How we are using Jupyter Notebooks* in the Northeast U.S. Shelf (NES) LTER

Stace Beaulieu (stace@whoi.edu),

NES-LTER Information Manager at Woods Hole Oceanographic Institution (WHOI)

Coordinator, WHOI Ocean Informatics initiative

Joe Futrelle (jfutrelle@whoi.edu),

Applications Development in WHOI Information Services

** And R Markdown, but our focus today is on Jupyter Notebooks with Python*

We are using Jupyter Notebooks in NES-LTER for:

- Data management:

  https://github.com/WHOIGit/nes-lter-notebooks

- Engaging scientists and students with NES-LTER data (e.g., data analysis and visualization):

  https://github.com/WHOIGit/nes-lter-examples

Example for data management

Challenge: How can we clean ship-provided data when something out-of-the-ordinary occurred during a cruise?

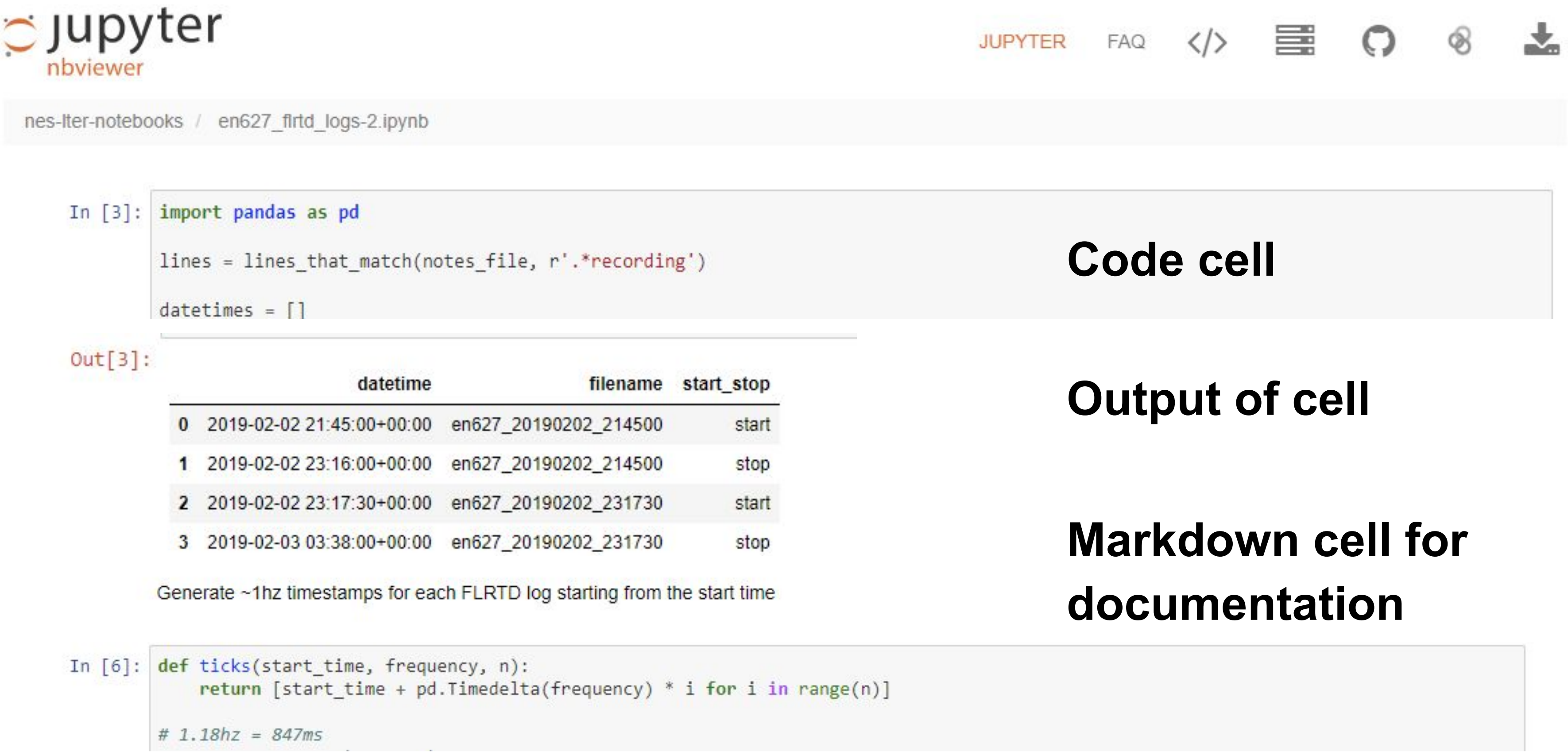
e.g., underway data, CTD data

Why a notebook?

- We do not want to include these specific fixes in our code library.
- We want to record the provenance for these specific fixes.

Example for data management

Need to add data from a particular instrument to the underway data



In [3]:

```
import pandas as pd

lines = lines_that_match(notes_file, r'.*recording')

datetimes = []
```

Code cell

Out[3]:

	datetime	filename	start_stop
0	2019-02-02 21:45:00+00:00	en627_20190202_214500	start
1	2019-02-02 23:16:00+00:00	en627_20190202_214500	stop
2	2019-02-02 23:17:30+00:00	en627_20190202_231730	start
3	2019-02-03 03:38:00+00:00	en627_20190202_231730	stop

Output of cell

Generate ~1hz timestamps for each FLRTD log starting from the start time

Markdown cell for documentation

In [6]:

```
def ticks(start_time, frequency, n):
    return [start_time + pd.Timedelta(frequency) * i for i in range(n)]

# 1.18hz = 847ms
```

https://nbviewer.jupyter.org/github/WHOIGit/nes-lter-notebooks/blob/master/en627_flrtd_logs-2.ipynb

Example for engaging scientists and students with NES-LTER data*

Challenge: How can we help compare their post-cruise analyses of samples with ship-provided data?

