Using Spatial Landscape Variables to Refine the Range Map of a Migratory Songbird

Andrew M. Cameron, Virginia Commonwealth University

INTRODUCTION

Reliable and accessible maps that accurately represent areas of likely species occurrence are important for a range of stakeholders, including researchers, conservationists, policymakers, and concerned citizens. Effective conservation work ultimately depends upon knowing where individuals of a given species live. Modeling species-habitat relationships requires being able to quantify environmental attributes associated with areas where species occur (Rotenberry & Balasubramaniam, 2020). Similarly, ascertaining species richness values with high confidence cannot be done without a clear understanding of the areas that individuals actually inhabit (Hurlbert & White, 2005).

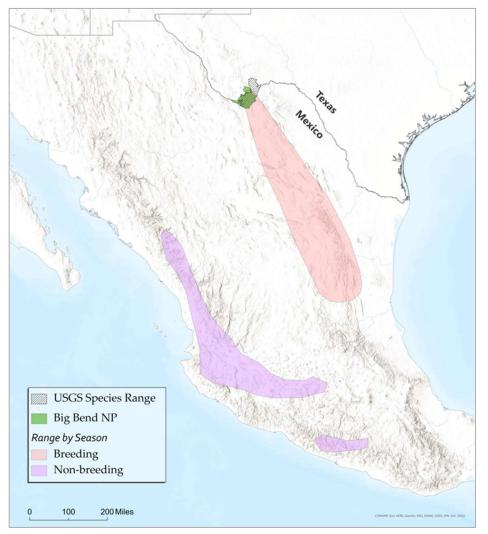
Despite their importance, range maps often display similar shortcomings. Most range maps are the products of subject-area experts who apply their knowledge to produce 'best estimates' of species distribution. These maps, often as a result of the scale at which they must be presented, tend to represent species occurrence with a relatively coarse grain. They typically display polygons that reflect occurrence in a binary way. Such maps suggest that a species occurs within the polygon and does not occur outside of it (Mainali et al., 2020). While such maps can provide a broad indication of the area outside of which it is unlikely or rare to encounter a species, they often fail to reflect the heterogenous distribution of said species within its occurrence polygon. Thus, expert range maps can fail to truly reflect where a species is likely to be found, with maps frequently overestimating species occurrence (Rotenberry & Balasubramaniam, 2020). As anyone who has ever used a field guide will likely note, there are invariably large swathes within a species'

recognized range that are unlikely if not wholly implausible areas in which to ever expect to see that species.

The disjunct between species' ostensible ranges and the far patchier and uneven distributions that are their underlying reality can be explained, at least in part, by the failure to account for spatial heterogeneity within the area bounded by a species range polygon (Rotenberry & Balasubramaniam, 2020). Through the use of geospatial technology, relevant environmental variables associated with species occurrence, such as vegetation cover and elevation, can supplement



Figure 1. Male Colima Warbler on breeding grounds in the Chisos Mountains in Big Bend National Park.



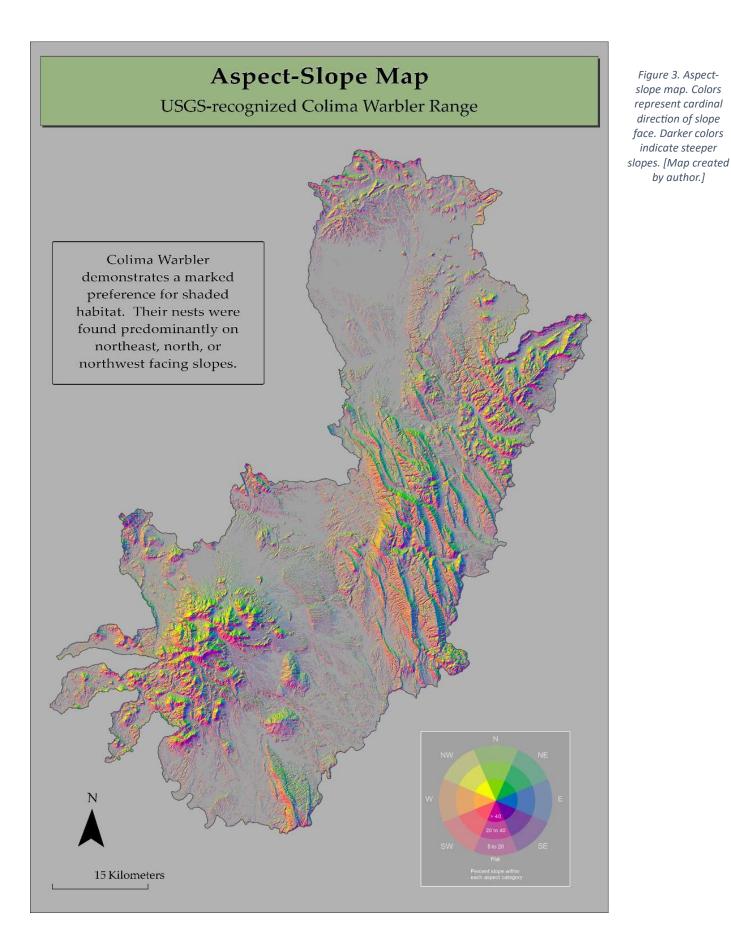
traditional range maps and thereby achieve a more accurate understanding of species distribution (Ocampo-Peñuela et al., 2016).

