EVALUATION OF ENSEMBLE HABITAT MAPPING TO SUPPORT NATIONAL PARK SERVICE DECISIONS ON FIRE MANAGEMENT ACTIVITIES AND INVASIVE PLANT SPECIES CONTROL

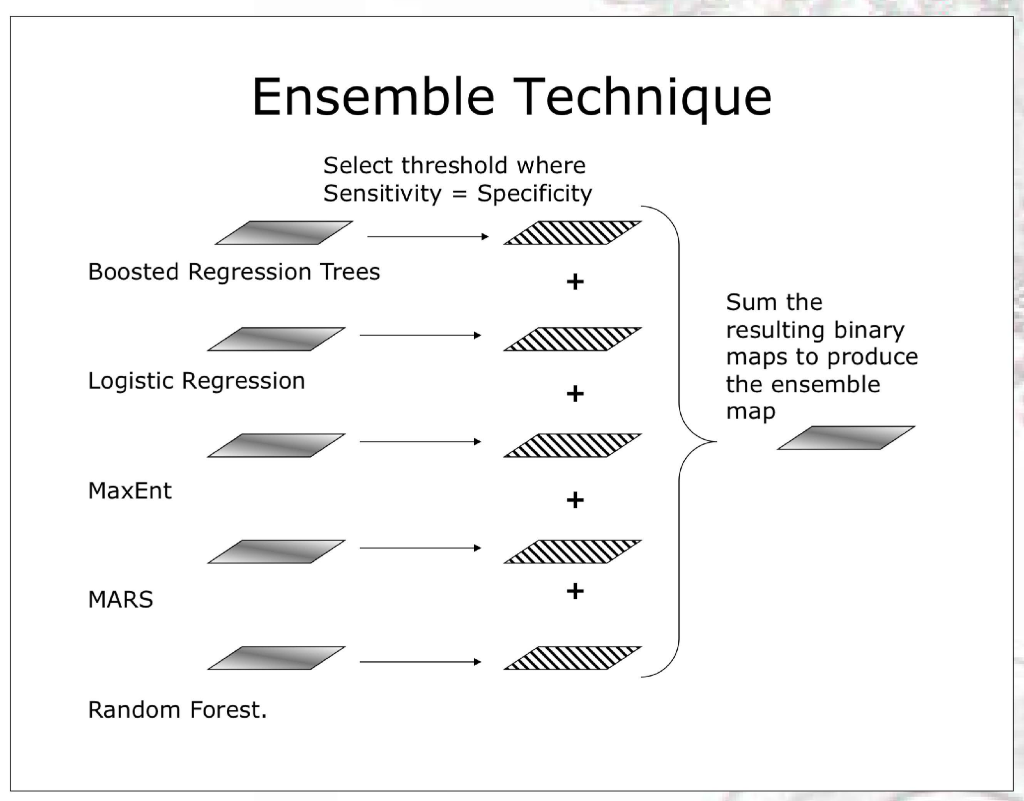
JEFF MORISETTE, USGS, NATE BENSON, NPS, KARA PAINTNER, NPS, PETE MA, NASA GSFC/INNOVIM, MONIQUE ROCCA, CSU, SUNIL KUMAR, CSU, ALAN SWANSON, YERC, CATHERINE JARNEVICH, USGS, TOM STOHLGREN, USGS, KELLYMCCLOSKEY, NPS, PAT LINEBACK, NPS, ATHENA DEMETRY, NPS, JENNIFER ALLEN, NPS, ANN RODMAN, NPS, CRAIG MCCLURE, NPS, TONY CAPRIO, NPS, JOEL SILVERMAN, NPS, NEAL MOST, NASA GSFC/INNOVIM

ABSTRACT

The USGS, NASA, Colorado State University, and the Yellowstone Ecological Research Center are working with NPS to develop robust habitat maps for invasive plant species. The models are based on field data observations of presence/absence provided by the parks combined with climate and vegetation predictors derived from satellite data. The work focused on Yellowstone, Grand Teton, Sequoia and Kings Canyon, and areas of interior Alaska. At least one focal invasive species has been selected for each park. We used an ensemble approach to combine five state-of-the-art models (logistic regression, boosted regression trees, random forest, multivariate adaptive regression splines, and maximum entropy) to produce one habitat map per species. This poster describes: The methods to produce the ensemble maps; how the parks utilize this information; our sampling strategy to evaluate the maps; and the initial results from field validation data collected in the summer of 2008. The ultimate objective for this work is to decide if such ensemble maps would be useful to NPS invasive species managers.