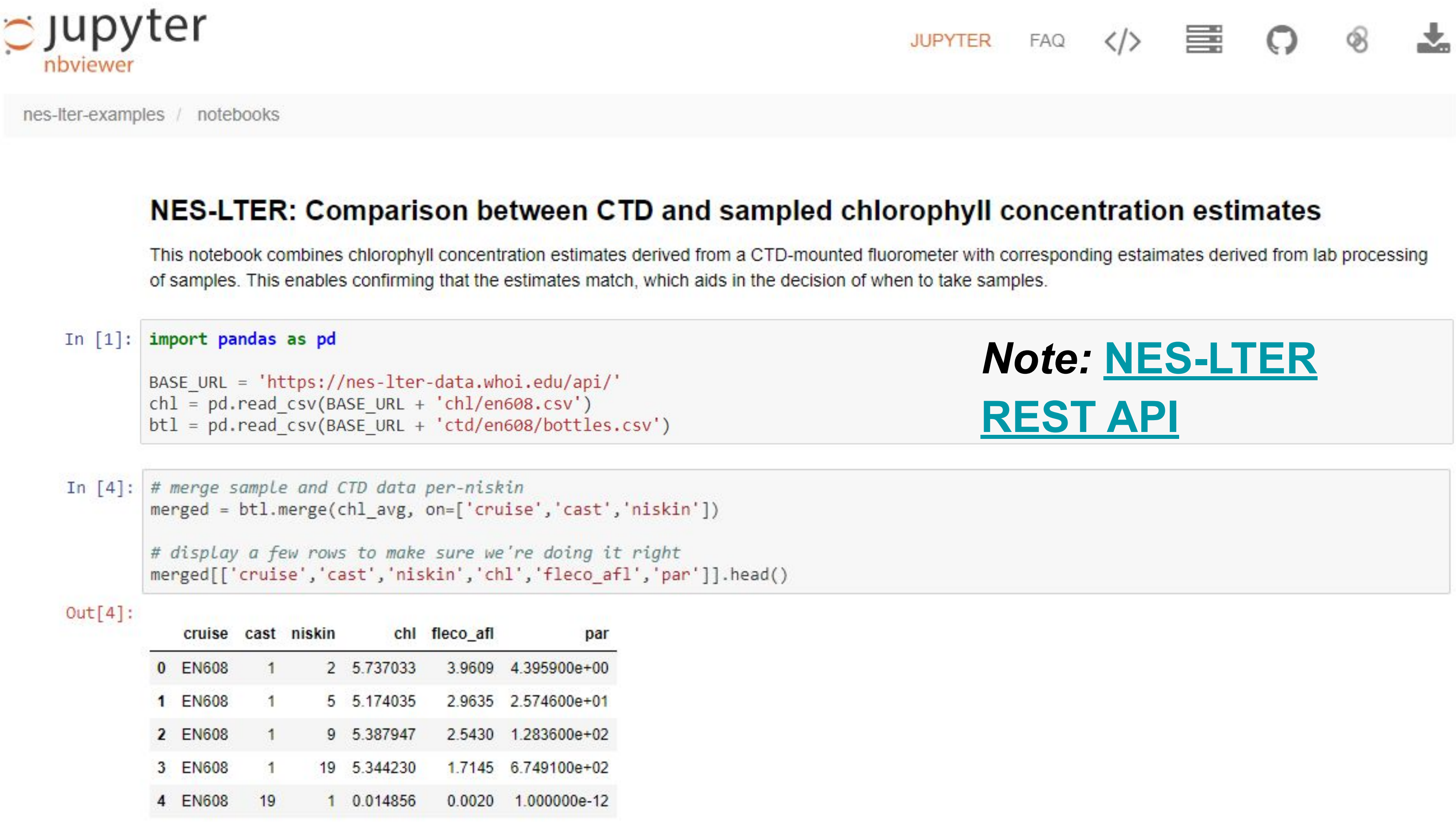
e.g., underway data, CTD data

Why a notebook?

- Nice way to provide code, visualization, and some documentation in a single interface.
- Jupyter Notebooks in particular render nicely in GitHub.

** With R Markdown, we are addressing additional quick visualizations per cruise for PI-provided data. Our EDI Fellow and one of our REUs this summer are developing reproducible workflows.*

Example for engaging scientists and students with NES-LTER data



NES-LTER: Comparison between CTD and sampled chlorophyll concentration estimates

This notebook combines chlorophyll concentration estimates derived from a CTD-mounted fluorometer with corresponding