Colima Warbler (henceforth COWA) (Leiothlypis crissalis) (Fig. 1) is a neotropical migratory songbird from the Parulidae family. Colima Warbler breeds at sites throughout the drv. high elevation forests of Mexico's northern Sierra Madre Occidental range, and just barely enters the United States in the Chisos Mountains. Within the US, COWA's range lies entirely within the boundaries of Big Bend National Park (Fig. 2). The species is one of the least studied North American warblers. Little is known of

Figure 2. Colima Warbler range based on BirdLife International's species range map. [Map and seasonal range polygons produced by the author.]

their nonbreeding distribution and ecology (Beason & Wauer, 2020). Data about breeding birds are limited to those collected as part of the few studies that have been conducted in the Chisos Mountains, as well as from public observation databases like eBird and iNaturalist (Beason & Wauer, 2020; Lanning et al., 1990; Van Tyne, 1955). Lack of robust data on the distribution and abundance of COWA is due in no small measure to the rugged terrain, remote location, and frequent inaccessibility that characterizes areas that the species inhabits.

COWA occurrence and nesting location in the Chisos Mountains are highly correlated with four variables: elevation, vegetation, slope of terrain, and aspect (i.e., the cardinal direction in which a slope or land surface faces) (Beason & Wauer, 2020). Breeding birds inhabit areas dominated by oak, pinyon, juniper, and Arizona cypress, and demonstrate a clear preference for elevations above 1,500 m, with individuals most frequently observed at elevations \geq 1,800 m (Lanning et al., 1990; Van Tyne, 1955). COWA employs a ground-nesting strategy and prefers steep (\geq 35°), north-facing slopes (Fig. 3), and sites that are shaded from direct sunlight for 70% of daylight hours (Beason & Wauer, 2020).



The present study takes these four landscape variables (*elevation, slope, aspect,* and *landcover*), which have been identified as predictors of COWA breeding occurrence and uses them as parameters to refine the official COWA range polygon of the United States Geological Survey representing the species area of occurrence within the United SItates. Two discrete zones of habitat suitability within the area bounded by the USGS polygon are then compared with respect to the four variables.

METHODS

This analysis was conducted using ArcGIS Pro 3.0.2 and R 4.2.1.

Data and projections

The study area is the COWA breeding range within the state of Texas, as delineated by the United States Geological Survey official species range map polygon.

This analysis utilized three input datasets. Spatial data were referenced using the North American Datum of 1983 and projected using Albers Conical Equal Area as coordinate system.

The input data layers consisted of:

- Three digital elevation models (DEMs) with a spatial resolution of 1/3 arc seconds mosaicked together. The resulting raster was projected into Albers and resampled to a resolution of 10x10 m. Input DEMs were obtained from the USGS National Map data delivery service.
- 2) Landcover raster data obtained from the Texas Parks and Wildlife Ecological Systems Classification and Mapping Project (Elliot, 2014). The data use a more fine-grained system of vegetation classification and offer a higher resolution (10x10 m) than the USGS National Land Cover Dataset (30x30 m).
- 3) A vector data layer consisting of a single contiguous polygon representing the USGS species extent of occurrence for *Leiothlypis crissalis*. The polygon data were on-the-fly projected into Albers Conical Equal Area.

Additionally, a pre-existing COWA habitat raster model, created by USGS and derived using elevation and landcover data, was used at the backend of the analysis as a point of comparison with the current study's model output.

