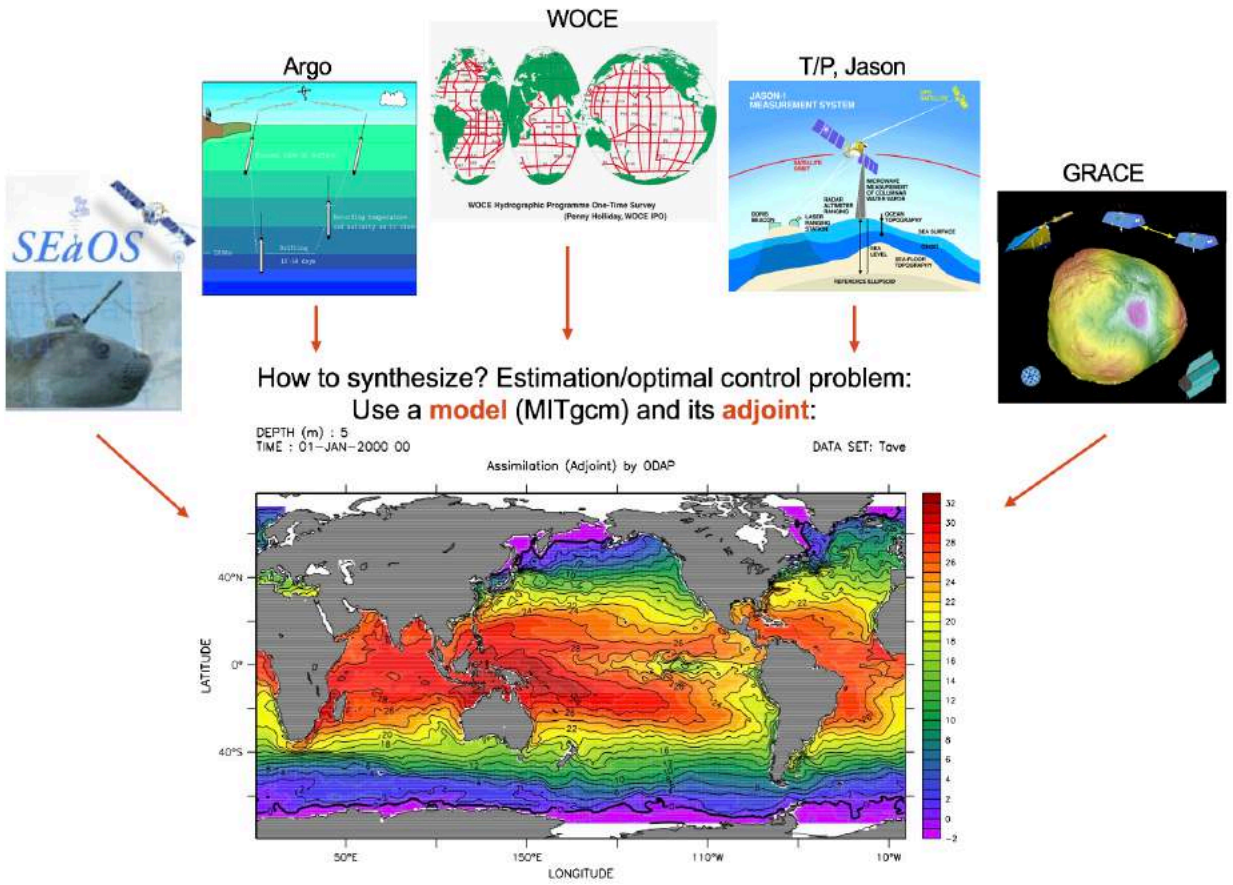
NASA Advancing Collaborative Connections for Earth System Science (ACCESS):

Data Access and the ECCO Ocean and Ice State Estimate



Patrick Heimbach

Institute for Computational Engineering and Sciences
The University of Texas at Austin

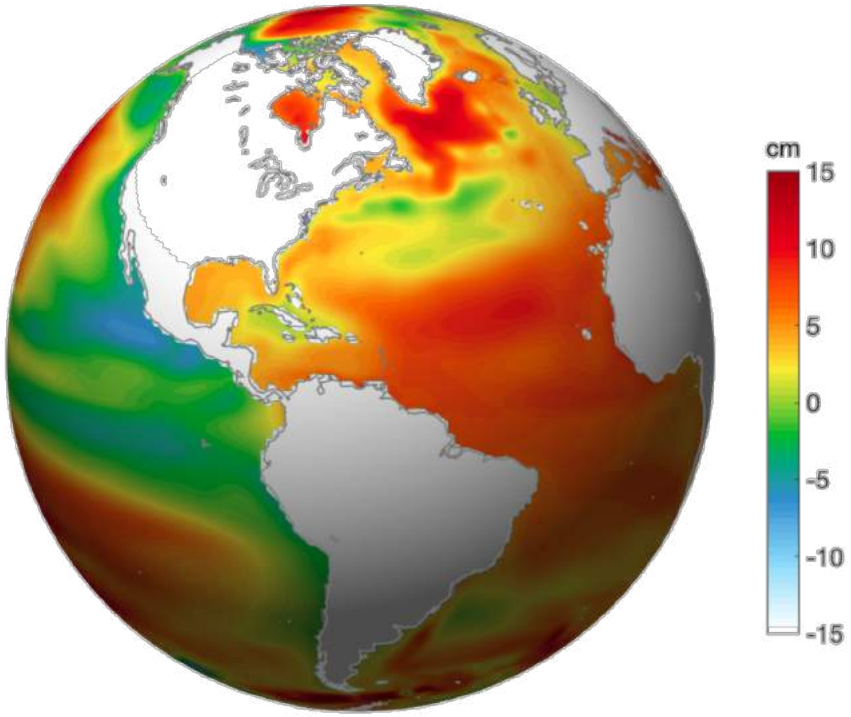
Ian Fenty & Thomas Huang

Jet Propulsion Laboratory
California Institute of Technology

ECCO synthesizes NASA Earth System Observations

The ECCO global ocean state estimation system is the premier tool for **synthesizing NASA's diverse Earth system observations** into a **complete physically-consistent description of Earth's time-evolving full-depth ocean and sea ice system.**

ECCO state estimates are of particular significance to NASA because on their own, all satellite observations, although global in coverage, *remain sparse in space and time relative to the inherent scales of ocean variability, and are blind to the ocean's interior.*



ECCO “Big Data” challenges

- Earth system data // ECCO Inputs
- Ocean State Estimate Product // ECCO Output
- Computational Infrastructure // ECCO Reproducibility

ECCO “Big Data” challenges

- **Earth system data // ECCO Inputs**
- Ocean State Estimate Product // ECCO Output
- Computational Infrastructure // ECCO Reproducibility

Earth system data // *Current ECCO Inputs*

Variable	Observations
Sea surface height	TOPEX/Poseidon (1993-2005), Jason-1 (2002-2008), Jason-2 (2008-2015), Geosat-Follow-On (2001-2007), CryoSat-2 (2011-2015), ERS-1/2 (1992-2001), ENVISAT (2002-2012), SARAL/AltiKa (2013-2015)
Temperature profiles	Argo floats (1995-2015), XBTs (1992-2008), CTDs (1992-2011), Southern Elephant seals as Oceanographic Samplers (SEaOS; 2004-2010), Ice-Tethered Profilers (ITP, 2004-2011) and other high-latitude CTDs and moorings
Salinity profiles	Argo floats (1997-2015), CTDs (1992-2011), SEaOS (2004-2010), and other high-latitude CTDs and moorings
Sea surface temp.	AVHRR (1992-2013)
Sea surface salinity	Aquarius (2011-2013)
Sea-ice concentration	SSM/I DMSP-F11 (1992-2000) and -F13 (1995-2009) and SSMIS DMSP-F17 (2006-2015)
Ocean bot. pressure	GRACE (2002-2014), JPL MASCON Solution
TS climatology	World Ocean Atlas 2009
MDT	DTU13 (1992-2012)
GM SSH & OBP	AVISO, CSIRO, NOAA; GRACE

ECCO “Big Data” challenges

- **Earth system data // ECCO Inputs**
  - *Identifying new or updated datasets (discovery)*
  - *Acquiring / downloading datasets*
  - *Preparing datasets for ingestion into ECCO State Estimation System (preprocessing)*
  - *Recording all steps from acquisition to synthesis in the model (reproducibility)*

ECCO “Big Data” challenges

- Earth system data // ECCO Inputs
- **Ocean State Estimate Product // ECCO Output**
- Computational Infrastructure // ECCO Reproducibility

ECCOv4r3: Monthly and daily mean fields

Ocean + sea-ice

- *T, S, u, v, w, η, ρ, Φ*
- Sea-ice and snow *h* and *c*
- Lateral and vertical fluxes of volume, heat, salt, and momentum

Atmosphere

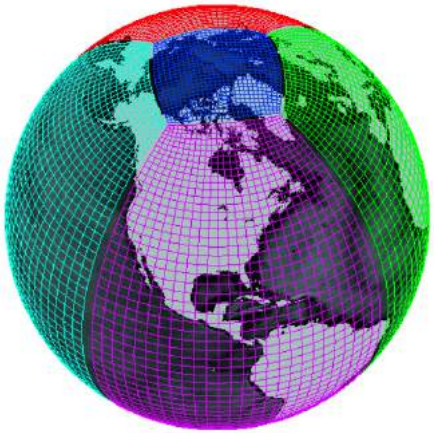
- T, q, |u|, τ, long- and radiative fluxes
- Air–sea-ice–ocean fluxes of heat, moisture, energy, and momentum

