podaac

Physical Oceanography Distributed Active Archive Center

CEOS Analysis Ready Data for the Oceans

Ed Armstrong
Jet Propulsion Laboratory, California Institute of Technology.
CEOS SST-Virtual Constellation Co-Lead

ESIP 2020 Summer Meeting
July 2020

These activities were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Dedicated funding for PO.DAAC activities is through a grant from NASA's ESDIS Project.

Introduction

- *CEOS (Committee on Earth Observation Satellites): International consortium to coordinate and foster earth satellite observations*
- ARD definition
- Current ARD activities within CEOS
  - CARD4L
  - CEOS COAST
- ARD for Oceans
  - Sea Surface Temperature (SST) and Ocean Color
  - Working example for L4 SST
- Closing thoughts

Analysis Ready Data definition

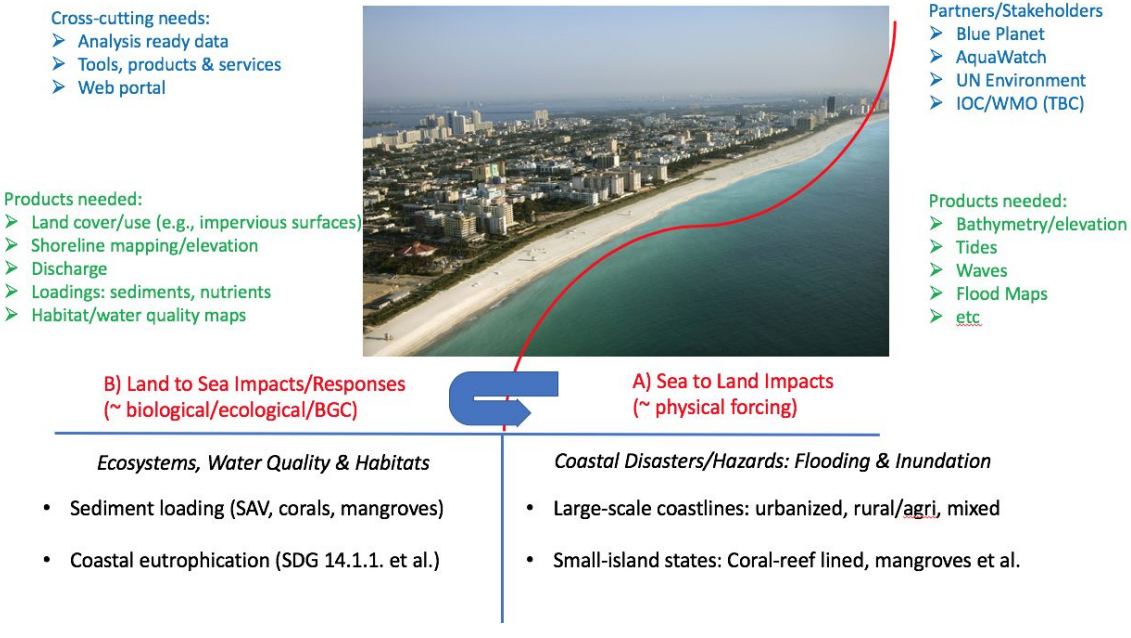
- Time series data pre processed for further scientific (and interdisciplinary) analysis
  - Ready for
    - Time series analysis
    - Subsetting
    - Regridding/Reprojection/Recombination/Reformating
    - New compute platforms (e.g., cloud, noSQL database)
  - Processed for
    - Environmental and physical corrections at pixel level
    - Big Data format
  - Documented for
    - Data processing, quality, and error assessment
- ARD definition separate from deployment or services

CEOS ARD - CARD4L

- CARD4L: CEOS Analysis Ready Data For Land
- Specification documents for ARD land surface temperature, surface reflectance and radar
  - Requirements for contents
    - Necessary General, and Ancillary Metadata and Provenance
    - Geometric/radiometric calibration at pixel level
  - Published in 2019-2020
  - Working with stakeholders like Digital Earth Africa
- Separate product and peer ARD assessment process
- URL: http://ceos.org/ard/

CEOS COAST

- RS applications to the environmental management and response at the land/ocean interface
  - Land to sea: Coastal Ecosystems/Water Quality/Habitats
  - Sea to land: Coastal Disasters/Hazards/Flooding