Deriving parameters

To derive layers that could be used as final inputs to model habitat for COWA within the extant USGS range polygon, both the DEM and the vegetation raster datasets were first filtered using the *Extract by Mask* tool with the USGS polygon used as the feature mask. The resulting masked DEM

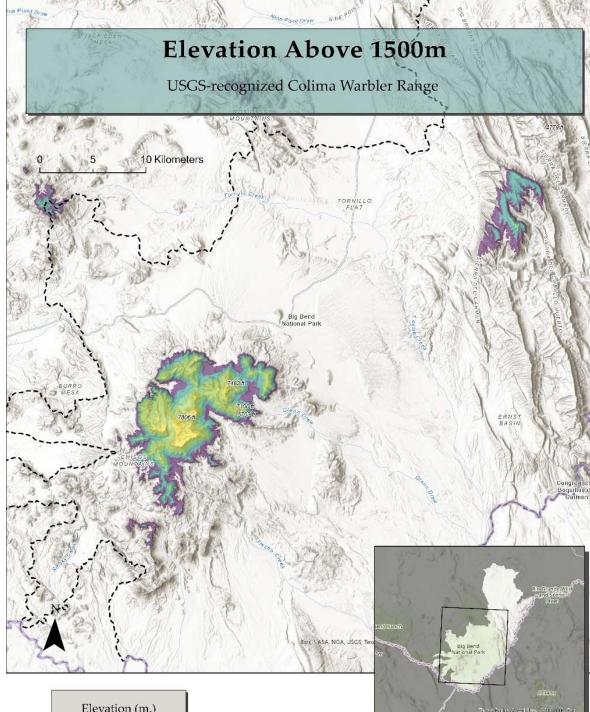


Figure 4. Areas within COWA USGS species range polygon containing elevation values of 1,500 m and above [Map created by author.].

E	llevation (m.)
	1,500 - 1,600
	1,600 - 1,700
	1,700 - 1,800
	1,800 - 2,100
	2,100 - 2,381
[]	USGS Species Range

Colima Warblers breed within a narrow altitudinal range. They are rarely found at elevations below 1500 m, and tend to nest between 1800-2100 m sea level.

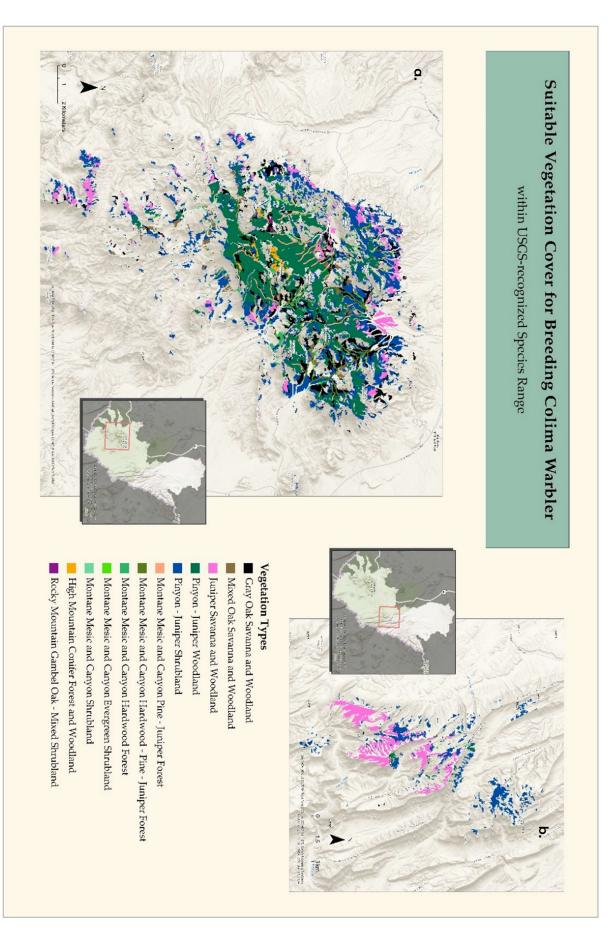


Figure 5. Areas within USGS-recognized species range containing suitable vegetation, symbolized by vegetation type. (a) Chisos Mountains (b) Sierra del Carmen/Sierra del Caballo Muerto. [Map created by author.] was used to derive both the slope of the terrain and the aspect of the slopes within the USGS polygon.

