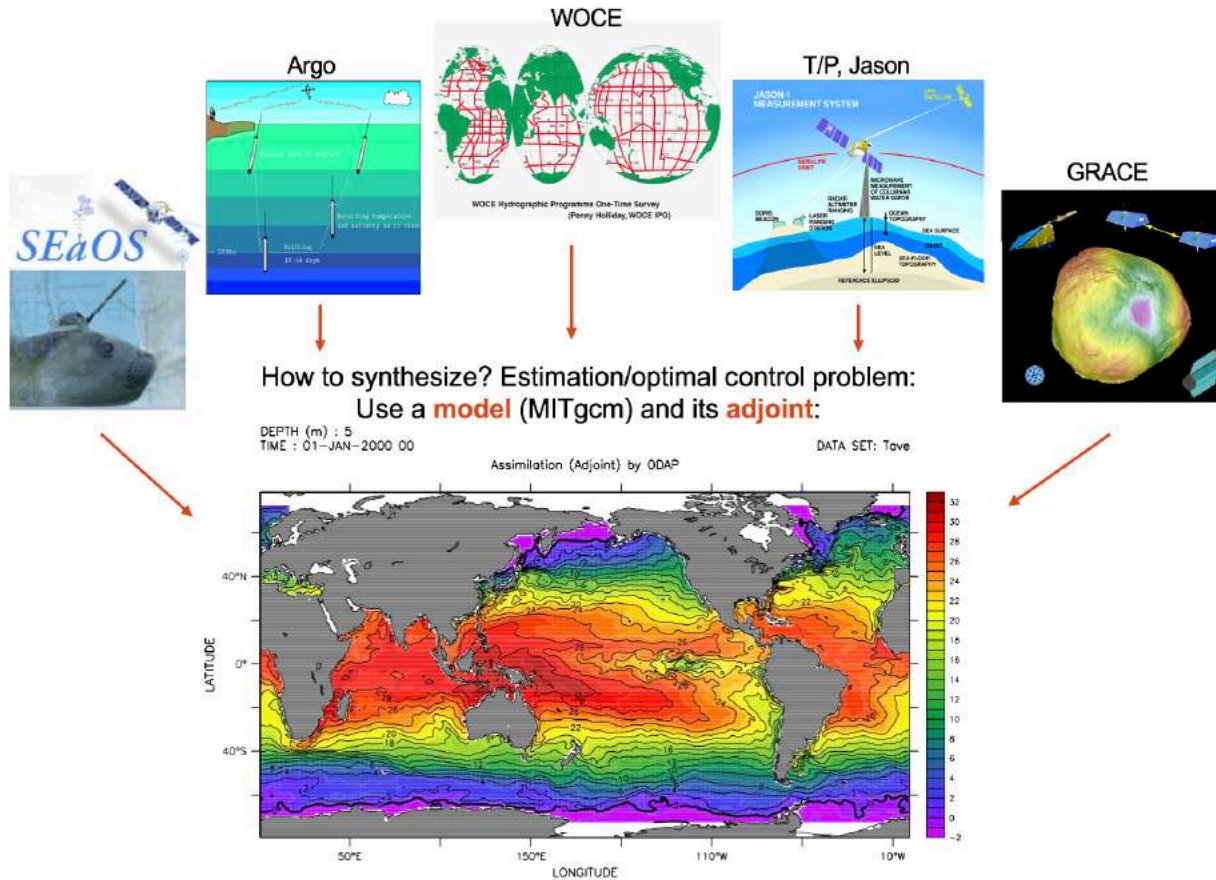


NASA Advancing Collaborative Connections for Earth System Science (ACCESS): *Data Access and the ECCO Ocean and Ice State Estimate*