ENSEMBLE MODELING

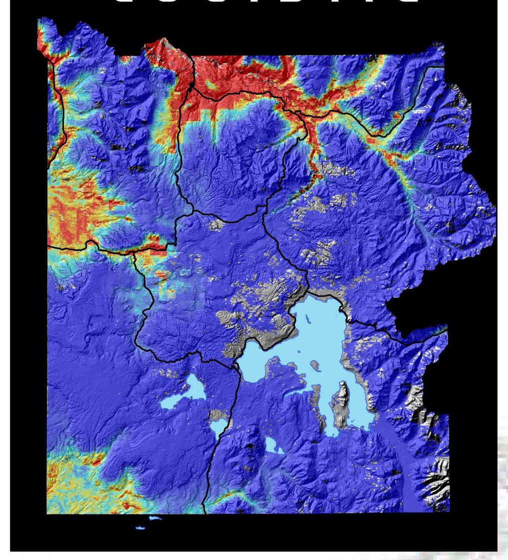
For this project we explored five different species distribution models (SDM) (Guisand and Thuiller, 2005). These included: Boosted Regression Trees, Logistic Regression, Maximum Entropy (MaxEnt), Multivariate Adaptive Regression Splines (MARS), and Random Forest. An overview of the first four models are presented in Elith et al. (2006) and Random Forest is presented in Peters et al. (2007). While it is academically interesting in a statistical sense to explore the nuances of these five (or other) modeling techniques, it is helpful to park managers to have a simple and robust summary that optimizes the modeling techniques to provide them with the best possible estimate of suitable habitat for a given invasive species. For this reason, we have generated one, easy to use, easy to interpret habitat map.



Ensemble maps were produced by combining the maps resulting from the five models through a frequency histogram ensemble approach (Arau´ jo and New. 2007). The method involves creating a binary map from each of the five modeling techniques; where the map is labeled 0 for unlikely habitat and 1 for suitable habitat. Initially, each of the five modeling methods produces output of continuous values.

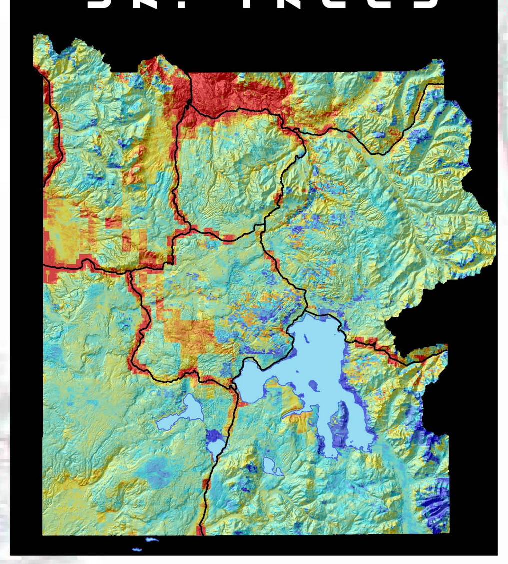
To convert these continuous maps into binary maps we selected a threshold such that values above that threshold were assigned a one and values at or below that threshold were assigned a zero. The field data used to construct the models were randomly divided in half. The first half was used to fit the models and the second half was used to evaluate the performance of the models. Using this second half, or "validation" data set, we were able to select a threshold at which the specificity was equal to the sensitivity. That is, we selected a threshold at which there were as many false positives as there were false negatives. Once a threshold was selected the model output was converted to a binary map. Then, all five binary maps were stacked (or added together) to produce a map with integers ranging from zero to five. This output can be considered a "vote" from the five modeling techniques. A score of zero indicates that none of the modeling techniques assigned that area as suitable habitat, while a value of five indicates that all modeling techniques assigned that the area as suitable habitat.

