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

National Center for Supercomputing Applications, Univ. of Illinois

Prof. Kaiyu Guan

Sheng Wang

Qu Zhou

Chenhui Zhang

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

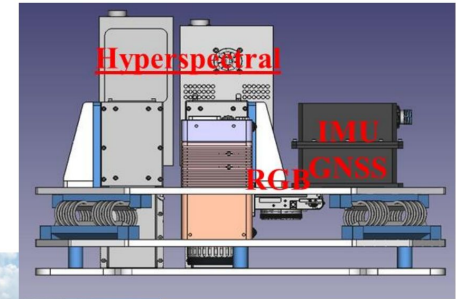| Human	UAV	Aircraft	Satelite
cm-dm	dm-m	m-km	km
Leaf	Field	Landscape	Global

Data Collection

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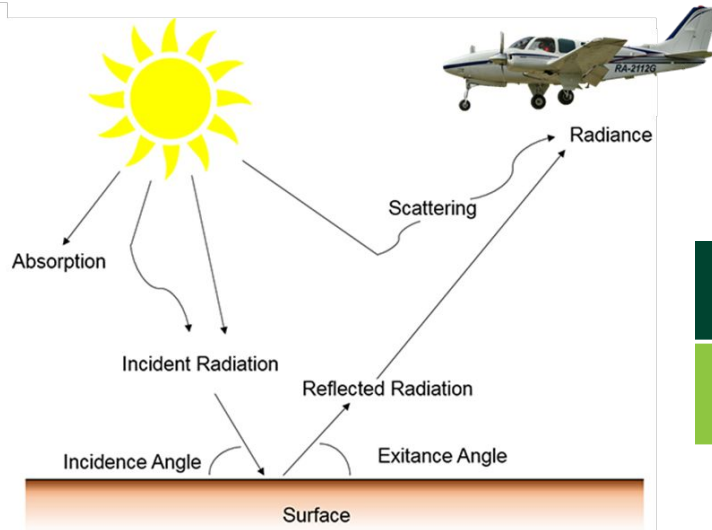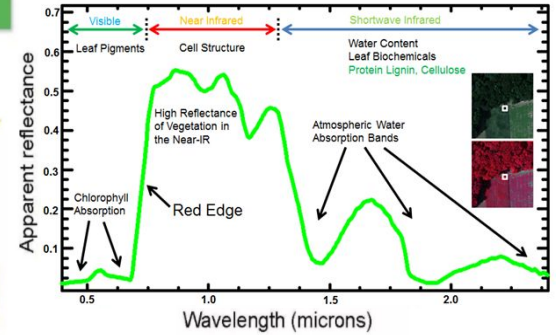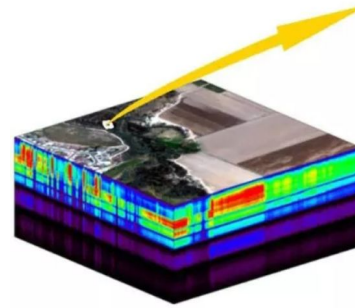

Hyperspectral payload

Aircraft:
Local vendor

Spectral Range	Spectral Resolution	Spatial Resolution	Data
400-2500 nm	3 - 5 nm	0.5 m	Hyperspectral reflectance

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
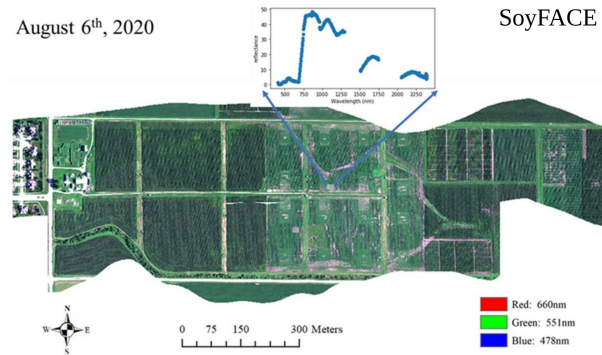 - Data volumes quickly get very large

