Using an extensively documented and validated habitat suitability model for conservation management of the Texas tortoise, *Gopherus berlandieri*, defining its status in the coastal prairies and eastern range

Submitted to TPWD as a final report

For project grant #T-101-1

November 7, 2014

By: Anjana Parandhaman, Shawn F McCracken, Michael R. J. Forstner,

and David Rodriguez.

Introduction:

The Texas tortoise, *Gopherus berlandieri* (Agassiz, 1857), is a state threatened reptile in Texas. It is listed in Appendix II under the Convention on International Trade in Endangered Species (CITES) (Groombridge, 1982; Rose and Judd, 1982). However, it is the only *Gopherus* species in the United States that is not federally listed under the United States of America Endangered Species Act. The historical range of the Texas tortoise includes the Tamaulipan thornscrub and coastal plains ecosystems that extends from Southern Texas to Coahuila, Nuevo Leon and Tamaulipas, Mexico (Rose and Judd, 1982). In Texas, the species occurs south of a line from Del Rio to San Antonio and Rockport (Figure 1) (Rose and Judd, 1982). This includes Val Verde, Kinney, Uvalde, Medina, Bexar, Karnes, Refugio, and Aransas counties and all counties southward in the state (Rose and Judd, 1982). However, there are also an array of Texas counties where the species has been detected outside this distribution. These specimens may have naturally migrated to some of these areas or simply been anthropogenically transported to them (e.g., Brewster, Tarrant, Coleman, Sutton, and Kimble counties) (Dixon, 2013).

Less is known about the Texas tortoise than the other North American tortoises, *Gopherus agassizii* and *Gopherus polyphemus*, even though it is the only species of tortoise in Texas. It is the smallest (maximum recorded size = 228 mm) and most sexually dimorphic of the four extant species of genus *Gopherus* (Judd and Rose, 1983; Bury and Smith, 1986). These long-lived herbivores are found in arid or semi-arid ecosystems, such as the thornscrub ecosystem of south Texas (Judd and Rose 1989). They are generally found in vegetation loosely characterized as coastal prairie or thornscrub (Rose and Judd, 1975; Rose and Judd, 1982). In coastal areas, Texas tortoises occur on 'lomas', which are clay to sandy ridges surrounded by the salt flats and marshes (Bury and Smith, 1986). The thornscrub ecosystem is mainly dominated by

Honey Mesquite (*Prosopis glandulosa*), Granjeno (*Celtis spinosa*), and Prickly Pear (*Opuntia* spp.) (Bury and Smith, 1986).

Texas tortoises are most active after short periods of rainfall, but if rainfall continues for a period of days then the tortoises become inactive again (Rose and Judd, 1975). Temperature and light also plays a role in activity of this species (Rose and Judd, 1975). The preferred body temperature for Texas tortoises is around 30°C, but never below 22°C (Rose and Judd, 1975). When temperatures reach 40°C, the tortoises become inactive (Rose and Judd, 1975). In April, tortoises are usually active at midday, but by August they have two 'diel' activity periods, morning and afternoon, avoiding high mid-day temperatures in the summer (Auffenberg and Weaver, 1969; Rose and Judd, 1975; Bury and Smith, 1986).

Due to changes in agricultural practices such as grazing of livestock and changes in land use including an increase in oil and gas exploration and extraction, there has been a reduction in available habitat for tortoises (Auffenberg and Weaver, 1969). Fracking (high-volume horizontal hydraulic fracturing) is a process used to obtain oil and gas and has become a large industry in Texas. Little is known about the effects of fracking on tortoises. Various effects of fracking on other species, especially those with small geographic ranges, are due to salinization and forest fragmentation caused by this practice of petroleum extraction (Gillen and Kiviat, 2012). Another concern is the development of access roads and significant increases in vehicular traffic that may contribute to increased road mortalities of Texas tortoises (Auffenberg and Weaver, 1969; Hellgren et al., 2000). Large-scale fracking is currently underway in the Eagle Ford Shale in South Texas, and will, or possibly has already affected the Texas tortoise. Also, the introduction of buffelgrass (*Pennisetum ciliare*), and continued agricultural practices including grazing and conversion to row crops has resulted in approximately 90% reduction of the brushland in the

Lower Rio Grande Valley (LRGV) of Texas since the 1900's (Jahrsdoerfer and Leslie Jr, 1988; Ramirez, 1986).

Habitat suitability modeling helps predict species distribution and therefore aids in conservation biology efforts. It is a tool for the management of endangered species, ecosystem reintroduction, and population viability analyses (Palma et al., 1999; Sanchez-Zapata and Calvo, 1999; Hirzel et al., 2001). Maximum Entropy (Maxent) modeling is a technique for making predictions from presence only data or from 'incomplete' information. Maxent estimates a target probability distribution by calculating the probability distribution of maximum entropy (i.e. most spread out) (Phillips et al., 2006). It is a method by which estimations of uniform distribution of sampling points are made as compared to background locations, given constraints in the data (Phillips et al., 2006; Phillips et al., 2004; Grendar and Grendar, 2001). The ability to use presence only data is useful when there is no absence data available, especially when available data is from museum or herbarium collection observations (Phillips et al., 2004). The purpose of this study is to assess the current habitat suitability for the Texas tortoise within its historic range, particularly in its eastern range, using Maxent (version 3.3.3k; http://www.cs.princeton.edu/~schapire/maxent/; (Phillips et al., 2004 and 2006) and ArcGIS (v 10.2; ESRI, 2013) modeling tools.

Methods and Materials:

Mendeley Reference Manager

An initial literature search pertaining directly to studies conducted on the Texas tortoise (*G. berlandieri*) was conducted using library resources at Texas State University, Google Scholar, and Web of Knowledge search engines. Citations of literature pertaining to the Texas tortoise were added to an online bibliography hosted at the free Mendeley Reference Management site (www.mendeley.com) under the group name: Gopherus berlandieri.

Texas Tortoise Surveys

Road surveys were conducted monthly from March through October of 2014, to search for live and dead tortoises along roads. Primarily, these surveys were focused on areas to the east of State Highway 16 and north of State Highway 285. Additional road surveys were conducted along the northern species boundary, roughly along a line from San Antonio to Del Rio. Surveys were conducted for approximately 12 hours a day, generally starting shortly after sunrise with a midday break and then continuing until sunset. Potential rain events in the study area were specifically prioritized for survey trips. Additionally, surveys were conducted by means of line transects on the James E. Daughtrey Wildlife Management Area, in McMullen County, on October 12, 2014. Potential burrows/pallets were observed but no tortoises were detected.

Data collected for each specimen observed included all data required by the TXNDD reporting form and additionally: weight, carapace length and width, plastron length and width, body depth, sex, and estimated age (scute annuli counts when possible). Each tortoise was photographed in dorsal, ventral, anterior, posterior, and lateral left and right aspects. Each live tortoise was assigned a unique ID at time of first encounter using a dremel tool to create a unique code using marginal scute positions. These ID numbers allow for recapture identification. From live tortoises, a small aliquot of blood (~1mL) was drawn from the femoral vein and placed in blood storage buffer. Muscle tissue was collected from the least-exposed area of the carcass and placed in 95% ethanol for tortoises found dead. Blood and tissue collections are stored at -80ºC in the MRJ Forstner Tissue Collection at Texas State University.

Texas Tortoise Observation Data

In addition to our survey data observation points, extensive database searches for Texas tortoise observational geodata were conducted using the following databases: MRJ Forstner Tissue Collection at Texas State University, Biodiversity Information Serving Our Nation (BISON), iNaturalist, Texas Natural Diversity Database (TXNDD) and unpublished field data of Dr. Francis Rose (Professor Emeritus; Texas State University). The number of observation points used in our modeling was reduced based on accuracy of the individual data records. We chose a cutoff accuracy of <5000 m. Coordinates will continue to be collected as more tortoises are observed this year.

Pre-processing of Data

Environmental variables, known to affect the presence of the Texas tortoise, were determined from the literature and expert consultation (Andersen and Beauvais 2013; Rose and Judd, 2014). The Texas tortoise model built by Andersen and Beauvais (2013) was used as a primary reference for this purpose. For their model, TXNDD biologists were consulted to determine environmental variables that affected the Texas tortoise. We used similar variables, and incorporated additional variables we deemed appropriate to the study. All processing was done using ArcMap (10.2). The layers used were as follows:

1. Climatic variables

Climate data was obtained from www.worldclim.org in the form of 30 arc-second cell size ESRI-rasters (Hijmans et al. 2005). Six predictor variables were selected based on the model built by Andersen and Beauvais, 2013.They were as follows:

Name of Variable	**Raster Name**	**Units**
Mean Diurnal Range (Mean of monthly(maximum temperature-minimum temperature)	Bio2	°C*10
Isothermality (Mean Diurnal Range/Temperature Annual Range)	Bio3	-
Minimum Temperature of Coldest Month	Bio6	°C*10
Mean Temperature of Warmest Quarter	Bio10	°C*10
Precipitation Seasonality (Coefficient of Variation)	Bio15	-
Precipitation of Warmest Quarter	Bio18	Millimeters

The variables were pre-processed using ArcMap and converted to ASCII files for use in Maxent. These variables were checked for multicollinearity using the SDM Toolbox available from http://sdmtoolbox.org/ (Brown 2014). All variables expressed high collinearity (>0.7), and with two variables (Bio15 and Bio6) showing the least collinearity (<0.8). The remaining two variables were used for further analyses (Figures 5-6).

2. Land Use and Land Cover

Data was obtained from the LANDFIRE existing vegetation cover dataset (http://www.landfire.gov/datatool.php) and National Land Cover Database (NLCD) (http://www.mrlc.gov/nlcd2011.php). The LANDFIRE (2010) data was used to create four variables, whereas the NLCD data was used for one variable (Figures 7-10).

Name of Variable	**Source**	**Raster Name**	**Units**
NLCD 2011 Percent Tree Canopy	National Land Cover Database	nlcd	Binary
LANDFIRE Shrub Canopy Cover	LANDFIRE	shrub	Percentage

LANDFIRE Forest Canopy Cover	LANDFIRE	Forest_2	Percentage
Agricultural Land	LANDFIRE	ag	Binary
Development	LANDFIRE	dev	Percentage

The NLCD layer was created by reclassifying the layer as follows:

NLCD Type	**Value**
Deciduous Forest/Evergreen Forest/Mixed Forest	1
sRemaining Types	0

The Agricultural Lands layer was created by reclassifying the Existing Vegetation Cover data from LANDFIRE as follows:

Existing Vegetation Cover Type	**Value**
NASS Row Crop-Close Crop/NASS Row Crop/ NASS Aquaculture/NASS Vineyard	1
Remaining Types	0

The Shrub Cover layer and Forest Cover layer were created by reclassifying based on the midpoints of the percent cover estimates for each level of vegetation (Andersen and Beauvais, 2013, as follows:

LANDFIRE Existing Vegetation Cover Type	**Shrub value**
Shrub Cover >= 10 and < 20%	15
Shrub Cover >= 20 and < 30%	25
Shrub Cover >= 30 and < 40%	35
Shrub Cover >= 40 and < 50%	45
Shrub Cover >= 50 and < 60%	55
Shrub Cover >= 60 and < 70%	65
Shrub Cover >= 70 and < 80%	75
Shrub Cover >= 80 and < 90%	85
Shrub Cover >= 90 and < 100%	95

LANDFIRE Existing Vegetation Cover Type	**Forest value**
Forest Cover >= 10 and < 20%	15
Forest Cover >= 20 and < 30%	25
Forest Cover >= 30 and < 40%	35
Forest Cover >= 40 and < 50%	45
Forest Cover >= 50 and < 60%	55
Forest Cover >= 60 and < 70%	65
Forest Cover >= 70 and < 80%	75
Forest Cover >= 80 and < 90%	85
Forest Cover >= 90 and < 100%	95

In addition, we were interested in looking at the effects of development on the tortoise, so a development layer was created as follows:

LANDFIRE Existing Vegetation Cover Type	**Reclassified Value**
Developed-Low Intensity	1
Developed-Medium Intensity	2
Developed-High Intensity	3
Developed-Roads	3

3. Soil

Soil layers were obtained from the Gridded SSURGO (gSSURGO) database (http://datagateway.nrcs.usda.gov/**).** Percentage Sand and Saturated Hydraulic Conductivity (ability of moisture to move through the soils) were chosen as the two predictor variables based on their use by Andersen and Beauvais, 2013. However, since the Texas tortoise uses pallets and other burrows that are relatively shallow (pallets less than their carapace length), we decided to not use the entire soil layer depth but rather just the surface soil (Auffenberg and Weaver, 1969). The dominant soil component was obtained for the surface soil of each soil map unit with help from NRCS staff (Amanda Bragg, pers. comm). These data layers represent the upper soil surface layer (<30 cm) of the soil and rainfall layer (Figures 11 and 12).