Subgrid-scale mixing parameters

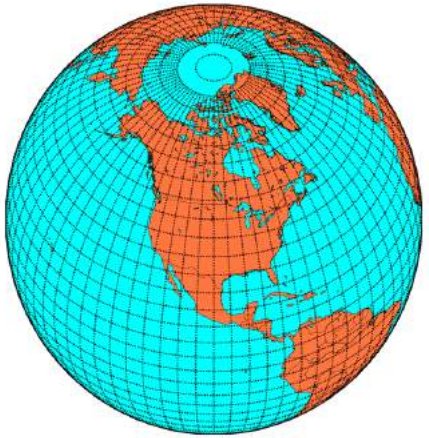
- 3D GM κ and Redi κ
- 3D vertical diffusivity

Fields are provided on two grids

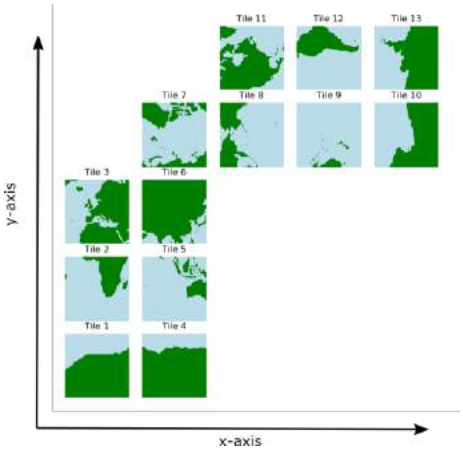
Curvilinear Cartesian
"lat-lon-cap 90"



Interpolated
0.5° lat-lon



13 tiles of 90x90x50



Ocean State Estimate Product // *Current ECCO Output*

Documentation

- Summary
- Analysis plots including climatology
- Instructions for re-running the model and calculating budgets

State estimate fields (NetCDF)

Observational data

Fields required to re-run the model

- Grid geometry
- Configuration files
- Model initial conditions
- Atmospheric and hydrological boundary conditions

Also mirrored at
https://web.corral.tacc.utexas.edu/OceanProjects/ECCO/ECCOv4/Release3/

ECCO “Big Data” challenges

- **Ocean State Estimate Product // ECCO Output**

ECCO v**4** llc90	**0.25 Tb**	1 deg, 50 nz
ECCO v**5** llc270	**3 Tb**	1/3 deg, 50 nz
ECCO v**6** llc1080	**80 Tb**	1/12 deg, 90 nz

  - How can we efficiently distribute ECCO to researchers?
  - How can we make ECCO products discoverable to researchers?

ECCO “Big Data” challenges

- Earth system data // ECCO Inputs
- Ocean State Estimate Product // ECCO Output
- **Computational Infrastructure // ECCO Reproducibility**

3 "Big Data" challenges

- **Computational Infrastructure // ECCO Reproducibility**

ECCO **v4** llc90	96 CPUs /	12 hr	dt = 60 min
ECCO **v5** llc270	787 CPUs /	36 hr	dt = 20 min
ECCO **v6** llc1080	10821 CPUs /	28 d	dt = 4 min
ECCO **v6** llc1080	10821 CPUs /	75 d	dt = 90 sec [tides]

  - How can we ensure reproducibility for researchers without access to large, dedicated supercomputer resources?

*680 hr = 28 days

ECCO “Big Data” challenges

- Earth system data // ECCO Inputs
- Ocean State Estimate Product // ECCO Output
- Computational Infrastructure // ECCO Reproducibility

Data Access and the ECCO Ocean and Ice State Estimate

2017 NASA ACCESS Program:

Advancing Collaborative Connections for Earth System Science

Data Access and the ECCO Ocean and Ice State Estimate

- NASA's Advancing Collaborative Connections for Earth System Science (ACCESS) Program develops and implements technologies to effectively manage, discover, and utilize NASA's archive of Earth observations for scientific research and applications.

- The program complements NASA's Earth Observing System Data and Information System (EOSDIS) by engaging external researchers and software developers in NASA's mission to:
  - "drive advances in science, technology, aeronautics, space exploration, economic vitality, and stewardship of the Earth"
  - "advance knowledge of Earth as a system to meet the challenges of environmental change and to improve life on our planet."

National Aeronautics and Space Administration

Earth Science Data Operations

ESMO Mission Operations

ESDIS Science Operations

Data Acquisition

Flight Operations, Data Capture, Initial Processing, Backup Archive

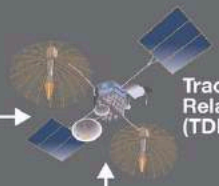
Data Transport

Science Data Processing, Data Management, Interoperable Data Archives, and Distribution

Discovery, Data Access, and Distribution

Earth Science Spacecraft

Tracking and Data Relay Satellite (TDRS)



Network

White Sands Complex (WSC)

EOS Data Operations System (EDOS) Raw Data Processing

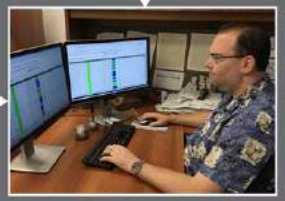
Direct Broadcast (DB)

EOS Polar Ground Stations

EOS Operations Center (EOC) Mission Control

Direct Broadcast/ Direct Readout Stations

EOSDIS Distributed Active Archive Centers (DAACs)



Instrument Teams and Science Investigator-led Processing Systems (SIPSs)

Internet (Search, Order, Distribution)

- Research
- Education
- Value-Added Providers
- U.S. Agency Partners
- Earth System Models
- International Partners
- Decision Support Systems

www.nasa.gov

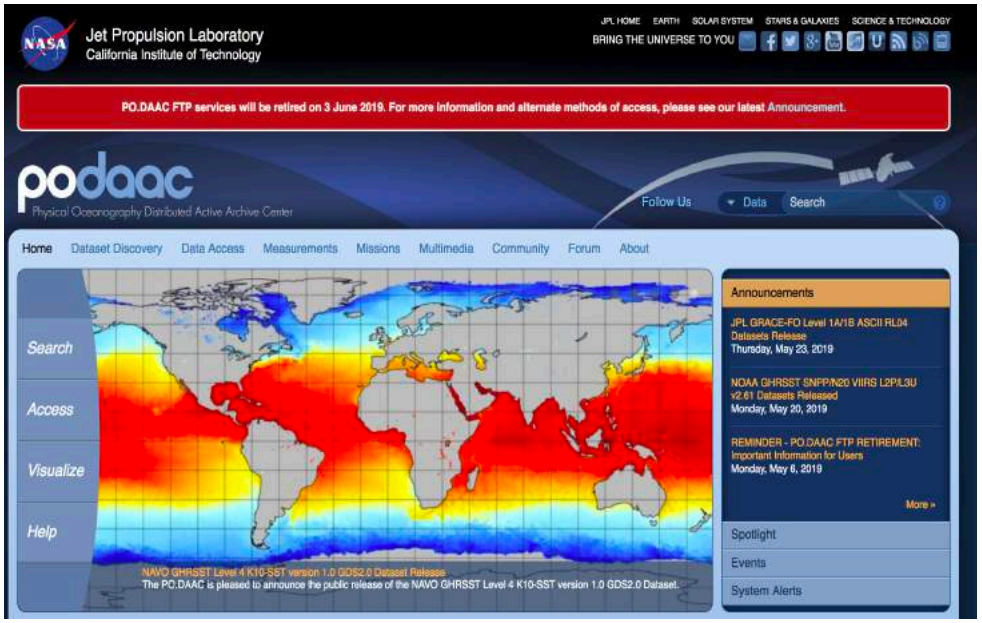
Data Access and the ECCO Ocean and Ice State Estimate

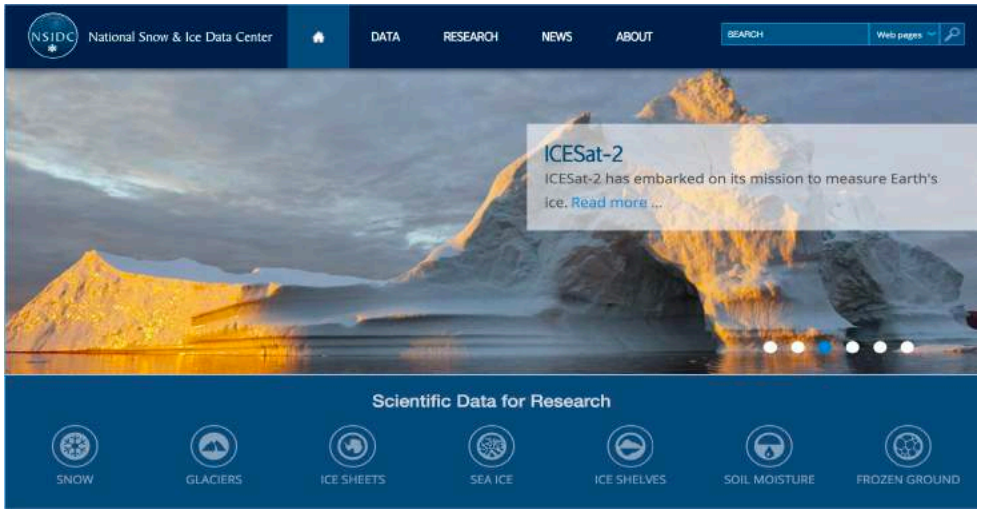
***Goal 1**: Expand and accelerate the integration of NASA Earth system data into the ECCO state estimate in a sustainable and scalable manner*

ECCO-CLOUD: Data Provisioning System (DPS)

- cloud-native storage and software system to automatically **update**, **preprocess**, and **transform** NASA ocean and sea-ice observations provided by EOSDIS close to the source of the data …
- ensures that ECCO state estimates always incorporate the most recent data streams, even as the number and volume of data streams continue to increase.

