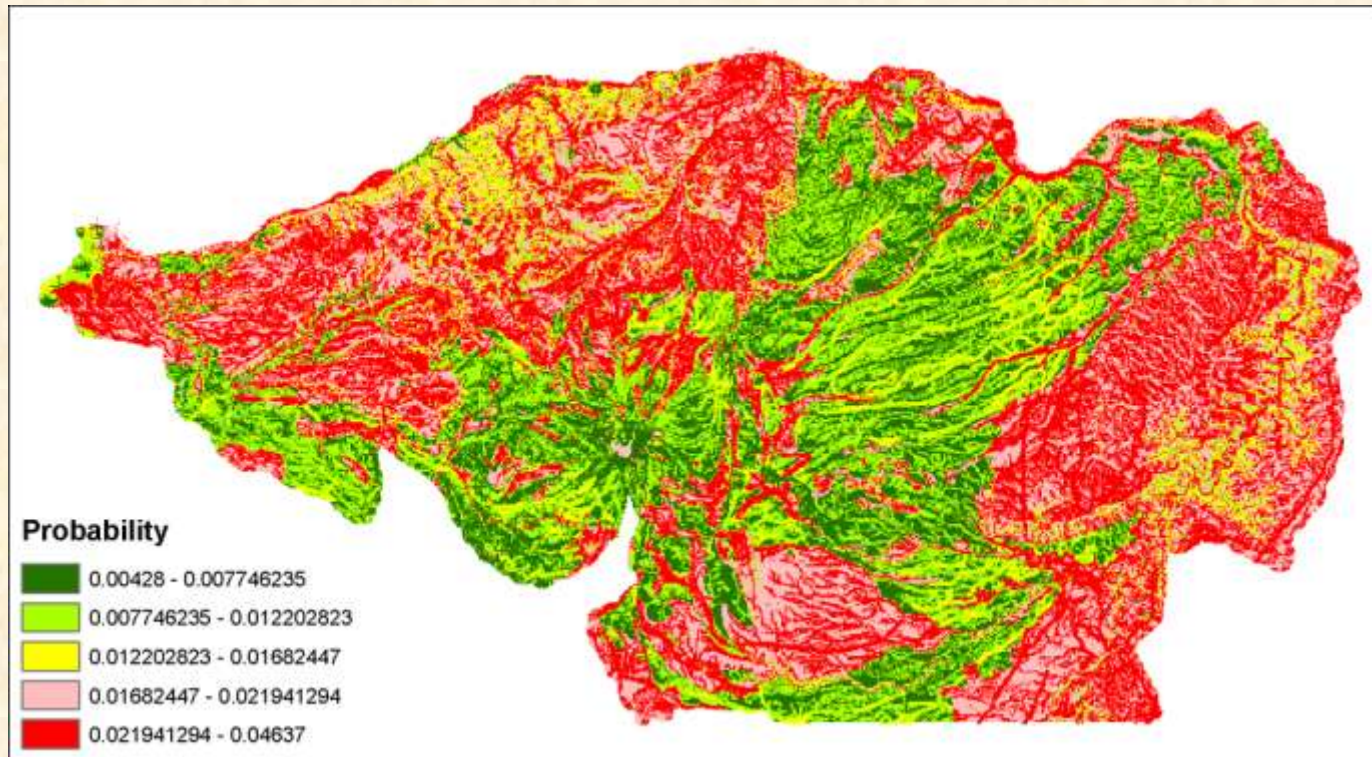


Predictive Modeling



Todd M. Ahlman
Historical Research Associates
May 24, 2011

What is predictive modeling

- Settlement patterning
 - Interaction of environmental characteristics and culture
- Predicting occurrence of archaeological sites (dependent variable) in an unsampled location using one or more independent variables (environment/locational characteristics)
- Suitability Modeling
 - Locations that exhibit the best characteristics for people to live
- Resource management
 - Tools for agencies to reduce level of effort for surveys
 - Tools for predicting areas where development should be avoided

Steps for developing a model

- Define study area
- Collect archaeological information
 - Site locations
 - Previous research
 - Cultural information – How did humans use the landscape
- Determine and Collect environmental Data
 - Most methods want binary data (yes or no)
- Determine statistical method and apply to data