Name of Soil Variable	**Raster Name**	**Units**
Saturated Hydraulic Conductivity	Ksat	Micrometers per second
Total Percent Sand	PercSand	Percentage

All of the above layers were projected in the USA Contiguous Albers Equal Area Conic USGS projection, converted to 30 meter pixel resolution, and then converted into ASCII format for input into Maxent.

Maxent Model Building

Maxent was used for modeling the species' distribution. Initially, two different models were built, one with a subset of location data just from the historic range (228 coordinates), and one with all the location data available (602 coordinates). The smaller model included sampling points that had an approximate accuracy of 15-30 m. The larger model included sampling points that had <5000 m accuracy, but over 80 % of points had 15-30 m accuracy. All 13 environmental layers were then run as a separate model using the smaller number of sampling points, but it was found that most variables were correlated with a Pearson's Correlation Coefficient > 0.7. Subsequent analyses were conducted using just 9 of the 13 variables, having removed the variables that were highly correlated. In summary, the three models in our final comparison were: 1) a model using a subset of coordinates (228) with all environmental variables, 2) a model, using a subset of coordinates (228) with a reduced set of environmental variables, and 3) a model using all coordinates (602) with a reduced set of environmental variables.

Model validation was performed by randomly selecting 25% of the sampling points (sample coordinates used) as a test dataset, to run separately from the remaining 75% of the dataset to be used as training data to build the final model. The Area Under the Curve (AUC) was calculated for the test dataset to predict model accuracy (Fielding and Bell, 1997). Also, Maxent creates background or 'pseudoabsence' points to model distribution of a species. These points help distinguish between areas 'used by' versus those 'available to' the species (Andersen and Beauvais, 2013). Around 10,000 random background points were created for this purpose representing all gradients available to the species.

A jackknife analysis was done on all environmental variables in the models, where one predictor was withheld and the model was refitted. The other model settings used were as follows: Feature types used: hinge linear quadratic; responsecurves: true; jackknife: true; writeclampgrid: false; writemess: false; randomtestpoints: 25; writebackgroundpredictions: true. The agriculture and NLCD-Forest cover layers were treated as categorical layers in Maxent. All output was logistic and in the form of ASCII files. These were then imported into ArcMap for further processing.

ArcGIS Model Building

In addition, a model was built using ArcMap based on a logistic regression run in R (insert R citation). Climatic variables (bio2, bio3, bio10 and bio 18) were found to be highly correlated and were removed from the model. The regression model statement used was as follows: 3.466+(shrub*0.9071)+(nlcd*13.22)+(forest*1.048)-(ag*15.16)-(dev*0.03895)+(bio6*0.02910)+(percsand*0.04335)+(ksat*0.008936)

The model built in ArcMap was not included as part of the study, as an accurate comparison between the two models could not be made and visual inspection of the map produced revealed obvious errors in species distribution predictability. Current literature demonstrates that Maxent is one of the most commonly used and best performing modeling techniques for presence only data (Warren et al., 2011).

Results:

Mendeley Reference Manager

We created a user Group (=Gopherus berlandieri) on the free online Mendeley Reference Manager (http://www.mendeley.com/groups/4224511/gopherus-berlandieri/) to produce a bibliography of all literature pertaining to the Texas tortoise. The Gopherus berlandieri Group currently contains 66 literature citations that directly pertain to the species and is continuously being updated as new literature is found.

Texas Tortoise Surveys

Approximately 17816 miles of roads were surveyed from March to October of 2014, representing the eastern, southeastern (coastal) and parts of the northern and northwestern range of the Texas tortoise (Figure 2). These surveys efforts represent a total of approximately 900-1000 person hours. In total, only 7 tortoises were found during these surveys (Figure 3). Four of these tortoises were road mortalities, while the other three were alive and were marked and released. Other tortoises were found during the same time period outside of the range being surveyed and were included in our final model. Most were suspected to be released pet tortoises or human translocated individuals outside their range, but a few were from the northeastern boundary for the species and may represent wild individuals (Figure 4).

Texas Tortoise Observation Data

Over 700 coordinates were identified from database searches and unpublished data obtained. This was subsequently reduced to a total of 602 coordinates available for modeling based on our accuracy cutoff of <5000 m. The majority of excluded data points were from TXNDD as we could not be certain all coordinates attained our accepted accuracy cutoff of within <5000 m.

Out of 228 points supplied as presence data for modeling, Maxent used only 193 points in the first two models. Similarly, out of 602 points supplied to Maxent in the third model, only 344 points were used for modeling. Out of the 193 points used in the smaller models, 48 points were used for testing and 145 for training and similarly for the larger model, out of 340 total points, 84 points were used for testing and 260 were used for training.

The optimal model chosen, out of the three models built, was decided based on the number of samples used in the analysis and also on the AUC values. Summary of model statistics below:

Model	**Variables included**	**Number of Training Points**	**Number of Test Points**	**Training AUC**	**Test AUC**
1. 228 presence locations with all 13 variables	Bio2, bio3, bio6, bio10, bio15, bio18, ksat, percsand, nlcd, forest, shrub, ag, development.	145	48	0.977	0.979
2. 228 presence locations with 9 variables	Bio6, bio15, ksat, percsand, nlcd, forest, shrub, ag, development.	145	48	0.971	0.970
3. 602 presence locations with 9 variables	**Bio6, bio15, ksat, percsand, nlcd, forest, shrub, ag, development.**	**260**	**84**	**0.950**	**0.947**

One way to measure fit or accuracy of the model produced through Maxent is by a Receiver Operating Characteristic Plots (ROC) (Baldwin, 2009). This ROC plot measures sensitivity and specificity of the data-sensitivity measures how well the data predicts presence, whereas specificity measures correctly predicted absences (Fielding and Bell, 1997). The ROC plot was developed by the use of separate training and testing data. The plot can be read by looking at the Area Under the Curve (AUC); in this case the high AUC of all three models indicates a perfect fit that is better than that expected by random (Baldwin, 2009). Model one, with all 13 variables, showed high multicollinearity and was excluded from further analysis, despite the highest AUC for training and test data amongst all three models. The variables included in model one, such as Mean Temperature of Warmest Quarter, had a high percent contribution, but this might have been skewed by the high multicollinearity. Model two, with a high AUC of 0.971 for training data and 0.970 for test data, was not selected as the final model because of the reduced number of sample points used in modeling and it did not perform significantly better than the more robust model three (Figure 14). Model three, with nearly twice as many sample points, did not have as high an AUC as the other two models and also had points with a higher location uncertainty (approx. <5000 m) (Figures 15 and 16). However, the number of samples used for training and testing was greater and had a more representative sample of the species historic range. It is for these reasons we selected model three as our optimal model for habitat suitability of the Texas tortoise.

The first output produced in Maxent is the analysis of omission/commission that evaluates model performance/bias, as a function of predicted occurrence. It displays the omission rate and predicted area at different thresholds (Young et al., 2011).

Environmental variable importance can be assessed in two different ways. First, Maxent provides the percent contribution and permutation of importance for each variable used in the model. These are calculated by determining the increase in gain by each variable in the model. This can be seen for the selected model in Appendix II. Variables bio15 and bio6 had maximum contribution and permutation importance followed by the developed land variable. The second is a jackknife analysis performed on the variables (Appendix II). This excludes a variable at a time while running the model to estimate performance of each variable, based on gain (Baldwin, 2009). Jackknives on test data, training data and also an AUC on test data were created. These jackknives show that the gain for bio6 is the maximum when looked at in isolation. Similar results can be seen in all jackknives performed. This indicates that bio6/minimum temperature of coldest month, followed by precipitation seasonality/bio15 and then developed lands have highest predictive power for determining distribution of Texas tortoise. The jackknife on test data also shows that percent sand has a reasonably high predictive power. However, when looking at the response curves one can see that more developed areas have higher probability of occurrence of tortoises. This indicates a slight skew in the data as many of our presence locations were collected near roads and inhabited areas due to inherent observer bias for these data.

As stated before, Maxent also produces response curve outputs (Appendix II). The first set of response curves was created by treating each variable in isolation and then averaging the rest of the variables. The second set of response curves was created by developing a Maxent model for each response variable separately. These curves indicate probabilities of occurrence of the tortoise in relation to each variable used in the model.

Discussion:

Mendeley Reference Manager

Mendeley has the potential to benefit future research and create a community of researchers. Currently, we have more than 66 Texas tortoise literature citations on Mendeley, with additional citations being added every day. This can now be used as a common platform to access studies on the species by subsequent researchers. We will continue to update the list as more literature on the species is obtained. One of the features of Mendeley is that Group users (=Gopherus berlandieri) can access and update the citations, allowing for perpetual updating of the literature and enabling networking between researchers involved with the Texas tortoise.

Texas Tortoise Surveys

In the time frame of our study we found a total of 7 tortoises, none in the eastern or northern range. In comparison, the number of iNaturalist observations we obtained for the same time period was just 9, despite there being more people contributing to the database compared to our study. This indicates that either the species is not as prevalent in the eastern range as other areas, or that more intensive surveying needs to be carried out. The results of our road surveys and the data obtained from iNaturalist is disconcerting when compared to historic road survey results from Hamilton (1944), when he observed 16 individuals along 2-3 miles of highway and a similar number just off the highway after a rain on August 3, 1938 in South Texas.

Habitat Suitability Modeling

Looking at the logistic output or the model itself (Figures 15 and 16), we can see areas of high suitability (probability of occurrence >0.6), in many parts of deep South Texas as well as areas further to the north. There are some areas of our model map in deep South Texas where we see linear features of very high suitability as well as in areas of higher human habitation that we interpret as artifacts of tortoises being observed in greater numbers where more people occur. It is interesting to note that there are suitable areas in the eastern region where tortoises are not encountered as frequently, and similarly to the northwest. There are areas outside the suitable habitat where tortoises were found, and this adds claim to the fact that these tortoises were most likely either relocated or longer term "pets" released into the wild (Figure 16).

In conclusion, we can state that despite not finding any tortoises in the far eastern or northern range of the species, our model indicates that suitable habitat patches remain present in these regions. This is also true further outside the range, where we can see areas of potential suitability just north of the generally accepted range in Gonzales, Dewitt, Victoria, Hays, and even Travis County (Figure 17). However, our environmental layers are not representative of ongoing landscape changes due to the high level of fracking operations and their supportive infrastructure in the Eagle Ford Shale of South Texas since 2008, particularly in the eastern range of the Texas tortoise (Figure 18 and 19). We are also unable to account for the effects of significant vehicular traffic increases to Texas tortoises within this same area. These areas will require extensive targeted survey efforts to more clearly validate the accuracy of the model, but extensive surveys are hampered by the large private land holdings that limit access to much of the potential tortoise habitat identified. Another limitation of our surveys was the relatively short amount of time we had for survey validation efforts, being limited to less than a single season of potential tortoise movement hampered detection. In addition, other potential areas of suitability delineated by the model can be explicitly targeted for future studies guided by the current model. This can help add to 'absence' data that in turn will help with reiterative model building and performance.

Our model differs from the model produced in the study by Andersen and Beauvais (2013). First, we used fewer environmental variables (9 versus 13) to reduce multicollinearity, and the values for our soil layers were restricted to the upper 30 cm of soil depth compared to the full horizon depth used by Andersen and Beauvais (2013). Second, we used more presence locations (344 sampling points versus 60 sampling points) and applied 25% of our points as test data to validate our model. Third, the sampling points we used were of higher accuracy at <5000 m versus <8000 m. Fourth, our model also indicates a slightly larger area of suitability, when contrasted to their final model, in the northwestern and eastern areas of the range (Figures 15 and

16). Finally, our selected model does not generate the additional areas of mid-level predicted suitability in areas of north and west Central Texas.

Future model development will include the use of more presence locations as more coordinates are obtained for the species from our data sources and future surveys. These models will subsequently incorporate the coordinates from TXNDD as part of an examination of observation point accuracy influence on the overall model performance. To do this, we plan to classify coordinates based on their accuracy and incorporate them into models, starting with highest accuracy points, iteratively.

