A Refined Approach for Examining *Phytophthora ramorum* Risk in the Eastern U.S.

Frank Koch
NC State University
USDA-FS FHM, RTP
In Support of the National SOD Survey Protocol…

- Two tasks:
  - Risk-based spatial database for SOD
    - Any relevant data layers
  - Use the database in a modeling context
I. Risk-based Spatial Database for SOD

• Basic strategy: Develop data for the three categories of the initial overlay analysis
• Still in progress, but have a number of examples…
Climate

• Previously using PRISM data for climate analyses
  – 2 km resolution, but expensive
  – Not updated to current

• Can we generate more up-to-date products?
  – NOAA daily and monthly station data

• Spatial Interpolation
  • Gradient plus inverse distance squared (Nalder & Wein 1998)
  • Compared favorably to several other methods, including ordinary kriging, detrended kriging, co-kriging, inverse distance squared
From Nalder and Wein (1998)

- GIDS Model (based on the 30 nearest neighbors)

\[
Z_k = \frac{\sum_{i=1}^{30} \left( Z_i + C_x (X_k - X_i) + C_y (Y_k - Y_i) + C_e (E_k - E_i) \right)}{\sum_{i=1}^{30} \frac{1}{d_i^2}}
\]

Where,
- \( Z_k \) = predicted value at unmeasured location \( k \)
- \( Z_i \) = measured value at location \( i \)
- \( X \) = x-coordinate
- \( Y \) = y-coordinate
- \( E \) = elevation
- \( D \) = distance from measured location \( i \) to \( Z \)
- \( C_x, C_y, \) and \( C_e \) are based on the ordinary least square solution of the following regression model using 30 nearest neighbors to location \( k \).
- \( Z = a + C_x X + C_y Y + C_e E + \epsilon \)

Where, \( a \) is the intercept and \( \epsilon \) is error.
Regression Model Selection

• Using three possible gradients, there are seven possible models
  1. \( Z = a + CxX + CyY + CeE + \varepsilon \)
  2. \( Z = a + CyY + CeE + \varepsilon \)
  3. \( Z = a + CxX + CeE + \varepsilon \)
  4. \( Z = a + CxX + CyY + \varepsilon \)
  5. \( Z = a + CxX + \varepsilon \)
  6. \( Z = a + CyY + \varepsilon \)
  7. \( Z = a + CeE + \varepsilon \)

• Test each model to examine if all the independent variables are significant
• If more than one model has all significant independent variables then the model with greatest \( R^2 \) is selected
• If no model has all significant independent variables then simple inverse distance square weighting is used (i.e. all coefficients are set to zero)
GIDS Surface of Temperature from 2003 NOAA Data

(4km² cells)

April 2003 Temperature (F)

High : 77

1
Temperature Model Selection
for each prediction point

Model
1. \( Z = a + CxX + CyY + CeE + \varepsilon \)
2. \( Z = a + CyY + CeE + \varepsilon \)
3. \( Z = a + CxX + CeE + \varepsilon \)
4. \( Z = a + CxX + CyY + \varepsilon \)
5. \( Z = a + CxX + \varepsilon \)
6. \( Z = a + CyY + \varepsilon \)
7. \( Z = a + CeE + \varepsilon \)
8. IDW
GIDS Surface of Precipitation from 2003 NOAA Data

April 2003 Precipitation (in)
Cross-Validation Results

Observed April 2003 Temperature (F)

Predicted Temperature (F) based on cross validation

Temperature
<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Mean error</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 2003 Temperature (F)</td>
<td>4994</td>
<td>52.5</td>
<td>0.007</td>
<td>1.468</td>
</tr>
<tr>
<td>April 2003 Precipitation (in)</td>
<td>6847</td>
<td>2.9</td>
<td>-0.004</td>
<td>0.868</td>
</tr>
</tbody>
</table>
GIDS & Poisson Regression

• GIDS seems to work well, but what about surfaces representing number of days with appropriate conditions?
• Count-based = Poisson distribution
• Like previous GIDS, seven possible models
• Best model chosen based on chi-squared significance (or log likelihood)
• Three-dimensional distance
• Still working out a few things…
Number of Days in 2004 With Optimal Temperatures (60-80°) and Precipitation > 0.05"
Number of Consecutive Days (Allowing One Day Off) in 2004 with Optimal Temperatures and Precipitation

allow1_04
Value

High : 26
Low : 0
Number of Consecutive Days (No Days Off) in 2003 with Optimal Temperatures And Precipitation – All Reporting Stations (n=4144)
Some Preliminary Results…