Elevation and vegetation were treated as filtering variables. Cells within the USGS polygon that were below 1,500 m or that did not contain appropriate vegetation were excluded from analysis (Fig. 4). Slope and aspect values, on the other hand, were included in the analysis if and only if they were contained in cells that passed through the elevation and vegetation filtering.

Search queries were executed using the *Extract by Attributes* tool to select for cells with values above 1,500 m (Fig. 4) and appropriate vegetation types (Fig. 5). The resulting vegetation and elevation rasters were designated *suitableVeg* and *elev_above1500*, respectively. A subset of high elevation cells lacked suitable vegetation (e.g., cliff faces, montane grassland), while some suitable vegetation within the range polygon fell below the 1,500 m threshold. Layers *suitableVeg* and *elev_above1500* were therefore further subset to generate *final_veg* and *final_elevation* layers, using one another as masks and extracting only overlapping cells.

The slope and aspect datasets were filtered to include only cells that overlapped with cell values above 1,500 m and corresponding to suitable vegetation types. This was done by executing the *Extract by Mask* tool and using the *final_elevation* raster as mask, resulting in layers named *final_slope* and *final_aspect*, respectively. Throughout the analysis, all tools that had a 'Snap Raster' environment option were assigned the original Texas Parks and Wildlife vegetation raster as snap raster to ensure cell alignment between the four data layers.

Reclassifying parameters and calculating habitat area

The four habitat predictors comprise both continuous (elevation, slope, aspect) and discrete (vegetation) variables. The topographical variable rasters contain cell values with incommensurate units and widely disparate value ranges. Overlaying these four datasets would result in output that is not easily interpretable and which over-weights variables with higher value ranges. To avoid this, the cell attributes of all four datasets were reclassified using the *Reclassify* tool to standardize their values and produce a final habitat raster model overlay with cell values ranging between 1 and 5.

Elevation data were reclassified such that cell values < 1800 received a value of 1, while cells values \geq 1800 were reclassified with a value of 2. All cells in the *final_veg* raster were assigned a value of one. Slope values were reclassified in the following way: < 15° = 1; < 30° = 2; < 60° = 3. Terrain with slope of 60° or greater was treated as unsuitable for ground nesting and assigned a value of zero. Aspect values between 0 and 112 and between 292.5 and 360 (corresponding to north, northeast, northwest, and east) were reclassified with a value of 2. All other cells were assigned a value of 1.

The four reclassified raster layers were overlaid and their values aggregated using the *Cell Statistics* tool to produce an output raster with cell values reflecting the sum of values from each of the four inputs. The output raster contained cells with values ranging from 3 to 7. That raster

was also reclassified to produce a final output habitat model raster containing values between 1 and 5, with five representing the best possible habitat suitability score.

Data analysis

Summary statistics were produced for both the current study model and the USGS habitat model by using the *Zonal Statistics as Table* and *Tabulate Area* tools in ArcGIS Pro. Tables were exported as .csv files and imported into R for manipulation. Data were manipulated using the *dplyr* package, and final tables were produced in R with the *kableExtra* package.

Cell values in each of the two zones of the final habitat model (see RESULTS below) were compared by testing for significant differences with a non-parametric Kruskal-Wallis test, using the *kruskal.test* function in R. Prior to testing, the final habitat model raster was converted to point data using the *Raster to Point* tool in ArcGIS. The points were selected by zone to create a new, separate layer for each group. The attribute tables were exported as .csv files and imported into R.

Both datasets (CM and SCCM) were too large (n=583,262; n=96,595) for Shapiro-Wilk normality testing, which can take as input a maximum of 5,000 data points. Each dataset was randomly sampled (n=500) 100 times and a Shapiro-Wilk normality test was conducted on each sample with resulting p-values stored in a vector. Mean p-values were 2.26 x 10⁻¹⁹ for CM and 7.33 x 10⁻²⁶ for SCCM, indicating that the null hypothesis of a normal distribution should be rejected for both datasets, and pointing to the need for a non-parametric analysis such as Kruskal-Wallis.

RESULTS

The final habitat model resulted in a reduction of the area of likely occurrence by 98.24% as compared with the original USGS species range polygon. The total area of cells contained in the final model output raster was 68.94 km², compared with an original range polygon area of 3,908.76 km² (Table 1). Nearly half of cells (48.98%) at or above 1,500 m lacked appropriate vegetation, while 28.44% of cells representing suitable vegetation were located at elevations below 1,500 m.