LOGISTIC



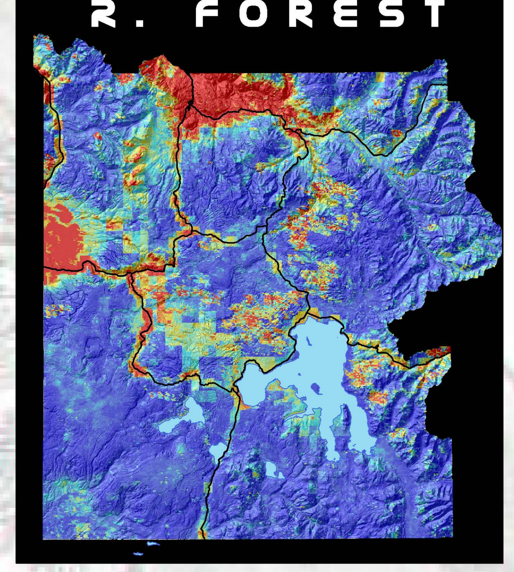
+

BR. TREES



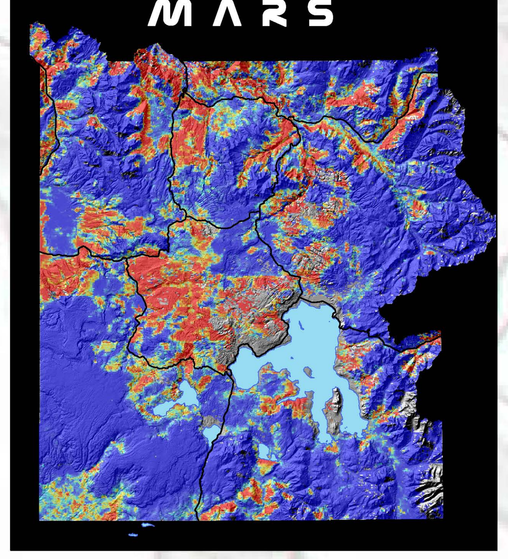
+

R. FOREST



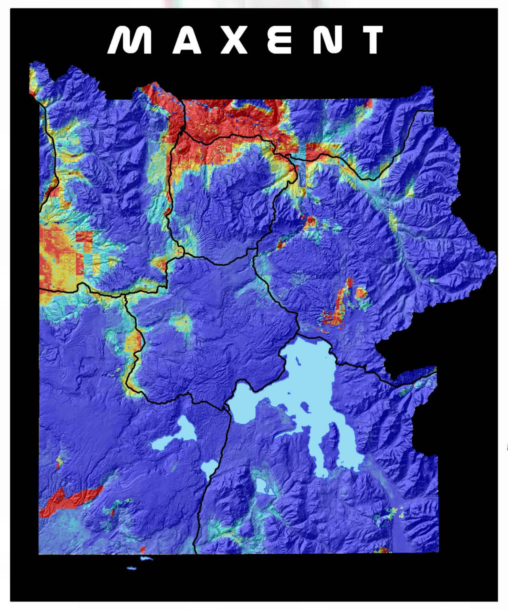
+

MARS



+

MAXENT



ENVIRONMENTAL DATA AND MODELING

NASA's MODIS land surface phenology product defines seasonal patterns of variation in vegetated land surfaces from satellite observations. This new data was combined with bioclimatic layers and multiple types of land surface data sets such as elevation, slope, aspect, percent tree cover, and incoming solar radiation. The environmental layers (covariates) considered exceeded 70 covariates for each study site, so co-linearity tests were conducted to reduce the contribution of similar information. In the end, each of the five ensemble models was constructed to only consider the most informative variables. Data sets that contributed the most to each habitat suitability map included: Annual mean temperature, annual precipitation, precipitation of driest quarter, date of vegetation green up, rate of vegetation green up, rate of vegetation brown down, elevation, impervious surface areas, and percent tree cover.