Earth system data // NASA Distr. Data Archive Centers





Sea surface topography
- Topex/Poseidon
- Jason 1/2/3

Sea surface salinity
- Aquarius
- SMAP

Sea surface temperature
- AVHRR
- Aqua/Terra MODIS

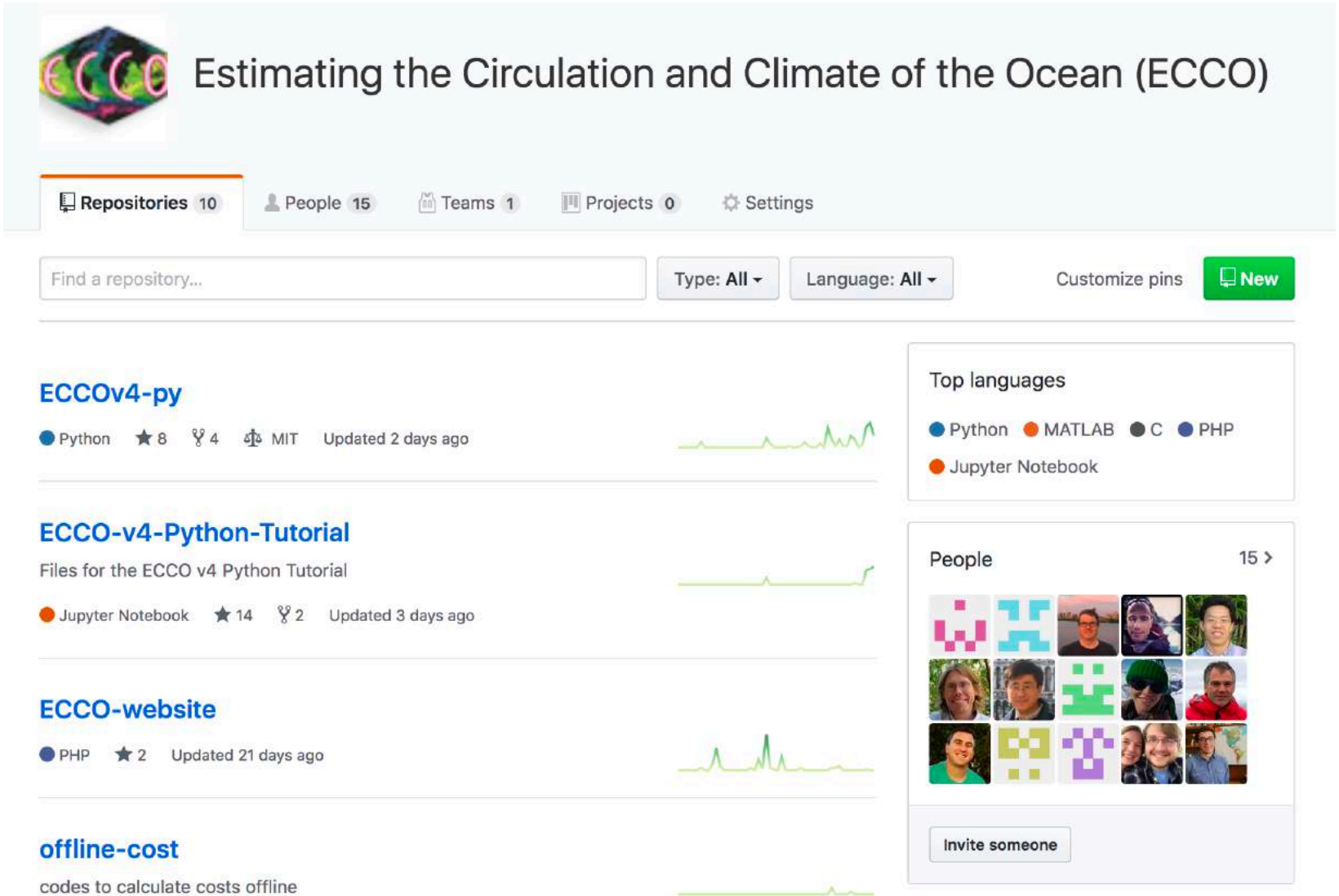
Ocean bottom Pressure
- GRACE, GRACE-FO

Sea ice concentration
- SSMR, SSM/I, SSMIS

Sea ice thickness
- Cryosat-2
- ICESat-1, ICESat-2

Sea ice motion
- AMSR-E,
- AVHRR, DRIFTING BUOYS,
- SMMR, SSM/I, SSMIS

ECCO: Observational data preprocessing, and **transformation codes on GITHUB**



Data Access and the ECCO Ocean and Ice State Estimate

***Goal 2**: facilitate and expand the scientific utilization of NASA remote sensing data integrated in ECCO by the growing community of interdisciplinary researchers*

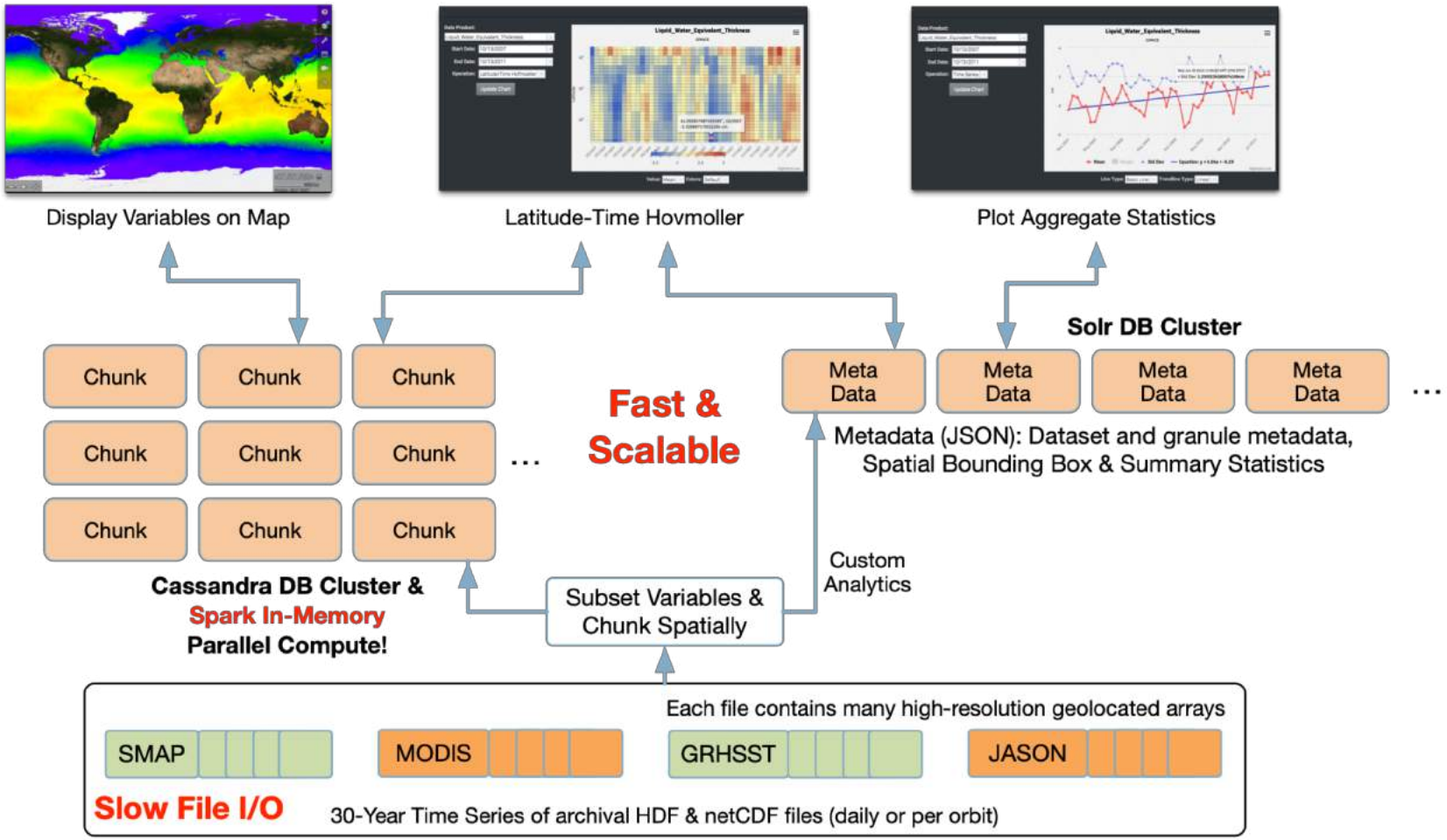
ECCO-CLOUD: Data Distribution System (DDS)

- accelerate the distribution of ECCO products to NASA Data Archiving Centers (PO.DAAC) and other distribution channels and thereby improve its accessibility to the research community.