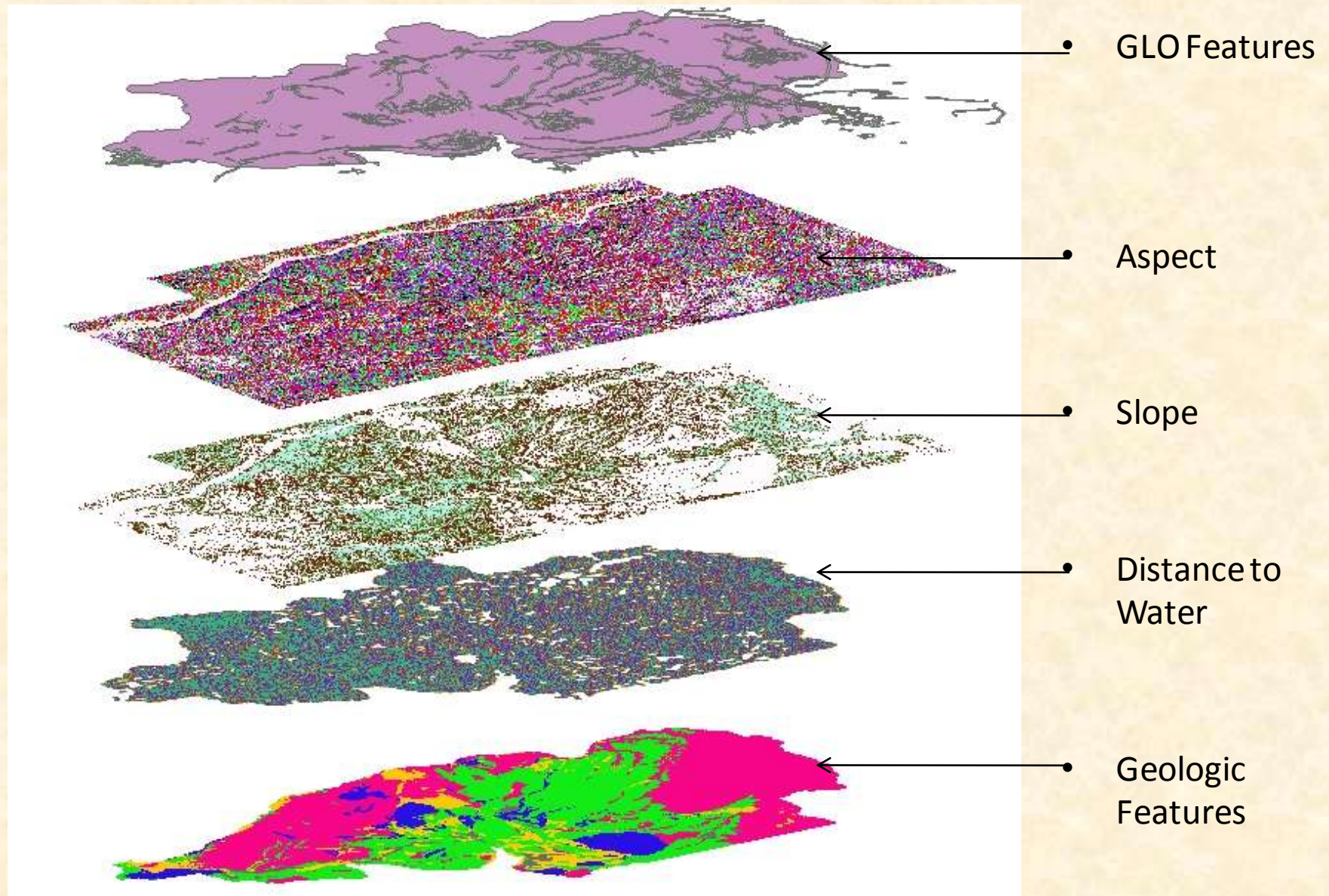
Predictive Models

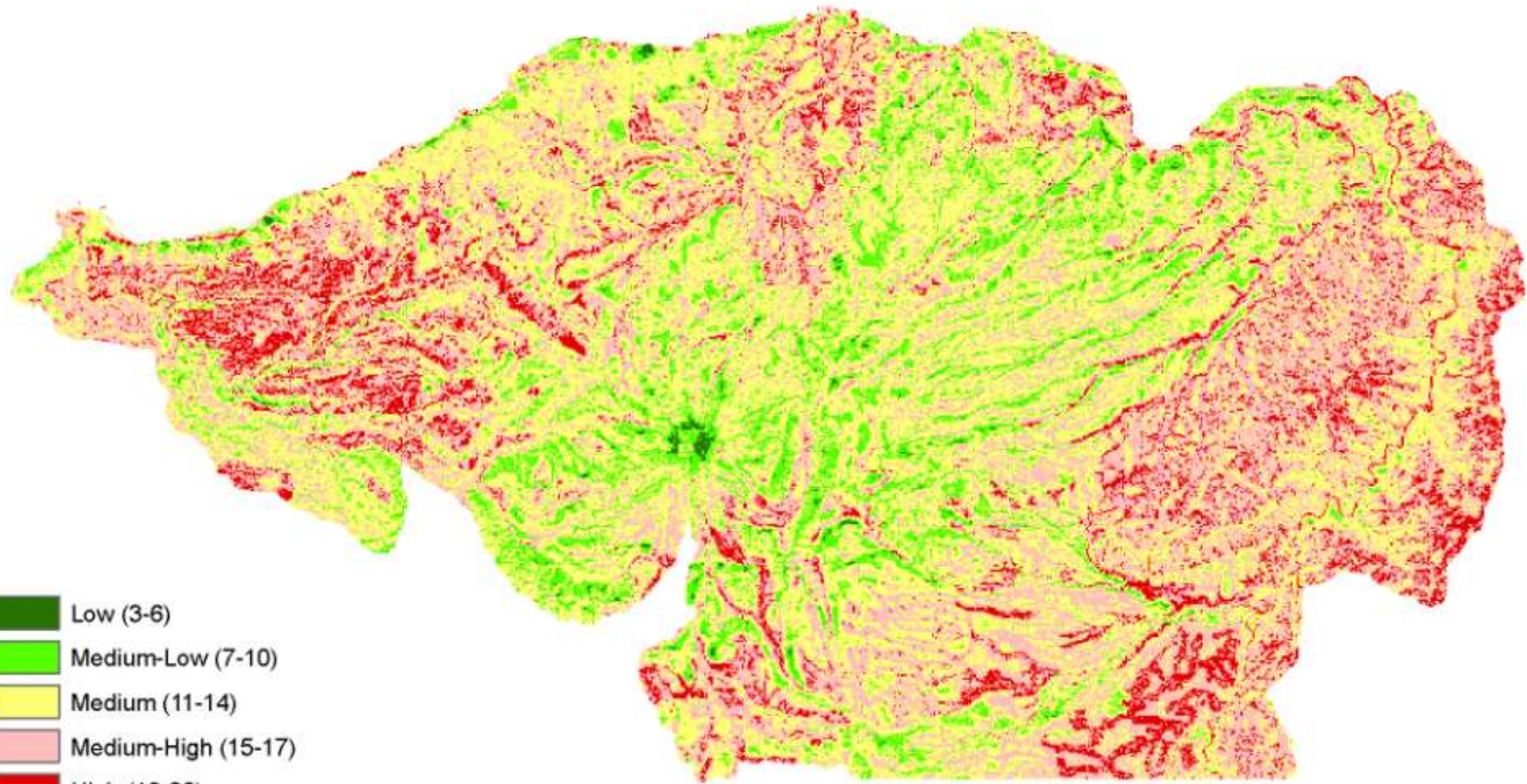
- Vector/Raster Intersection
- Logistic Regression
- Bayesian methods (Weights of evidence)

Vector/Raster Intersection

- Relies on previous research (or researcher's personal bias)
- Simple map algebra
- Score environmental variables based on previously know (imagined, believed) suitability for archaeological site occurrence
 - Importance of variable in predicting site locations
- Intersect maps and total up score
 - $\text{Map 1} + \text{Map 2} + \text{Map 3} \dots \text{Map X} = \text{Prediction Map}$