FEATURE	AREA (km2)	PERCENT RANGE POLY
USGS species range	3908.76	100%
Elevation above 1500m	135.27	3.46%
Suitable vegetation	88.65	2.27%
Elevation + Vegetation	69.02	1.77%
Final habitat model	68.94	1.76%

Table 1. Area, in sq. km, of original range polygon and derived landscape features, and the percent of the original USGS range polygon represented by each feature.

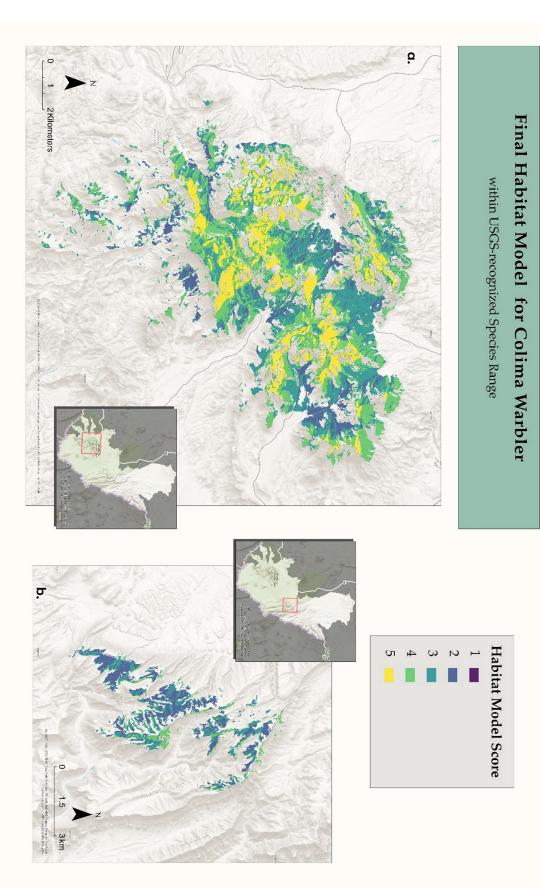


Figure 6. Final habitat model results. All colored cells represent suitable habitat, with higher scores indicating closer alignment with demonstrated species preferences and occurrence data. (a) Chisos Mountains (b) Sierra del Carmen/Sierra del Caballo Muerto.[Map created by author.] Suitable habitat was restricted to two narrowly circumscribed zones situated entirely within the boundaries of Big Bend National Park, (1) the Chisos Mountains zone (CM) to the southwest and (2) a much smaller zone spanning the Sierra del Carmen and Sierra del Caballo Muerto ranges (SCCM), roughly in the center of the range polygon (Table 2, Fig. 6). The habitat area within the CM zone is 6.14 times larger than the area within the SCCM zone, with total areas of 59.28 km² and 9.66 km², respectively. The mean cell value was higher in the CM zone (3.63) than in the SCCM zone (2.67). The value of the cell representing the 90th percentile was five in the zone CM zone and four in zone SCCM.

ZONE	AREA (km2)	MAX CELL VAL	MEAN CELL VAL	PCT90	
Chisos Mountains	59.28	5	3.63	5	
Sierra del Carmen/de Caballo Muerto	9.66	4	2.67	4	

Table 2. Comparison of zones of suitable habitat identified within the study area. PCT90 is the value of the cell representing the 90th percentile of each dataset.

Vegetation

The types and relative amounts of vegetation present in each zone differ substantially (Table 3, Fig. 6). Juniper-savanna and woodland ecosystem accounts for 74.02 % of the vegetation in zone SCCM, with pinyon-juniper shrubland accounting for another 21.74%. Oak ecosystems of any kind account for 0.21% in zone SCCM. In contrast, oak ecosystems comprise 16.12% of zone CM, with 40.74% of vegetation consisting of pinyon-juniper woodland and 27.02% consisting of pinyon-juniper shrubland. Mesic habitat is nearly absent from zone SCCM, while in zone CM, mesic habitat of various kinds make up 10.82% of extant ecosystems.