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Max.</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Days 2003</td>
<td>4144</td>
<td>64.84</td>
<td>191</td>
<td>13.36</td>
</tr>
<tr>
<td>Consecutive Days 2003</td>
<td>4144</td>
<td>6.03</td>
<td>21</td>
<td>2.15</td>
</tr>
</tbody>
</table>
Other Climatic Factors

- Relative Humidity
  - 100% humidity important for *P. ramorum* and other aerials (?)
  - Fewer stations
- Microclimate
  - Topographic Relative Moisture Index (TRMI)
  - Represent local conditions (e.g., hollows)
Estimated Percent Basal Area in Red Oaks and Live Oaks

- Generated by kriging of FIA plot basal area data - 1 km² cells
- High-frequency pattern to FIA data, so not well predicted by smooth interpolators (poor RMSE for validation or cross-validation)

Data from R. Morin, USDA-FS
What About Understory Hosts?

• Already have estimate of understory hosts for NE U.S.
  – Again, kriged from plot data…
• Archival understory data also exists for SE U.S.
  – Strange format
  – Don’t know how many plots
• Given limitations of interpolated FIA data…
  – County-level distribution maps from PLANTS national database, other sources

Probability of Understory Host Presence

- 0
- 0 - 0.2
- 0.2 - 0.4
- 0.4 - 0.6
- 0.6 - 0.8
- 0.8 - 1

Data from R. Morin, USDA-FS

Kalmia latifolia
Pathways

• Number of potential areas for refinement
• For example, how to get at landscapes where people plant (potentially infected) nursery stock?
  – Percentage of low density residential from land cover data
    • Existing neighborhoods where people may be planting
  – Changes in road density = areas of new construction
  – Nighttime lights expansion (Imhoff et al. 1997)
    – Broad estimate of suburbanization/expansion
Percentage of Low-Density Residential Land Cover (from NLCD)

low-density resid % Value
- 0-10
- 10-20
- 20-30
- 30-40
- 40-50
- 50-60
- 60-70
- 70-80
- 80-90
- 90-100

1 km² cells
Change in Road Density 1996-2005

road_dens_diff_95-04
<VALUE>
- 0 - 383.2547607
- 383.2547608 - 1,533.019043
- 1,533.019044 - 3,257.868588
- 3,257.868587 - 5,940.440792
- 5,940.440793 - 10,347.878554
- 10,347.878555 - 18,396.22652
- 18,396.22652 - 49,056.60339

2 km² cells
Nighttime Lights Expansion (Using 1992 and 2000 Imagery)

nighttime_lights

Value

-11 - 8
8 - 12
12 - 27
27 - 52
52 - 100

1 km² cells
Spread via Roads

- Freight Analysis Framework
  - Volume of cargo moved by trucks
  - Nursery stock movement
Nurseries

- ReferenceUSA database
  - 13 million U.S. businesses
  - Have geographic coordinates
  - Search for primary descriptions

- Comprehensive lists
• When add retail locations, including home improvement centers, data layer gets quite large
• Can calculate grids of number of nursery locations within distance of a point
  – # possible exposures to (potentially) positive nurseries

<table>
<thead>
<tr>
<th># Nurseries within 20k</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>5 - 17</td>
</tr>
<tr>
<td>17 - 36</td>
<td>36 - 59</td>
</tr>
<tr>
<td>59 - 88</td>
<td>88 - 122</td>
</tr>
<tr>
<td>122 - 163</td>
<td>163 - 215</td>
</tr>
<tr>
<td>215 - 316</td>
<td></td>
</tr>
</tbody>
</table>
Other Pathway Factors

- Housing / population density
- Road proximities
- Wildland-urban interface
  - Compiled for fire risk, but relevant for *P. ramorum*, other insects and pathogens
II. SOD Risk Modeling

- Build cost surface(s) incorporating all of these spatial data layers
  - Ideal spatial resolution?
- Model SOD movement after hypothetical introduction
  - Cellular automaton approach
  - Transition probabilities = interaction between cost and infection rate
SOD Risk Modeling (cont.)

• Many uncertainties, so try several different scenarios
  – Test range of infection rates (low, moderate, high)
  – Different cost surfaces

• Through repeated runs, can develop per-pixel risk ratings