RESULTS

Table 1 shows the Area Under the Curve (AUC) values for the different models and parks, calculated using the ROC-AUC program (Schroeder, 2006). From the results below, we conclude that the ensemble models did comparatively well. For each park/species the ensemble models' AUC scores were among the top three. It is interesting to note that some models that have a high AUC score for one park/species have a relatively low score for another park/species (e.g. MARS scores highest for Grand Teton but lowest for Sequoia and Kings Canyon). These results indicate that the ensemble map approach adds substantial robustness and consistency of performance across the four park/species considered.

It is important to note caveats for these results. In three of the four parks there were very few presence points relative to the number of absence points in the validation datasets. While we did not calculate the confidence intervals for these results (Schisterman et al., 2001), it is unlikely that there would be strong statistical significance in the observed differences. However, we believe the trend in the numbers indicates that the ensemble approach is worth further consideration.

SAMPLING STRATEGY

The partners at the national parks conducted surveys in the summer of 2008 to help establish the accuracy of the ensemble maps. These surveys followed a stratified random distribution where the strata were based on the six classes (zero through five) from the ensemble maps. The samples were set up such that there were 30 points in each strata.

TABLE 1.	AK	GRTE	SEKI	YELL
Ensemble	0.842 (2)	0.714 (3)	0.474 (1)	0.94 (1)
Logistic	0.784	0.615	0.406	0.746
Random Forest	0.847 (1)	0.615	0.453 (2)	0.902 (2)
Boosted Regresssion Tree	0.837 (3)	0.516	0.425	0.83 (3)
MaxEnt	0.379	0.802 (2)	0.442 (3)	0.822 (3)
MARS	0.835	0.859 (1)	0.345	0.587
Presence /Absence	224 to 101	2 to 96	8 to 191	3 to 93

Table 1

1 = highest AUC score
2 = second highest AUC score
3 = third highest AUC score.

RELATED TO FIRE

One of the primary objectives of this project is to provide fire managers with a useful tool for invasive plant species habitat mapping in order help identify areas that may be susceptible to exotic plant invasion as a result of potential changes caused by fire. This type of tool would be both useful for planning prescribed fires, managing wildland fires, and helping with post-fire rehabilitation and restoration. One of the criteria for invasive plant species selection was its potential to invade after fire.

