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*

Follow this and additional works at: [https://scholarworks.sfasu.edu/crhr](https://scholarworks.sfasu.edu/crhr)

Part of the [Archaeological Anthropology Commons](https://scholarworks.sfasu.edu/crhr)

Tell us how this article helped you.

**Repository Citation**

[https://scholarworks.sfasu.edu/crhr/281](https://scholarworks.sfasu.edu/crhr/281)

This Poster is brought to you for free and open access by the Center for Regional Heritage Research at SFA ScholarWorks. It has been accepted for inclusion in CRHR: Archaeology by an authorized administrator of SFA ScholarWorks. For more information, please contact cdsscholarworks@sfasu.edu.
Addressing Survey Bias: Maxent Models and Public Archaeology at the Davy Crockett National Forest

In February and March of 2018, a Passport-In-Time (PIT) project was conducted to test the first iteration of a suite of niche models designed with the intent of predicting historic and prehistoric archaeolo-
gical 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 re-
finenew 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 intro-
duced throughout 30+ years of linear, block, and compartment-level surveys, achieved through shovel-test-
ing a stratified random sample of locations throughout the DCNF. With this tool, 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
included scientometric 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 package used to generate the DCNF
models (inexact, ENAIval, 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.

Over the past 30+ years, compliance-based archaeological projects conducted on the
Davy Crockett National Forest have yielded an impressive contribution to the archeol-
ogical record. However, those efforts were limited to specific survey areas based upon spe-
cific needs. To begin address survey bias and the impact that it can have on site prob-
ability 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 sec-
ond 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.

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

In addition to the niche models, a geometric morphometric study is underway that asks whether the shape of
Gary and Kent data points differs at those sites where only lithics were recovered (assumed to date to the
Archical period) versus those where lithics and ceramics were recovered on the DCNF (assumed to
date to the Woodland period), and a hunter-gatherer-to-horticultural and/or agriculturalist niche model
affords additional depth to that discussion. Using the ASCFI 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.

Among those challenges associated with creating the models has been the identifi-
cation of deficiencies in the data. One of those deficiencies occurred in the stream
(as 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 meth-
ods 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 Rangeing
(LIDAR) data for the Neches River basin. The NFCT negotiated a higher resolution far
the Angelina, Davy Crockett, Sams 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
(datal models differ significantly from the first iteration, and (2) whether these data
contribute to the production of a more accurate model.

Acknowledgments: We would like to thank the over 30 volunteers and those who help with the logistics for the survey.
Production of the site models to the Davy Crockett National Forest and funded by a grant to LSU from the United States Forest Service.
The scientometric study of predictive modeling literature was funded by a gift to LSU from the National Center for Preservation Technology and Training.

Data sources for validation (dendro and shell work) associated with predictive modeling in archaeology. Analogies from
archaeology in the software and methodological approaches used in this study.

Locations used to begin testing the model were generated as a stratified random
sample (described using DCNF compartment). After shovel tests were placed at each
location, using the mapped site layers as constraints and make additional shovel tests in a cadastre and earliest locations from the contract.

mean - prehistoric
standard deviation - prehistoric
minimum - prehistoric
median - prehistoric
maximum - prehistoric

Value terms increase from each location using the rep-
resentative distribution of location

Locations used to begin testing the model were generated as a stratified random
sample (described using DCNF compartment). After shovel tests were placed at each
location, using the mapped site layers as constraints and make additional shovel tests in a cadastre and earliest locations from the contract.