

# unidentified-emission-lines

May 11, 2018

## 1 Catalog of Unidentified Cometary Emission Lines

**Authors:** Simoneva et al.

**Reviewed by:** Michael S. P. Kelley

**For:** NASA Planetary Data System Small Bodies Node

**Date:** 2018 May 15

### 1.1 Data set summary

- Lists of unidentified lines or molecular bands for 12 comets based on a literature search.
- References span from 1944 to 2007.
- Comets are a mix of Oort cloud, Halley-type, and Jupiter-family.
- Source spectra span a wide range of properties: resolving powers from ~2500 to ~60,000.
- Wavelengths are primarily optical, but some UV and near-IR.

### 1.2 Working with the data

#### 1.2.1 Plot one set of lines

```
In [1]: from pds4_tools import pds4_read
```

```
data = pds4_read('unid-emis/109pswift_tuttle_1.xml', quiet=True)
data.info()
```

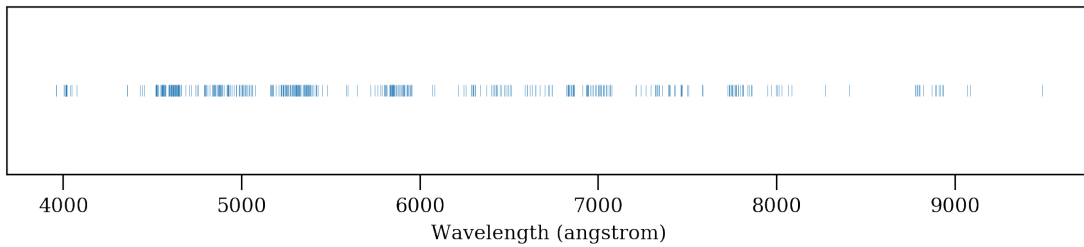
```
0  Header          'HEADER_0'  UTF-8 Text
1  Table_Delimited 'TABLE_0'   1 fields x 438 records
```

```
In [2]: %matplotlib inline
```

```
import numpy as np
import matplotlib.pyplot as plt

tab = data['TABLE_0']
lines = np.array(tab.fields[0])
unit = tab.label.find('Record_Delimited/Field_Delimited/unit').text
plt.figure(1, figsize=(8, 2), dpi=200).clear()
plt.scatter(lines, np.zeros_like(lines), marker='|', lw=0.2)
```

```
plt.setp(plt.gca(), xlabel='Wavelength ({}).format(unit)', yticks=[])
plt.tight_layout()
```



### 1.3 Find common lines

Find lines common to multiple comets. For the purposes of this task, we will assume features within of 0.05 Å are the same line.

#### 1.3.1 Load all files

```
In [3]: from glob import glob
```

```
lines = []
n_files, n_lines = 0, 0
for fn in glob('unid-emis/*.xml'):
    if fn != 'unid-emis/collection.xml':
        data = pds4_read(fn, quiet=True)
        comet = data.label.find('./Target_Identification/name').text
        this_set = data['TABLE_0'].fields[0]
        lines.extend(this_set)
        n_lines += len(this_set)
        n_files += 1
        print('{}: {} lines'.format(comet, len(this_set)))
print('\n', n_files, 'files,', n_lines, 'lines read.')
```

```
C/1980 Y1 (BRADFIELD): 27 lines
21P/GIACOBINI-ZINNER 1: 20 lines
C/1996 B2 (HYAKUTAKE): 18 lines
153P/IKEYA-ZHANG: 1862 lines
C/1973 E1 (KOHOUTEK): 5 lines
122P/DE VICO 1: 242 lines
19P/BORRELLY 1: 116 lines
109P/SWIFT-TUTTLE 1: 438 lines
23P/BORSEN-METCALF 1: 193 lines
C/2001 A2 (LINEAR): 15 lines
C/1995 O1 (HALE-BOPP): 141 lines
C/1942 X1 (WHIPPLE-FEDTKE-TEVZADZE): 14 lines
```

12 files, 3091 lines read.

## 1.4 Merge lines

```
In [4]: # Angstrom, delta- to use to merge lines
blend = 0.05 # R~50,000 at 4500 Å
merged = np.zeros(n_lines, bool) # to start, nothing is merged

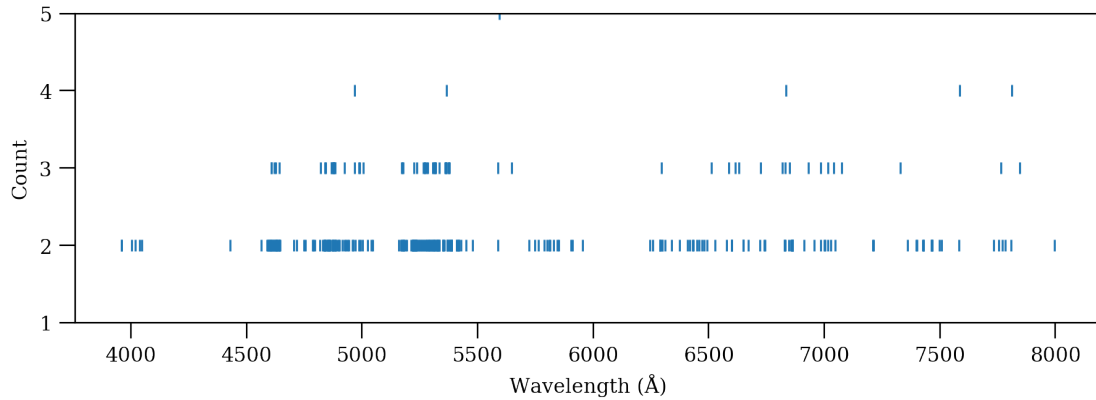
# sort
i = np.argsort(lines)
lines = np.array(lines)[i]

common_lines = []
for i in range(n_lines):
    if merged[i]:
        continue
    j = (lines >= lines[i]) * (lines <= (lines[i] + blend))
    n = np.sum(j)
    if n == 1:
        continue
    common_lines.append((np.mean(lines[j]), n))
    merged[j] = True
print(merged.sum(), 'lines merged into', len(common_lines), 'line groups.')
```

