2018

Addressing Survey Bias: Maxent Models and Public Archaeology at the Davy Crockett National Forest

Juanita D. Garcia
National Forests and Grasslands in Texas, United States Forest Service, jdgarcia@fs.fed.us

Robert Z. Selden Jr.
Center for Regional Heritage Research, Stephen F. Austin State University, zselden@sfasu.edu

David A. Foxe
National Forests and Grasslands in Texas, United States Forest Service, dafoxe@fs.fed.us

United States Forest Service
National Forests and Grasslands in Texas

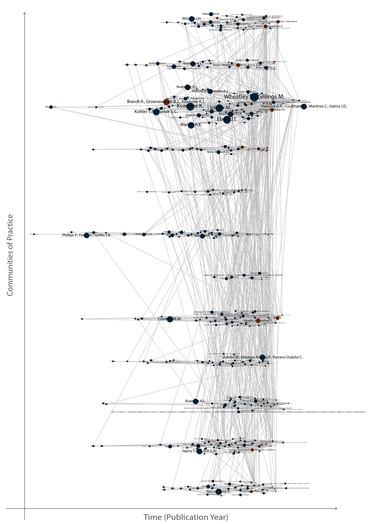
Juanita D. Garcia; Robert Z. Selden, Jr.; and David A. Foxe

Addressing Survey Bias: Maxent Models and Public Archaeology at the Davy Crockett National Forest

Over the past 30+ years, compliance-based archaeological projects conducted on the Davy Crockett National Forest have yielded an impressive contribution to the archaeological record. However, those efforts were limited to specific survey areas based upon specific needs. To begin to address survey bias and the impact that it can have on site probability models, a stratified random sample of 50 locations was generated for the forest. In early 2018, a public archaeology (Passport-in-Time) project was used to engage interested volunteers who assisted in testing these sample locations. These results, as well as other new data, were added to the maxent site probability model in advance of running a second iteration. A second call for volunteers in early 2019 will aid with site relocation and delineation of known resources, as well as an additional test using a new stratified random sample.

In February and March of 2018, a Passport-in-Time (PIT) project was conducted to test the first iteration of a suite of new niche models designed with the intent of predicting historic and prehistoric archaeological site locations in the Davy Crockett National Forest (DCNF). The PIT project leveraged collections data to produce a suite of niche models for specific time periods based upon the presence of temporally-diagnostic artifacts. Using a series of data layers associated with environmental variables, site locations, and other information, the current suite of models provides a data-driven method that can be continually refined as new data becomes available, and as we continue to generate and test novel hypotheses on the forests. While a valuable addition to the management and protection of these important resources, the model is imperfect. Additional work is needed to test and refine the model through mitigating bias introduced through 30+ years of linear, block, and compartment-level surveys, achieved through shovel-testing a stratified random sample of locations throughout the DCNF. With this test, that undertaking begins; however, to fully address survey bias, additional testing—beyond this initial effort—will be required.

Production of the DCNF model follows a recent systematic study of predictive modeling literature that enlisted scientometrics to identify communities of practice in peer reviewed research articles that use archaeological applications of predictive modeling or predictive modeling techniques. Similar networks were constructed that are focused on applications of the three R packages used to generate the DCNF models (maxent, ENMeval, and ENMTools), and are not limited to archaeology. These networks inform the continued development of the DCNF models, and have aided in the iterative refinement of niche models produced throughout this research program.



Citation network for publications (maroon) and cited works (blue) associated with predictive modeling in archaeology. Findings from that study aided in the selection of the software and methodological approaches used in this study.

Locations used to begin testing the model were generated as a stratified random sample (stratified using DCNF compartments). Nine shovel tests were placed at each location, using the assigned lat/long coordinate as the centroid and eight additional shovel tests 20m in cardinal and ordinal directions from the centroid.

Volunteers discovered a previously unrecorded multicomponent site during the first day of testing the niche models.

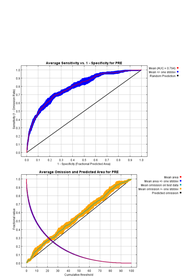
Shovel tests ranged in depth from 10-100cm.