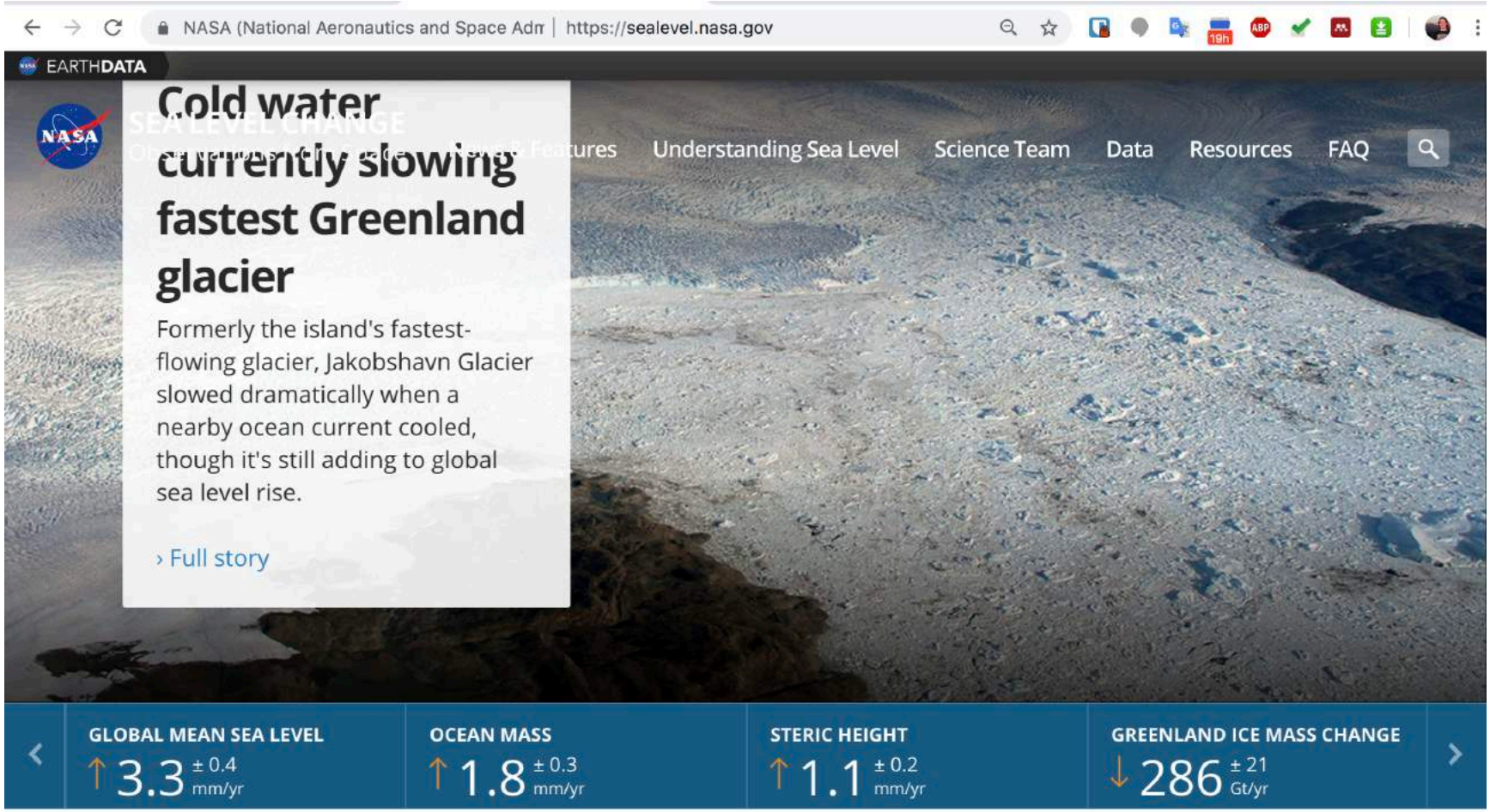
ECCO-CLOUD: Data Analysis System (DAS)

- increase access, exploration, and use of ECCO products and NASA data through web-based tools and web-service interfaces

NASA/JPL "Data Analysis Tool"



"Data Analysis Tool" on NASA Sea Level Change Portal



NASA (National Aeronautics and Space Adm | https://sealevel.nasa.gov/data/data-analysis-tool

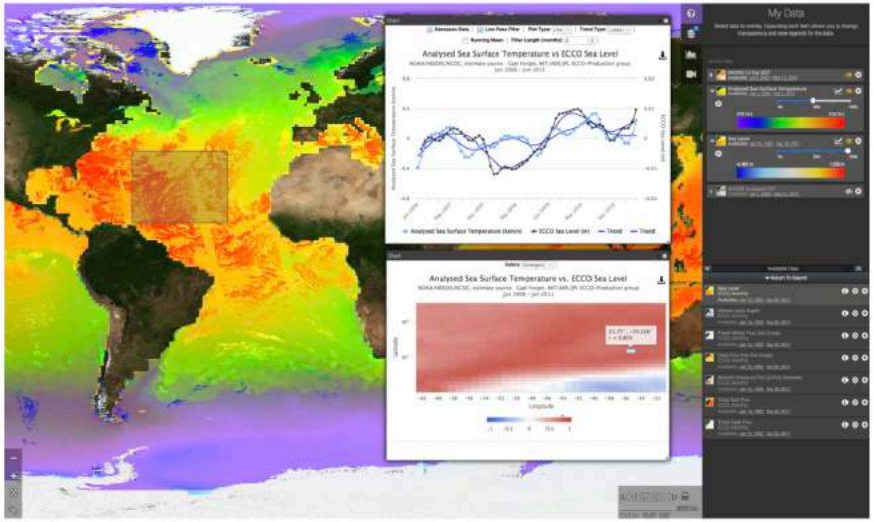
EARTHDATA

SEA LEVEL CHANGE
Observations from Space

News & Features | Understanding Sea Level | Science Team | Data | Resources | FAQ

DATA

Data Analysis Tool - Beta Version



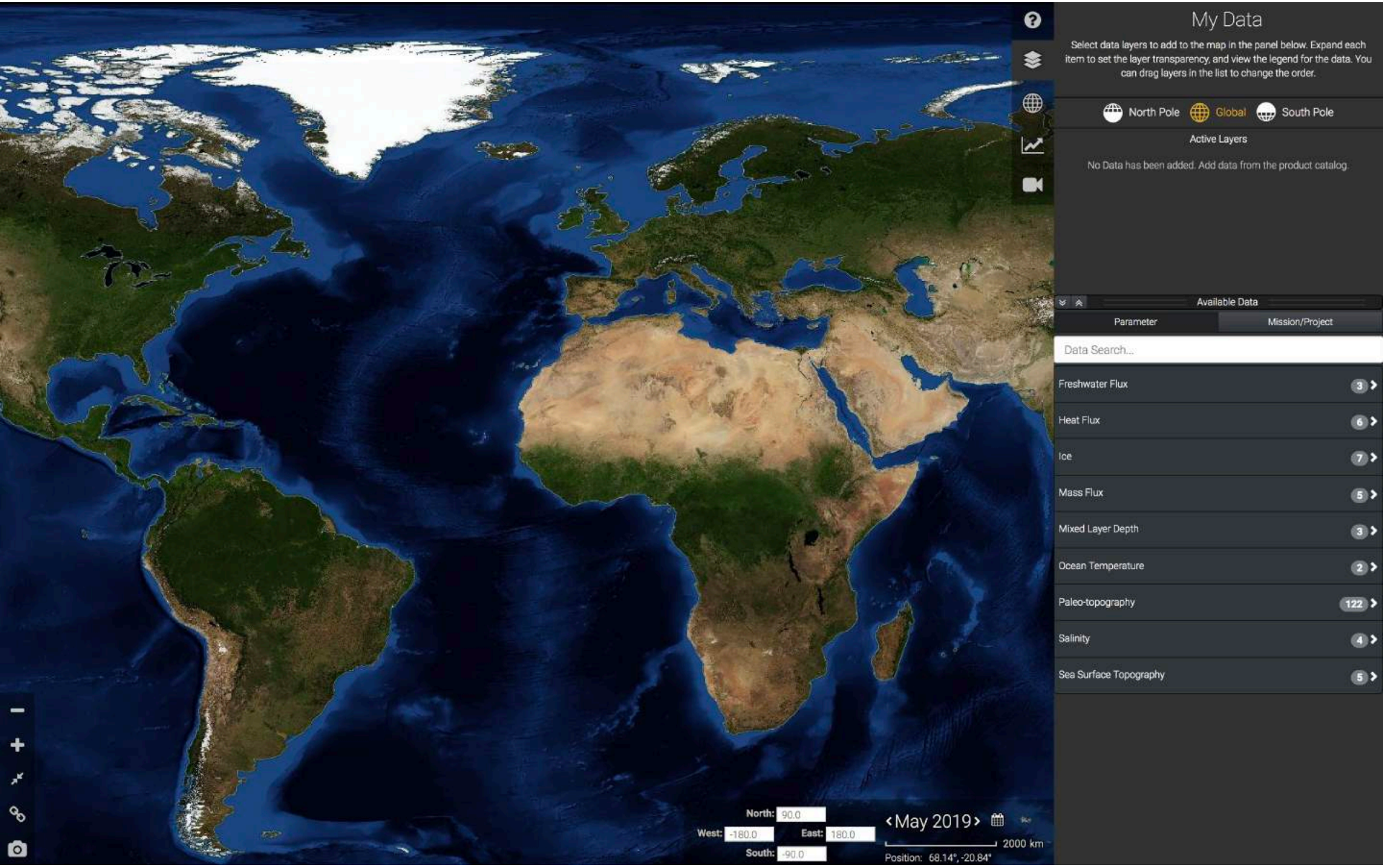
LAUNCH TOOL

The NASA Sea Level Change Data Analysis Tool (DAT) has been designed to allow for quick-look comparisons and analysis of NASA datasets of sea level change. The datasets range from sea level observations, to ice observations, to model output to quickly study anomalies and get