ARD for the Oceans

- Likely candidates: SST and/or Ocean Color
  - Varying processing levels (2/3/4) and data complexity
  - Varying levels of pre processing, quality and uncertainty information
- GHRSST (https//ghrsst.org) data sources
  - Represent an resource of over 100 international SST datasets (L2/L3/L4)
  - Packaged in self describing netCDF formats with pixel uncertainty and data quality
  - Question: How amenable are these datasets to ARD from the perspective of their data and metadata structure and contents ?
    - Answer: In at least one case there is a positive answer
      - AWS Level 4 Multiscale Ultrahigh Resolution (MUR) Open Data Registry Dataset (next slide) in Pangeo ecosystem
    - Survey of popular GHRSST dataset will help answer the question: are the available quality, uncertainty and other information enough to develop ARD?
      - Could practical examples be built ?

Cloud based AWS MUR Zarr

- Created by Chelle Gentemann (Farallon Institute) et al.
- Part of the Amazon Web Service (AWS) Open Data Registry on their cloud platform
  - https://registry.opendata.aws/mur/
- GitHub: https://github.com/pangeo-gallery/osm2020tutorial
- Provides a complete ecosystem based on Pangeo, Xarry, Dask, Jupyter and Binder to manipulate the MUR Zarr dataset on a few or >100 CPUs
- Compute credits provided by Amazon
- Supported by the NASA IMPACT program

AWS MUR Zarr analysis

- Existing Jupyter notebook
  - https://github.com/pangeo-gallery/osm2020tutorial
- Demonstrate SST time series, climatology and SST anomaly analyses and more on the complete MUR dataset in minutes vs hours

2. Explore the data

Let's explore the data

- look at all the SST data
- look at the SST data masked to only ocean and ice-free data
- With all data, it is important to explore it and understand what is contains before doing an analysis.
- The ice mask used by MUR SST is from NSIDC and is based on satellite passive microwave estimates of sea ice concentration
- The satellite data isn't available near land, so the is no estimate of sea ice concentration near land
- For this data, it means that there are some erroneous SSTs near land, that is likely ice and this is something to be aware of.

```
In [ ]: sst = ds_sst['analysed_sst']

cond = (ds_sst.mask==1) & ((ds_sst.sea_ice_fraction<.15) | np.isnan(ds_sst.sea_ice_fraction))

sst_masked = ds_sst['analysed_sst'].where(cond)

sst_masked
```

Using `.groupby` and `.resample`

Xarray has a lot of nice build-in methods, such as .resample which can upsample or downsample data and .mean. Here we use these to calculate a climatology and anomaly.

Create a daily SST anomaly dataset

- Calculate the daily climatology using `.groupby`
- Calculate the anomaly

Create a monthly SST anomaly dataset

- First create a monthly version of the dataset using `.resample`. Two nice arguments for `.resample`: `keep_attrs` which keeps the metadata and `skipna` which ensures that only data that is always present is included
- Calculate the monthly climatology using `.groupby`
- Calculate the anomaly

```
In [ ]: %%time
#create a daily climatology and anomaly
climatology_mean = sst_masked.groupby('time.dayofyear').mean('time',keep_attrs=True,skipna=False)

sst_anomaly = sst_masked.groupby('time.dayofyear')-climatology_mean  #take out annual mean to remo
ve trends

#create a monthly dataset, climatology, and anomaly
sst_monthly = sst_masked.resample(time='1MS').mean('time',keep_attrs=True,skipna=False)

climatology_mean_monthly = sst_monthly.groupby('time.month').mean('time',keep_attrs=True,skipna=Fa
lse)
```

ARD for the Oceans continued

- Ocean Color ARD
  - Challenging due to increased complexity (flags), uncertainty and error sources
  - But could benefit from a similar approach as SST
  - Number of sensors is less (likely candidates are SeaWiFS and MODIS Aqua)
- It would be fantastic if SST and OC could interoperate with minimal user programming
- See the datasets developed for the CEOS COVERAGE activity

Closing thoughts

- CEOS COAST will develop/leverage some type of land/ocean ARDs
  - Working examples exist in the CEOS Interoperability Lab
  - Interested in developing workflows and tutorials for ocean ARD datasets
- Co-covening an ARD Oceans session at next CEOS SIT Technical Workshop (Sept 2020) meeting
- CARD4L presents roadmap for ARDs in other disciplines
  - A recommendation strategy and requirement list for building ARD for GHRSST datasets in a first step
- For ARDs what kind of community assessment process is needed?