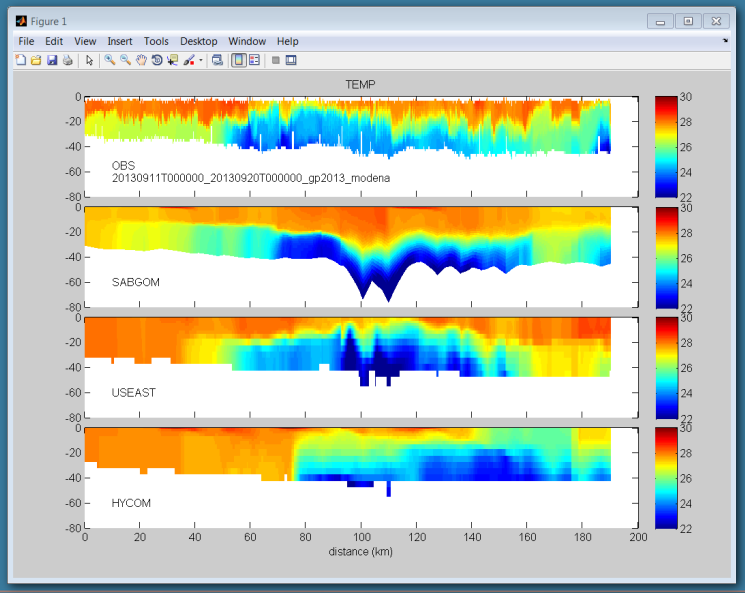
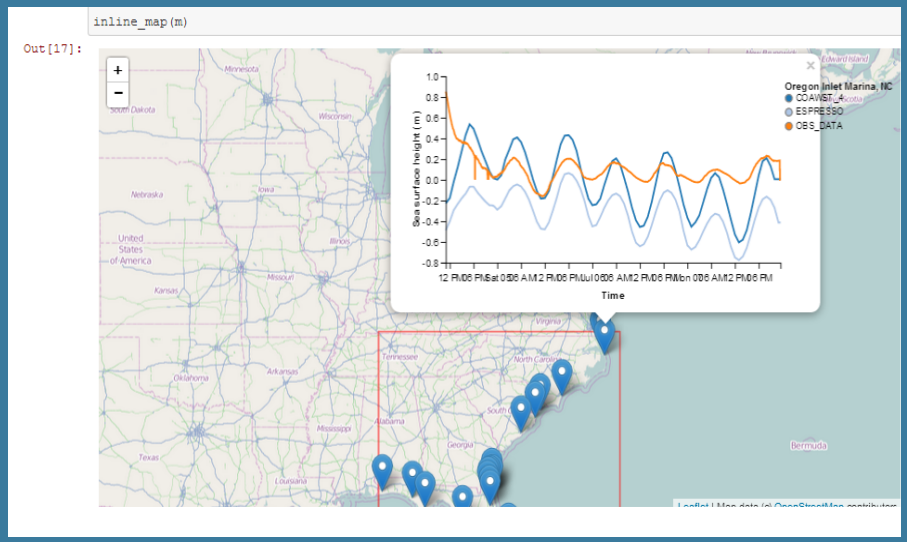
Tools for Ocean, Atmosphere & Climate Model Data Interoperability

Rich Signell (USGS-CMG)
USGS Woods Hole, MA

COAWST Training Workshop

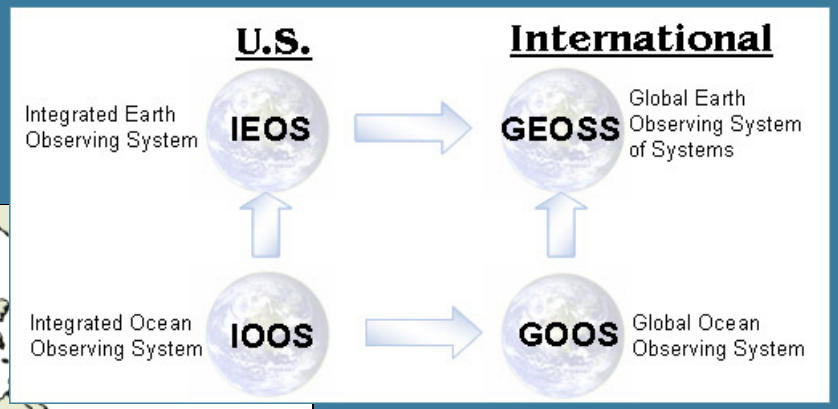
Woods Hole, MA: Aug 15-19, 2016

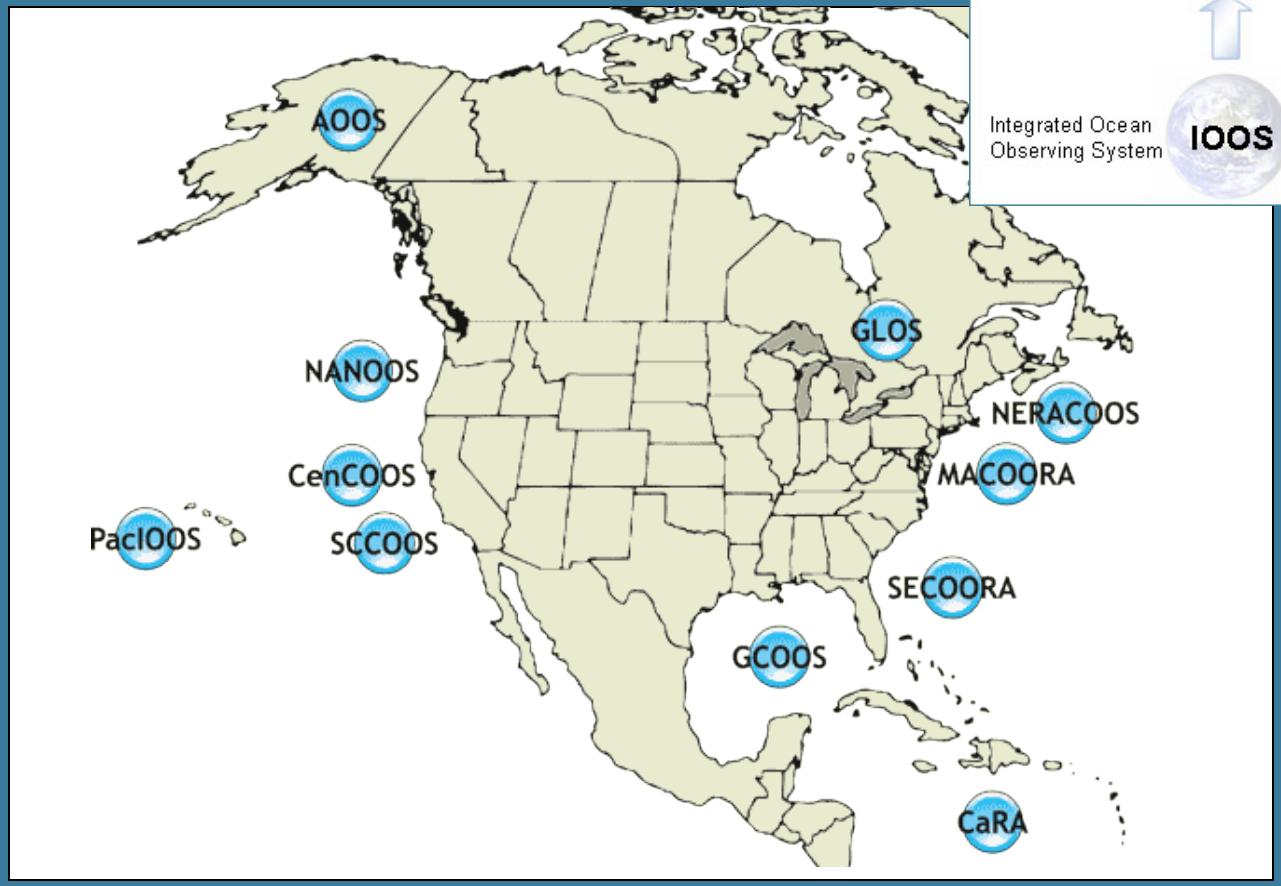




US Integrated Ocean Observing System (IOOS®)

- 17 Federal Agencies
- 11 Regional Associations





IOOS Core Principles

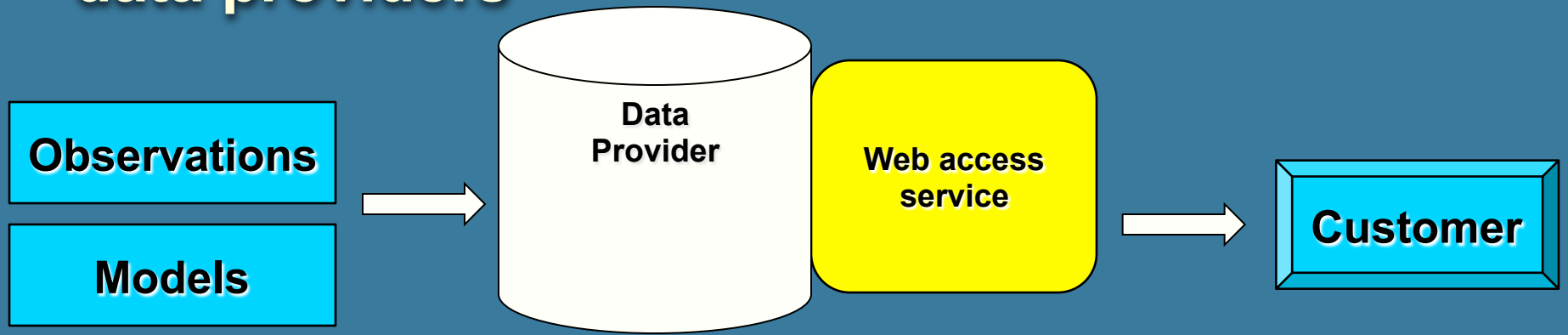
- Adopt open standards & practices

OGC® Open Geospatial Consortium, Inc.