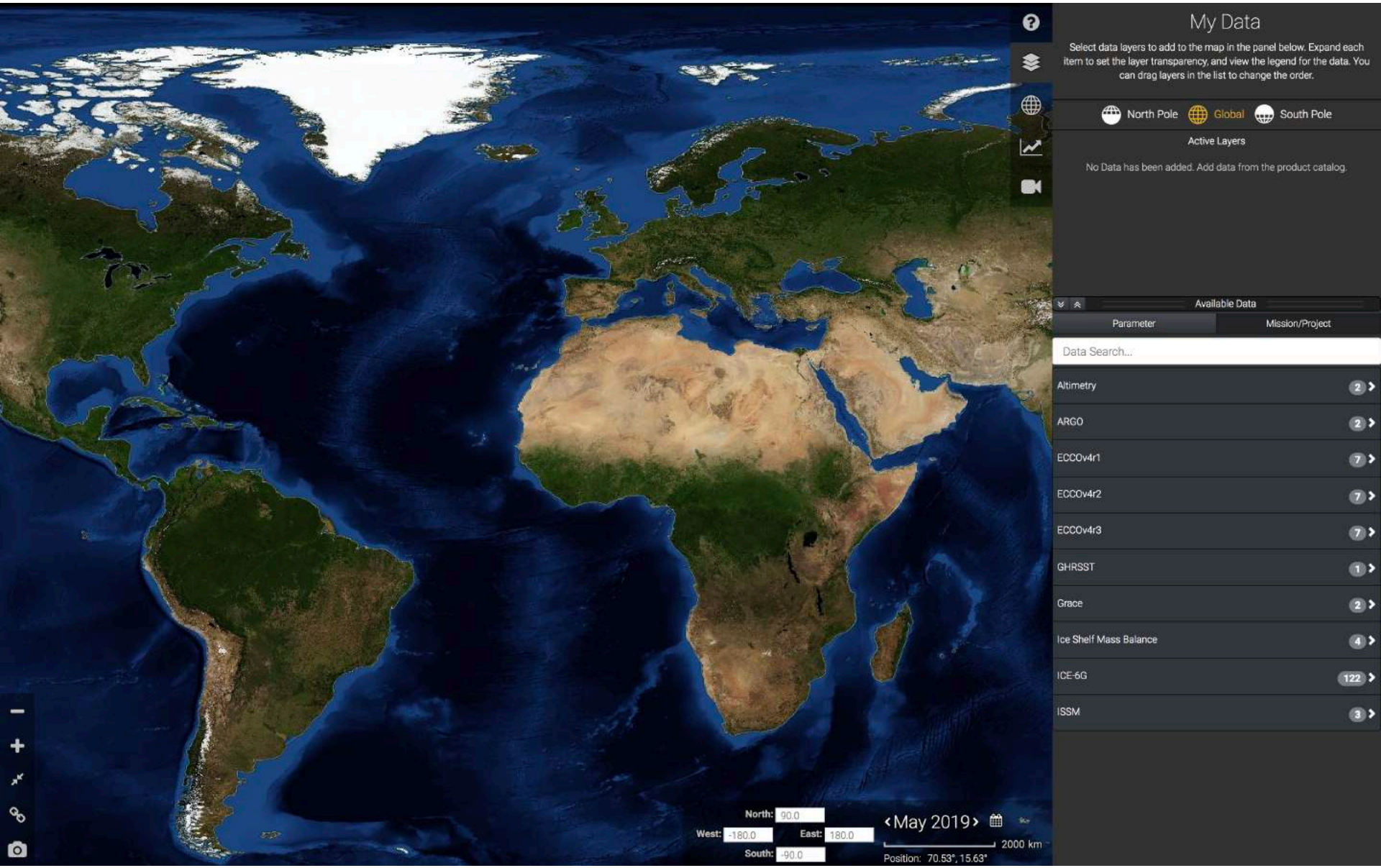
Analyze NASA datasets with the Data Analysis Tool

- **View and compare data layers**
  Select one or more layers and adjust visibility by controlling opacity.
- **Work with a region of interest**
  Draw a box or drop a pin with analysis tools to get statistics (min, max, standard deviation, trend, correlation) on data values.
- **Analyze and compare time series**
  The area averaged time series can be plotted for up to two datasets for comparison. Filters

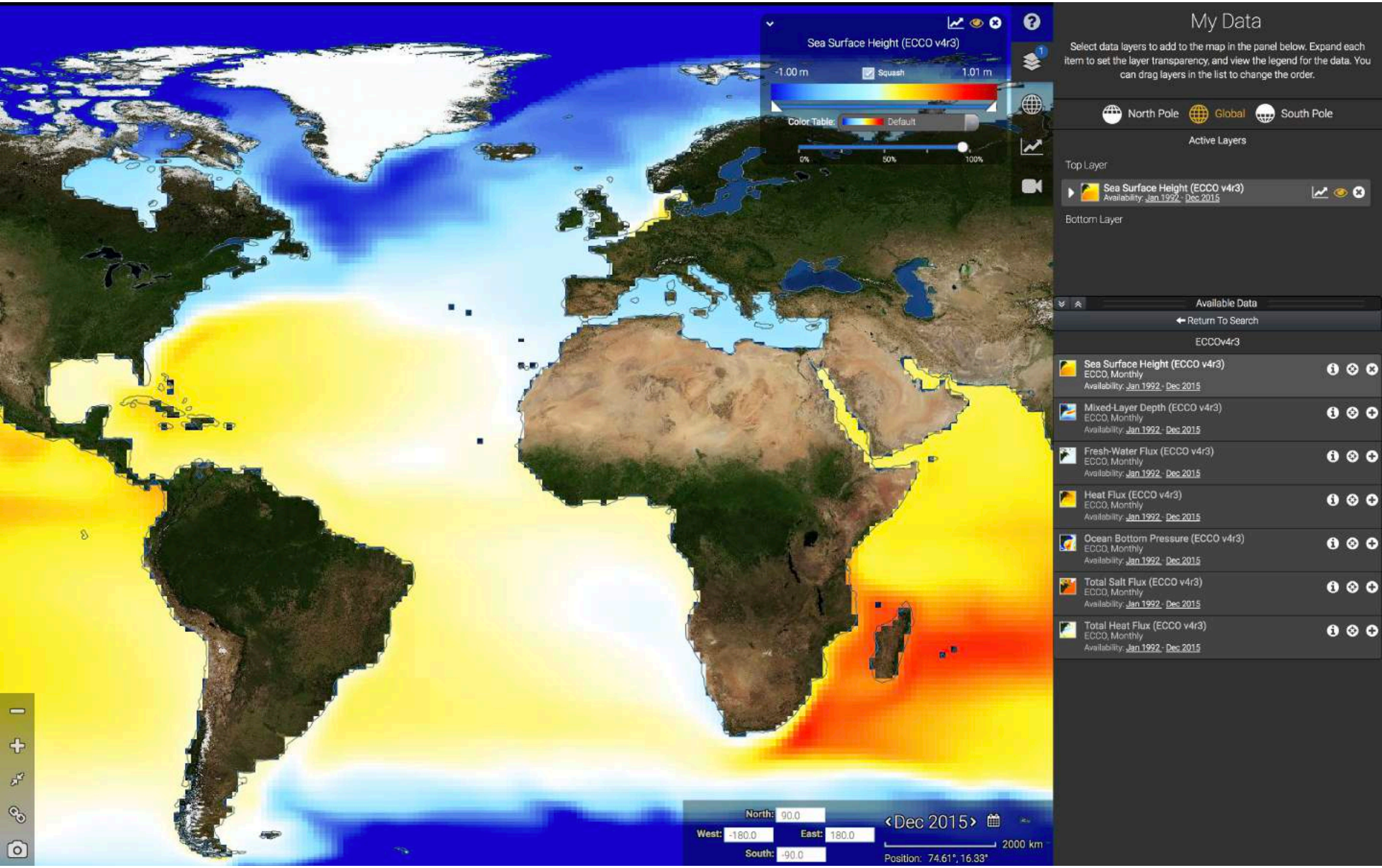
“Data Analysis Tool" on NASA Sea Level Change Portal



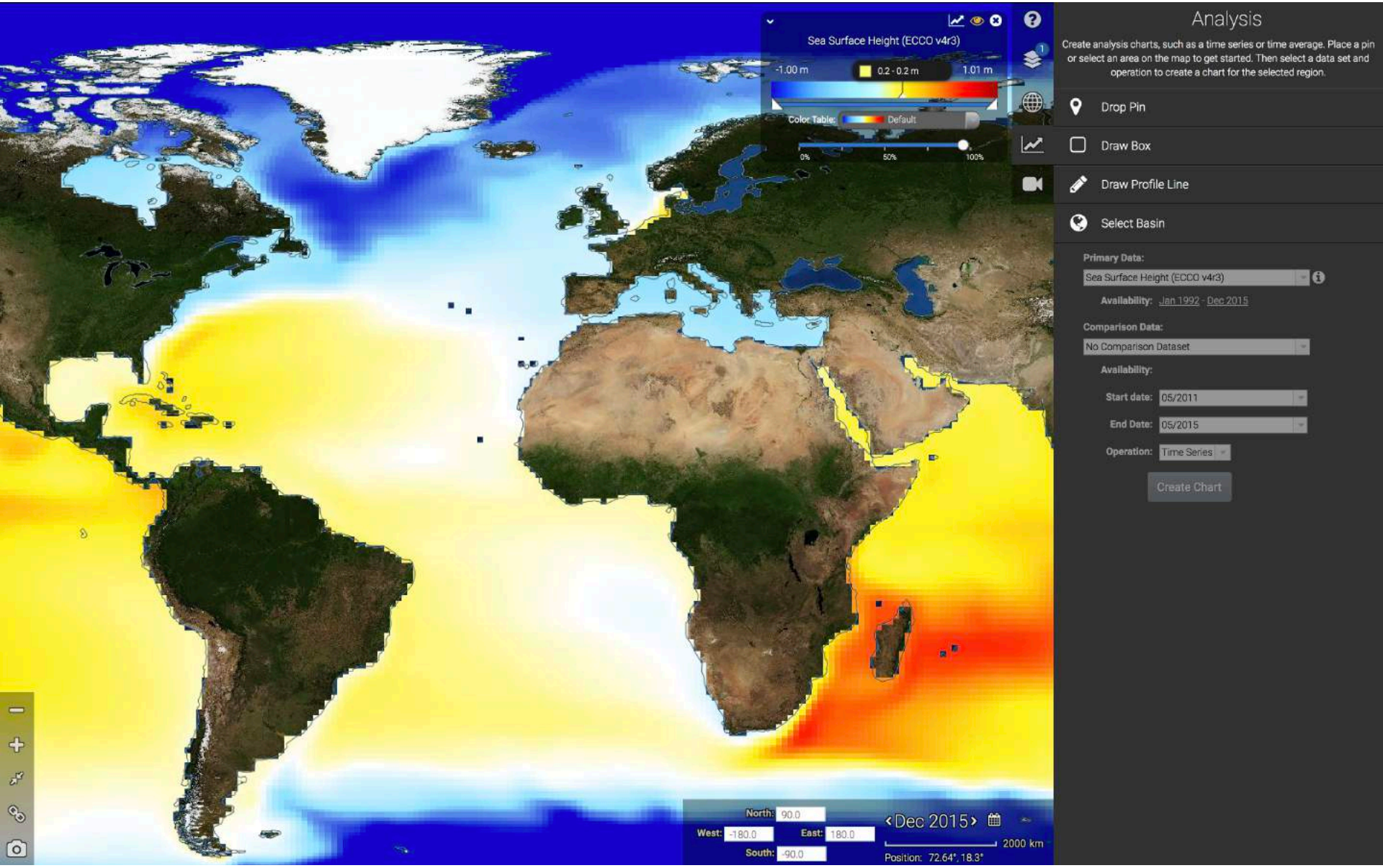
“Data Analysis Tool" on NASA Sea Level Change Portal