We have identified the Ecological Modeling Systems (EMST) of Texas layers, available from Texas Parks and Wildlife, as potential vegetation type data that can be used as an environmental predictor. We also plan to work on identifying other suitable soil data layers from the extensive SSURGO dataset. In addition, additional model selection methods will be used to determine the best-fit model with the future data we acquire.

The greatest success of our model is in its ability to enable detection for areas of greatest concern and the factors impacting the species so that conservation management efforts can be engaged in the remaining suitable habitat patches. However, caution must be exercised when using such graphical depictions of habitat suitability due to contemporary impacts on the landscape that are not reflected in these models. We have illustrated several concurrent anthropogenic landscape alteration activities within the range of the Texas tortoise that are rapidly changing the habitat of South Texas. Until we are able to gain a better understanding of the effects of ongoing and increasing disturbances in the region, and considering our very low number of observations under extensive survey effort, it is imperative that continued progress be made toward the protection of the Texas tortoise.

Literature Cited:

Andersen, M. D. and Beauvais G. P.. 2013. Predictive Distribution Modeling of Species of Greatest Conservation Need in Texas. Report prepared by the Wyoming Natural Diversity Database, Laramie Wyoming, for the Texas Natural Diversity Database, Texas Parks and Wildlife Department, Austin, TX. 31.

Auffenberg, W. and Weaver, W.G. Jr. 1969. *Gopherus berlandieri* in southeastern Texas. Bulletin Florida State Museum 13:141-203.

Agassiz, L. 1857. Contribution to the natural history of the United States of America. Boston Little, Brown and Co. 196.

Baldwin, R. A. 2009. Use of maximum entropy modeling in wildlife research. Entropy 11(4): 854-866.

Brown, J. L. 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods in Ecology and Evolution.

Bury, B., and Smith, E.L. 1986. Aspects of the ecology and management of the Tortoise *Gopherus berlandieri* at Laguna Atascosa, Texas. The Southwestern Naturalist 31: 387-394.

Dixon, J.R. 2013. Amphibians and reptiles of Texas: With keys, taxonomic synopses, bibliography, and distribution maps (W. L. Moody Jr. Natural History Series). Texas A&M University Press. Third Edition.

ESRI 2013. ArcGIS Desktop: Release 10.2. Redlands, CA: Environmental Systems Research Institute.

Fielding, A. H., and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation 24(01): 38-49.

Gillen, J. L., and Kiviat, E. 2012. Environmental reviews and case studies: Hydraulic fracturing threats to species with restricted geographic ranges in the Eastern United States. Environmental Practice 14: 320-331.

Grendár, M., and Grendár, M. (2001). Maximum entropy: Clearing up mysteries. Entropy 3(2): 58-63.

Groombridge, M. 1982. The IUCN Amphibia-Reptilia Red Data Book Part 1: *Testudines, Crocodylia, Rynchocephalia*. Gland, Switzerland IUCN. 426.

Hamilton, R.D. 1944. Notes on mating and migration in berlandier's turtle. Copeia 1944: 62.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hirzel, A. H., Helfer, V., and Metral, F. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145: 111-121.

Jahrsdoerfer, S. E., and Leslie Jr, D. M. 1988. Tamaulipan brushland of the Lower Rio Grande Valley of South Texas: Description, human impacts, and management options. Oklahoma Cooperative Fish and Wildlife Research Unit Stillwater.

Judd, F. W. and Rose, F. L. 1983. Population structure, density and movements of the Texas

Tortoise *Gopherus berlandieri.* The Southwestern Naturalist. 28: 387-398.

Judd, F. W. and Rose, F. L. 1989. Egg production by the Texas Tortoise, *Gopherus berlandieri*, in southern Texas. Copeia. 588-596.

Palma, L., Beja, P., and Rodrigues, M. 1999. The use of sighting data to analyze Iberian lynx habitat and distribution. Journal of Applied Ecology 36:812–824.

Phillips, S. J., Dudík, M., and Schapire, R. E. 2004. A maximum entropy approach to species distribution modeling. In Proceedings of the twenty-first international conference on Machine learning (p. 83). ACM.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190(3): 231-259.

Ramirez, J.R. 1986. Water development projects in the Rio Grande and their relationships to the Santa Ana and Rio Grande Valley National Wildlife Refuges. A report prepared for Santa Ana and Rio Grande Valley National Wildlife Refuges, Alamo, Texas; by U.S. Fish and Wildlife Service, Ecological Services, Corpus Christi Field Office, Corpus Christi, Texas.

Rose, F. L. and Judd, F. W. 2014. The Texas Tortoise: A Natural History. University of Oklahoma Press: Norman.

Rose, F. L. and Judd, F. W. 1982. Biology and status of Berlandier's tortoise *(Gopherus berlandieri*). In: Bury, R. B. (Ed.). North American Tortoises: Conservation and Ecology U.S.D.I. Fish and Wildlife Service Wildlife Research Report 12.

Rose, F. L. and Judd, F. W. 1975. Activity and home range size of the Texas Tortoise, *Gopherus berlandieri*, in south Texas. Herpetologica 31:448-456

Sanchez-Zapata, J.A., and Calvo, J.F. 1999. Raptor distribution in relation to landscape composition in semi-arid Mediterranean habitats. Journal of Applied Ecology 36:254–262.

Warren, D. L., and Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications 21(2): 335-342.

Young, N., Carter, L., and Evangelista, P. 2011. A MaxEnt Model v3. 3.3 e Tutorial (ArcGIS v10). Natural Resource Ecology Laboratory at Colorado State University and the National Institute of Invasive Species Science, Colorado, USA.

Appendix I - Figures

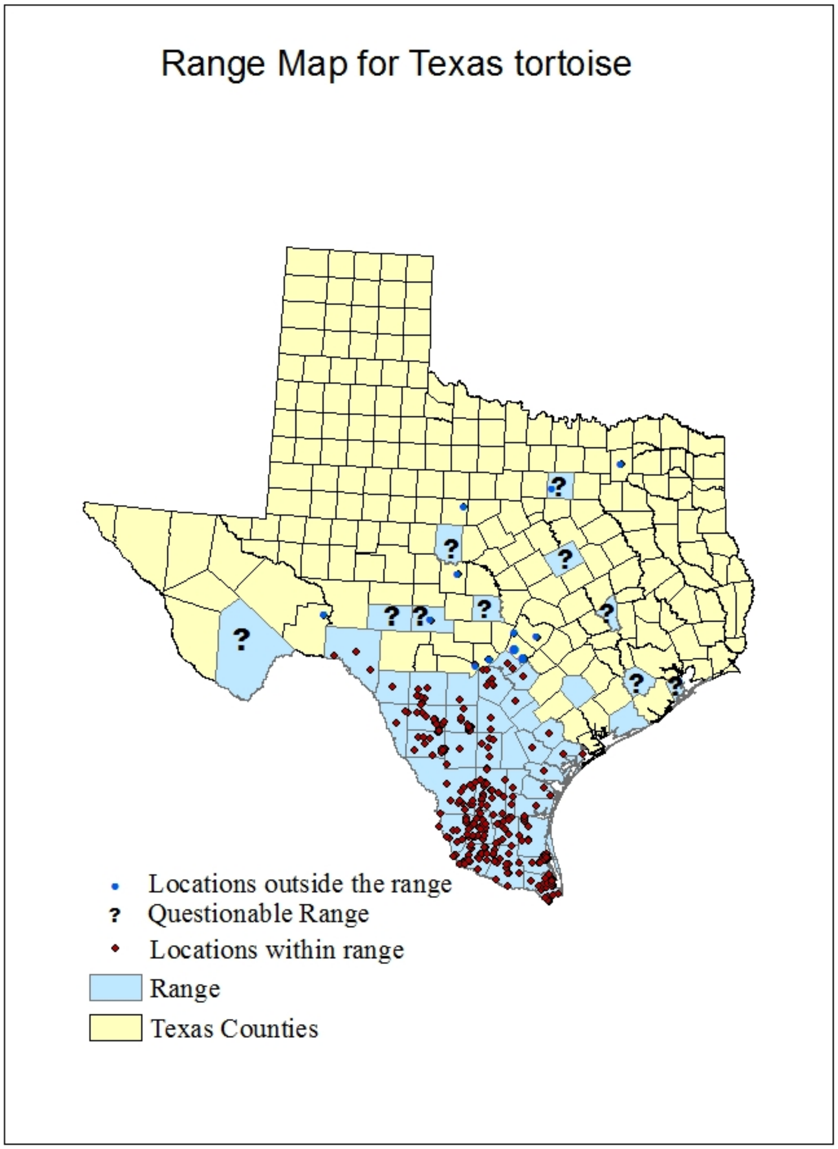


Figure 1: Texas map depicting the historic and current range for Texas tortoise, *G. berlandieri,* with location data within the range and outside. Locations outside might indicate potential relocations or pet tortoises released into the wild.

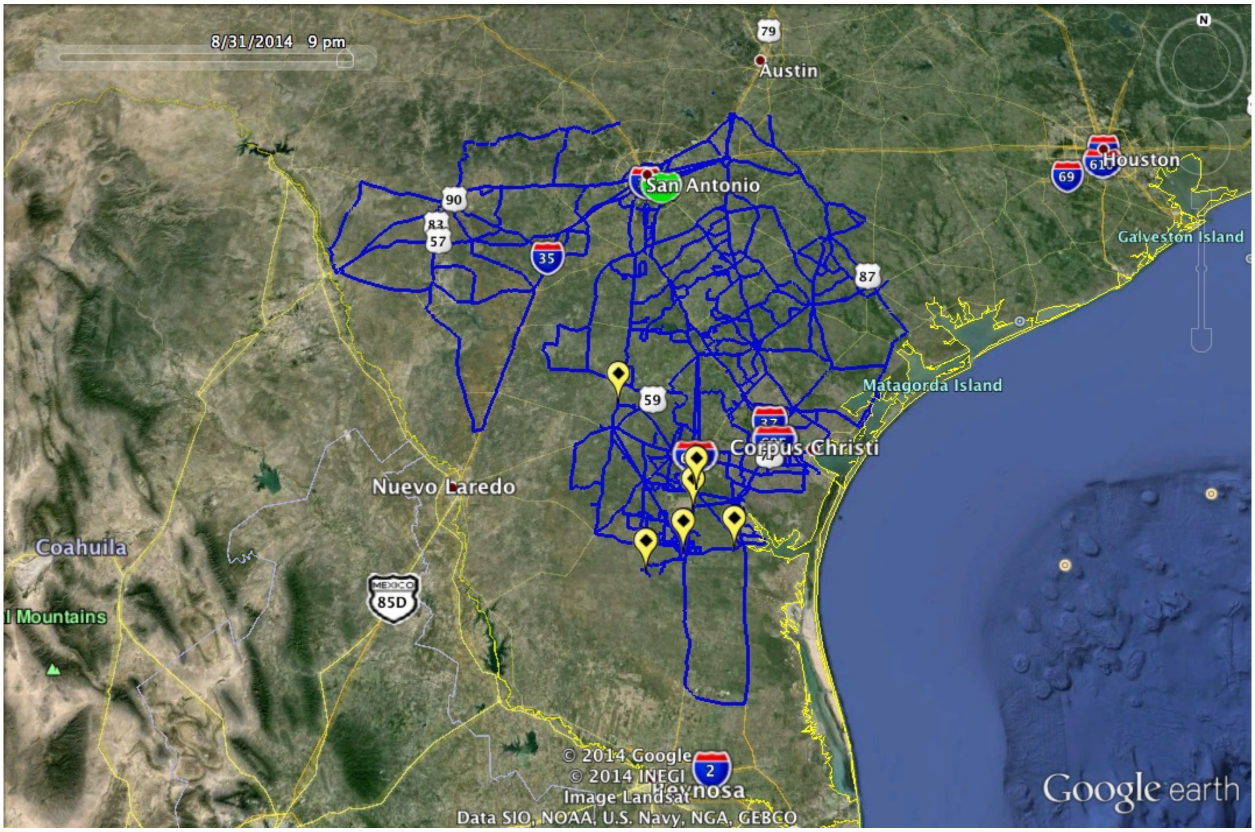


Figure 2: Tortoise road survey routes conduced from March-October 2014. Routes are displayed in blue and tortoises found during the surveys are depicted by yellow.

