Archaeological Predictive Model

• What is archaeological predictive modeling?

• Why do we want it?
Archaeological Predictive Model

• **What is a model?**
  • A simplified abstract view of a complex reality.
Archaeological Predictive Model

Archaeological Predictive Modeling?

• Native American Pre-Contact Sites
• Based on premise of patterned behavior
• Distribution of sites based on environmental factors
• Usually derived from logistic regression
• Implicit modeling as old as archaeology
Archaeological Predictive Model

• **Why do we want it?**

Some sites are hard to miss.
Archaeological Predictive Model

Some sites less so.
Archaeological Predictive Model

Same spot, 500 years ago.
Archaeological Predictive Model

Why Continued:

• **25,000+ Sites currently recorded in Pennsylvania (22,144 used in study)**

• **Represents 2-4% of all possible archaeological sites in Pennsylvania**

• **Large variations in survey coverage across state**

<table>
<thead>
<tr>
<th>Modeling Region</th>
<th>PASS Sites per Square Mile</th>
<th>Prehistoric Sites per Square Mile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>0.35</td>
</tr>
<tr>
<td>3</td>
<td>0.87</td>
<td>0.63</td>
</tr>
<tr>
<td>4</td>
<td>0.30</td>
<td>0.26</td>
</tr>
<tr>
<td>5</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>6</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>7</td>
<td>0.28</td>
<td>0.19</td>
</tr>
<tr>
<td>8</td>
<td>0.79</td>
<td>0.69</td>
</tr>
<tr>
<td>9</td>
<td>0.96</td>
<td>0.82</td>
</tr>
<tr>
<td>10</td>
<td>0.81</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Archaeological Predictive Model

• Two ways we (PennDOT) need to know where sites might be before survey:
  • For planning and Programming Purposes
  • To aid in Section 106 investigation
Archaeological Predictive Model

• **Before the current Predictive Model:**
  - The Archaeological Potential Scoring for Planning purposes was **based solely on distance** to known archaeological sites (but see caveats previously)
  - Predictive models were developed for EIS-level projects by the design firm, **each one done individually** and from scratch.
  - Archaeologists **used their intuition** (experience) in assessing whether a Phase I survey should be done, sometimes in conjunction with geomorphology studies.
Archaeological Predictive Model

• **Modeling Process typical for Archaeology**
  
  • **Data set = known archaeological sites**
  
  • **Sample bifurcated into build and test populations**
  
  • **Most likely environmental factors used**
  
  • **Various regression techniques used to find correlations between environmental factors and the presence of sites (in the build population)**
  
  • **Test population validates(?) model**
  
  • **Results divided into “High,” “Medium,” and “Low” potential**
  
  • **State map rasterized and assigned potentials to each raster, building high, medium, and low polygons**
Archaeological Predictive Model

• Predictive modeling part of archaeological conversation since early 1980’s

• Prior to 2015, exactly 1 state (Minnesota) had statewide operational model.
Archaeological Predictive Model

• Development of the PennDOT Archaeological Predictive Model
  • Initiated in 2013 and completed in 2015
  • Funded through FHWA Research Program
  • Completed by team led by Matt Harris of AECOM (formerly URS), with Noel Strattan, Ira Beckerman, and Christine Kula
Archaeological Predictive Model

Data Preparation
- PASS Database
- Archaeology Surveys
  - Site location and survey bias

Model Creation and Testing
Model Building
- Site & non-site data
  - Model Class 1: "Judgemental"
  - Model Class 2: "Proportional"
  - Model Class 3: "Statistical"
  - Output
    - Basis Functions

Testing -Validation
- K-Fold CV
- Split Sampling
  - Calibration and Optimization

Model Selection
- Spatial Analysis Model

Figure 2 - General organization of entire model building process.
Archaeological Predictive Model
Archaeological Predictive Model

• Not one model, but 132 models based on regions, watersheds, and topographic positions
Archaeological Predictive Model

8 Basins, 104 Watersheds
Archaeological Predictive Model

(91) Environmental variables covered:
- Elevation
- Slope
- Terrain roughness
- Distance to water

Statistical analysis and winnowing of most telling variables

Great variation in data quality from region to region, so model building toolkit used:
- Stepwise Logistic Regression
- Multivariate Adaptive Regression Splines
- Random Forest
Archaeological Predictive Model

• Testing and validation using K-fold Cross validation

• Resulting models applied to each region, divided into 10m rasters, which were later grouped into 30m rasters as mapped units

• High, Medium, and Low probabilities drawn with an eye toward reducing false negatives (model would over-predict sites)
  • Goal of finding 85% of the sites within a site-likely area of 15%
  • Kvamme Gain of 0.82
Archaeological Predictive Model

• When put into CRGIS
Archaeological Predictive Model

• CRGIS – Sensitivity Mapping
Archaeological Predictive Model

• Sensitivity Mapping with archaeological Sites
Archaeological Predictive Model

- Region around Melrose, Susquehanna County
Archaeological Predictive Model

• Sensitivity Mapping and Archaeological Sites (0)
Archaeological Predictive Model

- **Model is currently undergoing testing**
- **Will be “implemented” region by region**
- **Implemented means:**
  - Another tool in the Archaeologist’s toolkit
  - Validation could mean no statistical difference between human and machine predictions (Turing’s Imitation game?)
- **Look to Conducting Statistical Test in 2018 with data collected post 2014.**
Archaeological Predictive Model

• Entire 7-Volume Study on PennDOT’s CRM Site: www.penndotcrm.org

• Volume 7 is summary

Contact:

Ira Beckerman
PennDOT Cultural Resources Unit Head
ibeckerman@pa.gov
(717) 772-0830
Archaeological Predictive Model

Thank you/ Questions?