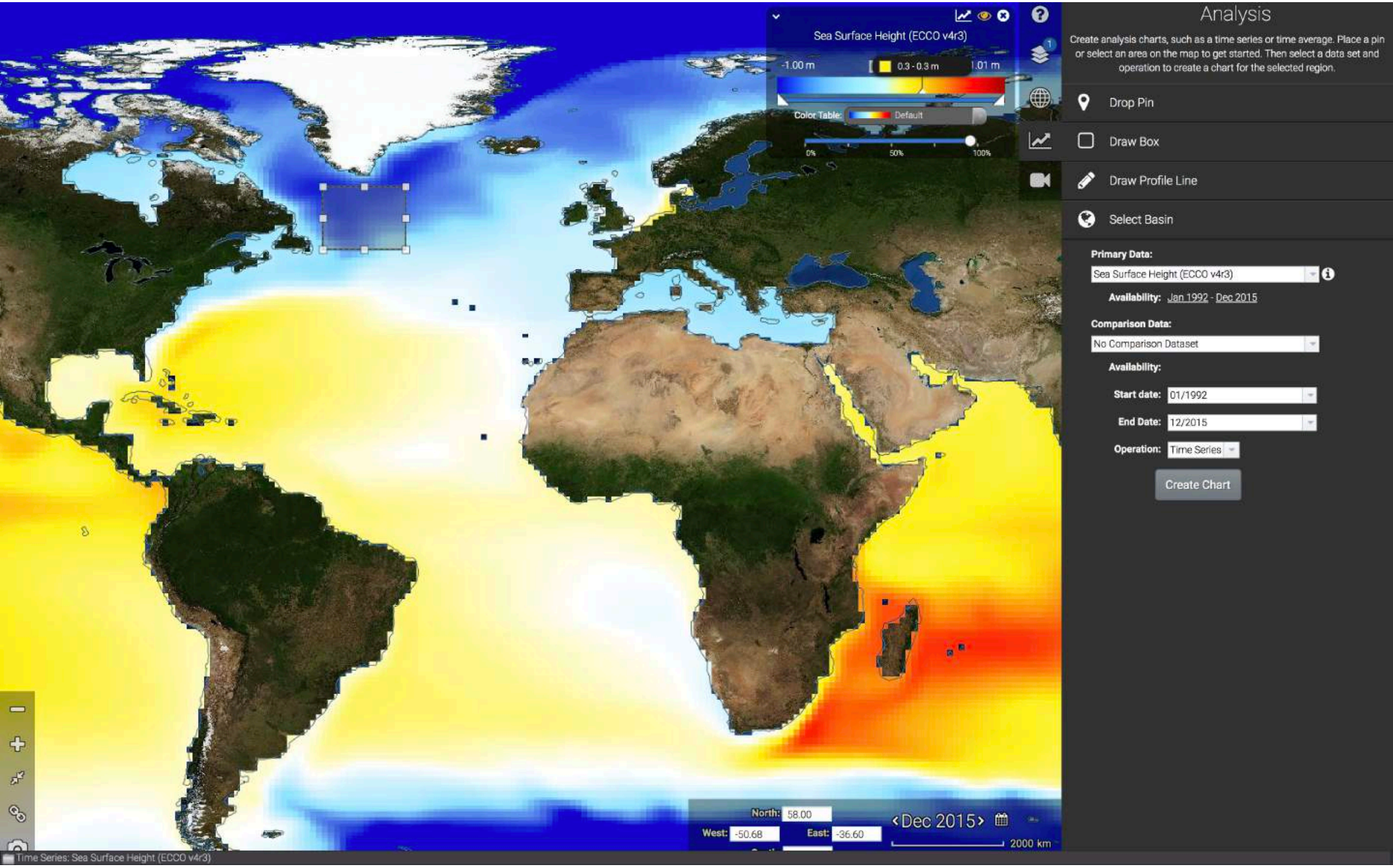
"Data Analysis Tool" on NASA Sea Level Change Portal



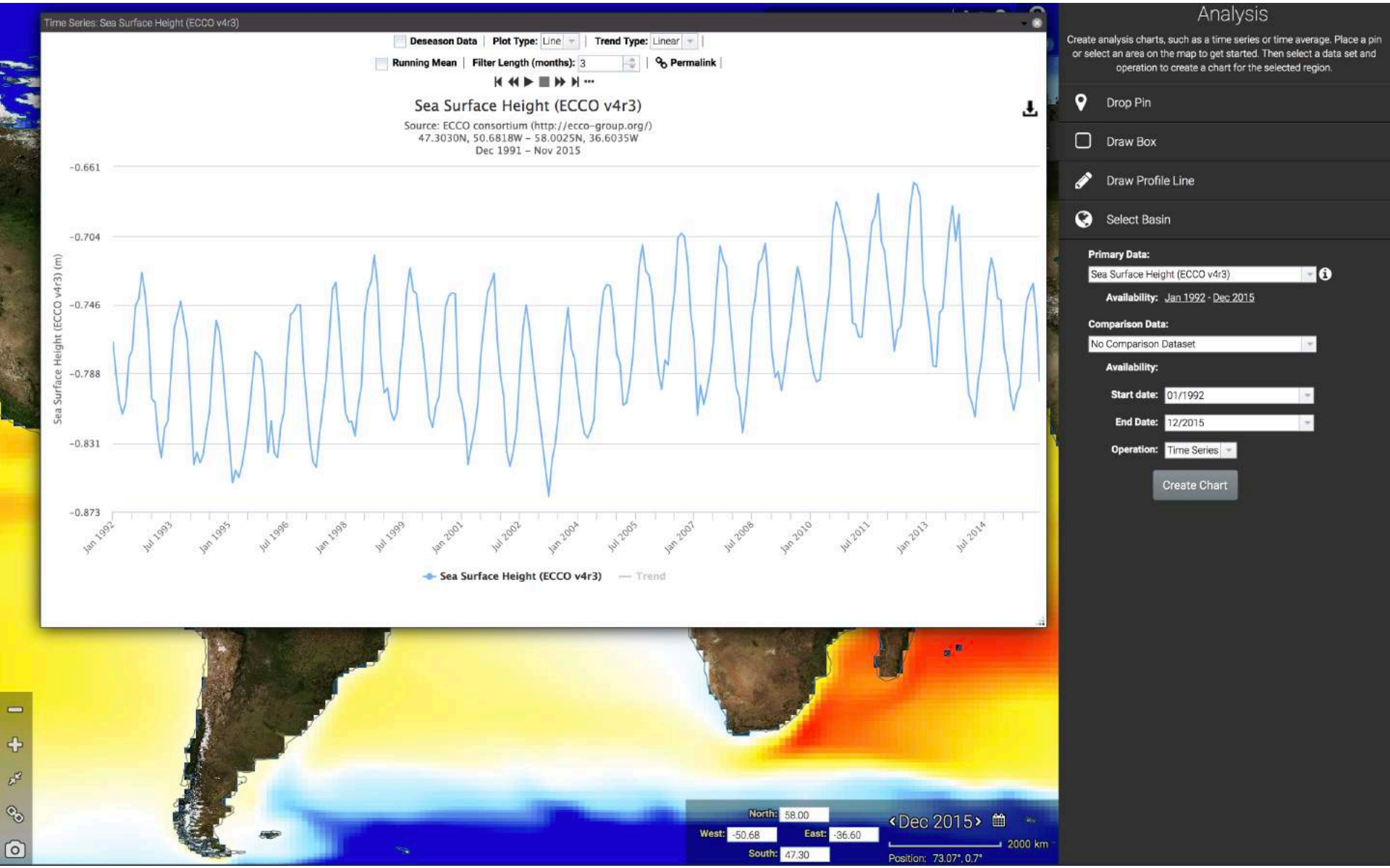
"Data Analysis Tool" on NASA Sea Level Change Portal