	Gray Oak Savanna &	Mixed Oak Savanna &	Juniper Savanna &	Pinyon- Juniper	Pinyon- Juniper	Mesic and Canyon Pine- Juniper	Canyon Hardwood – Pine - Juniper	Montane Mesic & Canyon Hardwood	Montane Mesic & Canyon Evergreen	Montane Mesic & Canyon	Rocky Mountain Gambel Oak - Mixed	High Mountain Conifer Forest &
CM	6.83	2.64	2.43	24.15	Shrubland	Forest	2.12	Forest	Shrubland	Shrubland	Shrubland	0.71

Table 3. Total area represented by ecosystem classes between Chisos Mountains (CM) and Sierra del Carmen/de Caballo Muerto (SCCM) zones. Areas given in square kilometers.

Elevation

Zone SCCM contains no cells with an elevation above 1,800 m, while 79.3% of the total area of suitable habitat in that zone lies at or below 1,650 m. Within zone CM, 23.37 km² of terrain lies at elevations above 1,800 m, accounting for 39.4% of the total area of suitable habitat. (Table 4)

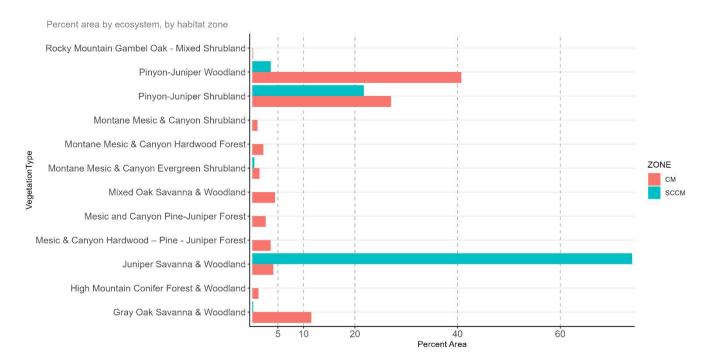


Figure 6. Percent area represented by ecosystem classes between the Chisos Mountains (CM) and Sierra del Carmen/del Caballo Muerto (SCCM) zones.

Slope

Among the two habitat zones, CM contains steeper terrain, with 78.63% of cells having a slope value \geq 16°, and 38.29% of cells \geq 31°. Within zone SCCM, 43.94% of cells have slope values \geq 16°, while 48.82% of cells reflect values < 16°. (Table 5) Area of cells (sq. km) by elevation (in meters), by habitat zone.

ZONE	1500-1650m	1651-1800m	1801-2000m	2001-2381m	
CM	20.35	15.57	12.06	11.31	
SCCM	7.67	1.99	0.00	0.00	

Table 4. Total area (km²) of cells by elevation (m.) band among habitat zones.

Aspect

Slopes within zone CM have a much more northerly exposure, with 48.26% of cells having a northern (N), northeastern (NE), or northwestern (NW) aspect. By contrast, only 23.71% of cells in zone SCCM have an aspect Percent of of suitable habitat by slope (in degrees) among habtiat zones.

ZONE	0-3°	4-9°	10-15°	16-30°	31-60°	61-90°
CM	0.34	7.74	13.02	40.34	38.29	0.27
SCCM	1.87	22.49	24.46	43.94	7.25	0.00

Table 5. Percentage of cells within defined slope ranges, grouped by habitat zone.

value corresponding to N, NE, or NW, while 41.07% of cells have aspect values corresponding to slopes with a southern (S), southeastern (SE), or southwestern (SW) orientation. (Table 6)

Comparison of habitat zone suitability scores

The Kruskal-Wallis rank sum test was used to compare the values of all cells in each of the two habitat zones. The test resulted in a chi-squared value of 3.3148 with 1 degree of freedom and a *p*-value of 0.06866.

F	Pecentage of cells by aspect among habitat zones.									
	ZONE	Ν	NE	Е	SE	S	SW	W	NW	
	СМ	17.03	14.32	8.96	6.25	5.60	8.89	13.50	16.9 <mark>1</mark>	
	SCCM	6.05	9.99	16.15	10.09	11.91	19.07	16.45	7.67	

Table 6. Percentage of cells corresponding to one of eight cardinal or ordinal directions.

DISCUSSION & CONCLUSION

While it is established that COWA nesting and occurrence outside of Mexico are restricted to the Chisos Mountain of southwest Texas (Beason & Wauer, 2020; Lanning et al., 1990; Van Tyne, 1955) (Fig. 7), publicly available range maps from authoritative sources, including BirdLife International and the USGS Species Data maps, fail to accurately convey this information. By leveraging geospatial data related to attributes associated with the areas where a species occurs, a more accurate understanding of potential species distribution can be achieved. Though habitat model cell values ranged from 1 to 5 with five being the 'best' score, it is important to bear in mind that any cell included in the final model represents broadly suitable habitat. This is because only cells with elevation and vegetation attributes associated with COWA habit preference were included in the final model.