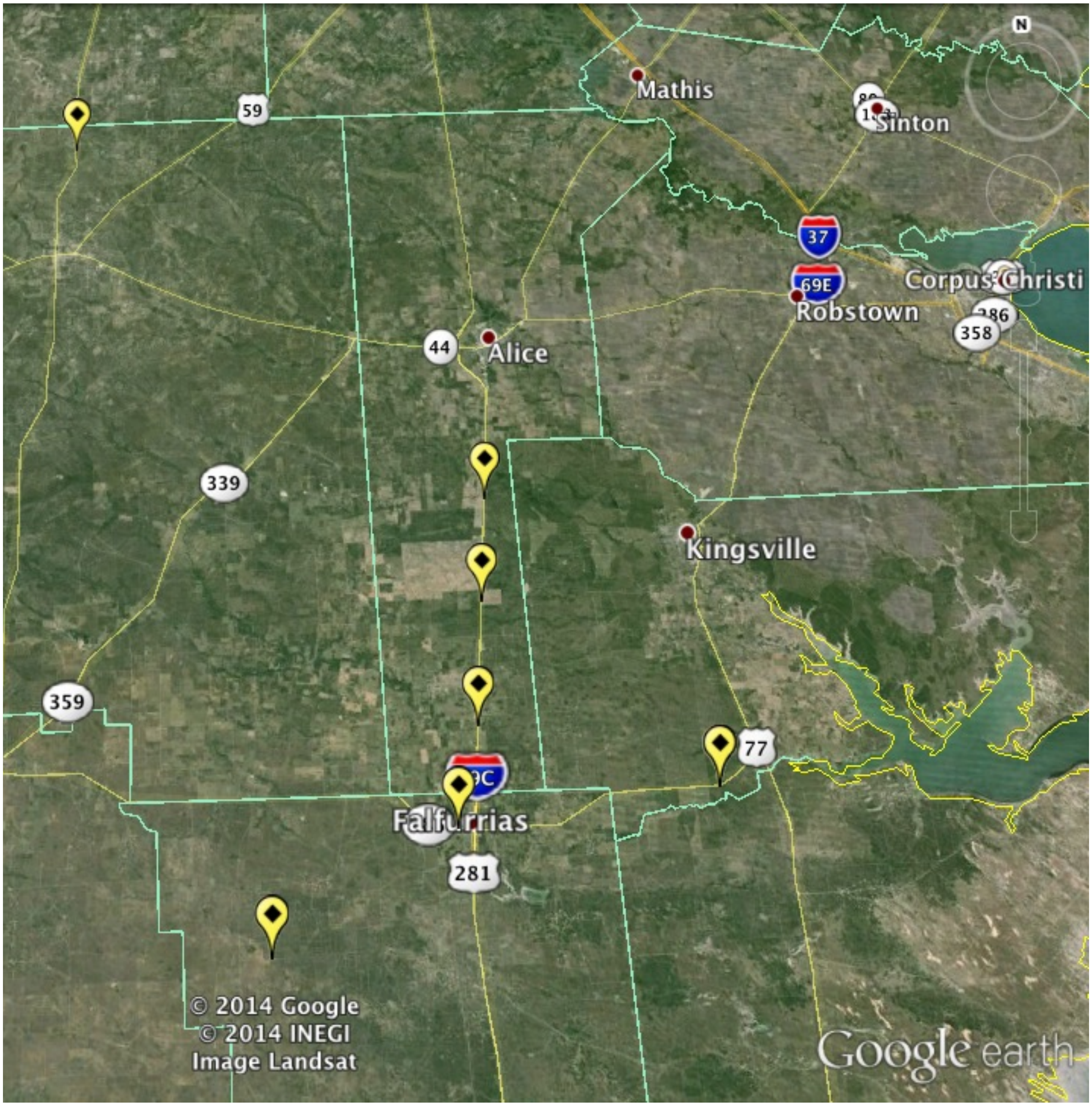
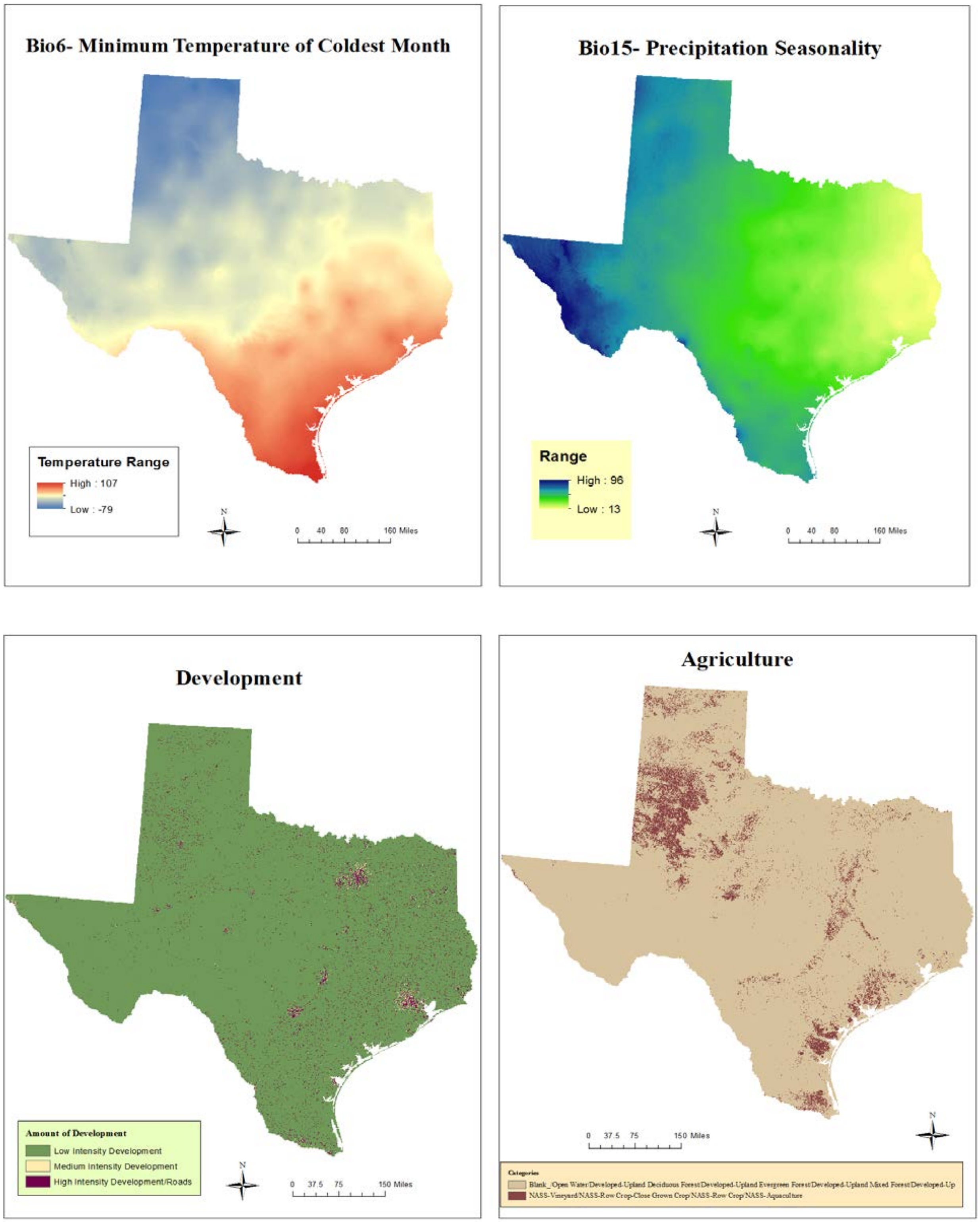
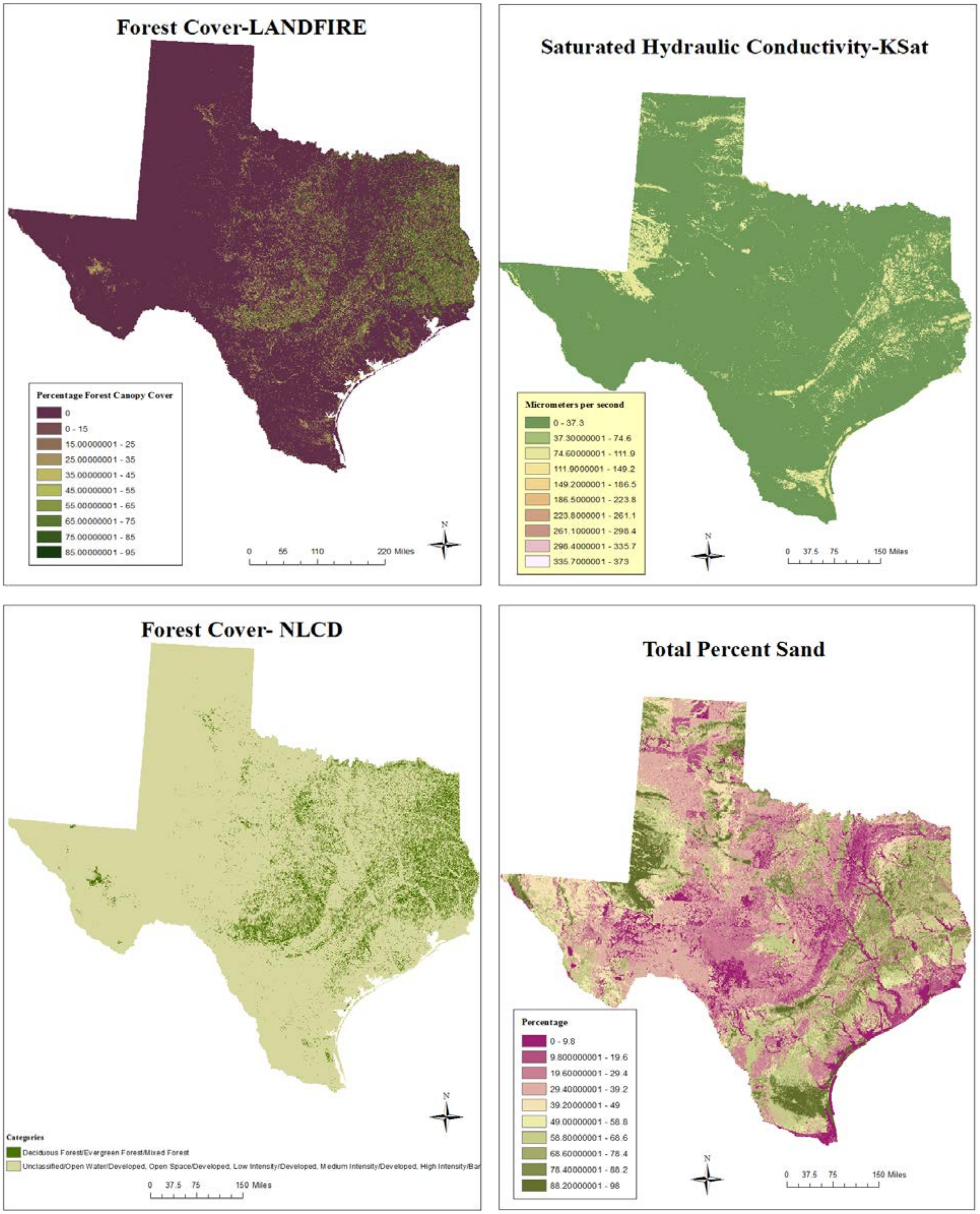


Figure 3: Tortoises found during road surveys conducted from March-October 2014. All tortoises were found east of Highway 16 and around or just north of Highway 285.

Figure 4: Tortoise found in a park in San Antonio. Markings and body condition indicate that the tortoise was a pet.



Figures 5-8: Final Bioclim variables and land use variables used as environmental layers in Maxent.



Figures 9-12: Final Land Cover variables and Soil variables used as environmental layers in Maxent.

