Modeling the natural distributions of aquatic species for biodiversity conservation

Yong Cao
Illinois Natural History Survey
Institute of Natural Resource Sustainability
University of Illinois at Urbana-Champaign
Why we need to know the natural distributions of species?

1) Assessing the conservation status of individual species
2) Identifying spatial patterns of biodiversity
3) Assisting for finding new populations
4) Guiding restoration and re-introduction
5) Assessing ecosystem health
6) Monitoring effects of global warming
Freshwater fish species diversity

Source: Revenga et al., World Resources Institute (WRI), Washington DC, 1998.
Freshwater fish diversity in Mississippi River Basin

What we normally know

Bluntnose darter (Etheostoma chlorosomum) in Illinois freshwaters (INHS)
Why collection data are not enough?

- Collection records are incomplete
- Distribution ranges of many species have reduced and fragmented
- Distributions of most species likely change due to global warming
Species distribution model (SDM)

A model that uses collection records and environmental variables to predict the distribution of species
What a SDM does

Modified from Elith et al. 2006 *Ecography*
A simple conceptual model

Watershed: Climate, geology, topography

- Flow regime
- Substrate & sediments
- Water chemistry

Biological factors:
- Food supply
- Competition
- Dispersal
- Diseases

- habitats
- Thermal regime

Species occurrence
A growing tool box

◆ Individual-species-based
  1) Generalized Linear Model (GLM)
  2) Generalized Additive Model (GAM)
  3) Genetic Algorithm for Rule-set Production (GARP)
  4) Random Forests (RF)
  5) Maximum Entropy (Maxent)

◆ Assemblage-based
  1) River InVertebrate Prediction and Assessment System (RIVPACS)
  2) Analysis of Nearest Neighbor Assessment (ANNA)
  3) Multivariate Adaptive Regression Splines (MARS)
  4) Generalized Dissimilarity Model (GDM)
A example based on Random Forests

Species: *Agnetina capitata*

Data: 194 sites in IL

Method: Random Forests
<table>
<thead>
<tr>
<th>Predicted</th>
<th>Observed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Present</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Absent</td>
<td>5</td>
<td>171</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>177</td>
</tr>
</tbody>
</table>

**Performances:**

- ✓ Correct Classification Rate: 0.94
- ✓ Sensitivity: 0.71
- ✓ Specificity: 0.97
- ✓ Kappa Value: 0.65
## Top 10 Important variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Absence</th>
<th>Presence</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual_mean_precip</td>
<td>3.35</td>
<td>10.91</td>
<td>4.04</td>
</tr>
<tr>
<td>Annual_mean_temp</td>
<td>3.16</td>
<td>9.22</td>
<td>3.65</td>
</tr>
<tr>
<td>Latitude</td>
<td>3.62</td>
<td>1.67</td>
<td>3.62</td>
</tr>
<tr>
<td>Annual_growth_degree_day</td>
<td>2.96</td>
<td>8.06</td>
<td>3.38</td>
</tr>
<tr>
<td>Mean_July_temp</td>
<td>3.00</td>
<td>6.12</td>
<td>3.28</td>
</tr>
<tr>
<td>Coarse_moraine (%)</td>
<td>2.26</td>
<td>10.10</td>
<td>3.14</td>
</tr>
<tr>
<td>Rock_River_Basin</td>
<td>2.13</td>
<td>9.21</td>
<td>2.74</td>
</tr>
<tr>
<td>Barren (%)</td>
<td>2.58</td>
<td>4.70</td>
<td>2.74</td>
</tr>
<tr>
<td>Alluvium_fluvial (%)</td>
<td>2.35</td>
<td>3.36</td>
<td>2.43</td>
</tr>
<tr>
<td>Bedrock (%)</td>
<td>1.51</td>
<td>6.07</td>
<td>2.31</td>
</tr>
</tbody>
</table>
Partial Dependence

(\text{Logit of probability})^2

Annual Precipitation (mm)

900 1000 1100 1200

Mean Annual Temp (°C)

8 9 10 11 12 13 14

-2.2 -1.8 -1.4 -1.0

-2.2 -1.8 -1.4 -1.0
Future Modeling Efforts

- Natural distribution of native fish, mussel, and stonefly species in IL and Midwest
- Impact of climate changes on aquatic species distributions
- Spreads of invasive aquatic species
Incorporating more data

- **Flow regime** (e.g., biologically-important descriptors, resolution, **historic flow**)

- **Geology data** – quantitatively summarized in a biologically meaningful way (e.g., water chemistry, substrates, sediments)

- **Climate Change** (e.g., max and min temp, seasonality, the number of growth-degree days)
Thanks!