The Kruskal-Wallis test for the comparison of cell values between the two habitat zone datasets resulted in a *p*-value (0.06866) that is not statistically significant at the conventional significance level of 0.05. It does, however, suggest a trend towards significance, indicating that the null hypothesis of no difference in habitat suitability between the two datasets cannot be conclusively accepted. This indication is further reinforced when considered in the context of the observed differences, discussed above, between zones CM and SCCM with respect to each of the four environmental variables included in this study. Research by Wauer (2020) suggests that COWA nests are disproportionately found on very steep slopes with northern aspects. The results of this study demonstrate that the Chisos Mountains are characterized by large swathes of steep slopes with northern exposure. Whether the observed trends in COWA nesting site location reflect something fundamental about the species' life history,

nesting strategy, and habitat preferences, or are better explained by the topographical reality of the range they inhabit remains an open question. Further field studies of COWA nesting and occurrence throughout their range are needed to determine if the landscape parameters used in the current model could effectively be applied to model species distribution at a larger scale.

It is important to note that in addition to the species range polygon used in this analysis, USGS also provides its own publicly available species distribution model (U.S. Geological Survey, 2018). The USGS habitat model (UHM) has a 30x30m cell resolution and overlaps much of the model produced in the current study. Like the current model, UHM uses elevation and vegetation as parameters. The total area of suitable habitat in UHM (94.38 km²) is 36.90% greater than the area generated by the model in this study (68.94 km²). Possibly owing to coarser resolution, UHM includes many cells that were excluded from the current model based on vegetation classification (e.g., mountain grassland, desert shrubland, deciduous chaparral).

BirdLife International is a consortium of conservation organizations that work to conserve birds and bird habitat. One of its key missions is to identify and prioritize species that are most at risk of extinction, and to develop conservation strategies to address the threats facing those species. According to BirdLife International,

[Colima Warbler] has a very large range, and hence does not approach the thresholds for Vulnerable under the range size criterion (Extent of Occurrence <20,000 km² combined with a declining or fluctuating range size, habitat extent/quality, or population size and a small number of locations or severe fragmentation). (BirdLife International, 2023).

This claim and resulting designation of COWA as a species of Least Concern may simply reflect a lack of adequate data and information about COWA life history throughout most of its range. The analysis in this study resulted in a range reduction of over

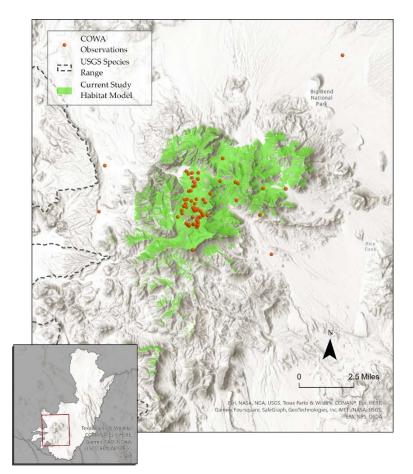


Figure 7. Extent of all reported COWA observations within the United States across all years. Data obtained from eBird and iNaturalist. Points do not represent precise latitude and longitude coordinates of observed bird, and in many cases may only be accurate to within ~1km or more.

98% within the study area. If similar modeling were applied to the rest of species' range, it is possible that the extent of occurrence for COWA could be reduced to a level approaching or below BirdLife International's threatened species threshold. While COWA's realized extent of occurrence may differ significantly from that indicated by BirdLife International's own calculation, there is no indication that the bird's range is undergoing contraction due to human activity. The Chisos

Mountains and therefore COWA's breeding range within the United States are under the protection of the National Park Service, while in Mexico, habitat associated with COWA is not seen as commercially valuable and is subject to minimal exploitation, e.g., logging and grazing (Beason & Wauer, 2020). Nevertheless, COWA's narrow breeding habitat requirements make it vulnerable to the impacts of forest fire in an area that is prone to them and at a time when the frequency and intensity of wildfires may be increasing due to climate change (Hurteau et al., 2014; Moir, 1982).