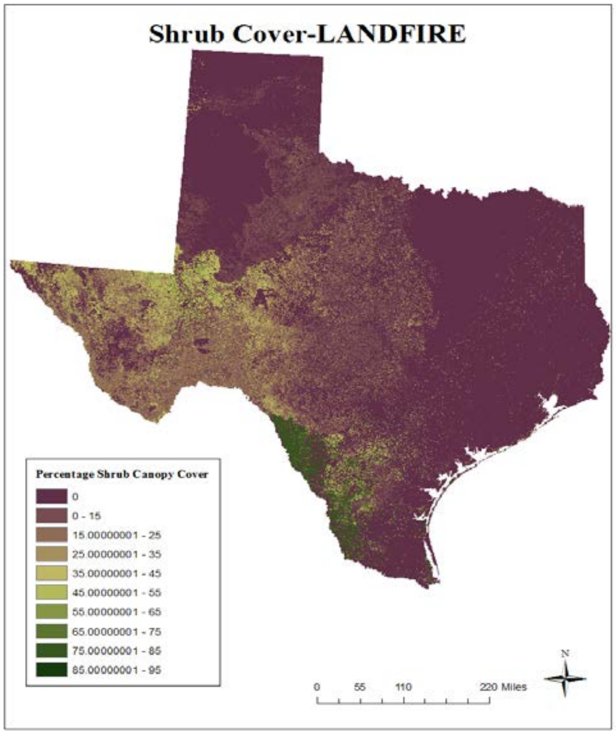


Figure 13: Final Land Cover variable used as an environmental layer in Maxent.

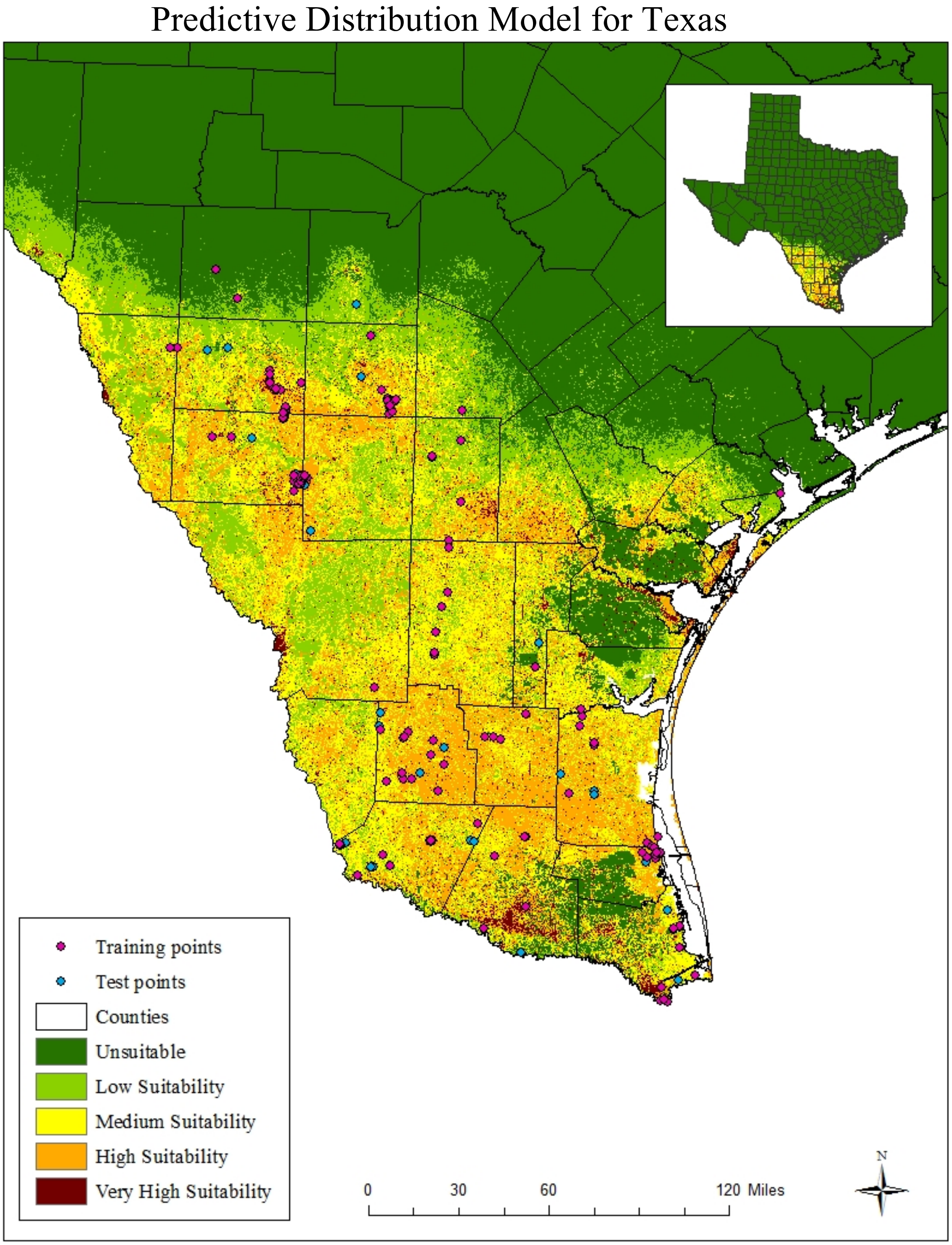


Figure 14: Map of Texas depicting areas of probable presence or suitability for Texas tortoise based on model 2 built in Maxent. 193 samples were used to build this model, representing a smaller subset of points as compared to the model selected with 344 points.

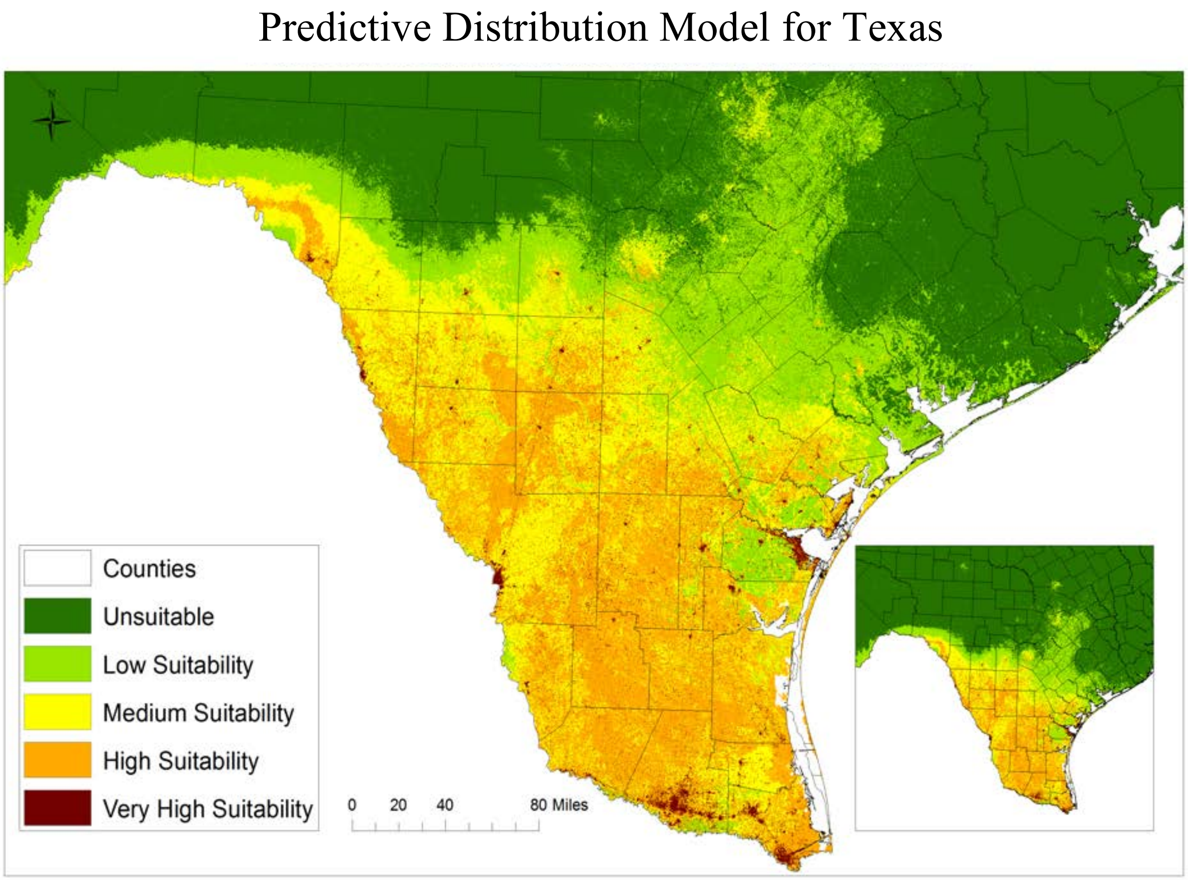


Figure 15: Map of Texas depicting areas of probable presence or suitability for Texas tortoise based on model 3 (selected model) built in Maxent. 344 sample points were used to build this model (260 points for training and 84 for testing and/or model validation). Areas of high suitability are in red/orange and can be seen in the south most parts of Texas as well as in some areas further north. Suitable areas are also present outside the range of the species.

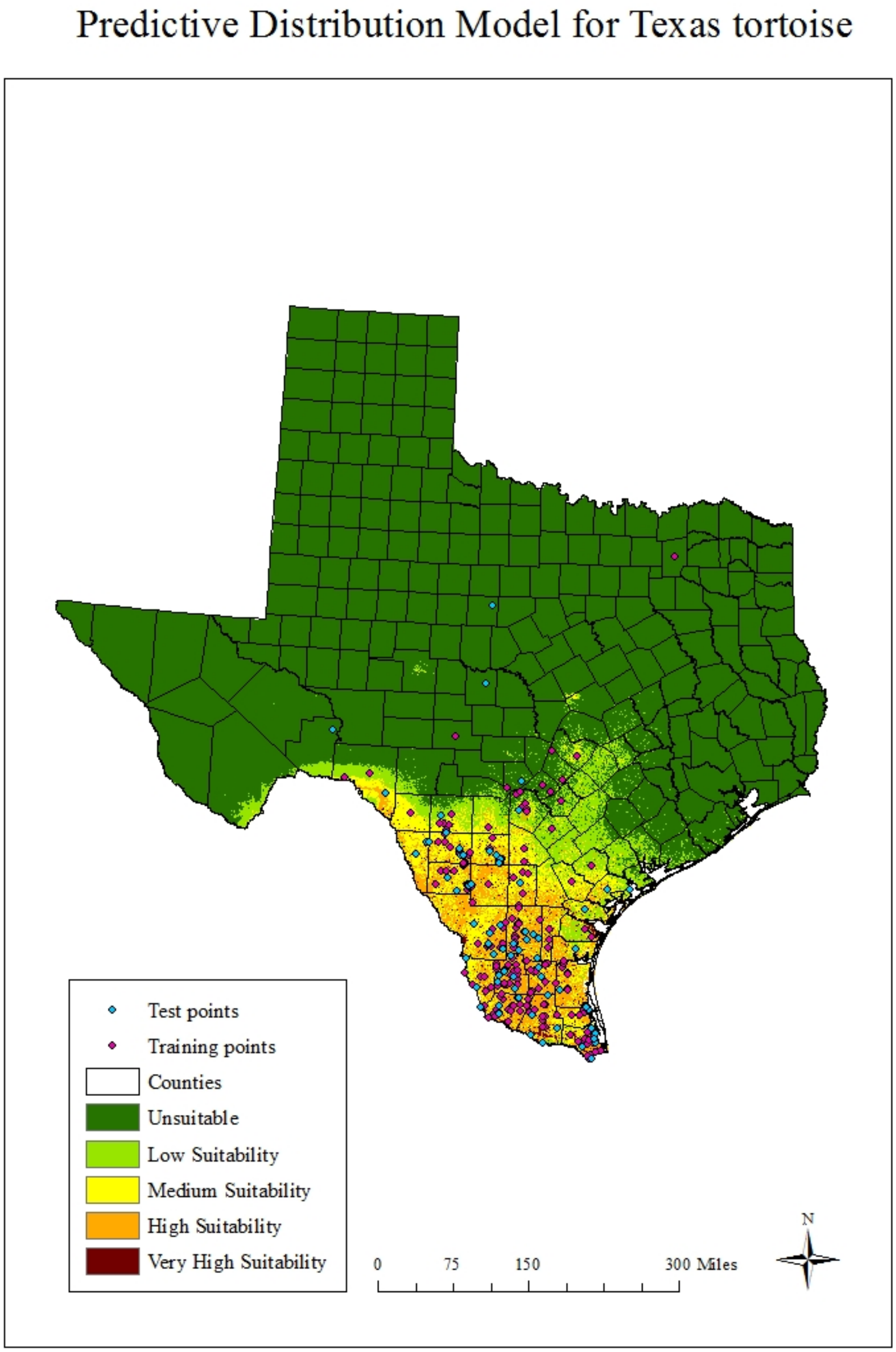


Figure 16: Map of Texas depicting areas of probable presence or suitability for Texas tortoise based on model 3 (selected model) built in Maxent. 344 sample points were used to build this model. 260 points (in magenta) were used for training and 84 (in blue) were used for testing and/or model validation.

