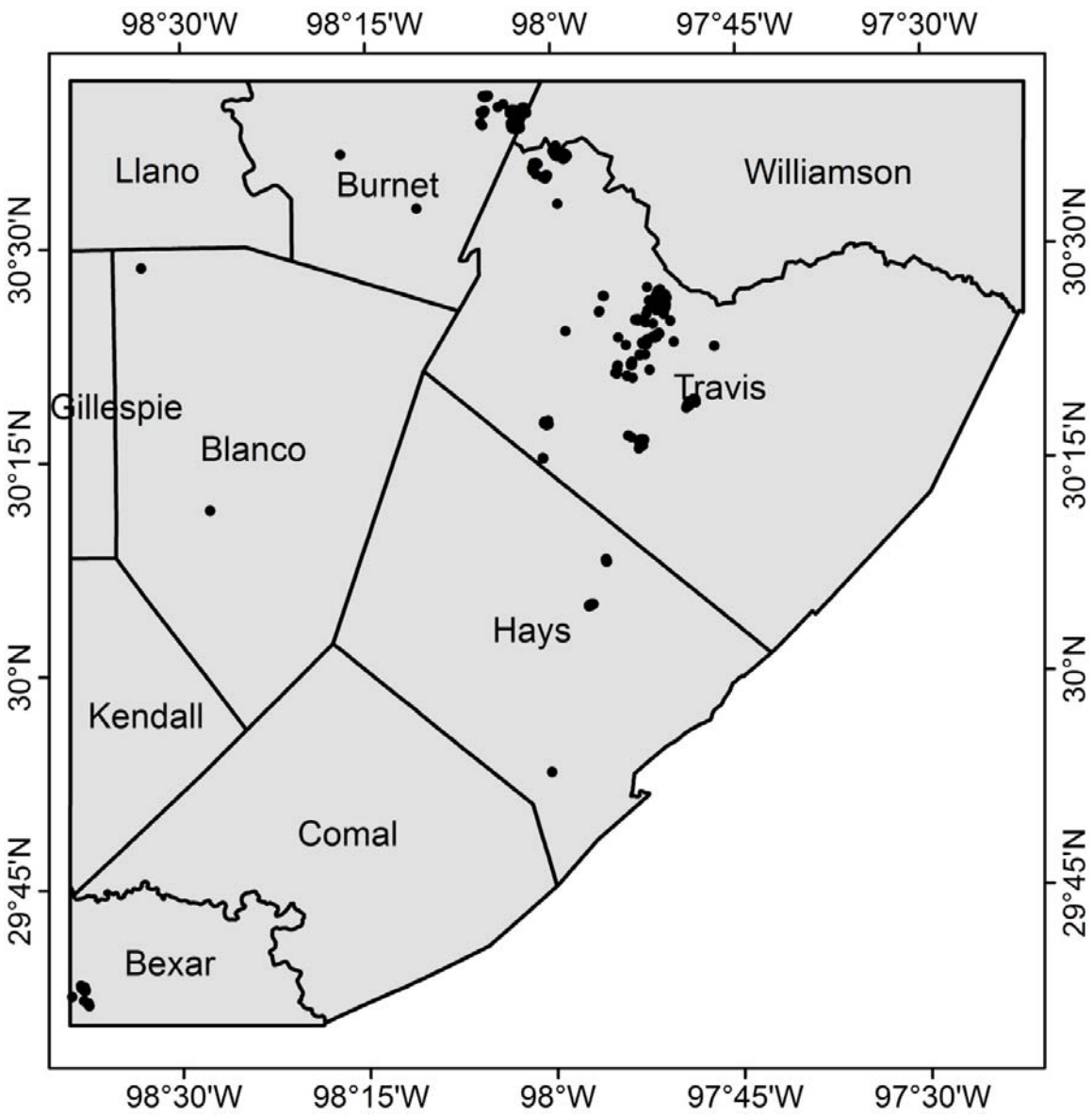


Figure 2

**Fundamental Niche of the Black-capped Vireo in the Study Area**

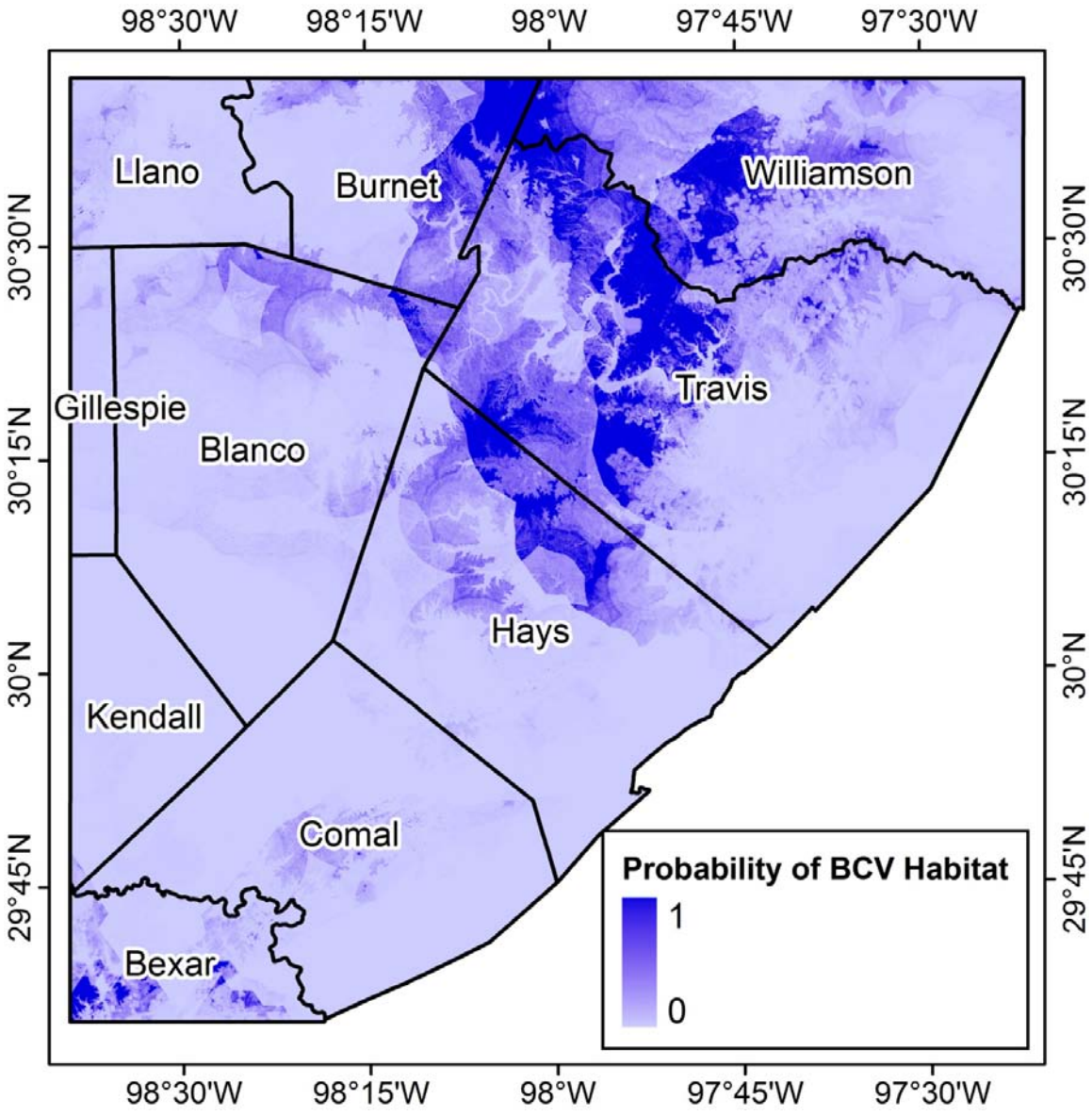


Figure 3

**Modified Fundamental Niche of the Black-capped Vireo in the Study Area**

Sites with a predicted probability less than 0.1 are excluded. (Areas shown in gray are either predicted by Maxent to be BCVI habitat with probability less than 0.1).