estaimates derived from lab processing of samples. This enables confirming that the estimates match, which aids in the decision of when to take samples.

```
In [1]: import pandas as pd

        BASE_URL = 'https://nes-lter-data.whoi.edu/api/'
        chl = pd.read_csv(BASE_URL + 'chl/en608.csv')
        btl = pd.read_csv(BASE_URL + 'ctd/en608/bottles.csv')
```

Note: **NES-LTER REST API**

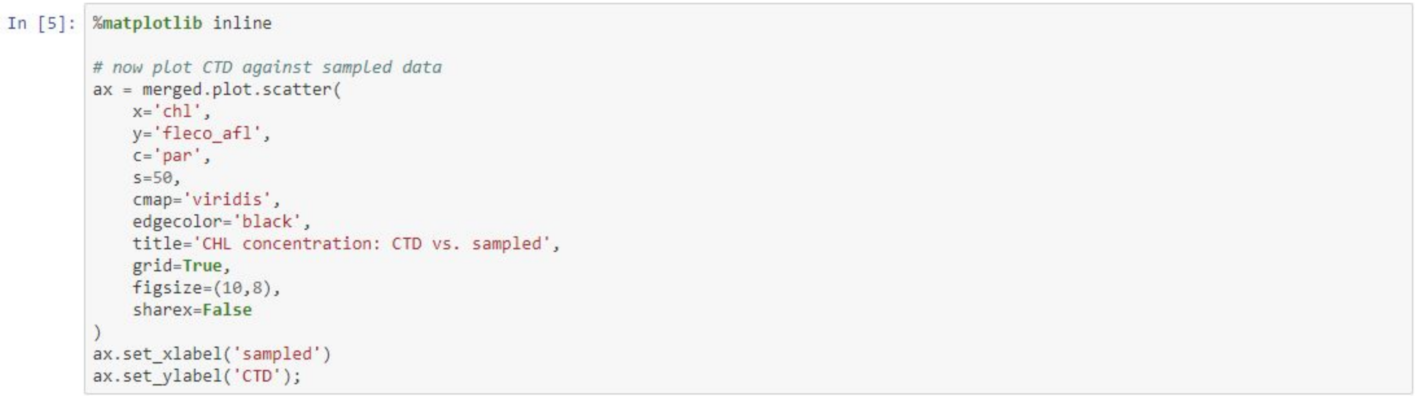
```
In [4]: # merge sample and CTD data per-niskin
        merged = btl.merge(chl_avg, on=['cruise','cast','niskin'])

        # display a few rows to make sure we're doing it right
        merged[['cruise','cast','niskin','chl','fleco_afl','par']].head()
```