ISO International Organization for Standardization

unidata

- Avoid customer-specific stovepipes
- Standardized access services implemented at data providers

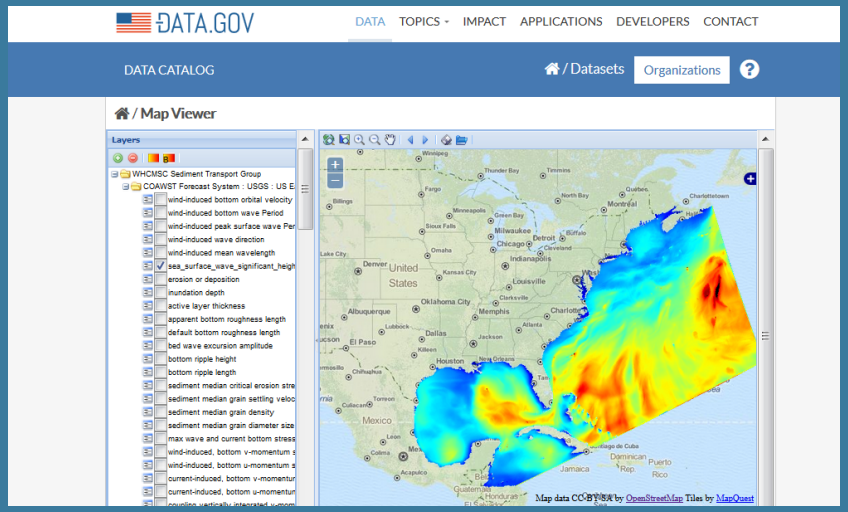


From pile of files to interoperability

Raw Dataset files

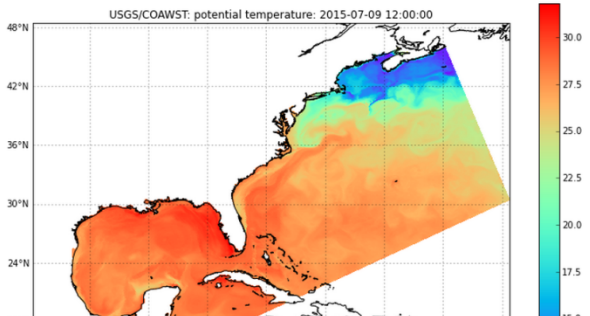
```
nganju@gam: /usgs/data0/bbleh/spring2012
nganju@gam:/usgs/data0/bbleh/spring2012$
nganju@gam:/usgs/data0/bbleh/spring2012$
nganju@gam:/usgs/data0/bbleh/spring2012$
nganju@gam:/usgs/data0/bbleh/spring2012$
nganju@gam:/usgs/data0/bbleh/spring2012$ ls
00_dir_roms.ncml   his_bbleh_0018.nc  his_bbleh_0036.nc  his_bbleh_0054.nc
his_bbleh_0001.nc  his_bbleh_0019.nc  his_bbleh_0037.nc  his_bbleh_0055.nc
his_bbleh_0002.nc  his_bbleh_0020.nc  his_bbleh_0038.nc  his_bbleh_0056.nc
his_bbleh_0003.nc  his_bbleh_0021.nc  his_bbleh_0039.nc  his_bbleh_0057.nc
his_bbleh_0004.nc  his_bbleh_0022.nc  his_bbleh_0040.nc  his_bbleh_0058.nc
his_bbleh_0005.nc  his_bbleh_0023.nc  his_bbleh_0041.nc  his_bbleh_0059.nc
his_bbleh_0006.nc  his_bbleh_0024.nc  his_bbleh_0042.nc  his_bbleh_0060.nc
his_bbleh_0007.nc  his_bbleh_0025.nc  his_bbleh_0043.nc  his_bbleh_0061.nc
his_bbleh_0008.nc  his_bbleh_0026.nc  his_bbleh_0044.nc  his_bbleh_0062.nc
his_bbleh_0009.nc  his_bbleh_0027.nc  his_bbleh_0045.nc  his_bbleh_0063.nc
his_bbleh_0010.nc  his_bbleh_0028.nc  his_bbleh_0046.nc  his_bbleh_0064.nc
his_bbleh_0011.nc  his_bbleh_0029.nc  his_bbleh_0047.nc  his_bbleh_0065.nc
his_bbleh_0012.nc  his_bbleh_0030.nc  his_bbleh_0048.nc  his_bbleh_0066.nc
his_bbleh_0013.nc  his_bbleh_0031.nc  his_bbleh_0049.nc  his_bbleh_0067.nc
his_bbleh_0014.nc  his_bbleh_0032.nc  his_bbleh_0050.nc  his_bbleh_0068.nc
his_bbleh_0015.nc  his_bbleh_0033.nc  his_bbleh_0051.nc  his_bbleh_0069.nc
his_bbleh_0016.nc  his_bbleh_0034.nc  his_bbleh_0052.nc  his_bbleh_0070.nc
his_bbleh_0017.nc  his_bbleh_0035.nc  his_bbleh_0053.nc
nganju@gam:/usgs/data0/bbleh/spring2012$
```

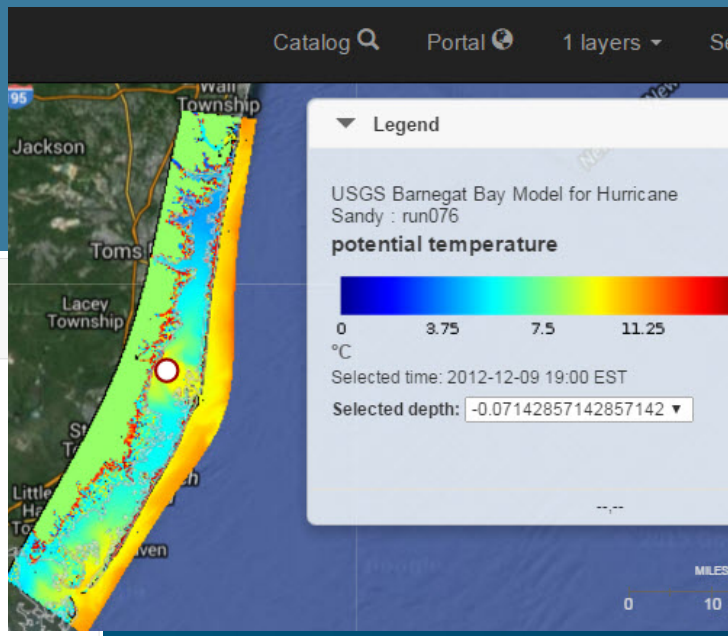
Standardized Search and Browse



Standardized Data Access

```
In [16]: model = 'USGS/COAWST'
         url = 'http://geoport.whoi.edu/thredds/dodsC/coawst_4/use/fmrc/coawst_4_use_best.ncd'
         var = 'sea_water_potential_temperature'
         lev = -1
         icube = var_lev_date(url=url, var=var, mytime=mytime, lev=lev, subsample=1)
         map_plot(icube, model=model)

slice retrieved in 48.000320 seconds
```