This study provides a detailed methodology for mapping and modeling habitat suitability for a breeding migratory bird with narrow habitat preferences and in a particular portion of its range. With sufficient data and understanding of life history and habitat preferences, the study's approach could be replicated in other geographic areas of COWA's range and for different species, thereby contributing to a better understanding of avifaunal species distribution, abundance, and habitat suitability, and contributing to maximally effective conservation efforts.

SUPPORTING MATERIALS

Supporting information, including documentation of R code used for data wrangling, analysis, and production of tables and plots, as well as a schematic workflow for the analysis in this study, are available with the included files.

REFERENCES

- Beason, R. C., & Wauer, R. H. (2020). Colima Warbler (Oreothlypis crissalis). *The Birds of North America Online*. https://doi.org/10.2173/bna.colwar.02
- BirdLife International (2023) Species factsheet: Leiothlypis crissalis. http://datazone.birdlife.org/species/factsheet/22721633
- Cornell Lab of Ornithology. (2023). eBird Basic Dataset. Version: EBD_relMar-2023. [Data set]. Ithaca, NY:Author.
- Elliott, L. F., Treuer-Kuehn, A., Blodgett, C. F., True, C. D., German, D., & Diamond, D. D. (2009-2014). Ecological systems of Texas: 391 mapped types. Phase 1-6, 10-meter resolution geodatabase, interpretive guides, and technical type descriptions. Texas Parks & Wildlife Department and Texas Water Development Board, Austin, Texas. Retrieved from http://www.tpwd.state.tx.us/gis/data/downloads#EMS-THurlbert, A. H., & White, E. P. (2005). Disparity between range map- and survey-based analyses of species richness: Patterns, processes, and implications. In *Ecology Letters* (Vol. 8, Issue 3). https://doi.org/10.1111/j.1461-0248.2005.00726.x
- Hurteau, M. D., Bradford, J. B., Fulé, P. Z., Taylor, A. H., & Martin, K. L. (2014). Climate change, fire management, and ecological services in the southwestern US. *Forest Ecology and Management*, 327. https://doi.org/10.1016/j.foreco.2013.08.007

- Lanning, D. V, Marshall, J. T., & Shiflett, J. T. (1990). Range and habitat of the Colima warbler. Wilson Bulletin, 102.
- Mainali, K., Hefley, T., Ries, L., & Fagan, W. F. (2020). Matching expert range maps with species distribution model predictions. *Conservation Biology*, *34*(5). https://doi.org/10.1111/cobi.13492
- Moir, W. H. (1982). A Fire History of the High Chisos, Big Bend National Park, Texas. *The Southwestern Naturalist*, 27(1). https://doi.org/10.2307/3671411
- Ocampo-Peñuela, N., Jenkins, C. N., Vijay, V., Li, B. V., & Pimm, S. L. (2016). Incorporating explicit geospatial data shows more species at risk of extinction than the current Red List. *Science Advances*, 2(11). https://doi.org/10.1126/sciadv.1601367
- Rotenberry, J. T., & Balasubramaniam, P. (2020). Connecting species' geographical distributions to environmental variables: range maps versus observed points of occurrence. *Ecography*, 43(6). https://doi.org/10.1111/ecog.04871
- Texas Parks and Wildlife Department. (2020). Ecological Mapping Systems [Data set]. Retrieved from https://tpwd.texas.gov/files/gis/EcologicalMappingSystems_10mRaster.zip
- U.S. Geological Survey (USGS) Gap Analysis Project (GAP). (2018). Colima Warbler (Vermivora crissalis) bCOLWx_CONUS_2001v1 Range Map [Data set]. U.S. Geological Survey data release.https://doi.org/10.5066/F7HH6J27.
- U.S. Geological Survey. (2019). USGS 1/3 Arc Second n30w102 [Data set]. ScienceBase Catalog. https://www.sciencebase.gov/catalog/item/606785e9d34edc0435c0a5c4
- U.S. Geological Survey. (2019). USGS 1/3 Arc Second n30w102 [Data set]. ScienceBase Catalog. https://www.sciencebase.gov/catalog/item/606785eed34edc0435c0a5c7
- U.S. Geological Survey. (2019). USGS 1/3 Arc Second n30w102 [Data set]. ScienceBase Catalog. https://www.sciencebase.gov/catalog/item/606785e9d34edc0435c0a5c4
- Van Tyne, J. (1955). Colima Warbler (Vermivora crissalis) in Colima. *The Auk*, 72(1). https://doi.org/10.2307/4081402