Out[4]:

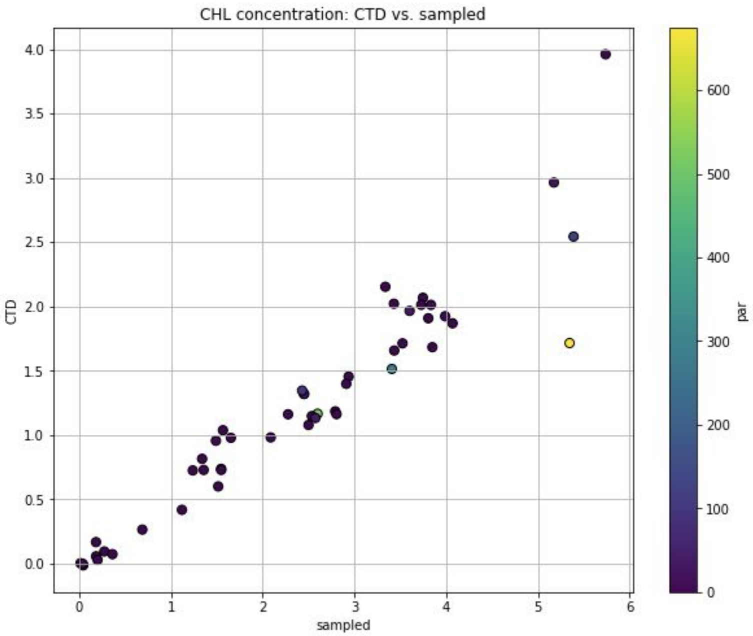
	cruise	cast	niskin	chl	fleco_afl	par
0	EN608	1	2	5.737033	3.9609	4.395900e+00
1	EN608	1	5	5.174035	2.9635	2.574600e+01
2	EN608	1	9	5.387947	2.5430	1.283600e+02
3	EN608	1	19	5.344230	1.7145	6.749100e+02
4	EN608	19	1	0.014856	0.0020	1.000000e-12

https://github.com/WHOIGit/nes-lter-examples/blob/master/notebooks/compare_ctd_chl_api.ipynb

Example for engaging scientists and students with NES-LTER data



```
In [5]: %matplotlib inline

        # now plot CTD against sampled data
        ax = merged.plot.scatter(
            x='chl',
            y='fleco_afl',
            c='par',
            s=50,
            cmap='viridis',
            edgecolor='black',
            title='CHL concentration: CTD vs. sampled',
            grid=True,
            figsize=(10,8),
            sharex=False
        )
        ax.set_xlabel('sampled')
        ax.set_ylabel('CTD');
```



https://github.com/WHOIGit/nes-lter-examples/blob/master/notebooks/compare_ctd_chl_api.ipynb

Strengths… and weaknesses for Jupyter Notebooks with Python

Strengths

- Data management: Keeps special cases out of our code library
- Engaging scientists and students: Proven success with this software with teams of ocean scientists.*