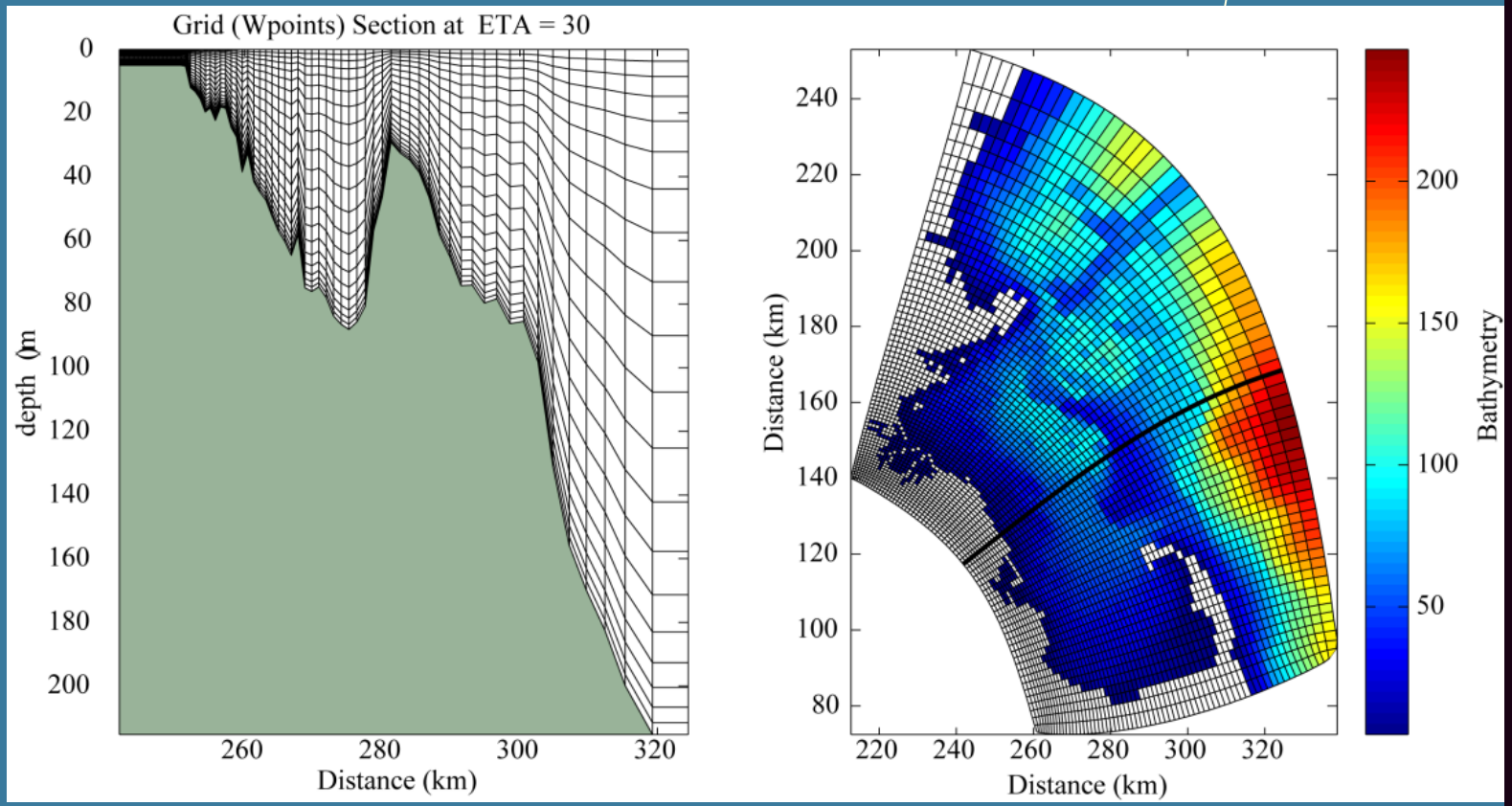




Ocean grids are often not regularly spaced!

Stretched surface and terrain following vertical coordinates

Curvilinear orthogonal horizontal coordinates



CF Conventions for met/ocean models

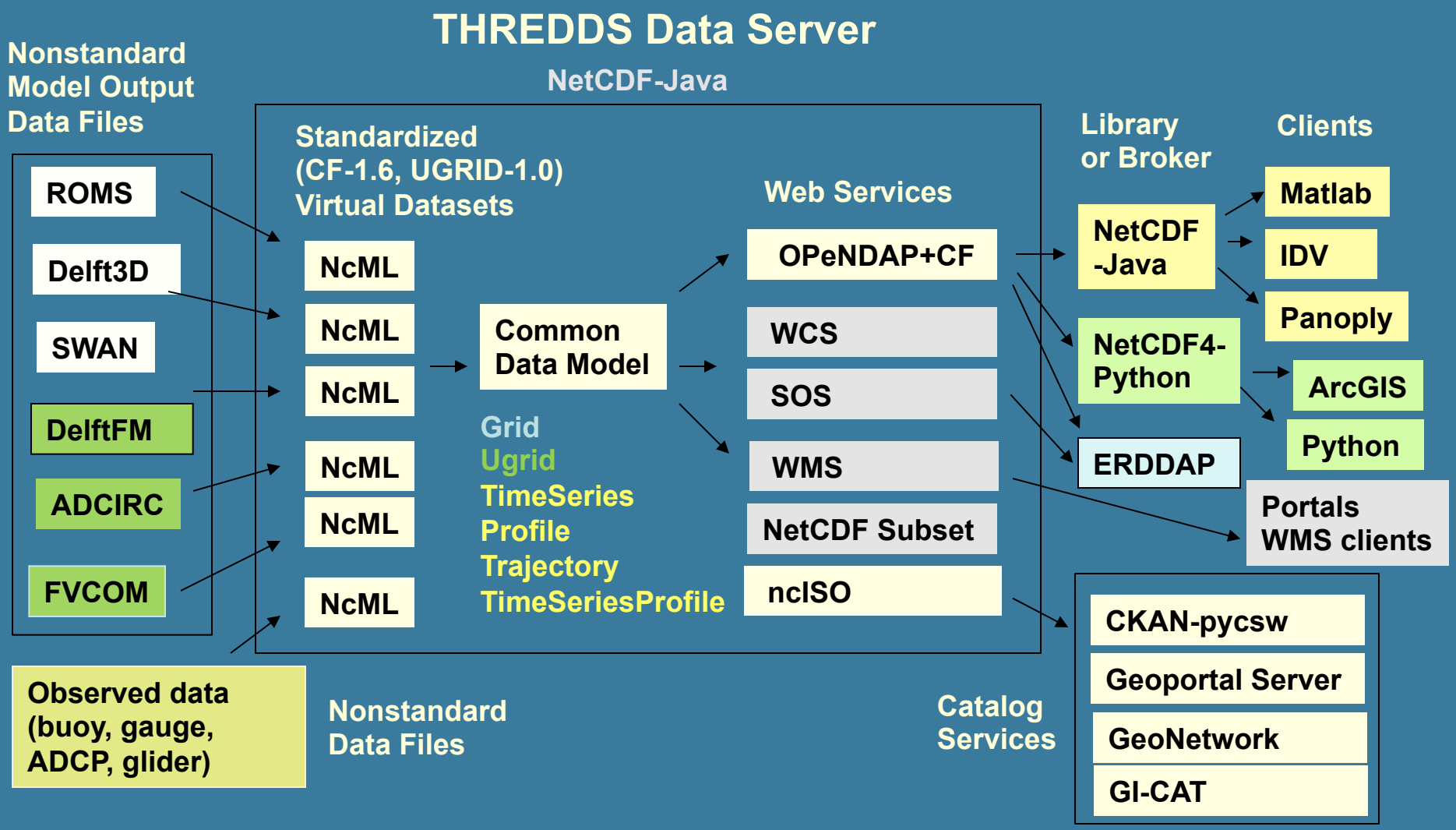
```
float temp(ocean_time, s_rho, eta_rho, xi_rho) ;
    temp: standard_name =  “sea_water_potential_temperature" ;
    temp: units = "Celsius" ;
    temp: coordinates = "lon_rho  lat_rho s_rho ocean_time" ;

double s_rho(s_rho) ;
    s_rho: long_name = "S-coordinate at RHO-points" ;
    s_rho: positive = "up" ;
    s_rho: standard_name = "ocean_s_coordinate_g1" ;
    s_rho: formula_terms = "s: s_rho C: Cs_r eta: zeta
                                    depth: h depth_c: hc "
```

Some of the groups using CF:

OGC: Open Geospatial Consortium
ODIP: Ocean Data Interoperability Platform
IOOS: Integrated Ocean Observing System
CMIP: Coupled Model Intercomparison Project (supporting IPPC)

IOOS

IOOS Model Data Interoperability Design



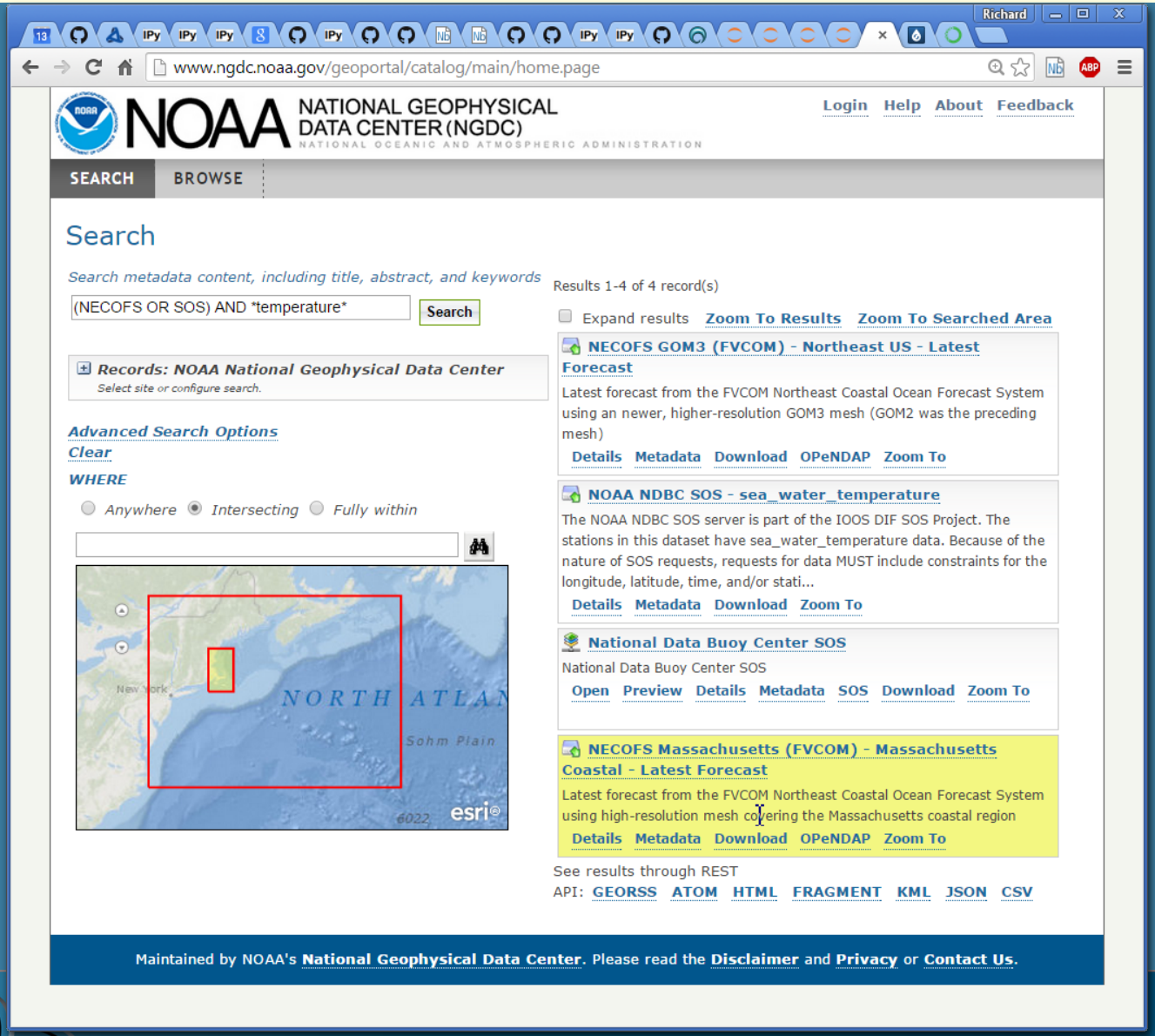
sudo apt-get docker
docker run –d axiom/docker-thredds