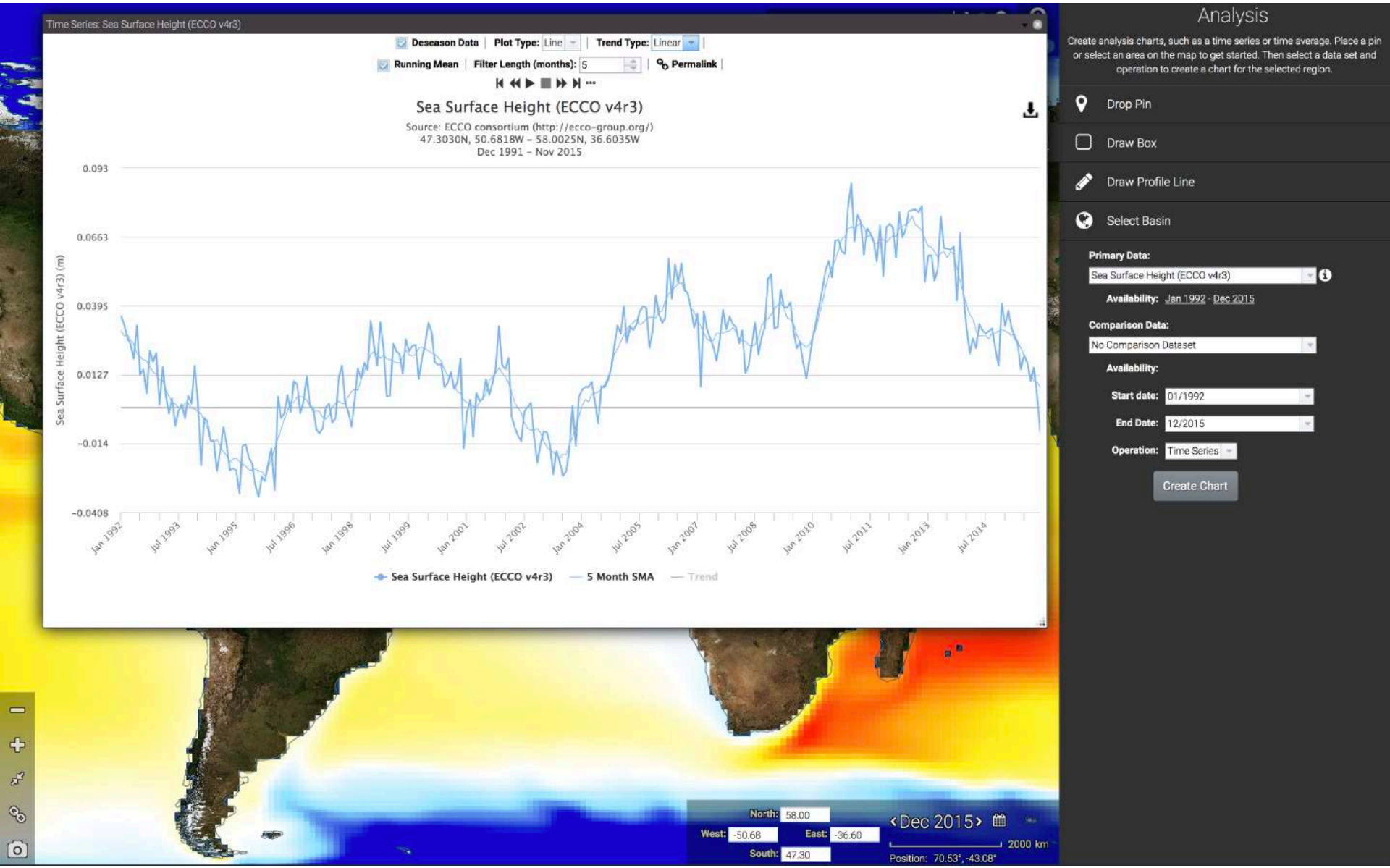
"Data Analysis Tool" on NASA Sea Level Change Portal



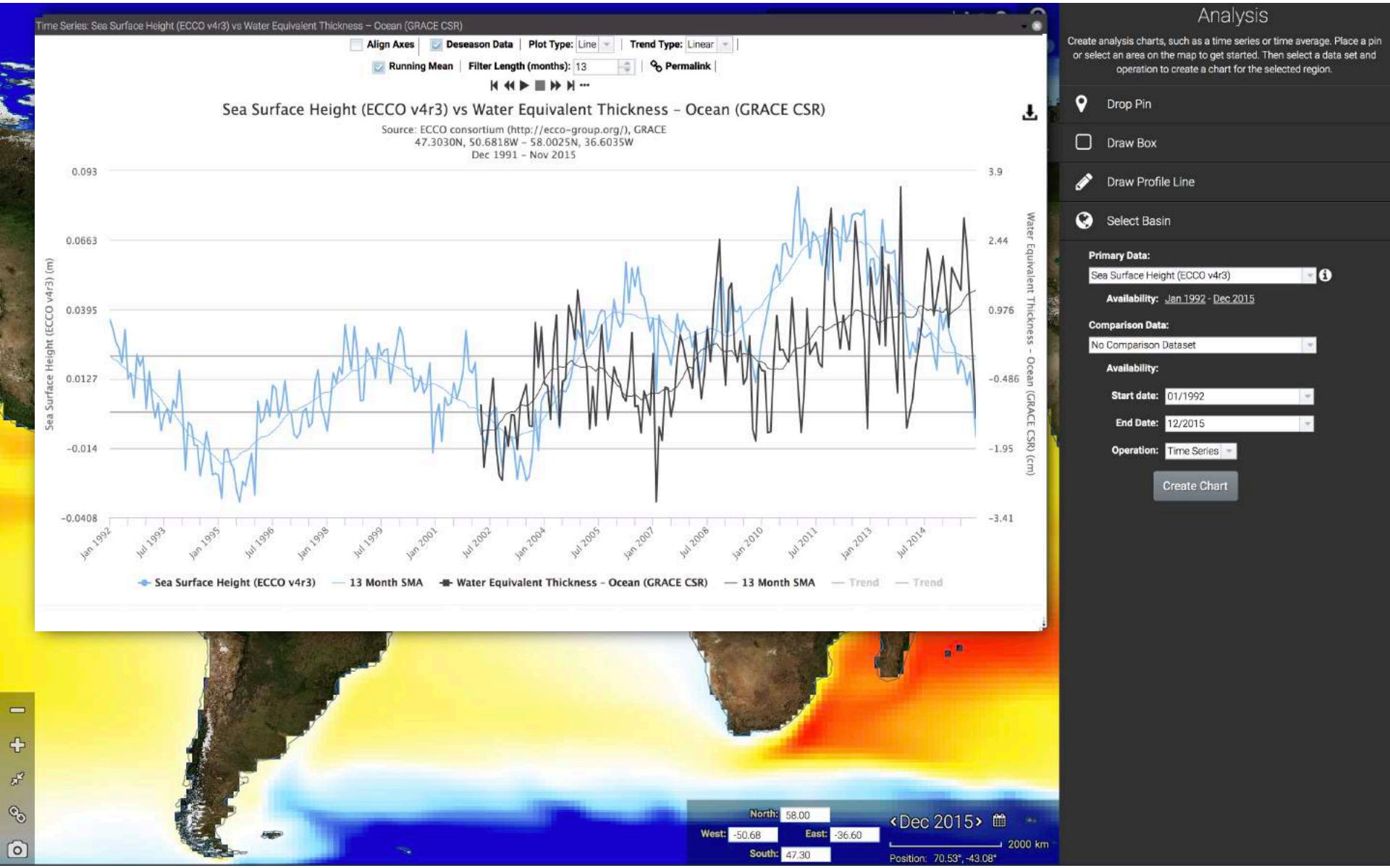
"Data Analysis Tool" on NASA Sea Level Change Portal



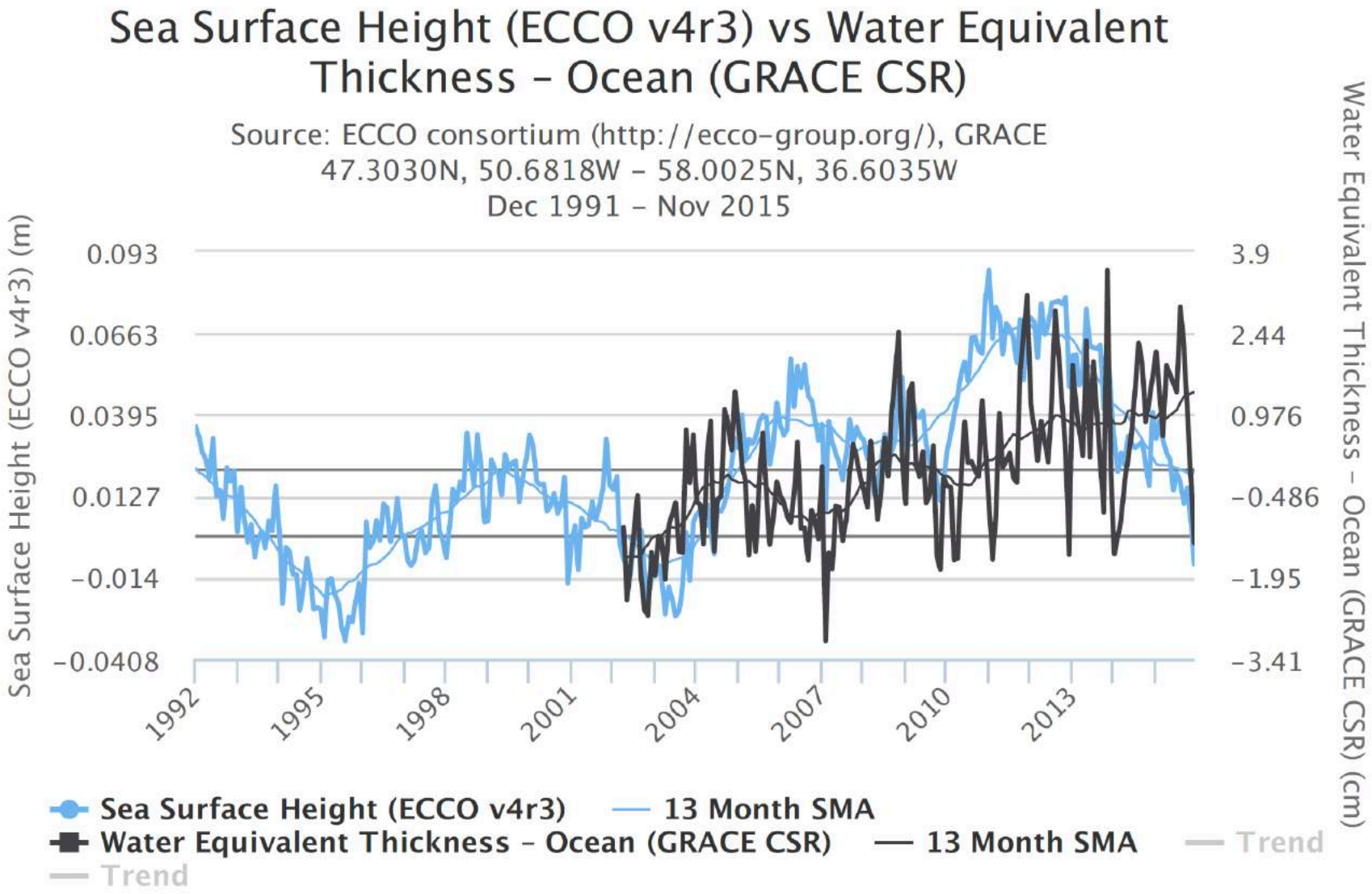
"Data Analysis Tool" on NASA Sea Level Change Portal