ENSEMBLE MAPS

Habitat Suitability

High

Low

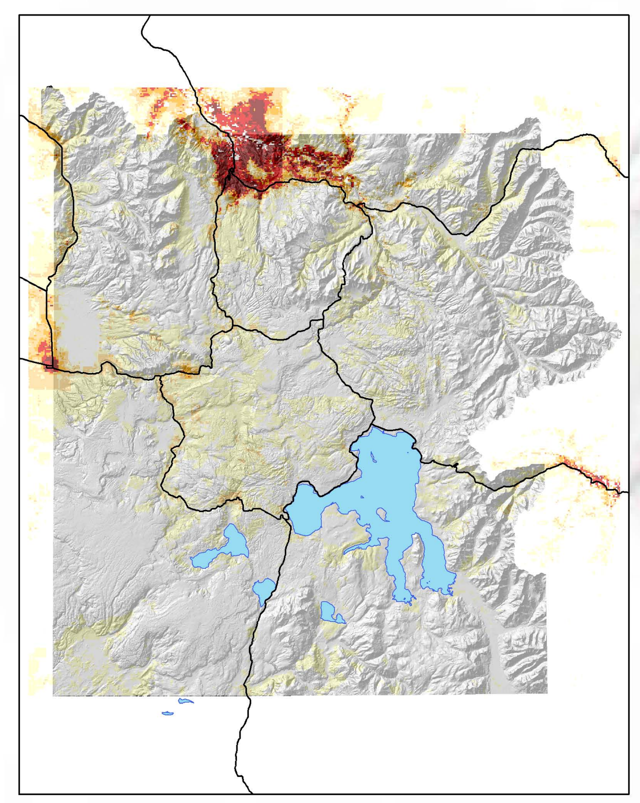
<-

=

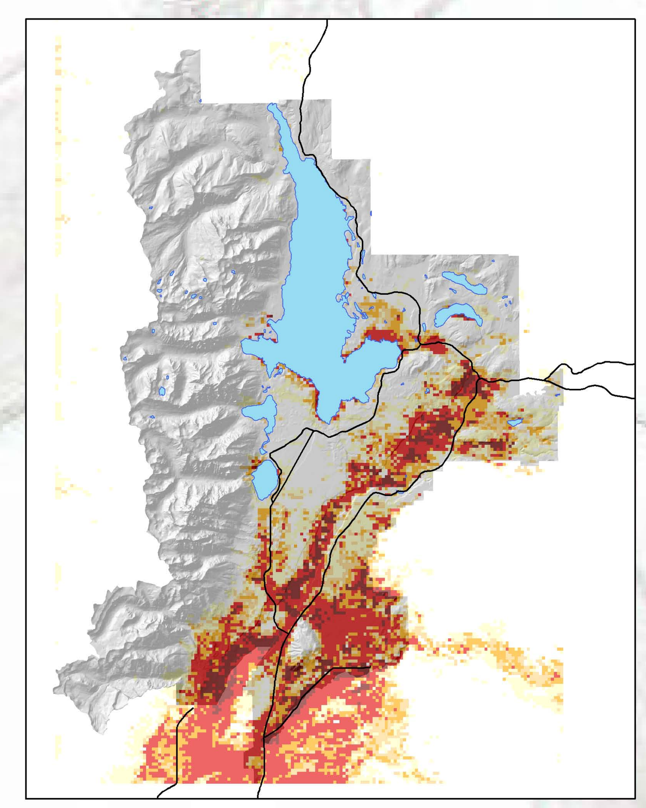
Ensemble Score

0
1
2
3
4
5

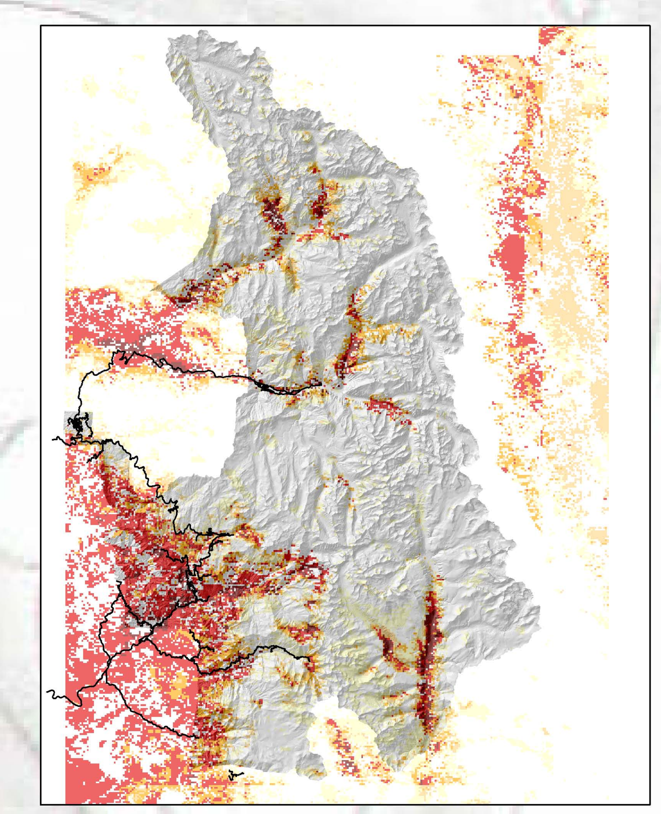
->



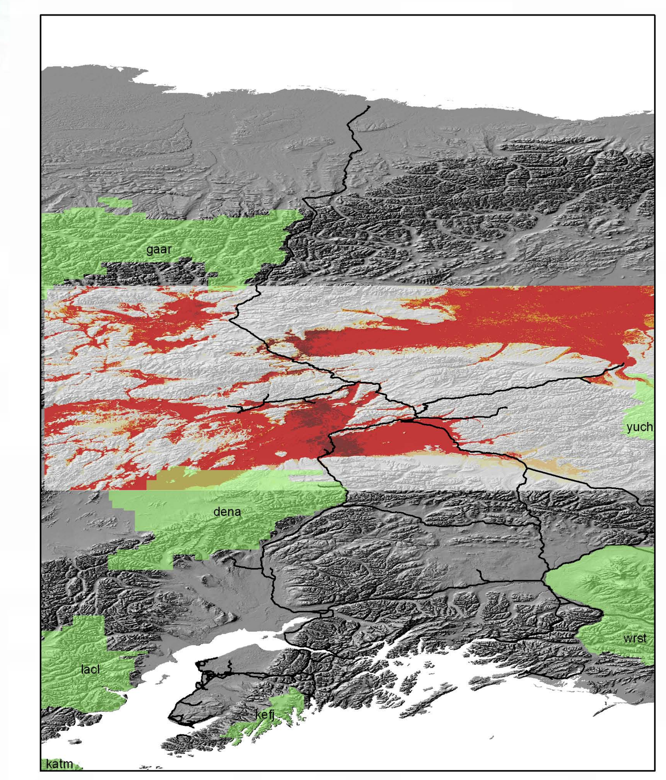
Yellowstone NP
Species: Dalmatian toadflax
(Linaria dalmatica)



Grand Teton NP
Species: Musk thistle
(Carduus nutans)



Sequoia and Kings Canyon NP
Species: Cheatgrass
(Bromus tectorum)



Alaska Interior (grey shaded area)
Species: White Sweet Clover
(Melilotus alba)

NOTES FROM YELLOWSTONE

In addition to the potential value of a species specific invasive plant model that can be used in the future to help direct staff to survey high probability habitats for priority species, the project provided the following immediate benefits.

The project facilitated pulling together invasive plant data from YNP and the Greater Yellowstone Area datasets that had been collected over the past 20 years.

The weed modeling project, which focused on the northern half of Yellowstone Park, enabled staff to survey over 300 kilometers of off-trial backcountry routes that had not been previously surveyed resulting in surveying about 600 acres.

While the surveys to 85 destination sampling points only detected one of the primary target species, dalmation toadflax (3 sites); three additional lower priority species were located including Canada thistle (3 sites), houndstongue (3 sites), and wooly mullein (2 sites). Neither new invaders nor any other high priority species were detected at any of the destination sampling points. However, while hiking to these points 10 different species, including three high priority species were located at 122 opportunistic locations. Seven of the ten species located are known to be distributed beyond roads and developed areas and have been found in the backcountry.

CONTRIBUTORS