Matlab NCTOOLBOX
https://github.com/nctoolbox

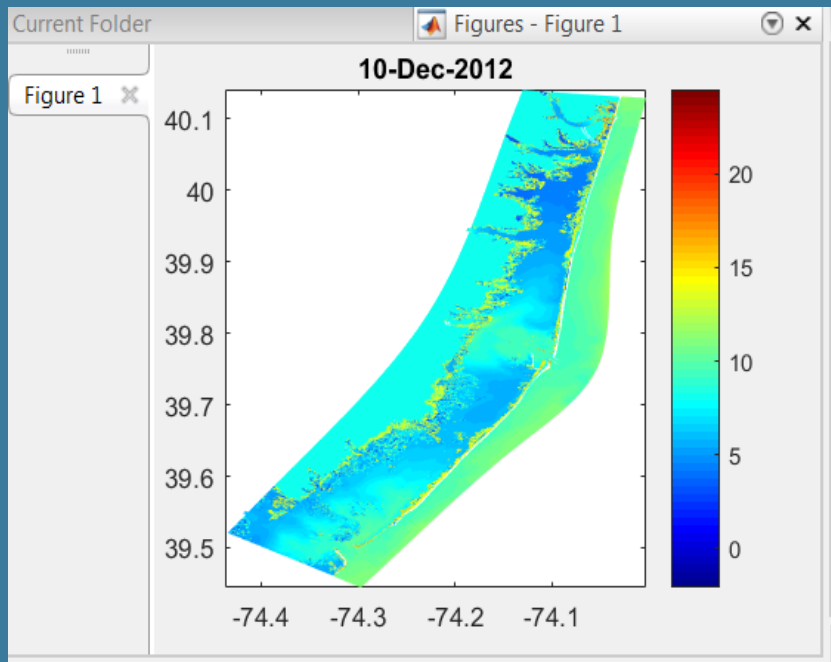
- Objective: Make it simple to access CF data
- Example function:
- [t, geo]=nj_tslice(URL,'temp',1);
- t = 7x800x160 single
- geo =
  - lat: [800x160 single]
  - lon: [800x160 single]
  - z: [7x800x160 double]
  - time: 735169.0208 (matlab datenum)
- nj_tslice works identically for COAWST/ROMS, WRF, HYCOM, Delft3D, Wavewatch3
- URL can be: local NetCDF, NcML, OpenDAP Data URL