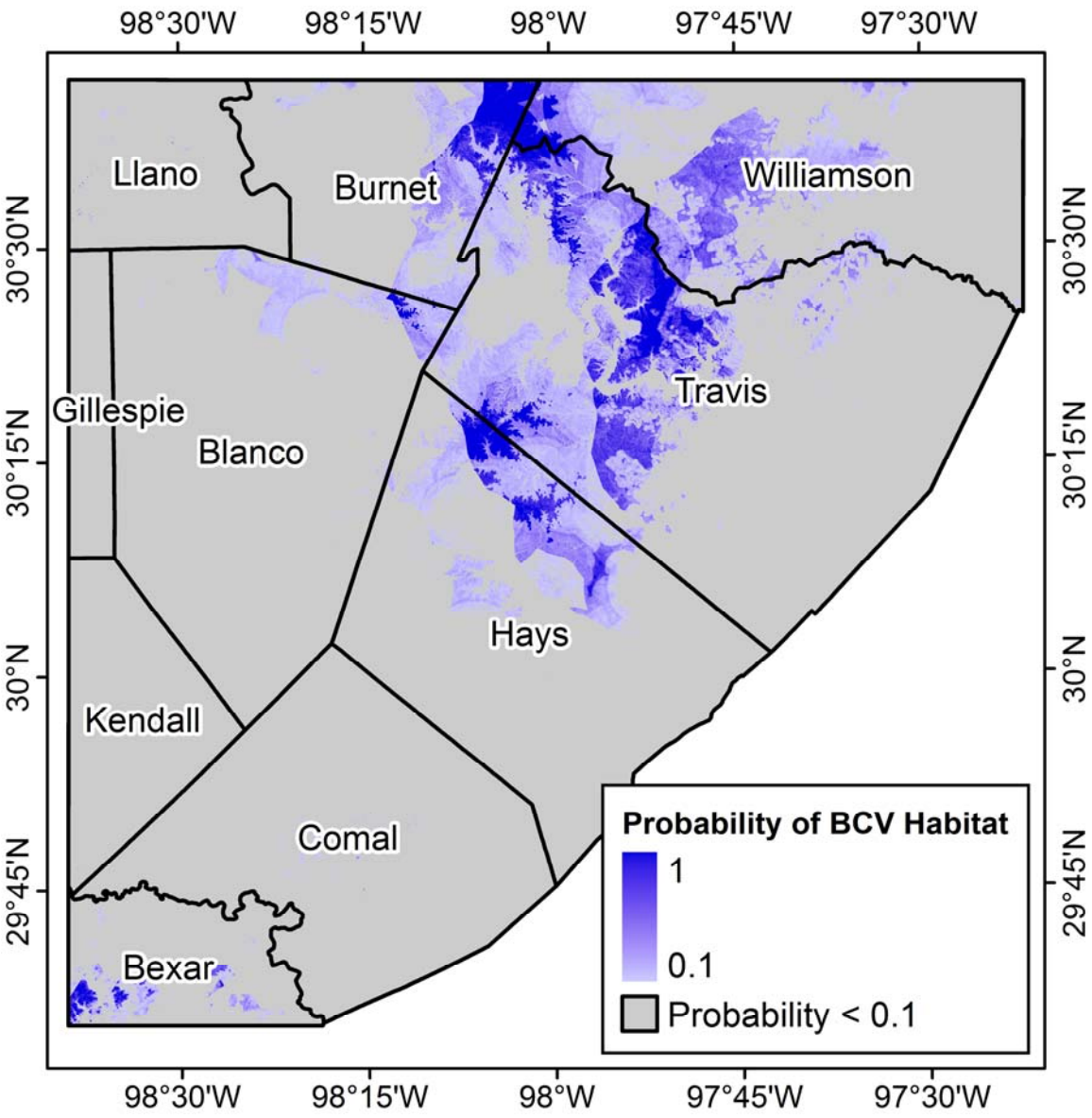


Figure 4

**Modified Fundamental Niche of the Black-capped Vireo in the Study Area Restricted to Shrubland**

Sites with a predicted probability less than 0.1 or lacking shrub vegetation are excluded. (Areas shown in gray are either predicted by Maxent to be BCVI habitat with probability less than 0.1 or lack shrubs.)