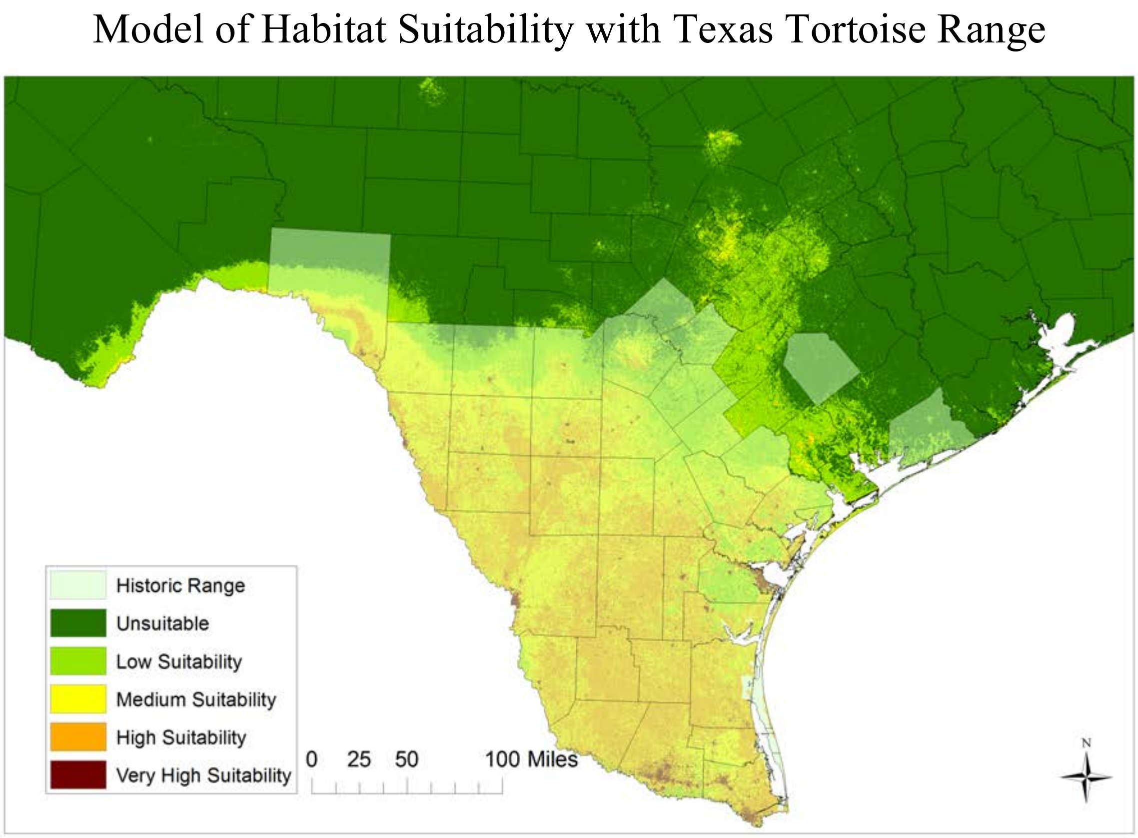
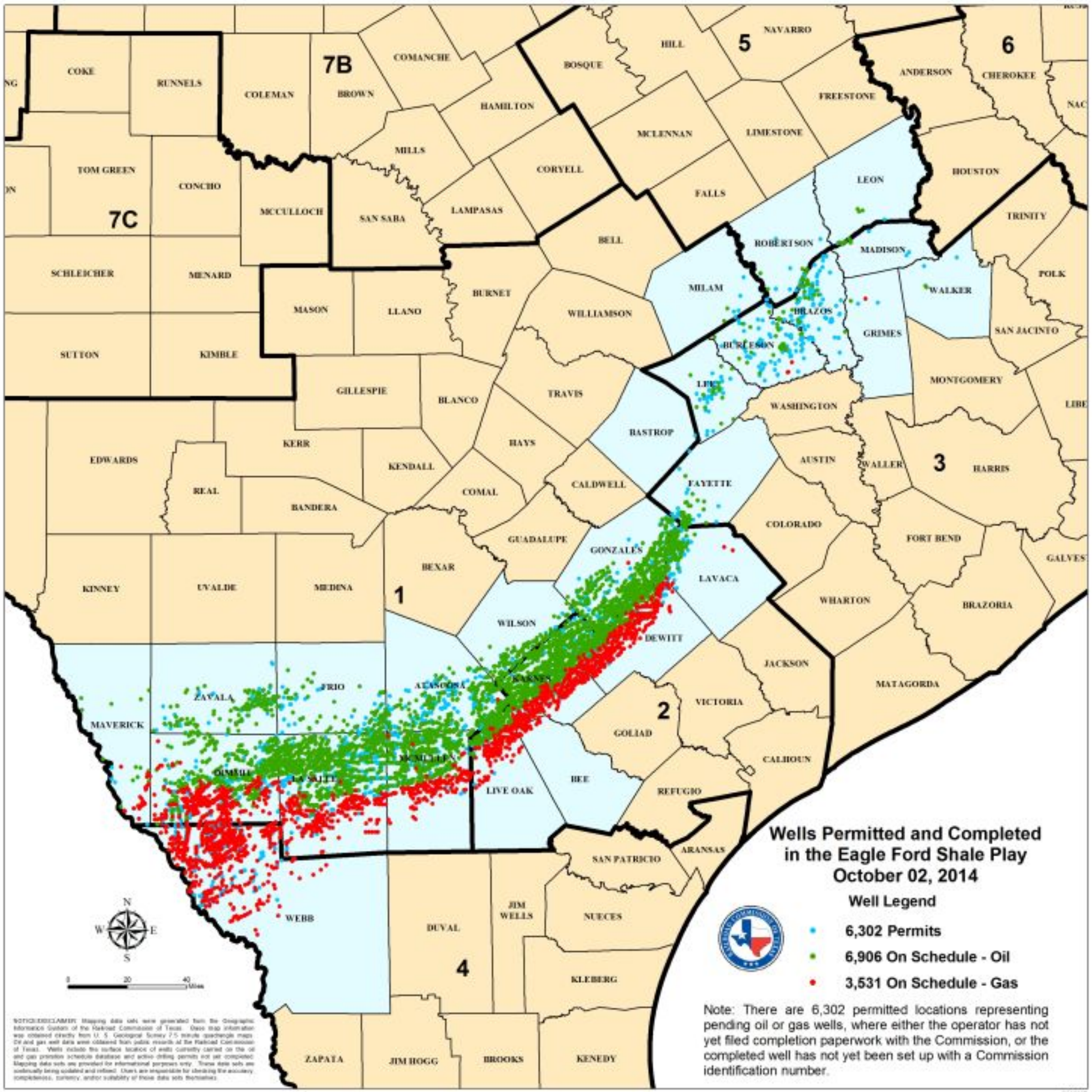


Figure 17: Map of habitat suitability with the Texas tortoise historical range overlaid. Areas of suitability outside this range are visible.



Wells Permitted and Completed in the Eagle Ford Shale Play October 02, 2014

Well Legend

- ● 6,302 Permits
- ● 6,906 On Schedule - Oil
- ● 3,531 On Schedule - Gas

Note: There are 6,302 permitted locations representing pending oil or gas wells, where either the operator has not yet filed completion paperwork with the Commission, or the completed well has not yet been set up with a Commission identification number.

Figure 18: Map of the Eagle Ford Shale hydraulic fracturing wells permitted and completed as of October 2, 2014. Map acquired from the Railroad Commission of Texas Eagle Ford Shale Information webpage (http://www.rrc.state.tx.us/oil-gas/major-oil-gas-formations/eagle-ford-shale/).

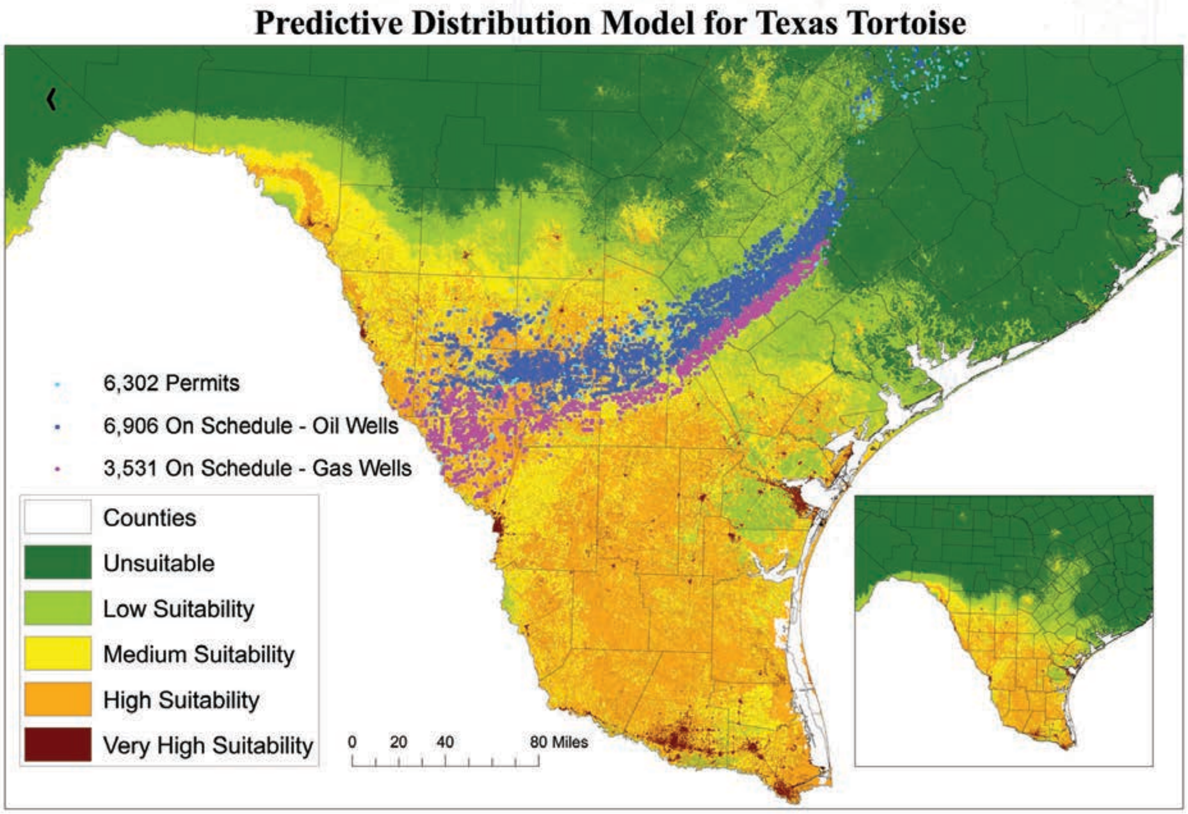
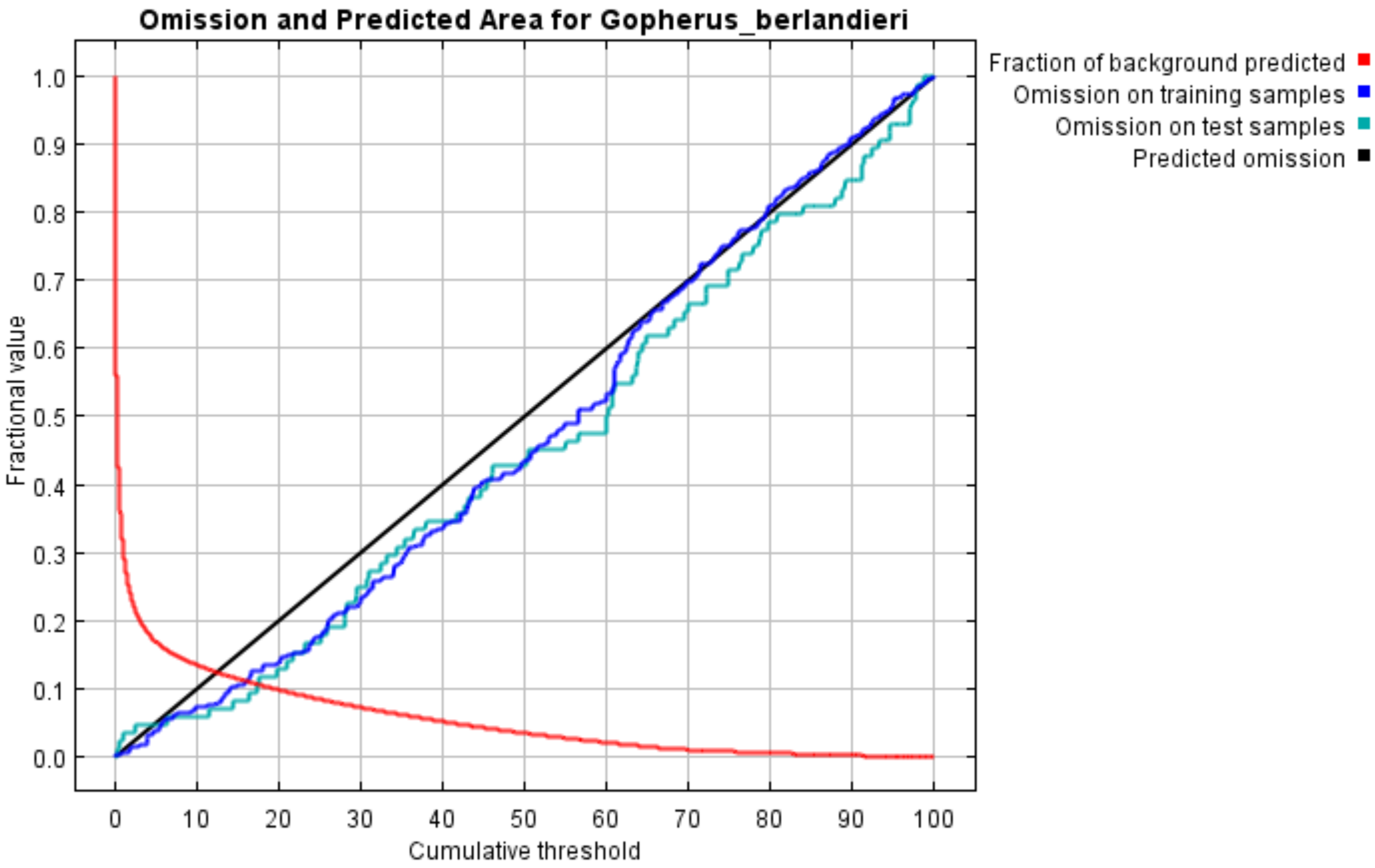