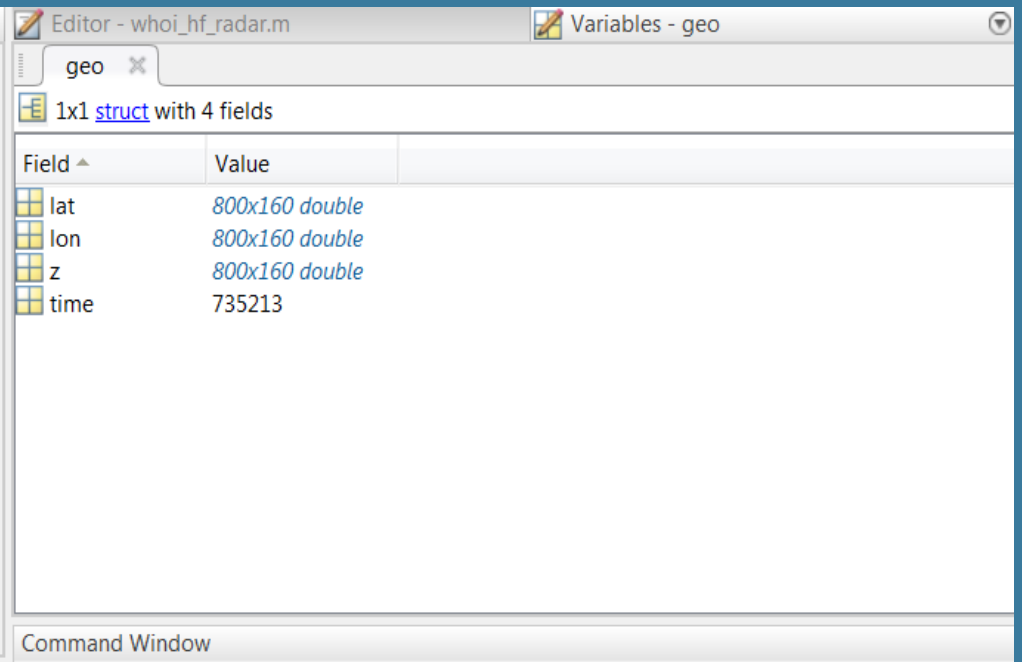
IOOS

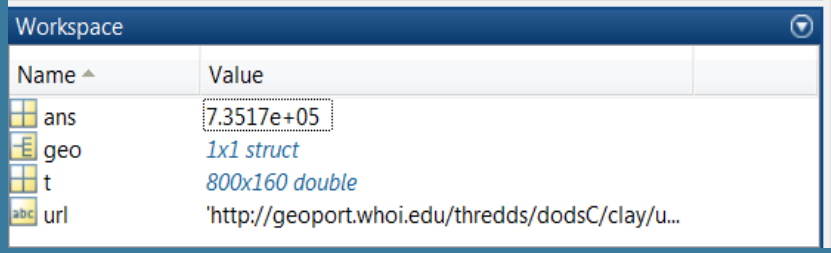
Catalog Search

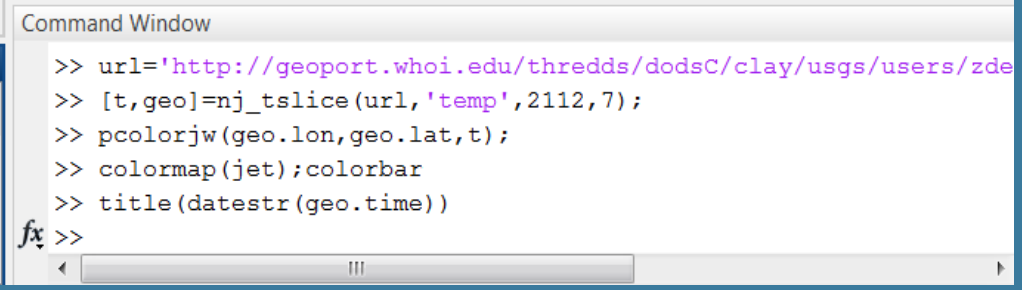


Matlab NCTOOLBOX
https://github.com/nctoolbox

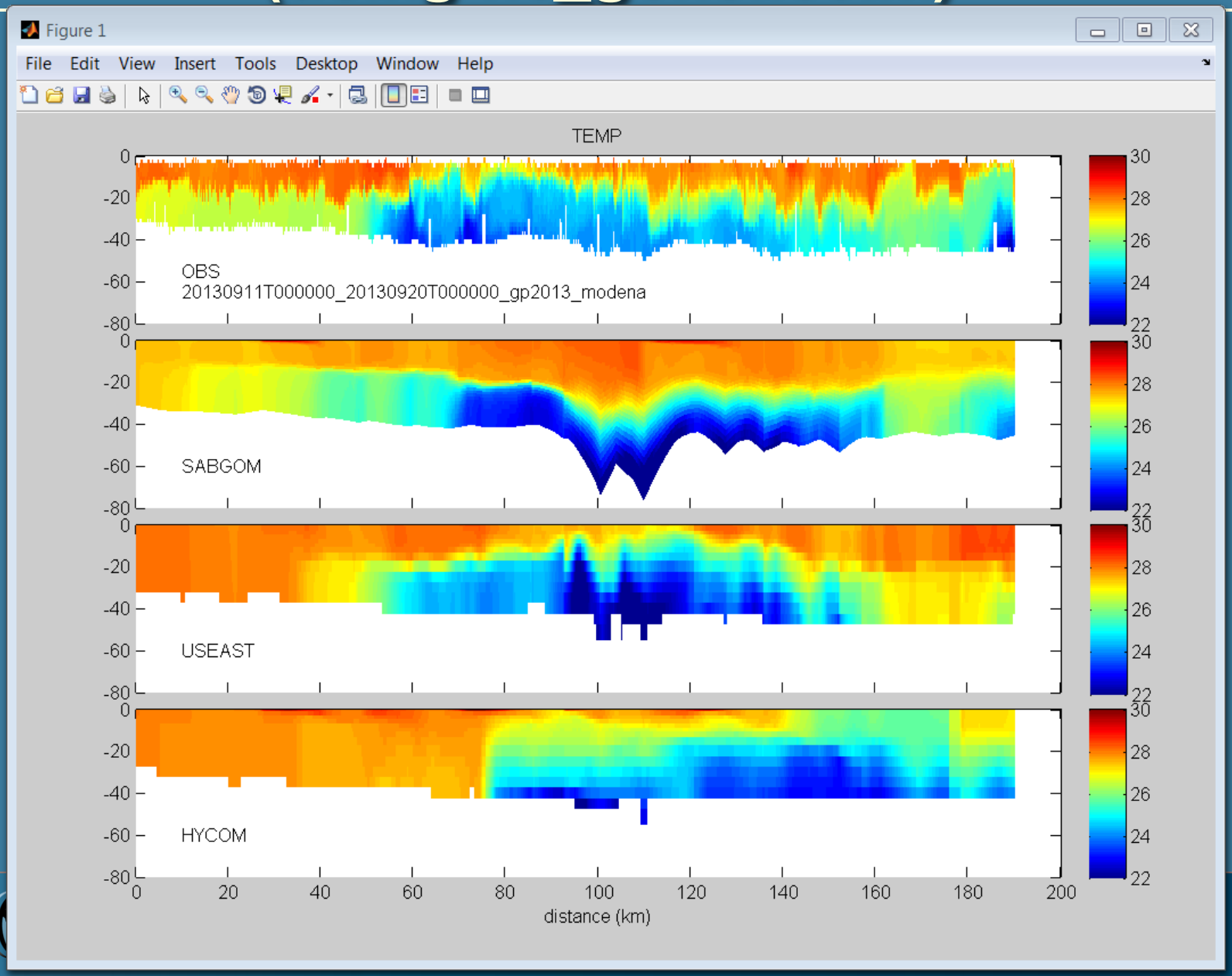






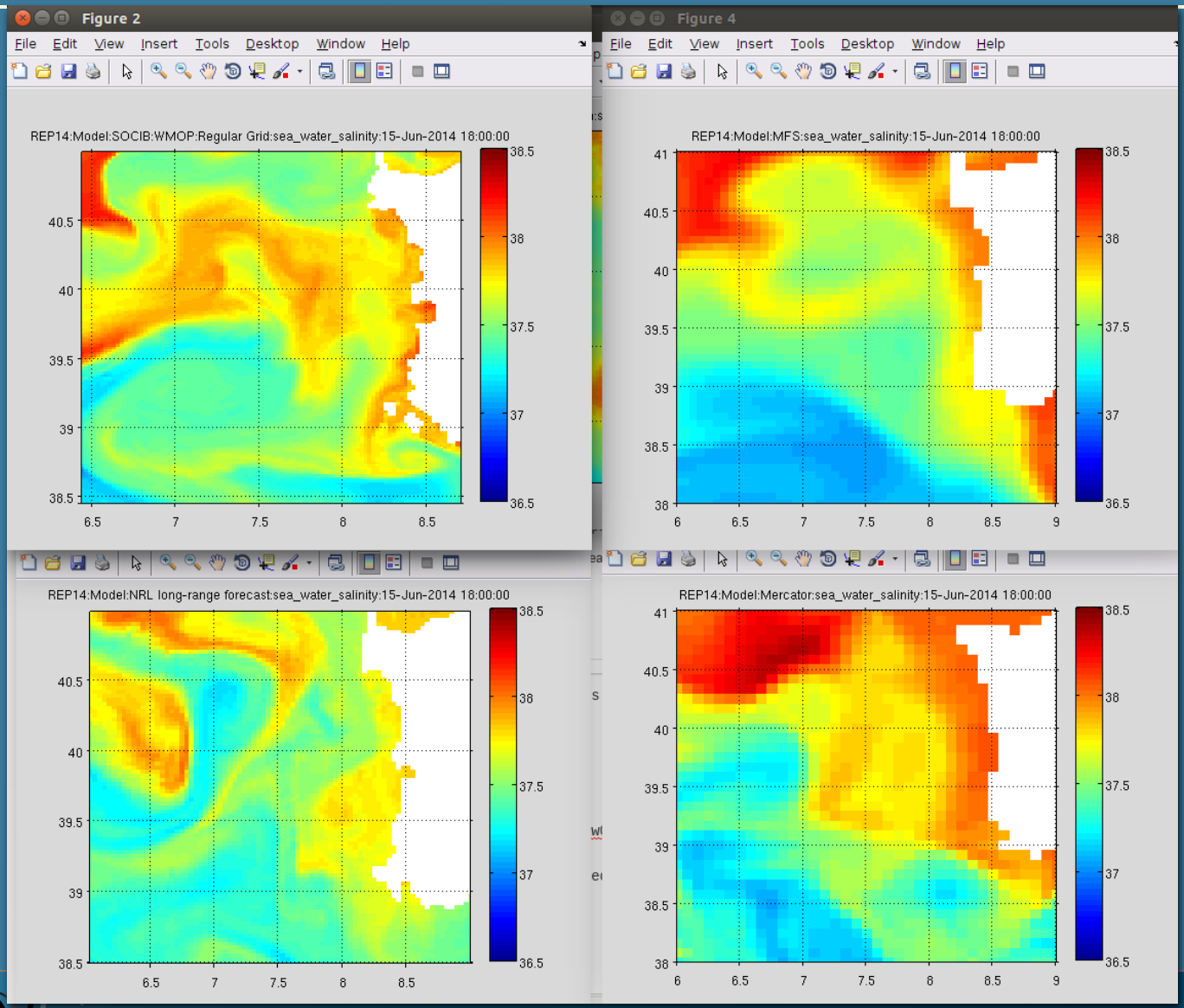


compare_secoora_model_sections.m (using nc_genslice.m)



Approach for Small Research Groups

Deployment at NATO



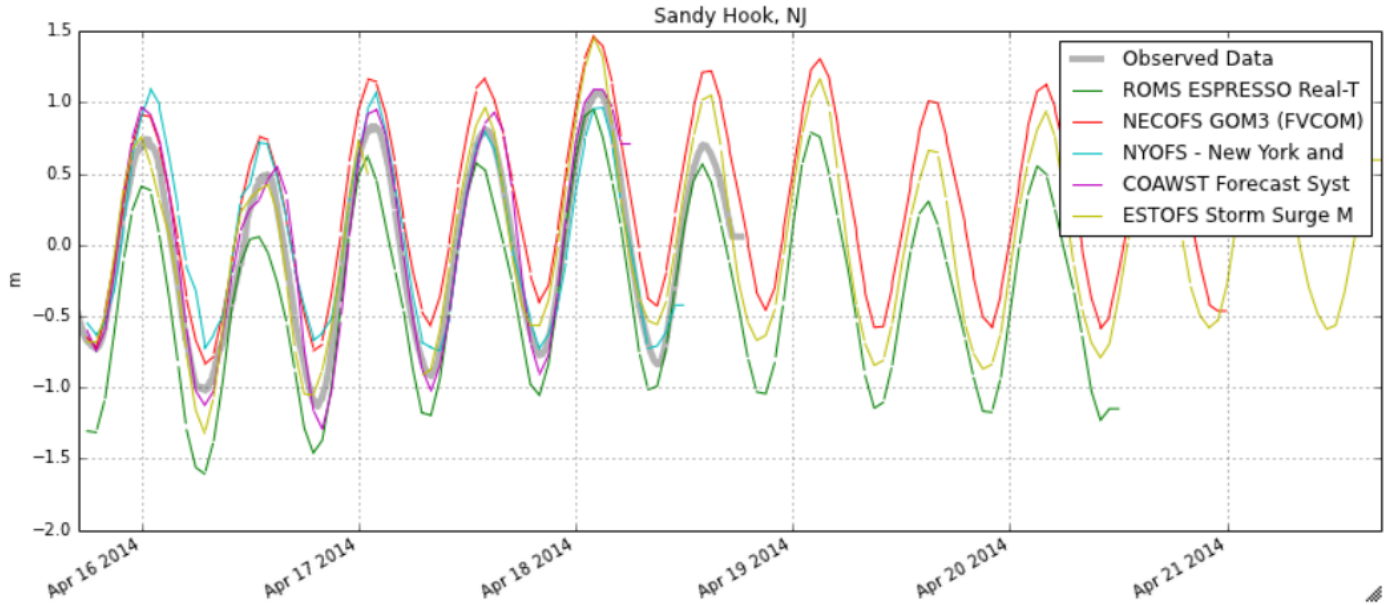
Automated model comparison

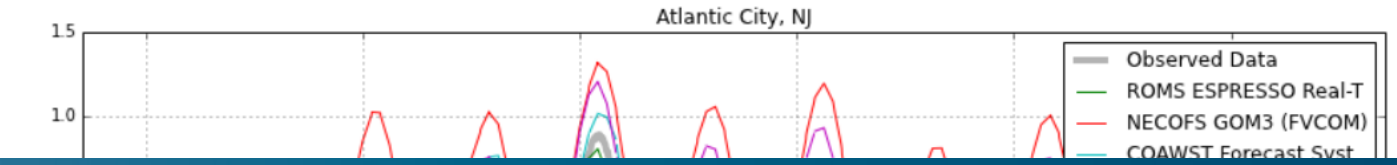
IP[y]: Notebook IOOS_inundation Last Checkpoint: Mar 15 16:08 (autosaved)