We wish to thank and fully credit the National Park Service collaborators from Yellowstone: Craig McClure, Ann Rodman and Eric Miller. Sequoia and Kings Canyon: Athena Demetry, Tony Caprio and Pat Lineback. Alaska: Jeff Heyes, Jennifer Allen and Brian Sorbel. And Grand Teton: Kelly McCloskey and Diane Abendroth.

We would also like to credit the Professionals, Volunteers, Plant Ecologists, Fire Ecologists, Biological Science Technicians, GIS Specialists and the many seasonal field technicians for their contributions of expertise, time, effort and data. More specifically, thanks to Christi Sorrell, Mike Stefancic, Roy Renkin, Gil Crain, Steve Miller, Bob Crabtree, Josh Harmsen, Dan Tyers, Susan Klein, Carolyn Richard, Carrie Guiles, Gary Muerle, Richard Thiel, Erik Frenzel, Karen Webster, Thomas McGinnis, Sylvia Haultain, Karen Folger, Jody Lyle, Thomas Liebscher, Trisha Wurtz, Katie Villano, Matt Macander, Blaine Spellman, Michele Young, Eric "Doc" Janssen, Klara Varga and The Greater Yellowstone Coordinating Committee.

REFERENCES

Arau´ jo, MB, and M New. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution Vol.22 No.1. 42-47.

Elith, J., C. Graham et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: Offering more than simple habitat models. Ecol. Lett. 8: 993-1009.

Peters, J., B. De Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. DeBecker, and W. Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling 207:304-318.

Schisterman, EF, D Faraggi, B Reiser and M. Trevisan, 2001. statistical inference for the area under the receiver operating characteristic curve in the presence of random measurement error. American Journal of Epidemiology Vol. 154, No. 2 : 174-179

Schroeder, B. 2006. ROC/AUC-calculation. http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download.html

USGS science for a changing world

Colorado State University

NASA

Poster Design: Pete Ma
0214.09