Figure 19: Hybrid map of Texas depicting areas of probable presence or suitability for Texas tortoise based on model 3 (selected model) with Eagle Ford Shale hydraulic fracturing wells permitted and completed as of October 2, 2014 superimposed.

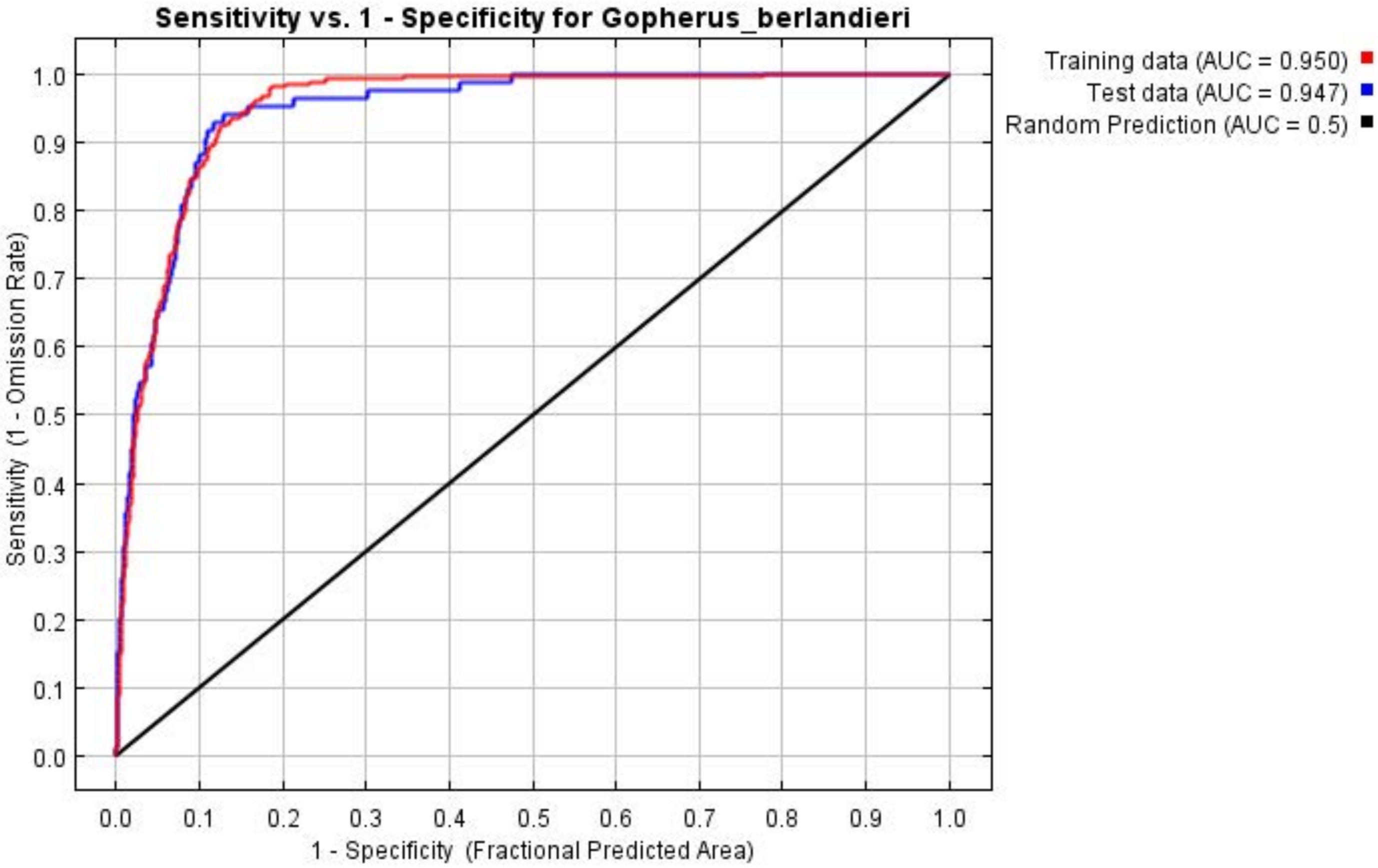
Appendix II – Maxent Output

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission. This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.942 rather than 1; in practice the test AUC may exceed this bound.



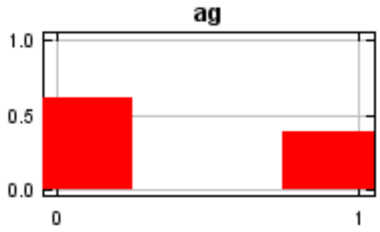
Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

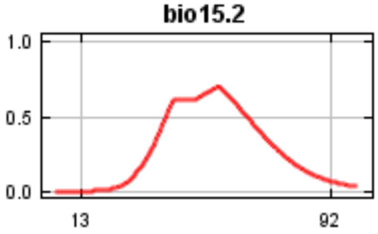
Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.013	Fixed cumulative value 1	0.300	0.008	0.036	1.668E-40
5.000	0.117	Fixed cumulative value 5	0.169	0.038	0.048	0E0
10.000	0.231	Fixed cumulative value 10	0.135	0.073	0.060	0E0

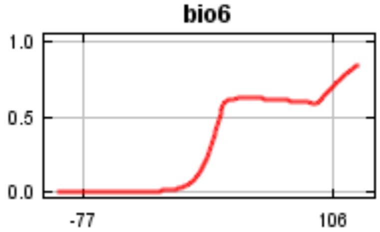
0.009	0.000	Minimum training presence	0.776	0.000	0.000	4.34E-7
14.169	0.285	10 percentile training presence	0.118	0.100	0.071	0E0
16.128	0.302	Equal training sensitivity and specificity	0.111	0.112	0.083	0E0
12.305	0.261	Maximum training sensitivity plus specificity	0.125	0.077	0.071	0E0
17.476	0.311	Equal test sensitivity and specificity	0.107	0.127	0.107	0E0
14.449	0.289	Maximum test sensitivity plus specificity	0.117	0.104	0.071	0E0
1.618	0.024	Balance training omission, predicted area and threshold value	0.251	0.008	0.036	1.213E-51
8.299	0.208	Equate entropy of thresholded and original distributions	0.144	0.065	0.060	0E0

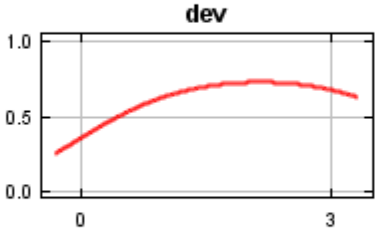
Response curves

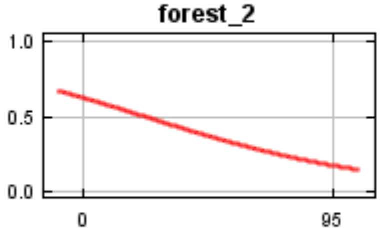
These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

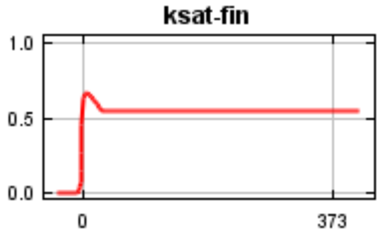

ag

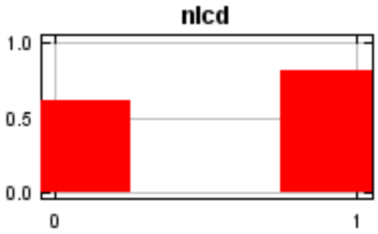

bio15.2

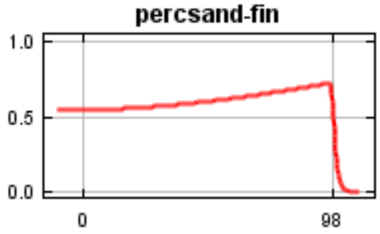

bio6

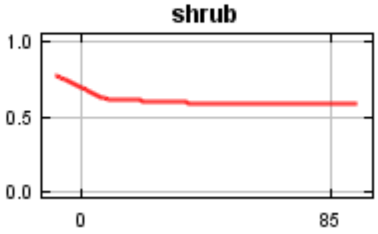

dev


forest_2

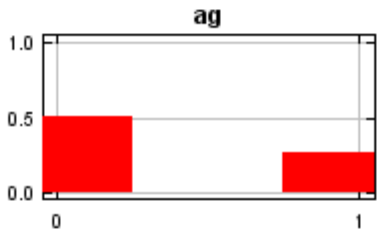

ksat-fin

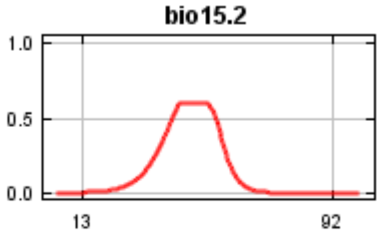

nlcd

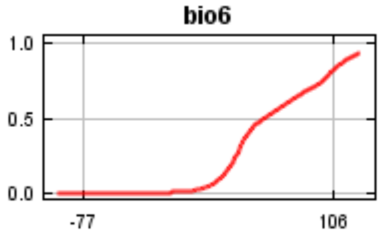

percsand-fin

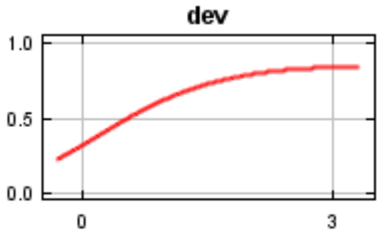

shrub

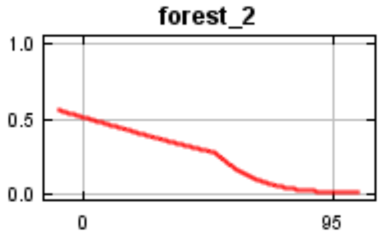
In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.