File Edit View Insert Cell Kernel Help

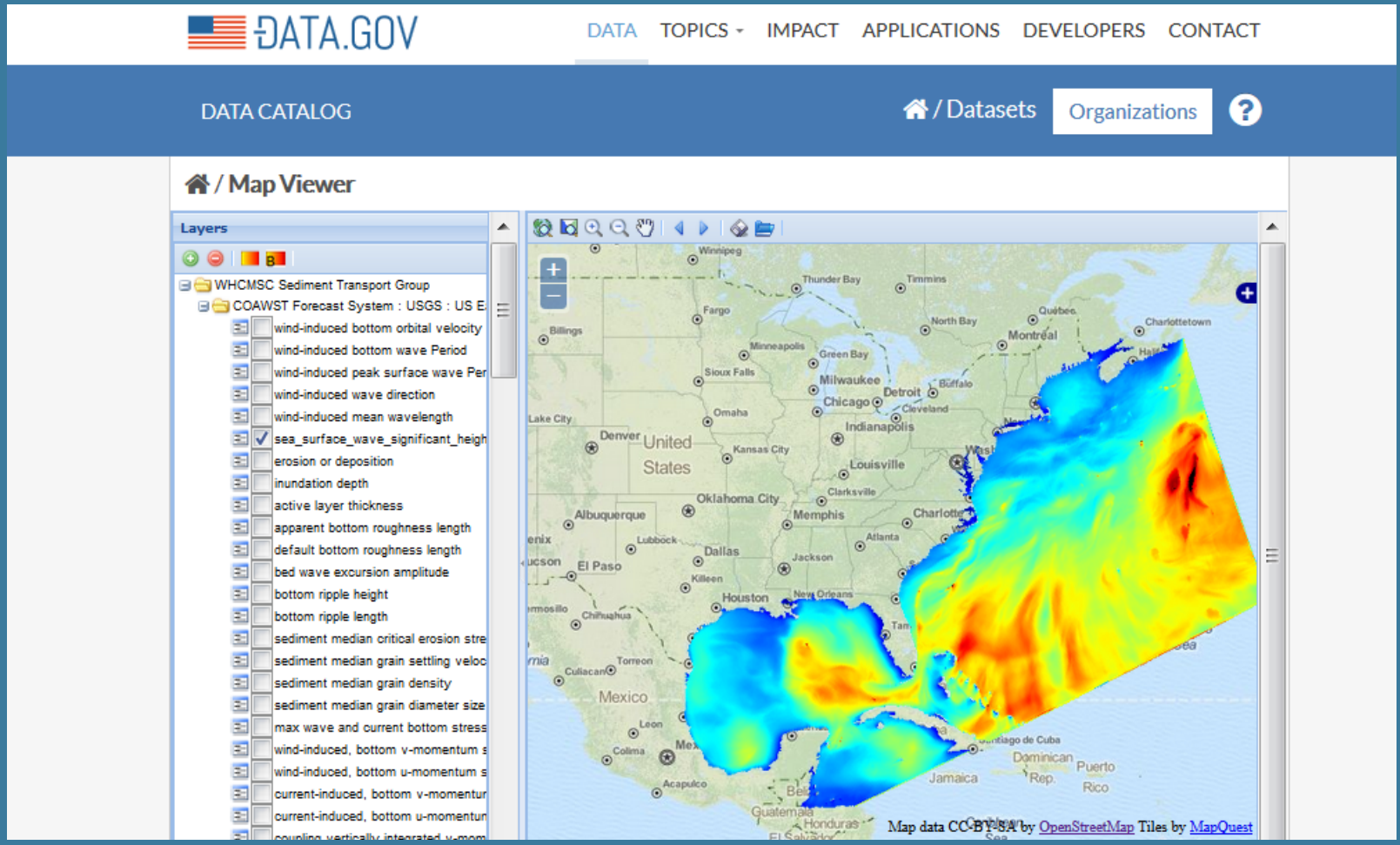
Code Cell Toolbar: None

In [45]:
```
for df in obs_df:
    df.plot(figsize=(14,6),title=df.name)
    ylabel('m')
```