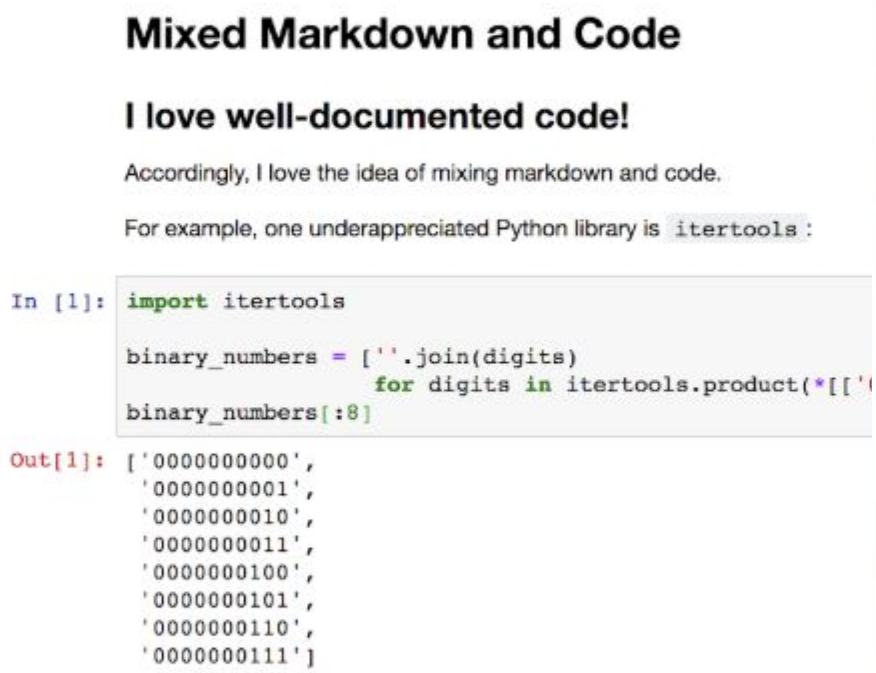
Weaknesses

- Data management: Now we have to manage notebooks!
- Engaging scientists and students: Learning curve not only for Jupyter Notebooks but also Python

** Beaulieu et al. (2017) Earth Science Informatics,* [*https://doi.org/10.1007/s12145-016-0280-4*](https://doi.org/10.1007/s12145-016-0280-4)

Strengths… and weaknesses
for Jupyter Notebooks in any language

I Don’t Like Notebooks
Joel Grus (@joelgrus)   #JupyterCon 2018



Mixed Markdown and Code

I love well-documented code!

Accordingly, I love the idea of mixing markdown and code.

For example, one underappreciated Python library is `itertools`:

```
In [1]:  import itertools

         binary_numbers = [''.join(digits)
                                   for digits in itertools.product(*[['

         binary_numbers[:8]

Out[1]:  ['0000000000',
          '0000000001',
          '0000000010',
          '0000000011',
          '0000000100',
          '0000000101',
          '0000000110',
          '0000000111']
```