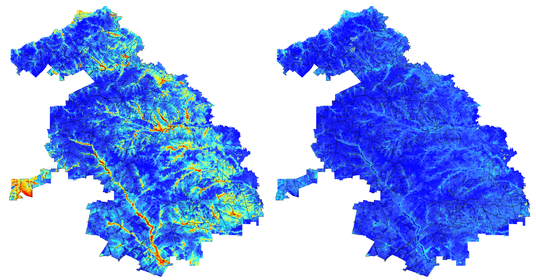
Volunteers learned how to test each location using the standard suite of protocols used in archaeological survey.

Among those challenges associated with creating the models has been the identification of deficiencies in the data. One of those deficiencies occurred in the stream (or blue-line) shapefile associated with the local waterways and drainages, which was off by up to 30 meters in some locations. To address this deficiency, a new stream shapefile was created using a digital elevation model coupled with the Strahler method to order the resulting stream layer. Another challenge is that the available digital elevation models capture the canopy of the forest; meaning that a freshly cut pine stand appears much lower in elevation than an adjacent mature stand. In late 2016, the Federal Emergency Management Agency collected Light Detection and Ranging (LiDAR) data for the Neches River basin. The NFGT negotiated a higher resolution for the Angelina, Davy Crockett, Sam Houston, and parts of the Sabine National Forests that would be covered by the survey. Bare-earth LiDAR is incorporated in the next iteration of the DCNF model, allowing us to test (1) whether the higher resolution (data) models differ significantly from the first iteration, and (2) whether these data contribute to the production of a more accurate model.

The current iteration of the DCNF models holds significant potential beyond the development of a heuristic niche model. Using diagnostic artifacts, the models can be further refined to investigate novel research questions. Might Archaic-era hunter-gatherer populations, for instance, have preferred a different suite of geographic and landscape-based site selection criteria than the horticulturalists and agriculturalists associated with Woodland and Caddo populations? There are also diagnostic artifacts assumed to transcend the hunter-gatherer-to-horticulturalist or agriculturalist transition in East Texas (Kent and Gary dart points in particular) that are relatively abundant in the NFGT collections.

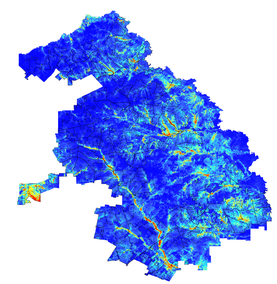
In addition to the niche models, a geometric morphometric study is underway that asks whether the shape of Gary and Kent dart points differs at those sites where only lithics were recovered (assumed to date to the Archaic period) versus those where lithics and ceramics were recovered on the DCNF (assumed to date to the Woodland period), and a hunter-gatherer-to-horticulturalist and/or agriculturalist niche model affords additional depth to that discussion. Using the ASCII files produced for the machine learning tools, we are beginning to test these hypotheses and others, providing a means of assessing complementary, and in some cases hierarchically-nested, research questions that were previously out of reach.



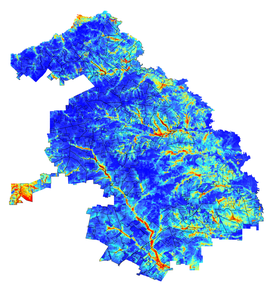


mean - prehistoric

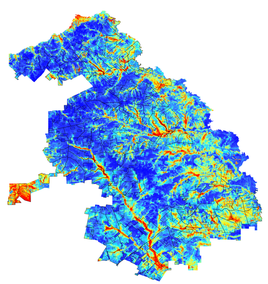
standard deviation - prehistoric



minimum - prehistoric



median - prehistoric



maximum - prehistoric

Acknowledgments. We would like to thank the many PIT volunteers for their help with testing this first iteration of the models. Production of the niche models for the Davy Crockett National Forest was funded by a grant to RZS from the United States Forest Service. The scientometric study of predictive modeling literature was funded by a grant to RZS from the National Center for Preservation Technology and Training.