Leads: Matt Cooper, Tabby Njung’e
Data Scientists: Anthony Arendt, Joe Hellerstein
Fellows: Cara Arizmendi, Mitch Goist, Krista Jones, Rob Shaffer
Approach
Environmental, Social, and Agricultural Data

- Environmental plot
  - Ground Cover: \( n \approx 1350 \)
  - Trees: \( n \approx 1350 \)

- Expenditures: \( n \approx 750 \)
- Assets: \( n \approx 750 \)

- Soil Lab
- Crop Yields Analysis

- Households
  - Events
  - Individuals: \( n \approx 2250 \)

- Fields
  - Infrastructure
  - Inputs
Data Pipeline
DETECTION-SUPPORT DASHBOARD
a small set of indicators for smart agricultural investment

ANALYSIS
mathematical models and algorithms are applied

MEASUREMENT
consistent metrics are gathered on the ground and remotely via satellites
Water Insecurity Model

VS Survey data
- Reported water insecurity at household level

Rasters
- Elevation
- Travel distance
- Demographics
- Climate

Watersheds
- % Forest
- % Cropland
- % Protected Area
- % Degraded Land

Survey data:

Reported water insecurity at household level

Rasters:

- Elevation
- Travel distance
- Demographics
- Climate

Watersheds:

- % Forest
- % Cropland
- % Protected Area
- % Degraded Land
Modeling Strategy

Why random forests?
- Large feature space
- Non-parametric
- Minimize prediction error

Problem: imbalanced classes
- False negatives
- Misleading accuracy

Regular Random Forest

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure</td>
<td>8722</td>
</tr>
<tr>
<td></td>
<td>254</td>
</tr>
<tr>
<td>Insecure</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

With SMOTE

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure</td>
<td>7625</td>
</tr>
<tr>
<td></td>
<td>45</td>
</tr>
<tr>
<td>Insecure</td>
<td>1097</td>
</tr>
<tr>
<td></td>
<td>209</td>
</tr>
</tbody>
</table>
Visualizing Water Insecurity
Water Insecurity Model

Water Insecurity = Rasters + Watersheds + Households
Incorporating Vital Signs Indicators

<table>
<thead>
<tr>
<th></th>
<th>True Positive</th>
<th>True Negative</th>
<th>F1 Score</th>
<th>Balanced Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>With VS Predictors</td>
<td>0.91</td>
<td>0.69</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td>Without VS Predictors</td>
<td>0.89</td>
<td>0.65</td>
<td>0.93</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Conclusion

Broader goals for the Vital Signs summer project:

- **Identify** social, environmental, and agricultural synergies and tradeoffs in rural Africa
- Answer these questions by **combining** data from a variety of sources

**Takeaway:** real-world policy questions require interdisciplinary data!