NASA and NEON Airborne Imaging Spectroscopy: Characterizing Fine-Scale Vegetation Function at the Continental Scale

Phil Townsend
University of Wisconsin-Madison
<table>
<thead>
<tr>
<th>Traits</th>
<th>$R^2$ (val)</th>
<th>% RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Carbon</td>
<td>0.63</td>
<td>13.06</td>
</tr>
<tr>
<td>Aluminum</td>
<td>0.55</td>
<td>10.4</td>
</tr>
<tr>
<td>Carotenoids (area)</td>
<td>0.57</td>
<td>11.92</td>
</tr>
<tr>
<td>Carotenoids (mass)</td>
<td>0.59</td>
<td>13.3</td>
</tr>
<tr>
<td>Cellulose</td>
<td>0.6</td>
<td>12.11</td>
</tr>
<tr>
<td>Chlorophyll (area)</td>
<td>0.64</td>
<td>11.14</td>
</tr>
<tr>
<td>Chlorophyll (mass)</td>
<td>0.58</td>
<td>13.85</td>
</tr>
<tr>
<td>Isotopic $\delta^{13}C$</td>
<td>0.63</td>
<td>13.12</td>
</tr>
<tr>
<td>EWT</td>
<td>0.62</td>
<td>14.13</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>0.53</td>
<td>14.5</td>
</tr>
<tr>
<td>Iron</td>
<td>0.51</td>
<td>15.03</td>
</tr>
<tr>
<td>Lignin</td>
<td>0.57</td>
<td>13.37</td>
</tr>
<tr>
<td>Leaf Mass per Area</td>
<td>0.78</td>
<td>10.2</td>
</tr>
<tr>
<td>Manganese</td>
<td>0.62</td>
<td>12.99</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>0.55</td>
<td>13.44</td>
</tr>
<tr>
<td>Phosphorous</td>
<td>0.5</td>
<td>12.93</td>
</tr>
<tr>
<td>Potassium</td>
<td>0.66</td>
<td>10.74</td>
</tr>
<tr>
<td>Starch</td>
<td>0.61</td>
<td>12.73</td>
</tr>
<tr>
<td>Sugar</td>
<td>0.45</td>
<td>16.37</td>
</tr>
<tr>
<td>Phenolics</td>
<td>0.82</td>
<td>8.4</td>
</tr>
<tr>
<td>Water (percent)</td>
<td>0.58</td>
<td>11.86</td>
</tr>
<tr>
<td>Zinc</td>
<td>0.54</td>
<td>16.12</td>
</tr>
</tbody>
</table>
ABoVE
Arctic-Boreal Vulnerability Experiment

AVIRIS-NG imaging + ground sampling

Figure: Ryan Pavlick, Morgan Dean and Kyle Kovach
Which airborne data sets do you use?

• AVIRIS-NG
• AVIRIS-Classic
• MASTER
• NEON AOP
• My lab generates its own imagery from NASA-funded research:
  • Cessna HySpex
  • UAV HySpex Mjolnir (heavy-lift)
  • UAV lidar
  • Various UAV multispectral and FLIR
• We also generate products that we wish to distribute (or are funded to):
  • ABoVE: BRDF corrections, trait maps
How do you access this data?

• In the past:
  • Receive hard drives
  • Direct download (AVIRIS from JPL)

• Currently:
  • Via API (NEON)

• Our own data are stored on local servers or in the cloud

• Issue with all of this:
  • Data storage costs
  • We largely work in code and only store final products

• We have not yet provided data to a DAAC to distribute
  • JPL connection
How are you using this data?

• We link field and image data to calibrate and validate models/maps of foliar functional traits (chemical, physiological, morphological), functional diversity, and other properties, e.g., related to stress, pests and pathogens.

• We have an extensive Python image processing workflow (HyTools) that we use to generate products (e.g., BRDF corrections, traits):
  • [https://github.com/EnSpec/hytools](https://github.com/EnSpec/hytools)
  • Has been adopted by quite a number of people for processing AVIRIS-NG imagery

• Airborne products to baseline/validate satellite products
  • Global scale functional trait mapping (SBG)
What works well for you

• We tend to work with “all” images
• Accessing data through API (NEON)
• Direct downloading is also fine
• High-throughput computing on our end
• Downloading datasets with the least amount of authentication, and without any GUI (batch processing)
What pain points do you find with data access or data use?

- The data volume is huge. It is difficult to download only a small spatial subset from a huge dataset. Sometimes, the product we need is bundled with other products, maybe for better organization. That makes us download many datasets we don’t need.
  - Need to download the entire tarred radiance data to get OBS/ORT files (which are small)
  - Data subsetting when accessing (cloud/on-demand processing would solve)
- Consistent data structure is preferred (all JPL stuff, may not be a DAAC issue)
  - Data type
  - Interleaving
  - Float vs integer
  - Nodata values
- Updates to data (versioning, notifications on recent or forthcoming updates)
  - Versioning issues: new processing implemented for one data set but older data not reprocessed
- Tracking processing workflows and provenance
  - Citations are not always complete
  - Code usually not open-source and therefore cannot be replicated
  - Not a DAAC problem, but data storage (hot, warm, cold) is costly with large volumes of data
  - Cloud storage
- On-demand processing / cloud computing options
What do you wish you could do but can’t?

• Inexpensive cloud storage
• On-demand processing
• Cloud computing
• We are well set up to process large volumes of data: my group has options
  • Others are not – there are barriers to access
Do you have any suggestions for improvement?

• Address the pain points 2 slides back
• Easily discovered documentation, especially for derived products
• Quality control information front and center (so we don’t waste our time, and exclude it when automating processing)
• We need to develop tools that lower barriers to access for non-expert users.
• Specifically, we need to distribute products that are more developed/mature for the non-experts (example BRDF)
Have you tried to use data in the cloud?

• Yes
Have you tried to use data in the cloud?

- Yes
- Well, no, not me personally in any meaningful way, but my lab does extensively
- Most of the data we use does not have this option
  - GEE gets modest use
- Note: development of GeoSPEC for on-demand processing
What support do you need from ADMG?

• I do not know what support you provide!
• Distribution of data we generate in support of NASA funded research
• Tutorials for data handling
  • Code repositories, jupyter notebooks
  • Readthedocs
  • Etc.
Thank you for asking!