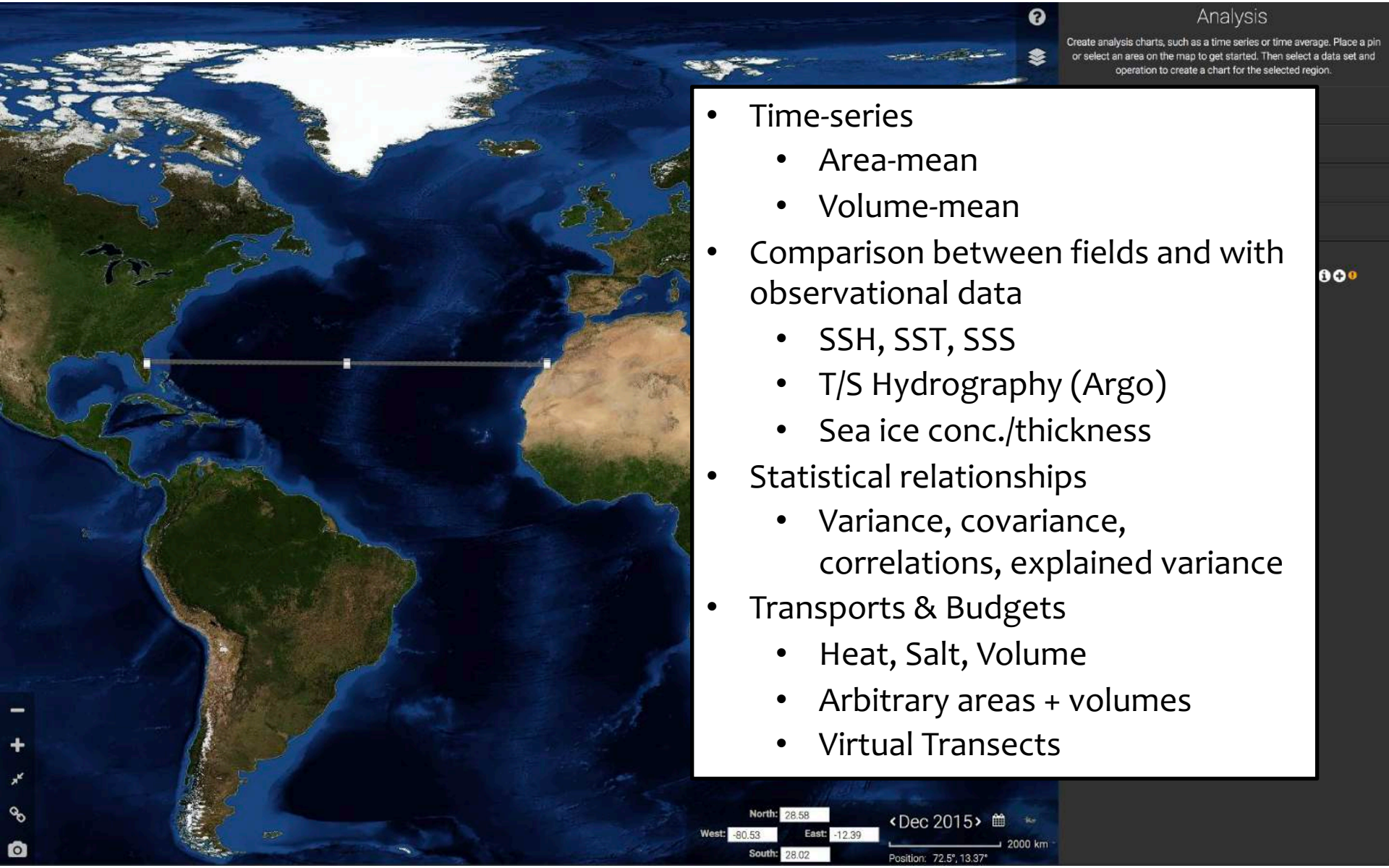
“Data Analysis Tool" on NASA Sea Level Change Portal



"Data Analysis Tool" on NASA Sea Level Change Portal



"Data Analysis Tool" on NASA Sea Level Change Portal



- Time-series
  - Area-mean
  - Volume-mean
- Comparison between fields and with observational data
  - SSH, SST, SSS
  - T/S Hydrography (Argo)
  - Sea ice conc./thickness
- Statistical relationships
  - Variance, covariance, correlations, explained variance
- Transports & Budgets
  - Heat, Salt, Volume
  - Arbitrary areas + volumes
  - Virtual Transects

"ECCO Jupyter Notebook"

```
# Request NEXUS to compute SST Time Series 2008/9/1 - 2015/10/1
# for the "blob" warming off Western Canada and plot the means
…
ds='AVHRR_OI_L4_GHRSST_NCEI'

url = … # construct the webservice URL request


# make request to NEXUS using URL request
# save JSON response in local variable
ts = json.loads(str(requests.get(url).text))

# extract dates and means from the response
means = []
dates = []
for data in ts['data']:
    means.append (data[0]['mean'])
    d = datetime.datetime.fromtimestamp((data[0]['time']))
    dates.append (d)

# plot the result
…
```

```
https://oceanxtremes.jpl.nasa.gov/timeSeriesSpark?spark=me
sos,16,32&ds=AVHRR_OI_L4_GHRSST_NCEI&minLat=45&minLon=-
150&maxLat=60&maxLon=-
120&startTime=1220227200&endTime=1443657600

It took: 2.9428272247314453 sec
```

User using Jupyter Notebook to call NEXUS' RESTUful API for time series generation. The NEXUS service is hosted at Amazon Cloud.

"ECCO Jupyter Notebook"

Questions that arise:

- Who is allowed access?
- Who pays for the egress costs and cloud computations?
  - Prevent attacks / spurious / malformed requests
- Where to do the calculations?
  - Locally → data must moved to local machine
  - Remotely → users submit chains of operations

$$V = \sum_{i,j,k} rA(i,j)\, drF(k)\, hFacC(i,j,k)$$

$$< T(t) >= \frac{1}{V} \sum_{i,j,k} T(i,j,k,t)\, drF(k)\, rA(i,j)\, hFacC(i,j,k)$$

Data Access and the ECCO Ocean and Ice State Estimate

***Goal 3**: enable reproduction of ECCO model solutions via elastic (on-demand) cloud-based computational resources*

ECCO-CLOUD: Elastic Reproducibility System (ERS)

- Provide an online front-end via *Amazon Elastic Compute Cloud (Amazon EC2)* for users so that they can (a) reproduce the full state estimate, and (b) formulate and conduct their own experimental simulations, based upon the ECCO configuration

ECCO-CLOUD: Elastic Reproducibility System (ERS

Cloud-based solutions for distributed climate modeling

Nadya Vinogradova[1], Mark Shiffer[1], Gael Forget[2], and Chris Hill[2]
[1]Cambridge Climate Institute, MA
[2]Massachusetts Institute of Technology, MA

Climate models integrate our best knowledge of the climate system behavior, its governing principles and ongoing changes, providing unique tools for studying the Earth's past, present, and future states. In addition to their widespread use by the research community, model-based solutions offer crucial guidance for decision-makers in their efforts to anticipate and mitigate hazards associated with climate change. The success of both efforts is often tied to the ability of a user to interpret model results and reproduce solutions in order to build on previous achievements. However, modeling capabilities remain limited in their accessibility, as re-running simulations created by other groups can require expertise and manpower. Furthermore, potential users may face challenges associated with limited on premise computational and storage resources. These common impediments slow down the overall progress of model development, diminish the general openness of modeling activities, and make collaboration between various groups less efficient.

In this respect, cloud-based approaches open up promising new avenues for widely collaborative and distributed climate modeling. Today, running climate models in the cloud has become a practical alternative to the use of conventional on premise or government-sponsored computing facilities. Here, we present a framework that leverages existing cloud services and enables researchers to easily develop, archive, re-use, and share modeling tools (Fig. 1).

https://dspace.mit.edu/bitstream/handle/1721.1/111605/eccoCloud.pdf?sequence=1

Users who may lack on-premise computational resources or IT support can use the included cloud computing recipe to leverage *Amazon Web Services*'s `cfncluster` technology. This recipe sets up a complete computational environment in the *AWS* cloud (hardware, software, model, and inputs). When this recipe was tested in January 2017, the 20-year *ECCO v4 r2* model run took under 36h using 96 vCPUs and *AWS spot instances* for a cost of about 40$.

Data Access and the ECCO Ocean and Ice State Estimate

***Goal 1**: Expand and accelerate the integration of NASA Earth system data into the ECCO state estimate in a sustainable and scalable manner*

***Goal 2**: facilitate and expand the scientific utilization of NASA remote sensing data integrated in ECCO by the growing community of interdisciplinary researchers*

***Goal 3**: enable reproduction of ECCO model solutions via elastic (on-demand) cloud-based computational resources*

ECCO-CLOUD: Data Provisioning System (DPS)
ECCO-CLOUD: Data Distribution System (DDS)
ECCO-CLOUD: Data Analysis System (DAS)
ECCO-CLOUD: Elastic Reproducibility System (ERS)

ECCO-CLOUD schematic