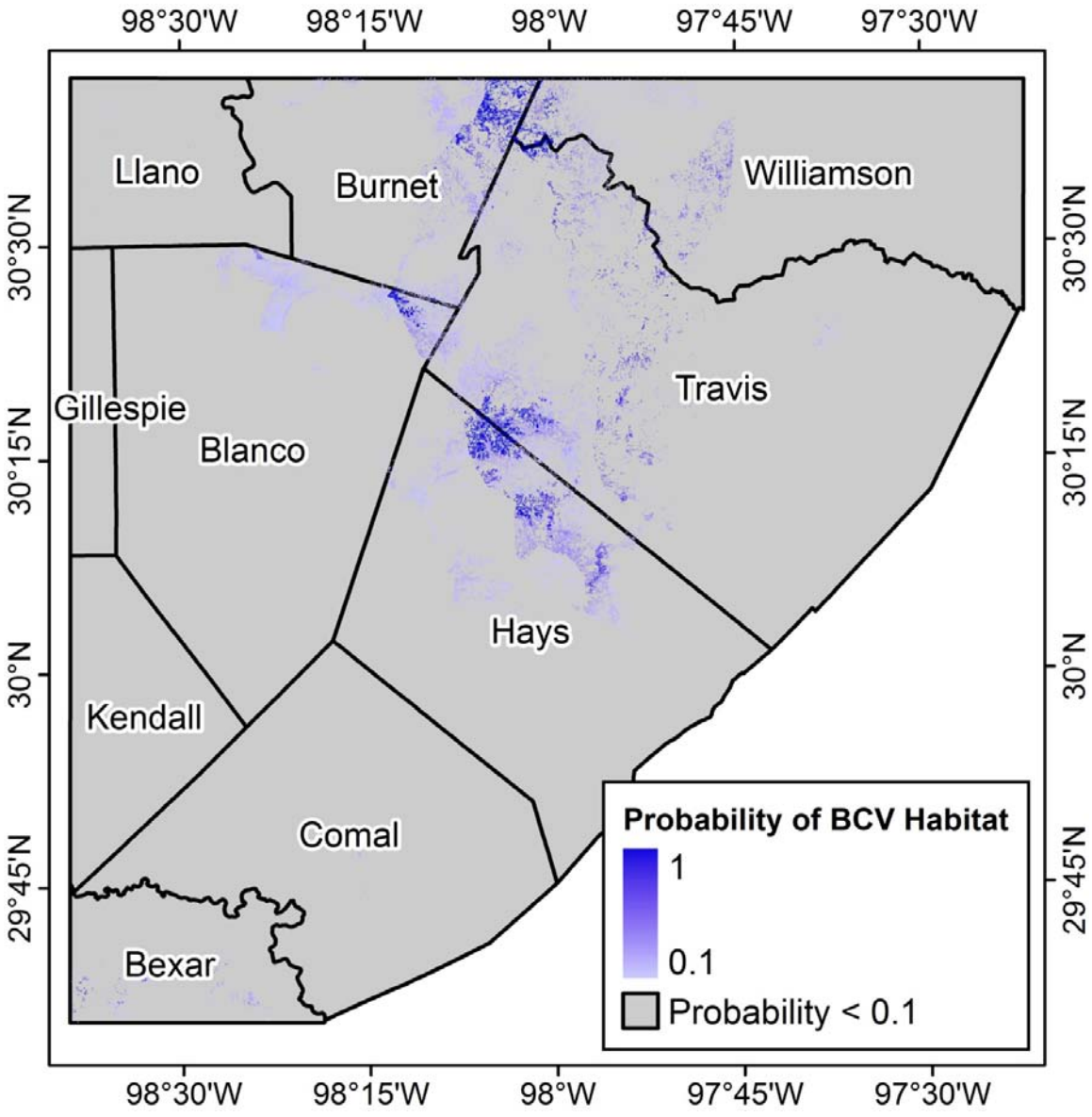


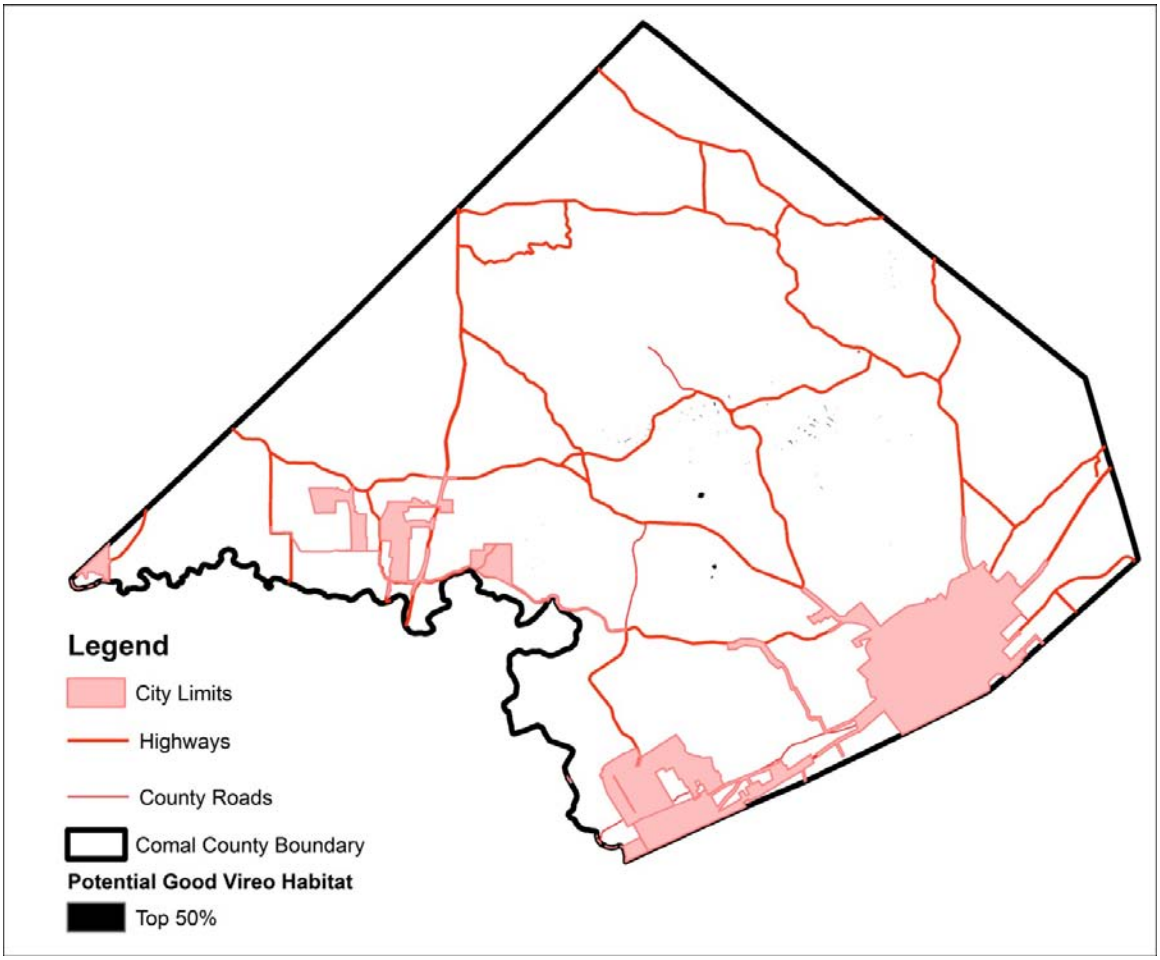
Figure 5

**Good and High-Quality Habitat for the Black-capped Vireo in Comal County**

Figures 5(a) and 6(b) are detailed habitat maps for Comal County. In both figures, good habitat is interpreted to be restricted to the top 50 % of the suitable sites; high-quality habitat was restricted to the top 10 %. The good habitat is clipped to exclude sites without shrubs.

