
Scott Smith, Eve Flager, Bahram Daneshfar, Grace Frank and Chuck Bulmer

Spatial Predictions Symposium, CSSS-ASSS annual meeting, San Antonio, TX, Oct 18, 2011
Presentation Outline

• Background – objectives, approach, team
• Method – techniques, predictors, knowledge base
• Results – hardened map, difference from legacy map
• Accuracy assessment
• Discussion – scaling-up issues, use of legacy maps, lessons learned.
Context

- Okanagan Basin is a semi-arid region with growing demand for a limited water supply

Objective

- Disaggregate existing legacy soil maps to provide raster-based soil attributes to modeling effort

Approach

Select a sub-watershed to test several methods of disaggregation

Assemble multidisciplinary team to undertake the work
Background – study area

- 75,000 ha sub-watershed of the Okanagan River, a tributary to the Columbia River
- Elevation range from 350m asl at Okanagan Lake to 2000 m on highest ridges
- Uplands are commercial forest land, valley floor is irrigated horticulture
- Annual precipitation ranges from 300 mm to 900 mm
- Mean annual temperature varies from 11°C to 3°C
- Xerolls to Cryods
Methods: Two main groups of methods for predictive mapping

- **Knowledge-Driven methods**
  Relationships between the target variable and predictor variables are defined based on expert opinion.

- **Data-Driven methods** *(quantitative empirical modeling)*
  Relationships between the target variable and predictor variables are quantified by the method, based on the data and then used for prediction.
Methods

- Fuzzy logic inference engine - ARC SIE (modified)
- Expert knowledge rule set (Scott’s folly)
- Logistic Regression
- Weights of Evidence (WOFE)
- Hybrid method using contrast values from WOFE to define rule curves in ARC SIE
### Methods - predictors

- **Target variable** – 23 soil series mapped in study area

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Dataset</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topographic</td>
<td>• Elevation&lt;br&gt;• Slope Percent&lt;br&gt;• Aspect&lt;br&gt;• Relative Heights Slope Position (RHSP)&lt;br&gt;• LandMapR Landform Classes&lt;br&gt;• Topographic Position Index (TPI)&lt;br&gt;• Topographic Position Index (TPI) Landform Classification&lt;br&gt;• Stream Network</td>
<td>• CDED DEM&lt;br&gt;• Terrain Resource Inventory Mapping (TRIM) Stream Network</td>
</tr>
<tr>
<td>Vegetation/Climate</td>
<td>• BEC Zones and Sub-Zones&lt;br&gt;• CIRCA 2000 Land Cover</td>
<td>• 1:50,000 polygons downloaded from Ministry of Environment&lt;br&gt;• 30m CIRCA 2000 raster downloaded from Geogratis</td>
</tr>
<tr>
<td>Parent Material</td>
<td>• Surficial Material</td>
<td>• 1:20,000 polygons provided by Ministry of Environment</td>
</tr>
</tbody>
</table>
Methods - Soil Polygon Refinement for Training Data Sampling

- Used soil polygons from the soil polygon coverage
- Used a polygon for a particular soil series if it consisted of at least 70% of that soil series
- Inverse buffered the polygons by 50m
- Refined the polygon using BEC zone, parent material, slope, and elevation
- For each soil series, 200 random training points were generated from the refined polygons for WOFE calculations
Weights of Evidence Terms

-_weights for patterns
  - $W_+$ weight for inside the pattern
  - $W_-$ Weight for outside the pattern
  - 0 Weights for areas of no data
- Contrast: a measure of the spatial association of pattern with sites
- Studentized Contrast: a measure of the significance of the contrast

Contrast: $C = W_+ - W_-$
Methods – Validation

• Stratified study area by ecological zone then selected forestry roads within each zone, field sampled (identified soil series) at landscape positions along the right of way whenever change occurred.

• Two years of field work, total of 300 field checks, half to build the expert knowledge and rule set, half to validate the predictions

• Represents more field work than was used to create the legacy map
Methods – Run analyses

• Ran 5 methods with various interventions and modifications related to:

  – input data to ARC-SIE run,
  – use of selected predictors used in the Hybrid method,
  – setting probability and certainty limits in the WOFE and LG
Results - Digital Soil Map vs Polygon Soil Map
Results - validation

Validation points buffered to 50 m and 100 m
## Results – % agreement observed vs predicted

<table>
<thead>
<tr>
<th>Buffer</th>
<th>ARC SIE</th>
<th>Ex Rules</th>
<th>Hybrid</th>
<th>WOFE</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 m</td>
<td>46</td>
<td>50</td>
<td></td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>100 m</td>
<td>60</td>
<td>56</td>
<td></td>
<td>48</td>
<td>44</td>
</tr>
</tbody>
</table>
Discussion – scaling up to 100 m cell
Discussion and Conclusions

• Use of legacy maps – problematic (correlation issues, difficult to identify taxonomic units in field)

• Keep it simple rule: in this case our simplest approach was not our best method, now where do we go?

• Lessons learned:
  – Predictors – a few good predictors better than many weak predictors
  – Expert knowledge – if exists for an area, then cost effective to capture, if not, very time-consuming to develop
  – No one method that satisfied all needs, but some cross over of data driven and knowledge driven systems produced our best results.
Thank you for your attention