What can predictive mapping tell us about the ecology of vector-borne diseases?

Michelle V. Evans

MPE 2013+ Workshop on Global Change and Vector-borne Disease
Mechanistic

- Based on a known relationship between environmental variables and species physiology
- Requires empirical data on physiological responses to temperature (or other environmental variable)
- Often used to extrapolate in the case of climate change or novel environments

Correlative

- Infers a relationship between data and environmental variables (black box)
- Requires species occurrence or disease incidence data, which is generally widely available
- Not appropriate for extrapolation (depending on who you talk to)
Data-driven predictions of vector-borne diseases

- Strong relationship between environmental variables (temperature, precipitation) and mosquito dynamics
- Remotely sensed data is increasing in both temporal and spatial resolutions all the time
- Disease data can be coarse (country or state level)
- Models can be updated in near real-time as new data comes in
Using bagging to deal with sparse datasets

Bagging = bootstrap aggregating

PROS:
- Reduces overfitting
- Very flexible for non-linear relationships
- Limits bias-variance trade-off
Take your average dataset...

<table>
<thead>
<tr>
<th>Presence</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.06</td>
<td>0.1</td>
<td>-2.34</td>
<td>0.27</td>
<td>0.99</td>
</tr>
<tr>
<td>1</td>
<td>2.72</td>
<td>-2.09</td>
<td>0.25</td>
<td>-2.86</td>
<td>1.55</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>1.67</td>
<td>1.83</td>
<td>0.42</td>
<td>-2.86</td>
</tr>
<tr>
<td>0</td>
<td>-0.24</td>
<td>0.84</td>
<td>-0.94</td>
<td>2.92</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>-0.98</td>
<td>0.34</td>
<td>0.81</td>
<td>0.27</td>
</tr>
</tbody>
</table>
Conventional bagging

<table>
<thead>
<tr>
<th>Presence</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.06</td>
<td>0.1</td>
<td>-2.34</td>
<td>0.27</td>
<td>0.99</td>
</tr>
<tr>
<td>1</td>
<td>2.72</td>
<td>-2.09</td>
<td>0.25</td>
<td>-2.86</td>
<td>1.55</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>1.67</td>
<td>1.83</td>
<td>0.42</td>
<td>-2.86</td>
</tr>
<tr>
<td>0</td>
<td>-0.24</td>
<td>0.84</td>
<td>-0.94</td>
<td>2.92</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>-0.98</td>
<td>0.34</td>
<td>0.81</td>
<td>0.27</td>
</tr>
<tr>
<td>Presence</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>0</td>
<td>1.06</td>
<td>0.1</td>
<td>-2.34</td>
<td>0.27</td>
<td>0.99</td>
</tr>
<tr>
<td>1</td>
<td>2.72</td>
<td>-2.09</td>
<td>0.25</td>
<td>-2.86</td>
<td>1.55</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>1.67</td>
<td>1.83</td>
<td>0.42</td>
<td>-2.86</td>
</tr>
<tr>
<td>0</td>
<td>-0.24</td>
<td>0.84</td>
<td>-0.94</td>
<td>2.92</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>-0.98</td>
<td>0.34</td>
<td>0.81</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Range bagging

Drake 2015
Bagging in action... for yellow fever
Bagging in action... for yellow fever

Reni Kaul
Yellow fever outbreak of 2017 in Brazil

About 80% of the confirmed cases and deaths are concentrated in Minas Gerais.
Yellow fever outbreak of 2017 in Brazil
What is the monthly risk of yellow fever spillover across Brazil? Can we predict it?

What are the drivers of spillover events?
Covariate creation based on *a priori* knowledge

Image Credit: Council on Foreign Relations
Covariate creation based on *a priori* knowledge

- Temperature
- Rainfall
- Vector Ranges
- NDVI
- Primate Ranges
- Population Density
- Fire Density
- Agricultural – NHP Overlap

Schmidt et al. 2017
Covariate creation based on *a priori* knowledge

Temperature
Rainfall
Vector Ranges

NDVI
Primate Ranges

Population Density
Fire Density
Agricultural – NHP Overlap

**Environmental anomalies that trigger spillover**

- Temperature
- Rainfall
- NDVI
- Fire Density

Scaled to maximum for that calendar month
Bagged logistic regression

Presence

Background

subset

10

100

Final model has 500 bags

logistic regression

all covariates
Multiple models to explain regional processes
Multiple models to explain regional processes
Low richness region nearly identical to national model

Seasonal trends differ across models

Regional model performs marginally better
Ranked Variable Importance

1. Mean Rainfall
2. NHP Richness
3. Mean Temperature
4. Fire Density
5. Scaled Mean Rainfall
6. Vector Occurrence
7. NHP Agriculture Overlap
8. NDVI

National
Ranked Variable Importance

1. Mean Rainfall
2. NHP Richness
3. Mean Temperature
4. Fire Density
5. Scaled Mean Rainfall
6. Vector Occurrence
7. NHP Agriculture Overlap
8. NDVI

National - Low Richness
Ranked Variable Importance

1. Mean Rainfall
2. NHP Richness
3. Mean Temperature
4. Fire Density
5. Scaled Mean Rainfall
6. Vector Occurrence
7. NHP Agriculture Overlap
8. NDVI

National, Low Richness, High Richness
Ranked Variable Importance

National

Low Richness

High Richness

Mean Rainfall

NHP Richness

Mean Temperature

Fire Density

Scaled Mean Rainfall

Scaled Fire Density

Vector Occurrence

Scaled Mean Temperature

NHP Agriculture Overlap

Population Density

NDVI

Mean Temperature

Scaled NDVI

Mean Rainfall

Population Density
Ranked Variable Importance

National

Low Richness

High Richness

Mean Rainfall

NHP Richness

NDVI

Fire Density

Scaled Fire Density

Scaled Mean Rainfall

Vector Occurrence

NHP Agriculture Overlap

Mean Temperature

Population Density

Scaled NDVI

Mean Rainfall
What did we learn?

- One model may not fit all over a large and varied geographic area
- Drivers of yellow fever differ in the Amazon basin and more populated coastal regions
  - Spillover in the Amazon is ‘triggered’ by encroachment events and environmental anomalies
  - Spillover in Southeastern region is driven by mosquito population dynamics
- Correlative models can work in tandem with empirical work and mechanistic models, highlighting future areas of study
I got 99 problems and data are all of them...

- Data wasn’t easily available (relatively), and only at municipality level
- Missing fine-scale data on vaccination
- Issue of scale-mismatch between environmental and disease data (temporal and spatial)

These can be even worse for other diseases in countries with less health reporting infrastructure
This bad boy can answer so many questions about emerging arboviruses.
Crowd-sourced Zika data for S. and Central America

- Organized by CDC researchers
- Data collected from Ministry of Health and PAHO documents
- Date range: 2016 – Present
- Fifteen countries
- Currently in the process of georeferencing
- One issue: lack of standardization across countries

Check it out: github.com/cdcepi/zika