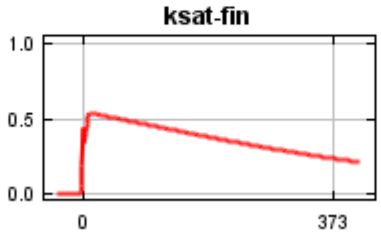


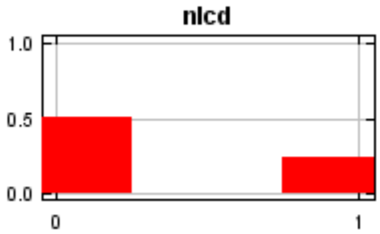


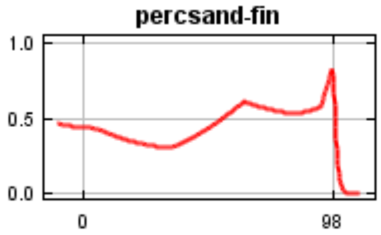


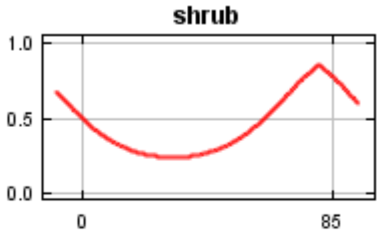










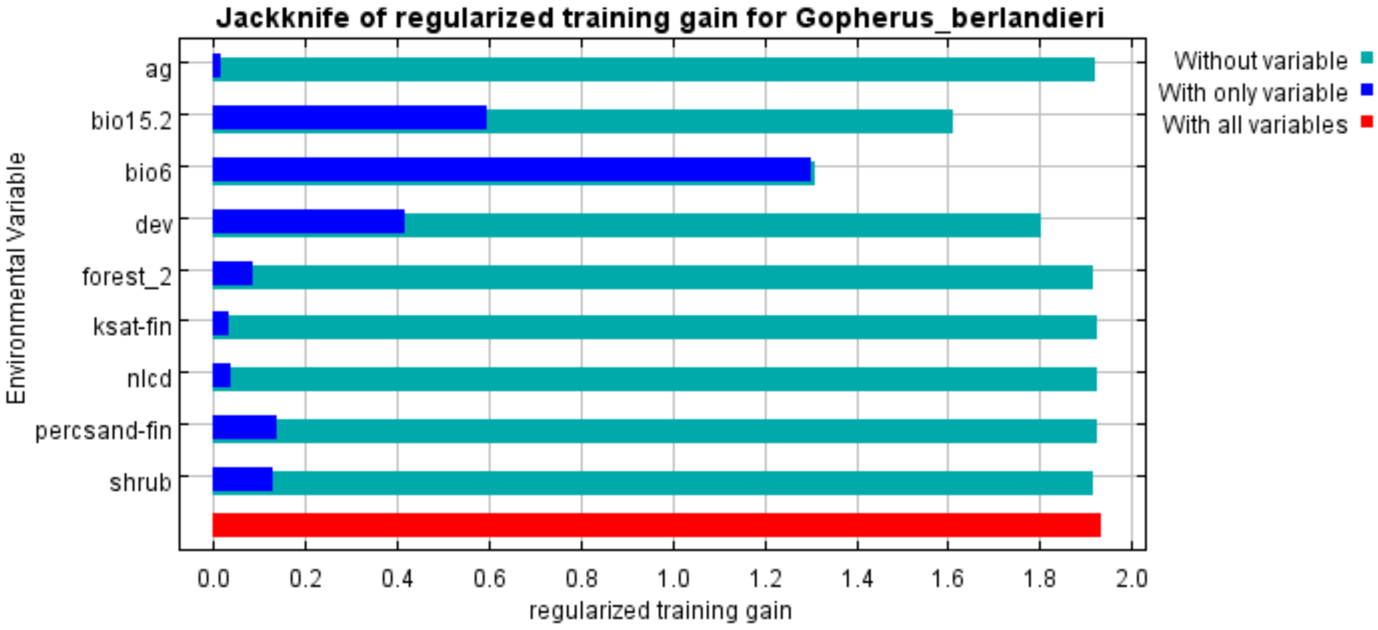


Analysis of variable contributions

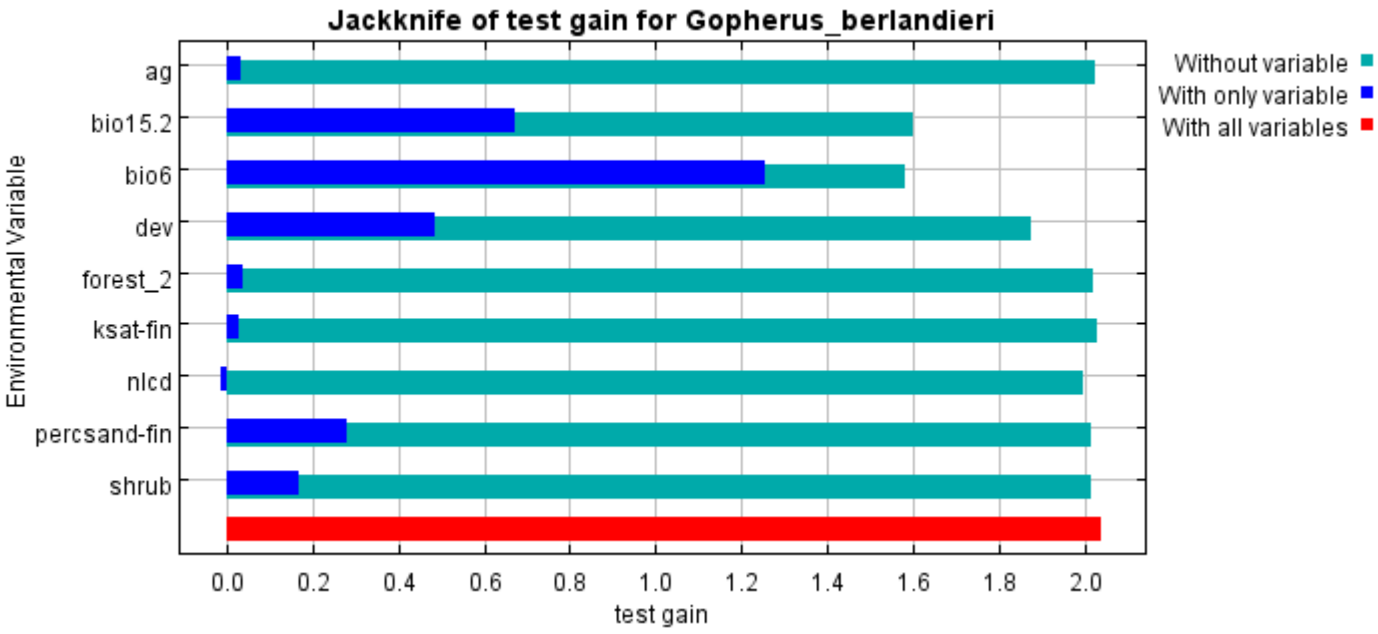
The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution	Permutation importance
bio6	64.1	66.5
bio15.2	19.7	26.5
dev	12.8	1.7
ag	1.1	0.8
ksat-fin	1	0.7
forest_2	0.6	1.4
shrub	0.4	0.5
nlcd	0.3	1
percsand-fin	0.2	0.9

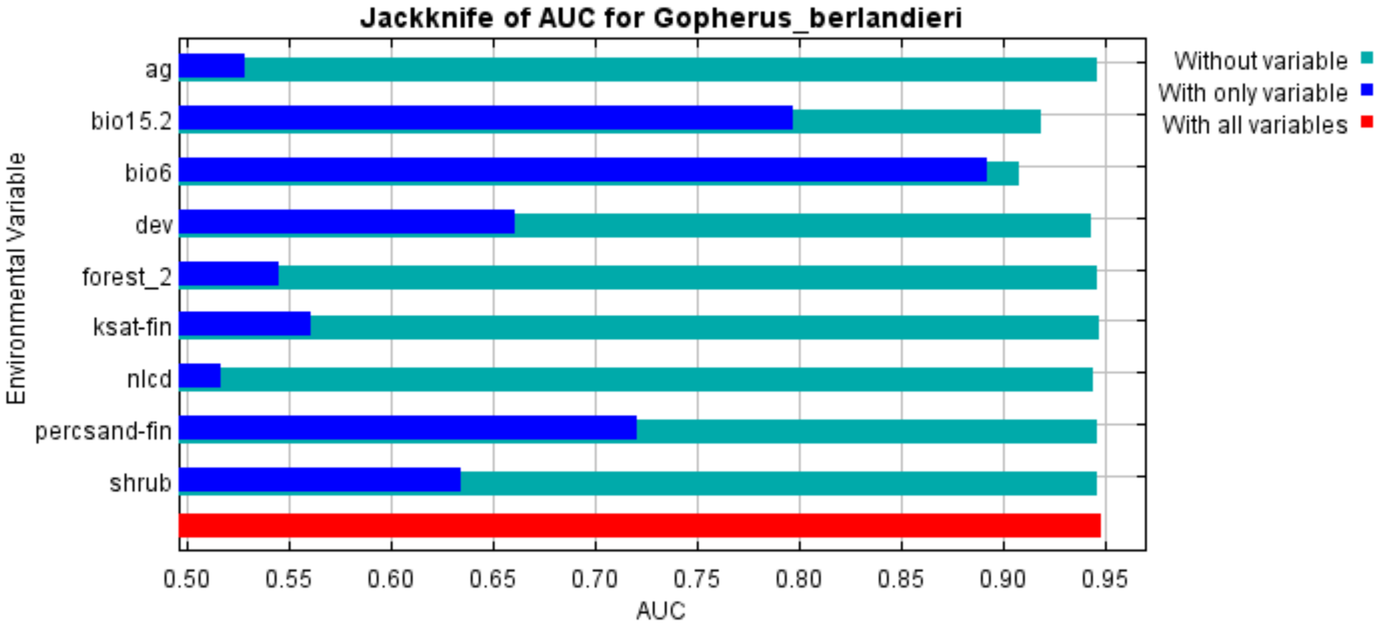
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is bio6, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is bio6, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

Regularized training gain is 1.931, training AUC is 0.950, unregularized training gain is 2.032.
Unregularized test gain is 2.034.
Test AUC is 0.947, standard deviation is 0.009 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2).
Algorithm terminated after 500 iterations (2 seconds).

The follow settings were used during the run:
260 presence records used for training, 84 for testing.
10243 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used: ag(categorical) bio15.2 bio6 dev forest_2 ksat-fin nlcd(categorical) percsand-fin shrub
Regularization values: linear/quadratic/product: 0.050, categorical: 0.250, threshold: 1.000, hinge: 0.500
Feature types used: hinge linear quadratic
responsecurves: true
jackknife: true
outputfiletype: bil
outputdirectory: C:\Users\Bufo Win\Desktop\Anjana\converted layers\FINAL\results\TOT-LOG-25%-156
samplesfile: C:\Users\Bufo Win\Desktop\Anjana\converted layers\FINAL\samples\total.csv
environmentallayers: C:\Users\Bufo Win\Desktop\Anjana\converted layers\FINAL\maxent.cache

```
writeclampgrid: false
writemess: false
randomtestpoints: 25
writebackgroundpredictions: true
product: false
threshold: false
writeplotdata: true
autofeature: false
doclamp: false
allowpartialdata: true
Command line used:

Command line to repeat this species model: java density.MaxEnt nowarnings noprefixes -E "" -E
Gopherus_berlandieri responsecurves jackknife outputfiletype=bil
"outputdirectory=C:\Users\Bufo Win\Desktop\Anjana\converted layers\FINAL\results\TOT-
LOG-25%-156" "samplesfile=C:\Users\Bufo Win\Desktop\Anjana\converted
layers\FINAL\samples\total.csv" "environmentallayers=C:\Users\Bufo
Win\Desktop\Anjana\converted layers\FINAL\maxent.cache" nowriteclampgrid nowritemess
randomtestpoints=25 writebackgroundpredictions noproduct nothreshold writeplotdata
noautofeature nodoclamp allowpartialdata -N ag2 -N bio10 -N bio102 -N bio15 -N bio18 -N
bio18.2 -N bio2 -N bio3 -N ksat2 -N percsand_301 -t ag -t nlcd
```