NOTEBOOKS ARE GREAT FOR ITERATIVE DEVELOPMENT
NOTEBOOKS ARE DANGEROUS UNLESS YOU RUN EACH CELL EXACTLY ONCE, IN ORDER

We are offering training to NES-LTER students this summer

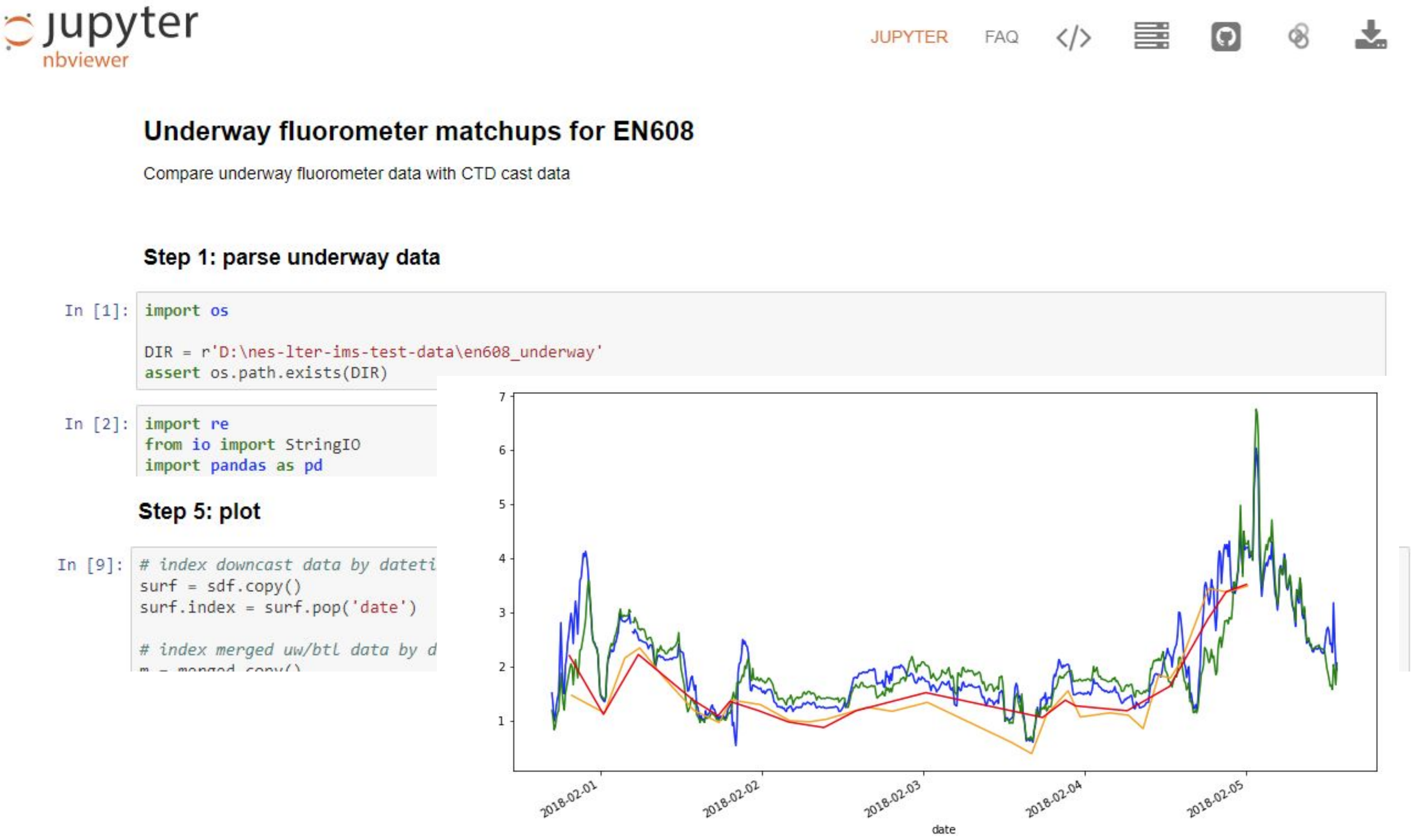
XSEDE Jetstream Education Allocation:

Navigating an Ocean of Data: Curriculum Development and Implementation | TG-OCE190011

Thanks!

Stace Beaulieu (stace@whoi.edu)

Extra: Next notebook we are building for ship-provided NES-LTER data



Underway fluorometer matchups for EN608

Compare underway fluorometer data with CTD cast data

Step 1: parse underway data

```
In [1]: import os

        DIR = r'D:\nes-lter-ims-test-data\en608_underway'
        assert os.path.exists(DIR)
```

```
In [2]: import re
        from io import StringIO
        import pandas as pd
```

Step 5: plot

```
In [9]: # index downcast data by dateti
        surf = sdf.copy()
        surf.index = surf.pop('date')

        # index merged uw/btl data by d
        m = merged.copy()
```

https://nbviewer.jupyter.org/gist/joefutrelle/09f2efba6b7629fbe19e13553a78f7cc

“Getting Stuff Done with R, Python and Jupyter Notebooks”

ESIP Summer Meeting, 17 July 2019

Moderator: M. Gastil-Buhl

Speakers: John Porter • Colin Smith • Chris Turner and Stace Beaulieu