Vector/Raster Intersection Environment and Cultural Themes

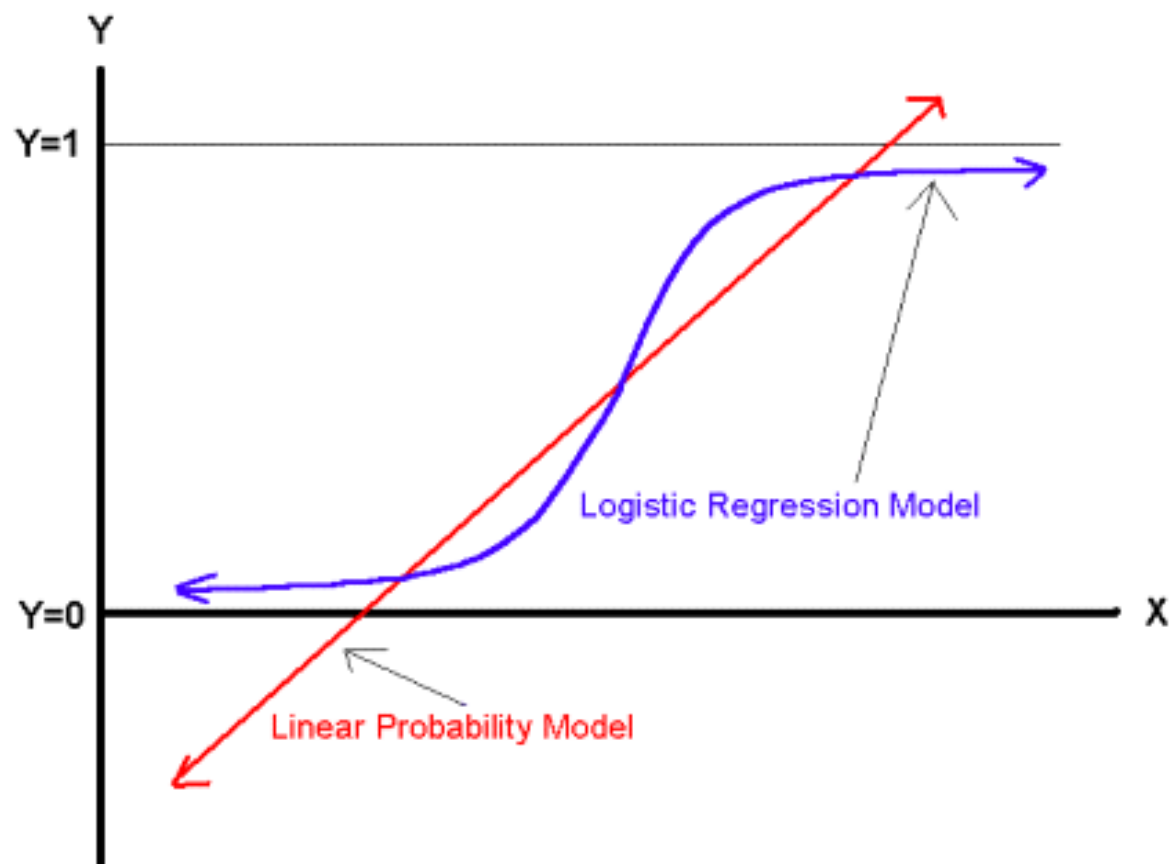




Logistic Regression

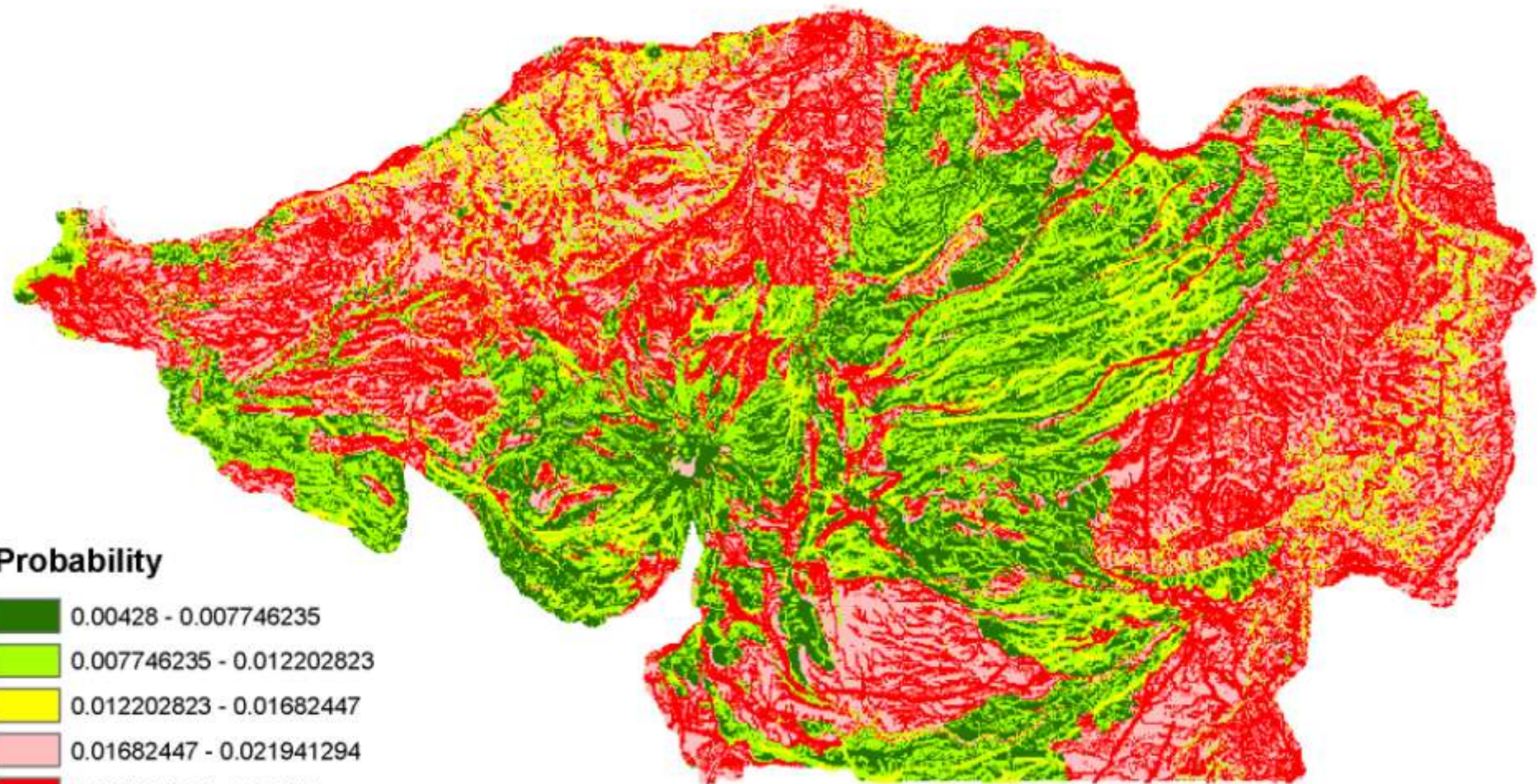
- Considered to be very powerful method
- Uses environmental variables to predict site and non-site locations
 - Presence or absence of sites with environmental variable combinations
- Interaction of environmental variables to determine site and non-site locations
 - Need non-site locations to associate environmental variables with negative information
- Creates S-shaped regression curve
- Does not require normal data

Comparing the LP and Logit Models



Logistic Regression

- First stage is data conversion
 - Converting data to raster
- Next stage is deriving environmental variables from the data
- Final stage is analyzing and evaluating the relationship between the archaeological database (the dependent variable) and the environmental (or independent) variables



Probability

- 0.00428 - 0.007746235
- 0.007746235 - 0.012202823
- 0.012202823 - 0.01682447
- 0.01682447 - 0.021941294
- 0.021941294 - 0.04637

Weights of Evidence

- Uses known archaeological site locations
 - Don't need non-site locations
 - Environmental variables
- Bayes theorem
 - *event A* conditional on another event *B* is generally different from the probability of *B* conditional on *A*.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$
$$\propto L(A|B) P(A)$$

- Calculates prior probability (a priori probability) or occurrence of sites across study area

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

- Calculates posterior probability
 - Posterior probability is proportional to the product of the prior probability and the likelihood

Weights of Evidence

- Known sites
- Environmental variables
- Occurrence of sites at specific environmental variables
- Intersection of these variables across landscape means prediction for site

Weights of Evidence

- Posterior probability
 - Closer to 1, better prediction
 - Value greater than prior probability indicates some predictive power
 - Value lower than prior probability indicates little predictive power
- Can scale posterior probability into different levels of probability (high, medium, and low)

