A Statistical Approach to Archaeological Surveys Using Remotely Sensed Data

Zachary Lubberts, Carey Priebe, Doug Comer, Daniel Sussman, Li Chen
Applied Math and Statistics, Johns Hopkins University

The Problem

Develop a statistically principled decision support tool for government agencies complying with the National Historical Preservation Act of 1966. In particular, we want to create maps to:

- isolate areas without archaeological significance that are suitable for development.
- differentiate archaeologically significant areas on the basis of amount of archaeological material and intactness of the site.

Method

Raw Data → Pre-processing → Image Processing → Feature Extraction → Classification

- Pre-processing: Apply atmospheric and radiance corrections to the raw data.
- Image processing: Calculate local image features that are some function of the band images for each location.
- Feature extraction: Reduce the dimensionality of the local image features. (Bias-Variance Tradeoff).
- Classification: Return the posterior probability that each site in the region has archaeological significance based on extracted features (a "heat map").

The latter three steps can all be trained on data.

Image Processing

We normalize the entries in the matrices storing the images and generate band difference ratios for each pair of bands $\mathcal{B}_i$, $1 \leq i \leq k$.

$$BDR_{\mathcal{B}_1,\mathcal{B}_2}(i,j) = \frac{\mathcal{B}_1(i,j) - \mathcal{B}_2(i,j)}{\mathcal{B}_1(i,j) + \mathcal{B}_2(i,j)}.$$

Feature Extraction

1. Calculate the Wilcoxon Rank-Sum Test statistic for each coefficient, and remove all but the n best (lowest p-value).
2. On selected coefficients perform Principle Components Analysis (PCA).
3. Dimension for PCA selected via cross validation.

Classification

Fisher's Linear Discriminant

Research funded by the Department of Defense Legacy Research Program and conducted in collaboration with Cultural Site Research and Management, Inc.

Data

- Fort Irwin and China Lake, Mojave Desert, CA
- 9 registered bands, $\approx 9000 \times 23000$ pixels
  - 8-band WorldView 2 imagery
  - Slope band derived from LIDAR imagery
- 25 known locations for Lithic sites,
  21 known locations for Habitation sites
- 100 random locations in areas with no sites

Results

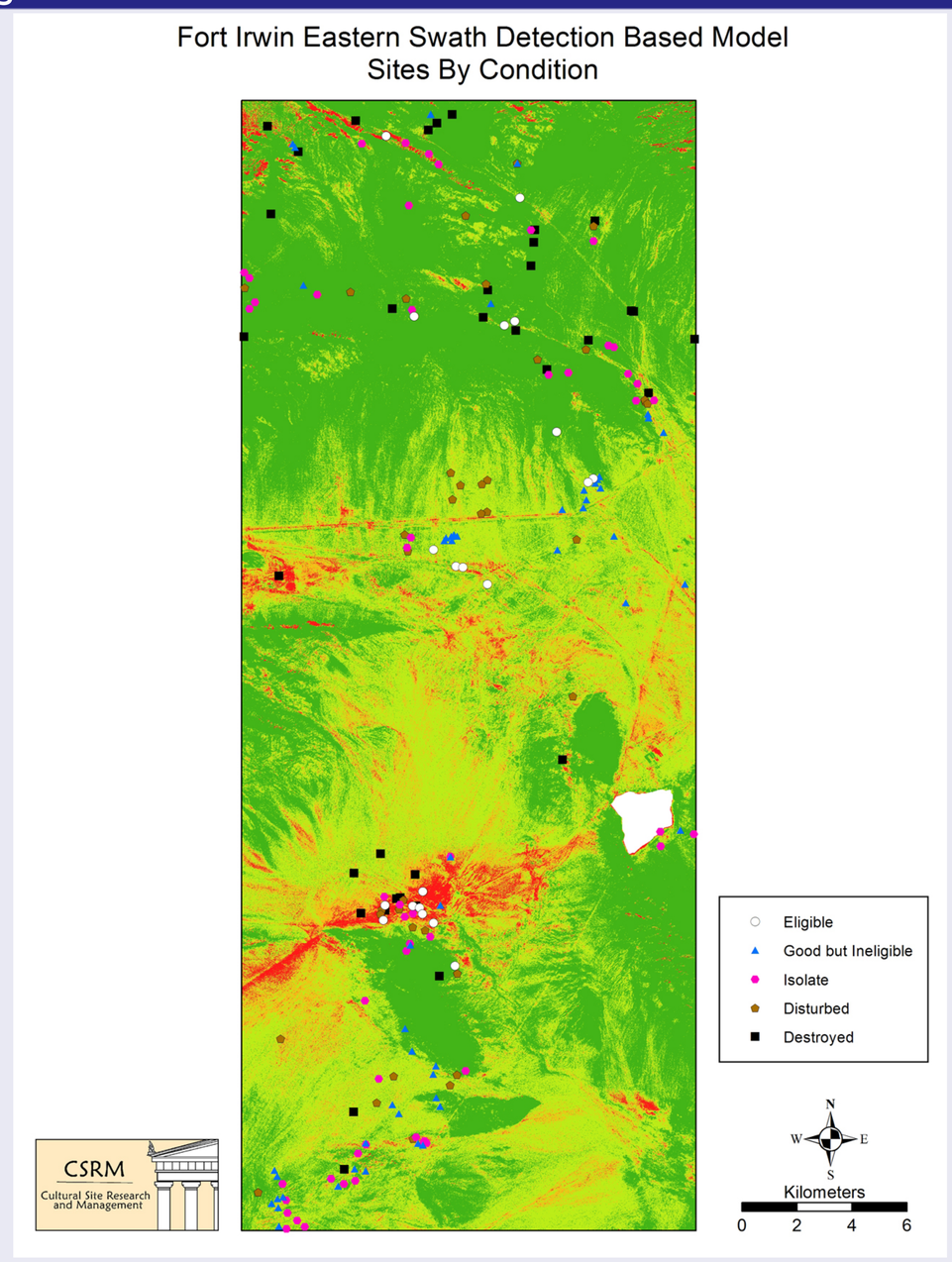

Figure: This is the heatmap generated from the classifier posterior probabilities for the East test region of Ft. Irwin overlaid with the locations and classifications of known sites.

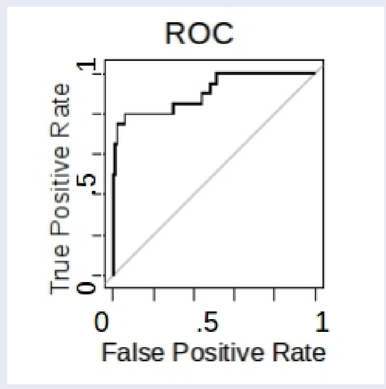

Figure: This is a receiver operating characteristic (ROC) curve generated for a classifier trained on the East Ft. Irwin data using the same process as was done to create the heatmap pictured. This figure results from the posterior probabilities of the training data points used to make the classifier.

Additional acknowledgements: Bryce A. Davenport: CSRM; James C. Tilton, Edoardo Pasolli: NASA Goddard Space Flight Center.