697 lines merged into 310 line groups.

```
In [5]: # Plot
wave = [line[0] for line in common_lines]
count = [line[1] for line in common_lines]

plt.figure(1, figsize=(8, 3), dpi=200).clear()
ax = plt.gca()
plt.scatter(wave, count, marker='|', lw=1)
plt.setp(ax, xlabel='Wavelength (Å)', ylabel='Count', ylim=[1, 5])
plt.tight_layout()
```



### 1.4.1 Comapre to C/1996 B2 (Hyakutake)

Use the on-nucleus spectrum of this comet by A'Hearn et al. 2015, as archived in PDS.

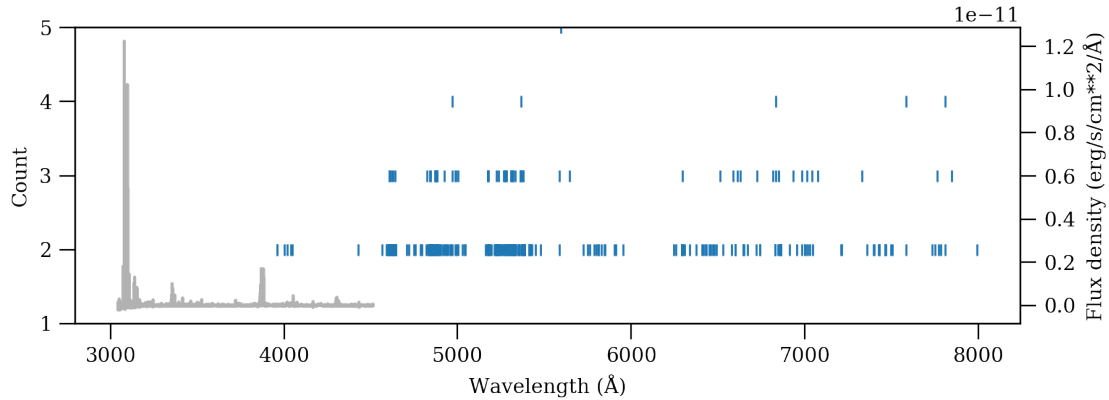
#### Download spectrum from PDS

```
In [6]: from urllib.request import urlretrieve
```

```
label_fn, headers = urlretrieve('https://pdssbn.astro.umd.edu/holdings/pds4-gbo-kpno:hya')
table_fn, headers = urlretrieve('https://pdssbn.astro.umd.edu/holdings/pds4-gbo-kpno:hya')
hya = pds4_read(label_fn)
hya.info()
```

```
0 Table_Character 'Spectrum' 2 fields x 22434 records
```

```
In [7]: plt.figure(1, figsize=(8, 3), dpi=200).clear()
ax = plt.gca()
ax.scatter(wave, count, marker='|', lw=1)
rax = ax.twinx()
plt.plot(hya['Spectrum']['Wavelength'], hya['Spectrum']['Flux Density'], color='k', alpha=0.5)
plt.setp(ax, xlabel='Wavelength (Å)', ylabel='Count', ylim=[1, 5])
plt.setp(rax, ylabel='Flux density ({}).format(hya['Spectrum']['Flux Density']).meta_data)
plt.tight_layout()
```



## 1.5 Review

- Such a collection of unidentified lines can enable their future identification, thus this is a worthwhile dataset to archive.
- The labels are brief but seem to be complete.
- The data are readable and straightforward to understand.
- The file headers have much of the needed metadata for interpreting the results (with some exceptions).

### 1.5.1 Questions and recommended liens

- The heliocentric distance is a relevant parameter that is missing from the files.
- The data set needs some documentation. Some recommended items to include:
  - Please define Reference code (tell the user that this should be ignored), Information bulletin, etc.
  - Please describe the method for the literature search and the criteria used to determine if a line is real and unidentified.
  - Why is SNR == N/A for most? Signal-to-noise ratio should always be applicable in this context.
  - Similar to SNR: Why are some resolving powers == N/A?

**Why is documentation needed?** The PDS archived atlas of Cochran et al. 2002 lists 4055 unidentified lines for comet 122P/de Vico between 3830 and 10192 Å, but this data set lists 242 lines between 3862 and 4187. Without documentation a user cannot determine the origin of the discrepancy.

The references searched span a wide range of time. Lines that were unidentified in 1980 may be identified in 2018. Documentation would also help the user understand if the unidentified lines are simply those that were unidentified at time, or if any currently understood lines have been removed from the lists.

### 1.5.2 PDS4 label questions

- Some of the line lists are in the UV and near-IR. Should the collection also include these wavelength ranges in the Primary\_Result\_Summary? If so, then the individual labels should include their own wavelength\_range to facilitate searching.
- Are PDS4 units case sensitive? angstrom vs. Angstrom
- Are XML entities OK? & found in the labels (within literature references).