(a)

**Good Habitat**



(b)

**High-Quality Habitat**

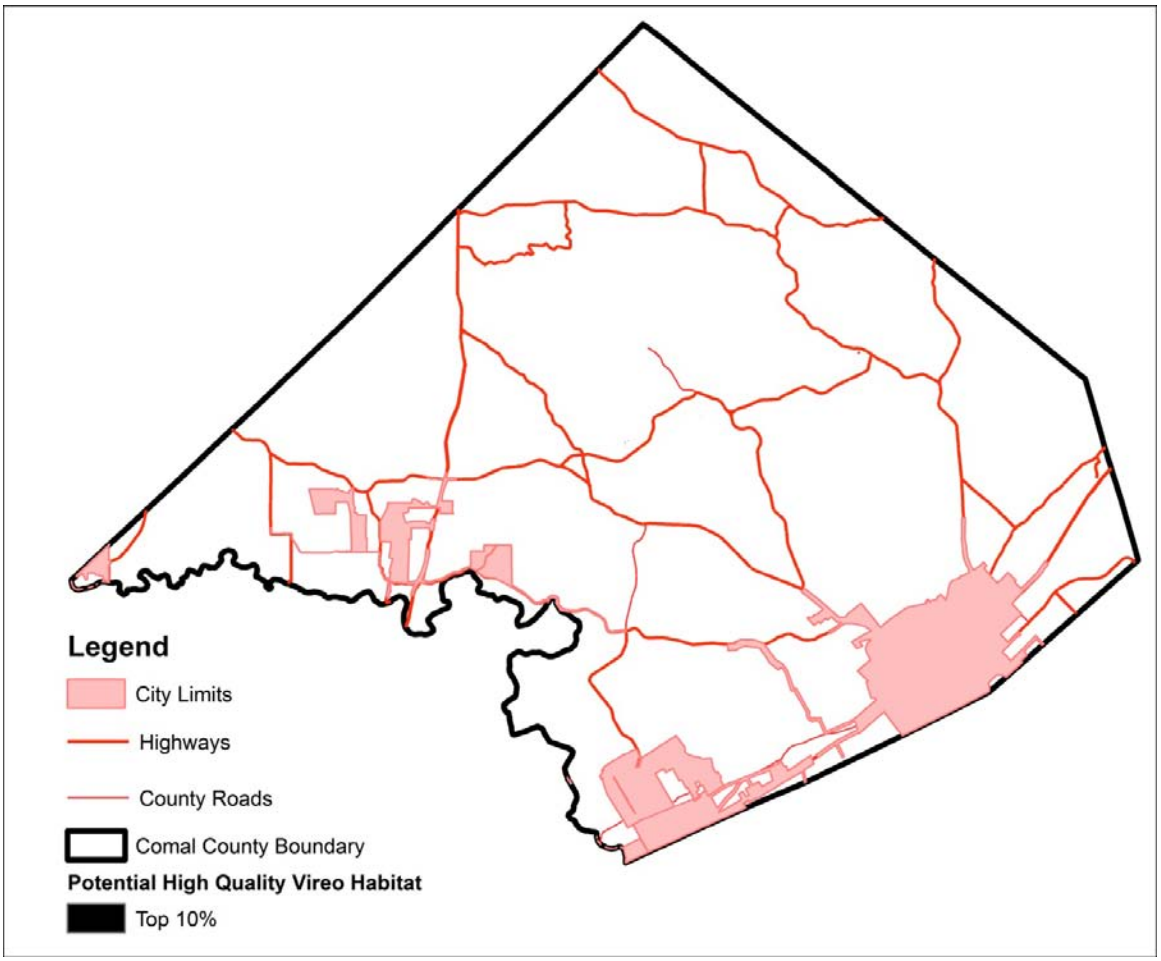


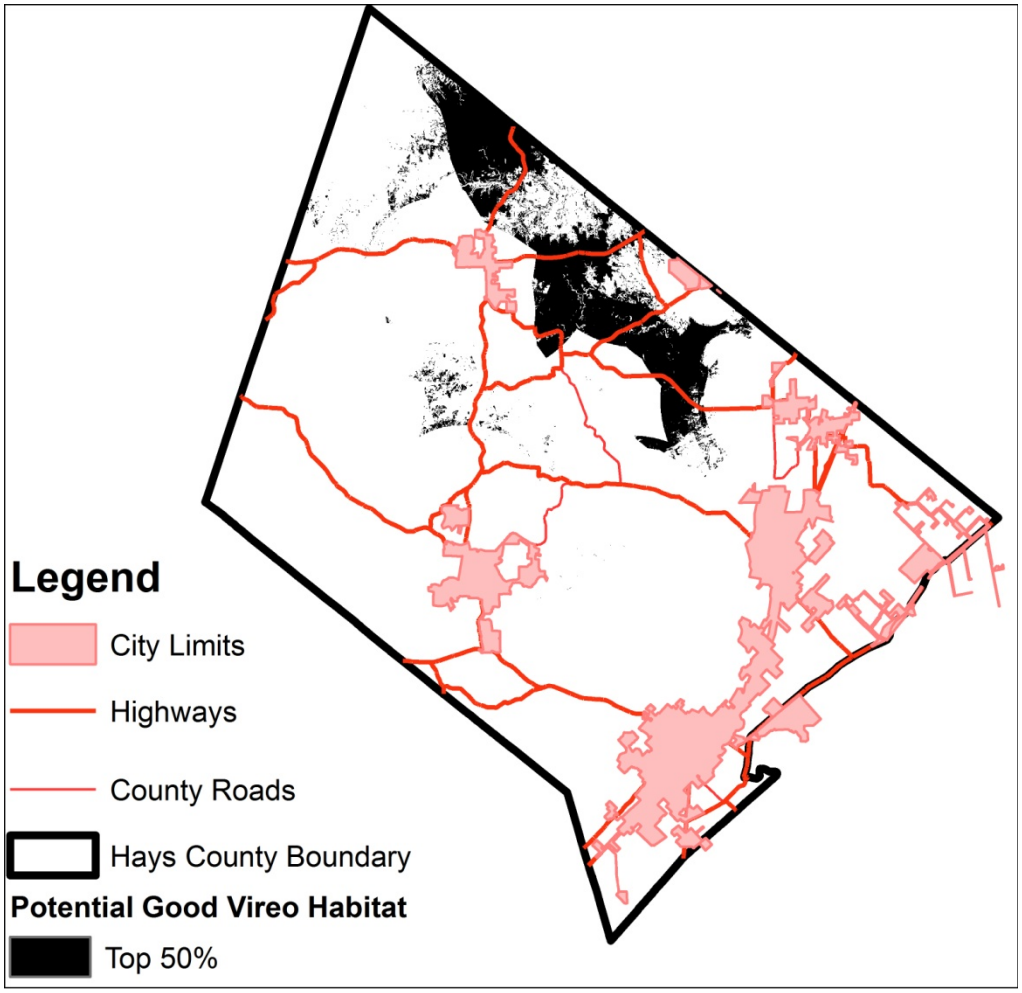
Figure 6

**Good and High-Quality Habitat for the Black-capped Vireo in Hays County**

See legend for Figure (5), but now referring to Hays County.

(a)

**Good Habitat**



(b)

**High-Quality Habitat**