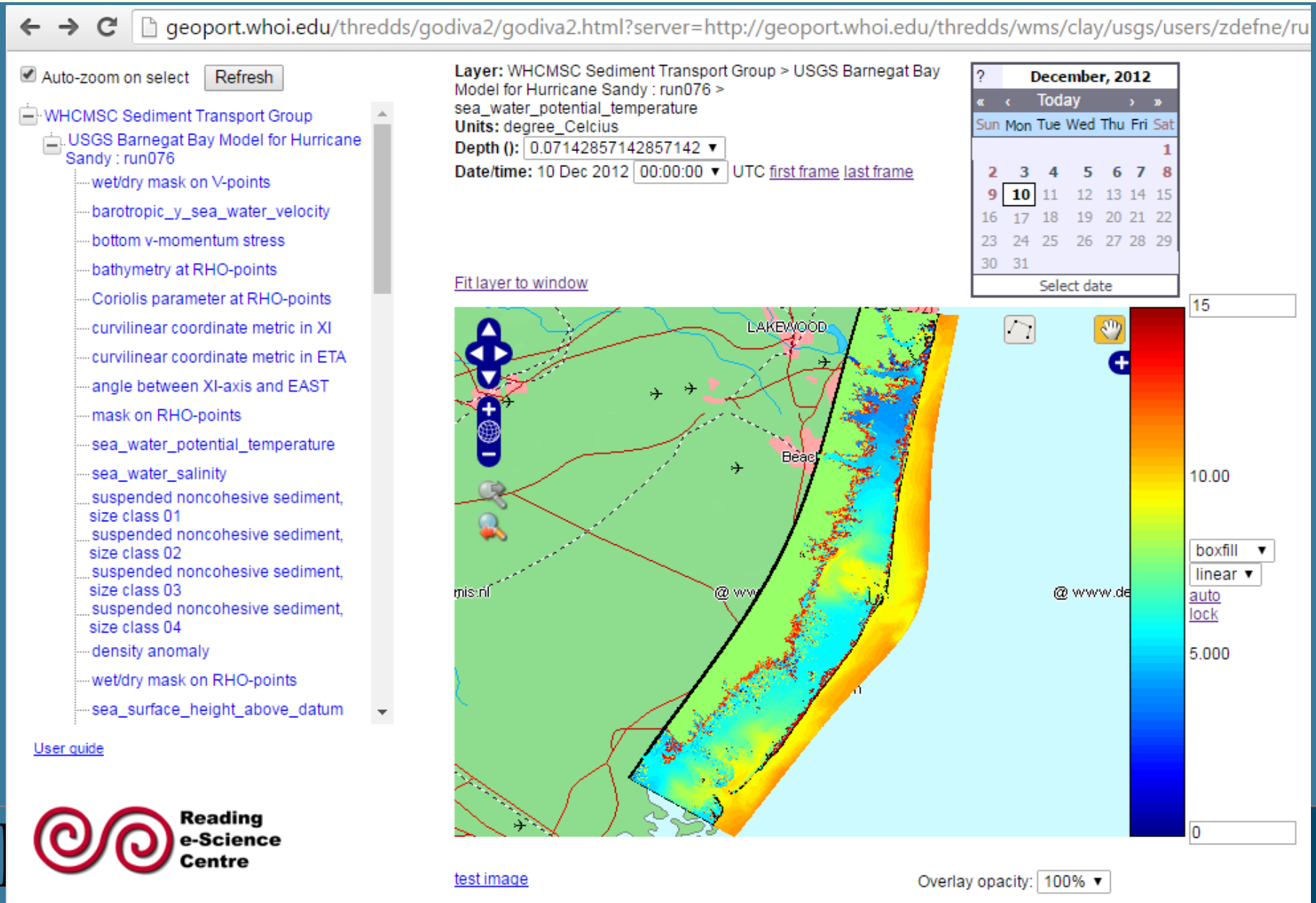


Catalog Search

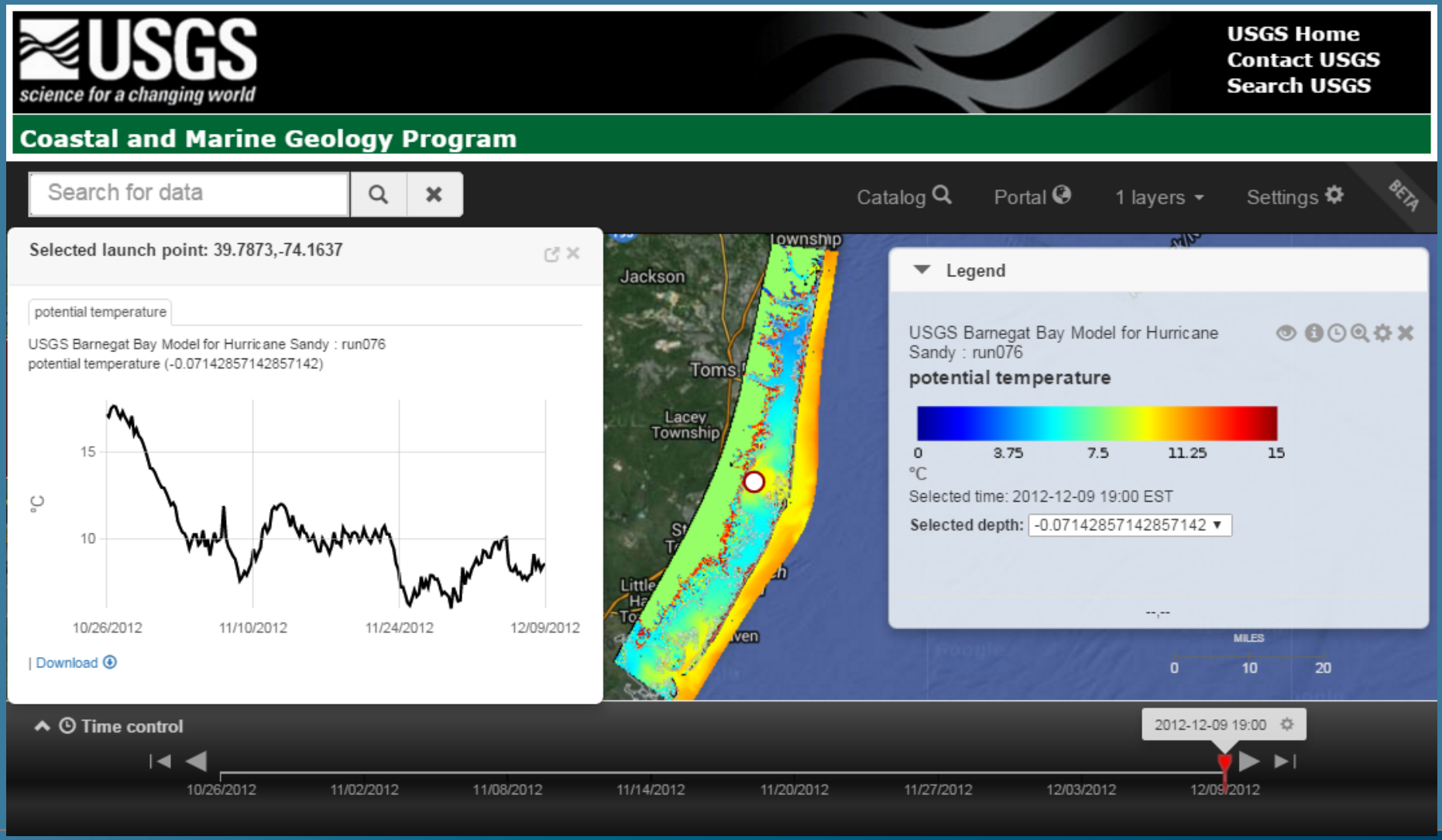


Catalog services can be federated via OGC CSW (Catalog Service for the Web)

WMS Browsing with THREDDS/ncWMS



WMS Browsing with CMG Oceanographic Portal



Demo

NetCDF Subset Service to CSV

WHCMSC Sediment Transport Group

THREDDS Data Server

Catalog http://geoport.whoi.edu/thredds/catalog/coawst_4/use/fmrc/catalog.html

Dataset: coawst_4_use/Best Time Series

- *Data format:* netCDF
- *Data type:* GRID
- *Naming Authority:* gov.usgs.er.whsc
- *ID:* coawst_4/use/fmrc/coawst_4_use_best.ncd

Documentation:

- **summary:** Best time series, taking the data from the most recent run available.
- **summary:** ROMS USE Output from COAWST
- Carolinas Coastal Change Program
- ReadMe.txt

Access:

1. **OPENDAP:** /thredds/dodsC/coawst_4/use/fmrc/coawst_4_use_best.ncd
2. **NetcdfSubset:** /thredds/ncss/grid/coawst_4/use/fmrc/coawst_4_use_best.ncd
3. **WMS:** /thredds/wms/coawst_4/use/fmrc/coawst_4_use_best.ncd
4. **ISO:** /thredds/iso/coawst_4/use/fmrc/coawst_4_use_best.ncd
5. **NCML:** /thredds/ncml/coawst_4/use/fmrc/coawst_4_use_best.ncd
6. **UDDC:** /thredds/uddc/coawst_4/use/fmrc/coawst_4_use_best.ncd

geoport.whoi.edu/thredds/ncss/grid/coawst_4/use/fmrc/coawst_4_use_best.ncd/pointDataset.html

Apps | Gmail - Inbox (... | Bank of Americ... | Enphase Energy... | Calendar | WHSC Intranet | Rich Signell's W... | Other bookmar

NCSS Grids As Point Data (Gridded Dataset)

unidata
THREDDS data server NetCDF Subset Service

Dataset: /thredds/ncss/grid/coawst_4/use/fmrc/coawst_4_use_best.ncd (Gridded Dataset Description)

Base Time: 2012-06-25T01:00:00Z

You must select at least one Variable and a Lat/Lon location.

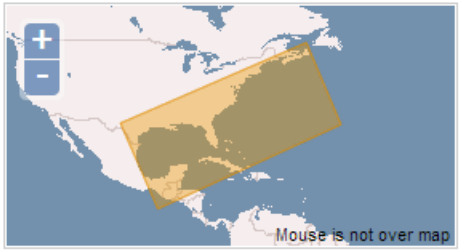
Select Variable(s):

- [ ] angle = angle between XI-axis and EAST
- [ ] f = Coriolis parameter at RHO-points
- [ ] h = bathymetry at RHO-points
- [ ] mask_psi = mask on psi-points
- [ ] mask_rho = mask on RHO-points
- [ ] mask_u = mask on U-points
- [ ] mask_v = mask on V-points
- [ ] pm = curvilinear coordinate metric in XI
- [ ] pn = curvilinear coordinate metric in ETA

Variables with Time coordinate time

- [ ] Dwave = wind-induced wave direction
- [x] Hwave = wind-induced significant wave height
- [ ] Lwave = wind-induced mean wavelength
- [ ] Pwave_bot = wind-induced bottom wave Period
- [ ] Pwave_top = wind-induced peak surface wave Period
- [ ] Uwave_rms = wind-induced bottom orbital velocity
- [ ] Uwind = surface u-wind component
- [ ] Vwind = surface v-wind component
- [ ] Zo_app = apparent bottom roughness length
- [ ] Zo_def = default bottom roughness length
- [ ] bed_wave_amp = bed wave excursion amplitude
- [ ] bedload_Usand_01 = bed load flux of sand in U-direction, size class 01
- [ ] bedload_Usand_02 = bed load flux of sand in U-direction, size class 02
- [ ] bedload_Usand_03 = bed load flux of sand in U-direction, size class 03
- [ ] bedload_Usand_04 = bed load flux of sand in U-direction, size class 04
- [ ] bedload_Usand_05 = bed load flux of sand in U-direction, size class 05
- [ ] bedload_Usand_06 = bed load flux of sand in U-direction, size class 06

Choose Lat/Lon Location:



Latitude: 39.5
Longitude: -69.5

Within Bounding Box:

north: 48.4639
west: -101.7498 -53.2530 east
11.8884
south

Choose Time Subset:

Time range **Single time**

Starting: 2014-08-27T01:00:00Z
Ending: 2014-08-28T00:00:00Z
reset to full extension

Choose Vertical Level:

Level:

Choose Output Format:

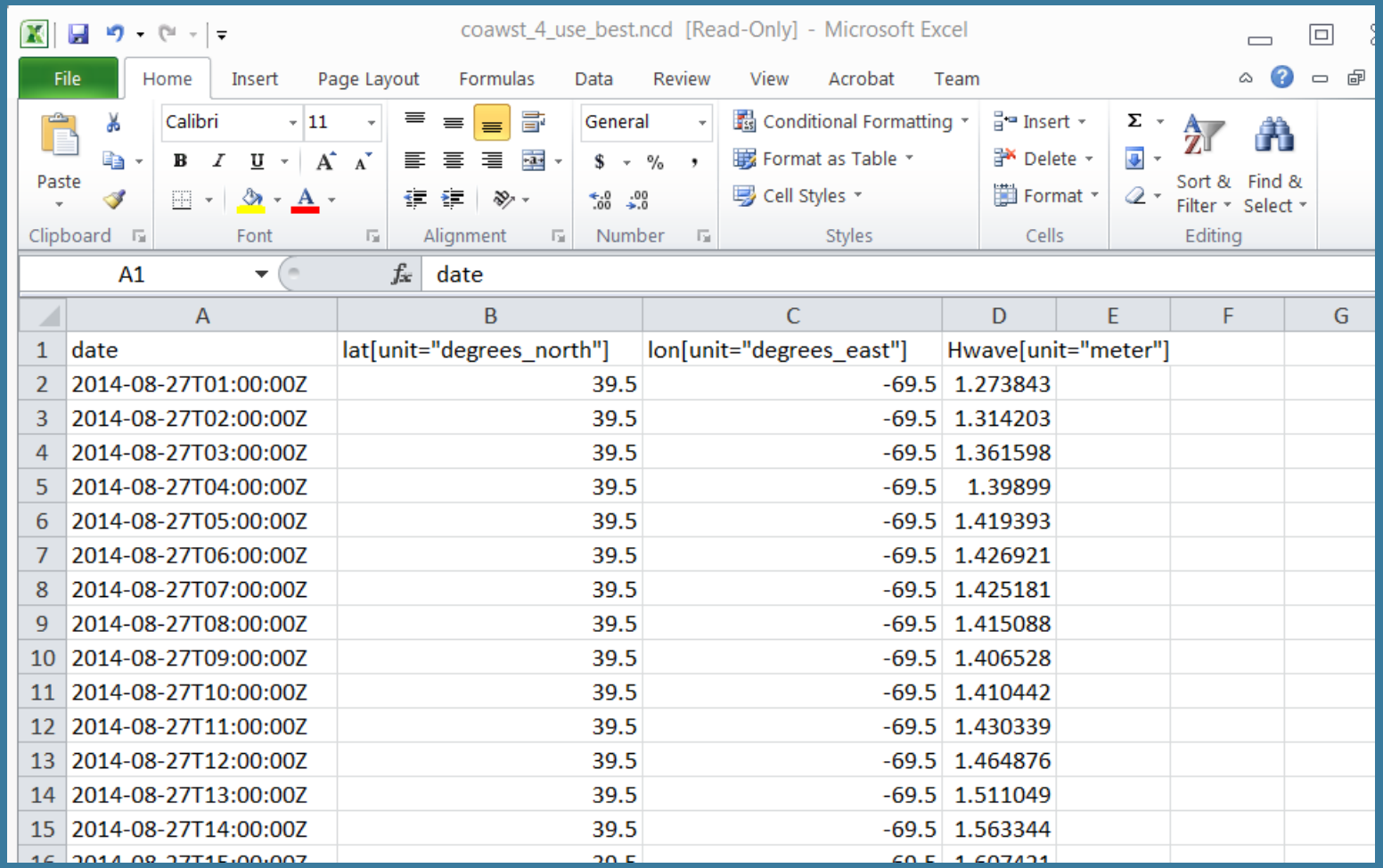
Format: csv

CSV Time Series from Subset Service

```
date,lat[unit="degrees_north"],lon[unit="degrees_east"],Hwave[unit="meter"]
2014-08-27T01:00:00Z,39.5,-69.5,1.273842692375183
2014-08-27T02:00:00Z,39.5,-69.5,1.3142027854919434
2014-08-27T03:00:00Z,39.5,-69.5,1.3615976572036743
2014-08-27T04:00:00Z,39.5,-69.5,1.398990273475647
2014-08-27T05:00:00Z,39.5,-69.5,1.419392704963684
2014-08-27T06:00:00Z,39.5,-69.5,1.426921010017395
2014-08-27T07:00:00Z,39.5,-69.5,1.4251813888549805
2014-08-27T08:00:00Z,39.5,-69.5,1.4150882959365845
2014-08-27T09:00:00Z,39.5,-69.5,1.406528115272522
2014-08-27T10:00:00Z,39.5,-69.5,1.4104422330856323
2014-08-27T11:00:00Z,39.5,-69.5,1.4303393363952637
2014-08-27T12:00:00Z,39.5,-69.5,1.4648756980895996
2014-08-27T13:00:00Z,39.5,-69.5,1.5110485553741455
2014-08-27T14:00:00Z,39.5,-69.5,1.5633442401885986
2014-08-27T15:00:00Z,39.5,-69.5,1.6074210405349731
2014-08-27T16:00:00Z,39.5,-69.5,1.6265686750411987
2014-08-27T17:00:00Z,39.5,-69.5,1.6184144020080566
2014-08-27T18:00:00Z,39.5,-69.5,1.5933209657669067
2014-08-27T19:00:00Z,39.5,-69.5,1.5608850717544556
2014-08-27T20:00:00Z,39.5,-69.5,1.527510643005371
2014-08-27T21:00:00Z,39.5,-69.5,1.5050616264343262
2014-08-27T22:00:00Z,39.5,-69.5,1.5094025135040283
2014-08-27T23:00:00Z,39.5,-69.5,1.5698728561401367
2014-08-28T00:00:00Z,39.5,-69.5,1.7087903022766113
```

CSV request is a "RESTful" URL:
http://geoport.whoi.edu/thredds/ncss/grid/coawst_4/use/fmrc/coawst_4_use_best.ncd?var=Hwave&latitude=39.5&longitude=-69.5&time_start=2014-08-27T01:00:00Z&time_end=2014-08-28T00:00:00Z&vertCoord=&accept=csv

Accessing the CSV data in Excel



	A	B	C	D	E	F	G
1	date	lat[unit="degrees_north"]	lon[unit="degrees_east"]	Hwave[unit="meter"]			
2	2014-08-27T01:00:00Z	39.5	-69.5	1.273843			
3	2014-08-27T02:00:00Z	39.5	-69.5	1.314203			
4	2014-08-27T03:00:00Z	39.5	-69.5	1.361598			
5	2014-08-27T04:00:00Z	39.5	-69.5	1.39899			
6	2014-08-27T05:00:00Z	39.5	-69.5	1.419393			
7	2014-08-27T06:00:00Z	39.5	-69.5	1.426921			
8	2014-08-27T07:00:00Z	39.5	-69.5	1.425181			
9	2014-08-27T08:00:00Z	39.5	-69.5	1.415088			
10	2014-08-27T09:00:00Z	39.5	-69.5	1.406528			
11	2014-08-27T10:00:00Z	39.5	-69.5	1.410442			
12	2014-08-27T11:00:00Z	39.5	-69.5	1.430339			
13	2014-08-27T12:00:00Z	39.5	-69.5	1.464876			
14	2014-08-27T13:00:00Z	39.5	-69.5	1.511049			
15	2014-08-27T14:00:00Z	39.5	-69.5	1.563344			
16	2014-08-27T15:00:00Z	39.5	-69.5	1.607421			

Access the CSV data in Python, R, etc

https://www.wakari.io/sharing/bundle/rsignell/RESTful_model_assessment

Apps | Gmail - Inbox (... | Bank of Americ... | Enphase Energy... | Calendar | WHSC Intranet | Rich Signell's W... | Other bookmar

Wakari

RESTful_model_assessment

10

View Other Bundles by rsignell | Download Entire Bundle | Download This Notebook | Run/Edit this Notebook

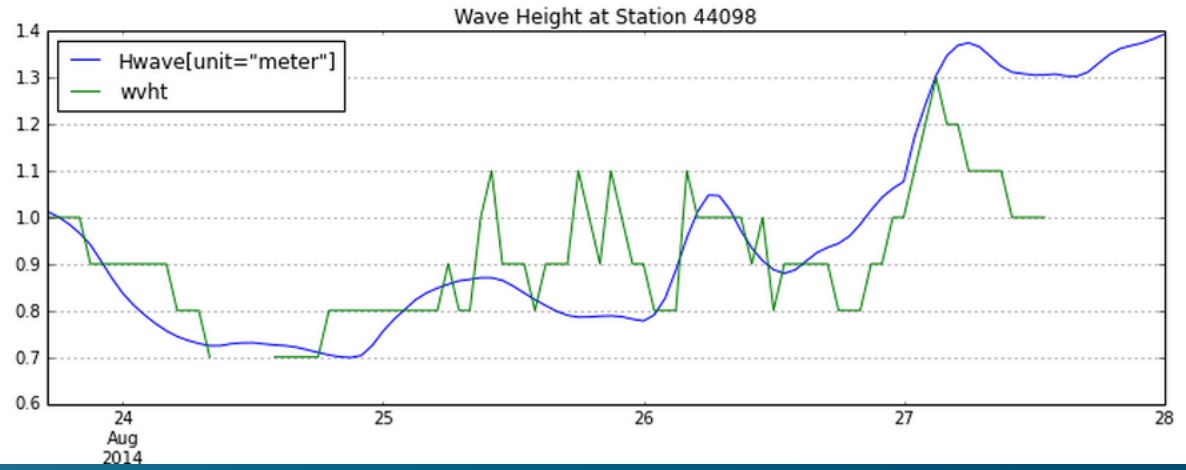
```
vertCoord=&accept=csv' % (var,lat,lon,start,stop)
print(url)
```

```
http://geoport.whoi.edu/thredds/ncss/grid/coawst_4/use/fmrc/coawst_4_use_best.ncd?var=Hwave&l
de=42.801000&longitude=-70.169000&time_start=2014-08-23T16:01:13Z&time_end=2014-08-28T16:01:1
rtCoord=&accept=csv
```

In [11]:
```
#load model data CSV into Pandas DataFrame
df_mod = pd.read_csv(url,index_col='date',parse_dates=True)
```

Plot the time series

In [12]:
```
fig, ax = plt.subplots(figsize=(12, 4))
ax = df_mod['Hwave[unit="meter"]'].plot(ax=ax, legend=True)
df_obs['wvht'].plot(ax=ax, legend=True)
ax.set_title('Wave Height at Station %s' % station);
```



Service-based Framework Components

- Common data models for “feature types” (structured and unstructured grids, time series, profiles, swaths) (Unidata CDM)
- Standard web data services for delivering these common data model “feature types” (OPeNDAP, SOS)
- Standard catalog services for the metadata
- Tools to search, access and process these services in common analysis environments: R, Matlab, Python, ArcGIS, JavaScript

Getting model data connected

- Find someone with a THREDDS Data Server or install your own
- Drop your files in a directory, and add an NcML file that starts with “00_dir” (e.g. “00_dir_roms.ncml”) to aggregate, standardize and describe the dataset: Sample ROMS NcML file
- If you want your data to end up in the portal, add “CMG_Portal” to the “project” attribute:
  `<attribute name="project" value="CMG_Portal"/>`
- If you want your datasets to be discoverable, edit the list of thredds catalogs on CMG github repo, submit PR
- Full instructions on the USGS-CMG Portal Github Wiki

Framework benefits

- Lots of choices for data access (Browser, Matlab, Python, R, Excel)
- Less time wasted messing with data, more time spent on science
- Less time responding to data requests
- Easy development of portals
- More usage of products
- Faster feedback to modelers, improved models

It's all on Github