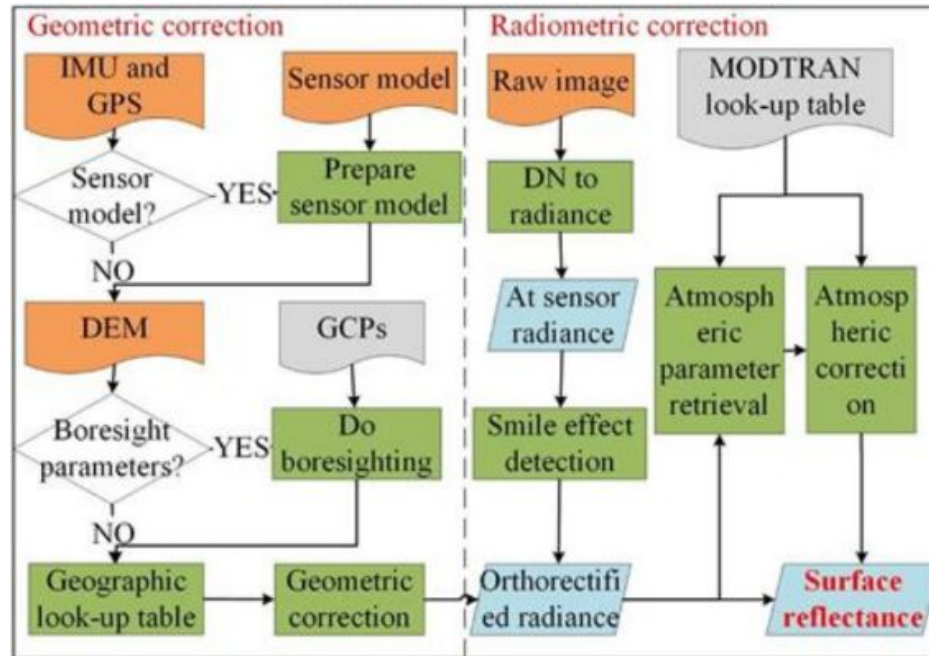Hyperspectral remote sensing

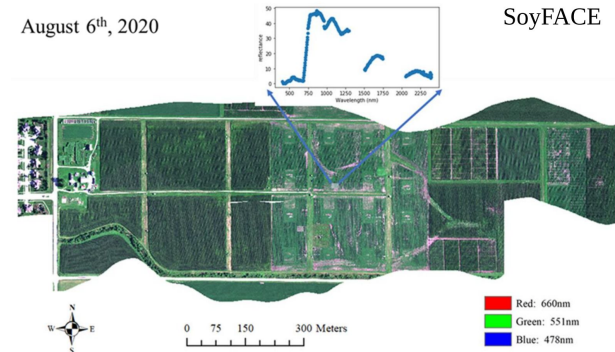Source: NEON

August 6th, 2020

SoyFACE

Red: 660nm
Green: 551nm
Blue: 478nm

0   75   150   300 Meters

Data Processing

August 6th, 2020   SoyFACE

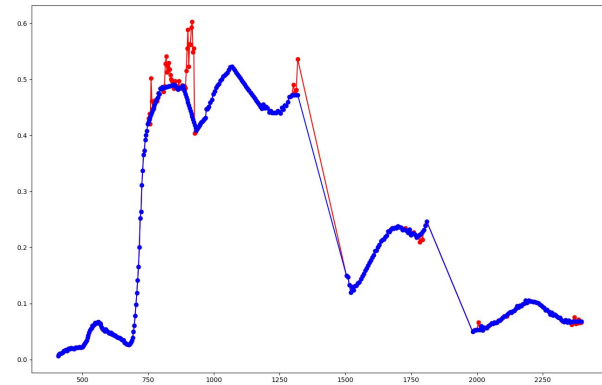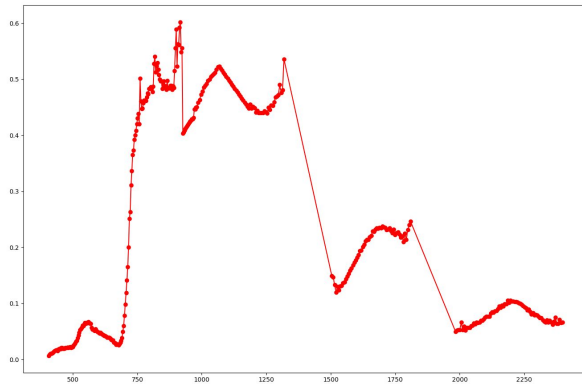0   75   150   300 Meters

Red: 660nm
Green: 551nm
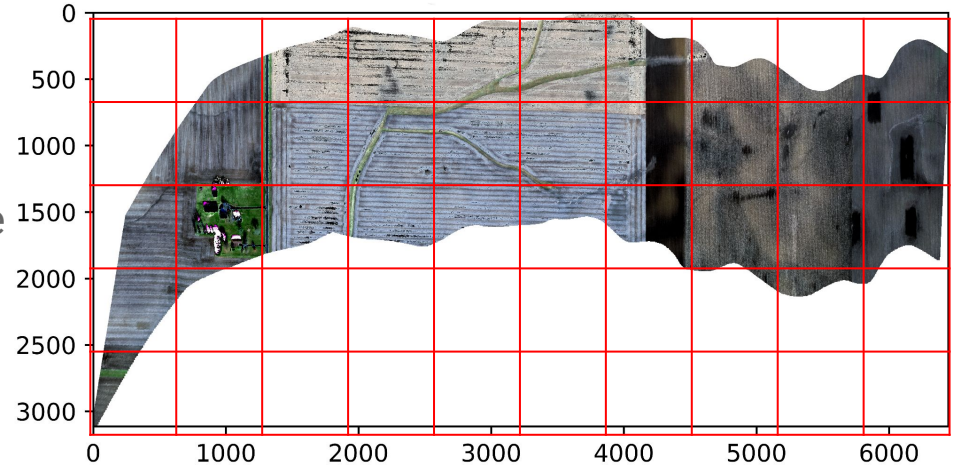Blue: 478nm

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