The Python Spectral Analysis Tool (PySAT): Powerful, Flexible, and Easy Preprocessing and Machine Learning with Point Spectral Data

Ryan Anderson¹, Nicholas Finch¹, Sam Clegg², Trevor Graff³, Dick Morris³, Jay Laura¹

¹USGS Astrogeology, ²Los Alamos, ³NASA JSC
What is PySAT?

◊ A Python library for spectral analysis

◊ Two parallel PySAT projects led by USGS
  ◊ Gaddis, Laura et al.: orbital spectrometers (e.g. M3, CRISM, etc.)
  ◊ Anderson et al.: point spectrometers (e.g. ChemCam, SuperCam, etc.)

◊ Shared back-end:
  ◊ https://github.com/USGS-Astrogeology/PySAT

◊ Point Spectra GUI:
  ◊ https://github.com/USGS-Astrogeology/PySAT_Point_Spectra_GUI
Why is PySAT Necessary?

worm Interpreting spectral data is vital, but difficult.
  ◇ Even for instrument team members, it is hard to test new processing and analysis methods.
worm Scientists end up writing their own code.
  ◇ Often reinvent the wheel
  ◇ Often end up using simple analysis methods out of necessity
worm Commercial options are often expensive, proprietary, inflexible.
worm We want to remove these barriers.
PySAT Point Spectra GUI

- Python-based tool for preprocessing and analyzing point spectra
  - Free
  - Open-source
  - Powerful
  - Flexible
  - User friendly

- We want to enable planetary scientists to process and analyze point spectra without specialized programming expertise.
  - PyQt5-based GUI
  - Leverage scikit-learn to provide machine learning methods
Data Format

- .csv files with dual-level column labels (read into a multi-indexed Pandas data frame)
  - The tool can read ChemCam “CCS” data on the PDS into this format.

- Standard top-level labels are:
  - “meta” = metadata
  - “comp” = compositional data
  - “wvl” = spectral data

- Others are added to record additional information (e.g. PCA scores, regression predictions, etc.)

<table>
<thead>
<tr>
<th>Sample</th>
<th>meta</th>
<th>meta</th>
<th>comp SiO2</th>
<th>comp TiO2</th>
<th>comp Al2O3</th>
<th>wvl 224.336</th>
<th>wvl 224.391</th>
<th>wvl 224.446</th>
<th>wvl 224.501</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>-23.17</td>
<td>-23.17</td>
<td>-23.17</td>
<td>3.83</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>11.17</td>
<td>11.17</td>
<td>11.17</td>
<td>-23.83</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>7.67</td>
<td>7.67</td>
<td>7.67</td>
<td>-7.33</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>7.94</td>
<td>7.94</td>
<td>7.94</td>
<td>-16.06</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>14.5</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>2.78</td>
<td>2.78</td>
<td>2.78</td>
<td>-2.22</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>19.11</td>
<td>19.11</td>
<td>19.11</td>
<td>10.11</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>-11.56</td>
<td>-11.56</td>
<td>-11.56</td>
<td>10.44</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>-3.61</td>
<td>-3.61</td>
<td>-3.61</td>
<td>11.39</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>-5.39</td>
<td>-5.39</td>
<td>-5.39</td>
<td>23.61</td>
</tr>
<tr>
<td>CP-5, Basalt / AREF019</td>
<td>JSC-0270</td>
<td>JSC1416</td>
<td>54.06</td>
<td>2.27</td>
<td>13.17</td>
<td>-6.28</td>
<td>-6.28</td>
<td>-6.28</td>
<td>23.72</td>
</tr>
</tbody>
</table>
Interface and “Workflows”

✧ The PySAT GUI is modular for maximum flexibility:
  ✧ User can specify different steps in whatever order desired (within reason)
  ✧ We call these steps “modules”
  ✧ We call a series of steps a “workflow”
  ✧ Modules can be inserted and deleted
  ✧ Information can be passed from one module to the next without having to run the workflow
  ✧ Workflows can be saved and restored using .json
Data Transformation

- Interpolate data onto new wavelengths
- Masking
- Normalization
- Derivative
- Multiply by vector
- Peak Area Binning
- Baseline removal (9 algorithms available)
- Dimensionality reduction
  - PCA, ICA, t-SNE, LLE
Data Manipulation and Grouping

- Data Manipulation
  - Remove rows
  - Split data
  - Combine data sets

- Outlier removal
  - Isolation Forest
  - Local Outlier Factor

- Clustering
  - K-Means
  - Spectral
  - More to come

- Stratified Folds
Regression

- K-fold cross validation over any parameters
- Multivariate regression:
  - Ordinary Least Squares (OLS)
  - Partial Least Squares (PLS)
  - Gaussian Process (GP)
  - Support Vector Machines (SVM)
  - Bayesian Ridge Regression (BRR)
  - Lasso
  - Elastic net
  - Orthogonal Matching Pursuit (OMP)
  - Least Angle Regression (LARS)
  - Automatic Relevance Determination (ARD)
- Blend “sub-models”
- Save/restore trained models
Conclusion

◊ PySAT point spectra tool is a powerful and flexible spectral processing and regression software
  ◊ Let scientists spend their time analyzing data rather than writing code.

◊ Future work
  ◊ Calibration transfer
  ◊ Additional clustering and classification
  ◊ Additional regression
    ◊ Ensemble methods (e.g. bagging, boosting, etc.)
    ◊ Local methods
  ◊ Better plotting
  ◊ Debugging and testing
  ◊ Documentation

◊ Although designed with LIBS in mind, it is flexible enough for most spectral work (and even suitable for some non-spectral data!)
  ◊ Interested, but not sure how to use PySAT for your work? Let’s talk!
  ◊ Is PySAT missing some capability you would use? Let me know!
PySAT Point Spectra Tool

Back end:
https://github.com/USGS-Astrogeology/PySAT

GUI:
https://github.com/USGS-Astrogeology/PySAT_Point_Spectra_GUI

Contact me:
Ryan Anderson
rbanderson@usgs.gov