Using Parsl to Handle Large Agroecosystem Data

Douglas N Friedel
National Center for Supercomputing Applications, University of Illinois
The Team

Dept. of Natural Resources and Env. Science, Univ. of Illinois
Prof. Kaiyu Guan
Sheng Wang
Qu Zhou
Chenhui Zhang

National Center for Supercomputing Applications, Univ. of Illinois
Daniel S. Katz
Douglas N. Friedel
# Agroecosystem Monitoring & Data

- Obtain high quality “ground truth” data on crop distribution and health
- Build system from small scale up to enable satellite-based monitoring
  - From cm to km size resolution
- Meter scale and larger resolution data are largely missing

<table>
<thead>
<tr>
<th>Human</th>
<th>UAV</th>
<th>Aircraft</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td>cm-dm</td>
<td>dm-m</td>
<td>m-km</td>
<td>km</td>
</tr>
<tr>
<td>Leaf</td>
<td>Field</td>
<td>Landscape</td>
<td>Global</td>
</tr>
</tbody>
</table>
Data Collection

Mounted a hyperspectral imaging system on a small plane and fly it across fields in Illinois

<table>
<thead>
<tr>
<th>Spectral Range</th>
<th>Spectral Resolution</th>
<th>Spatial Resolution</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>400-2500 nm</td>
<td>3 - 5 nm</td>
<td>0.5 m</td>
<td>Hyperspectral reflectance</td>
</tr>
</tbody>
</table>
Data

- ~470 Spectral channels
- ~1500 pixels in the imaging array
- Image every 0.5 m for as long as the flight lasts
- 40,000 field-acres per hour
- ~2.6 MB per scan or ~5.3 GB per km
  - Data volumes quickly get very large

Source: NEON
Data Processing

[Flowchart of Data Processing]

August 6th, 2020

SoyFACE
Removing Noise

The spectra should be smooth, but sometimes there is noise.
Removing Noise with Parsl

- Single spectrum takes a fraction of a sec
- But a small run produces a 27 GB file (6300x3200x325)
- Break data up, each chunk is processed by a different Parsl ‘job’
- Reduced runtime by ~80% (single core vs 24 core machine)
Acknowledgements

- Presentation template by SlidesCarnival
- Photographs by Unsplash