Patrick Heimbach

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

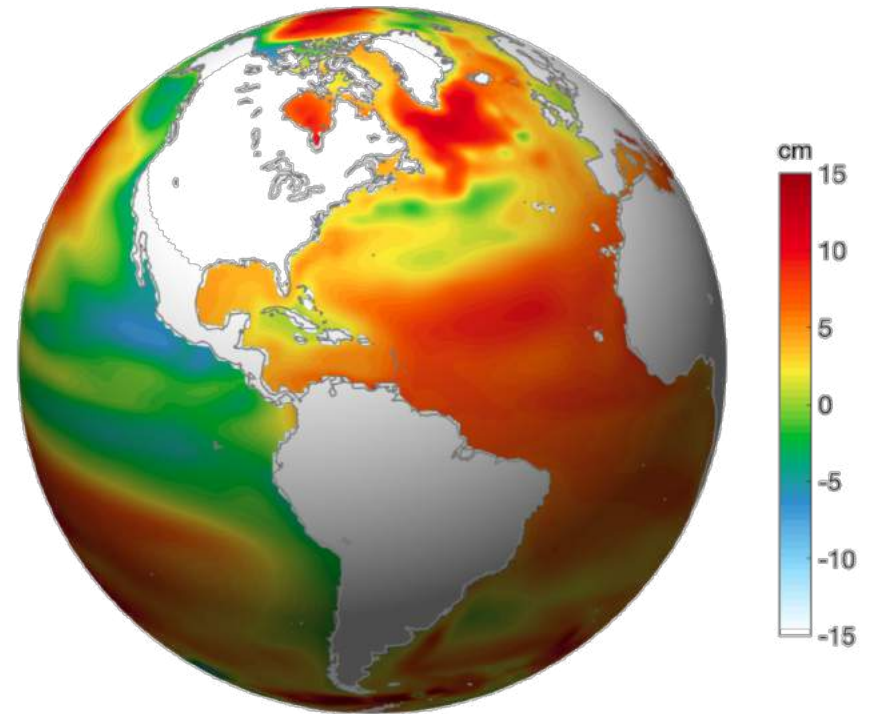
Ian Fenty & Thomas Huang

*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

- $T, S, u, v, w, \eta, \rho, \Phi$
- Sea-ice and snow h and c
- Lateral and vertical fluxes of volume, heat, salt, and momentum

Atmosphere

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

Subgrid-scale mixing parameters

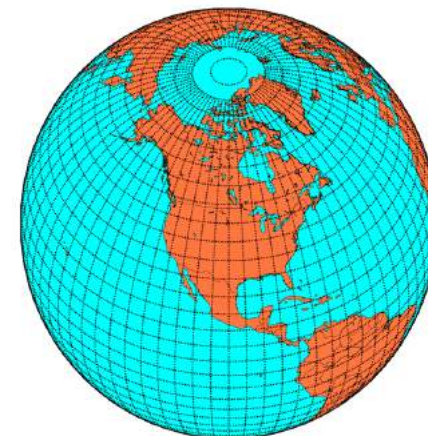
- 3D GM κ and Redi κ
- 3D vertical diffusivity

Fields are provided on two grids

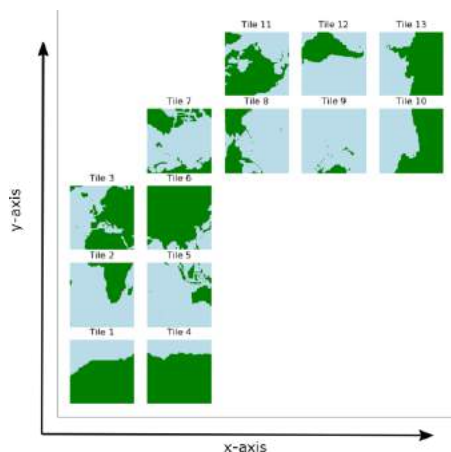
Curvilinear Cartesian
“lat-lon-cap 90”





Interpolated
 0.5° lat-lon



13 tiles of 90x90x50



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

-  README
-  doc/
-  input_ecco/
-  input_forcing/
-  input_init/
-  interp_monthly/
-  nctiles_daily/
-  nctiles_grid/
-  nctiles_monthly/
-  nctiles_monthly_snapshots/
-  other/
-  profiles/

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 v4 llc90 **0.25 Tb** 1 deg, 50 nz

ECCO v5 llc270 **3 Tb** 1/3 deg, 50 nz

ECCO v6 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

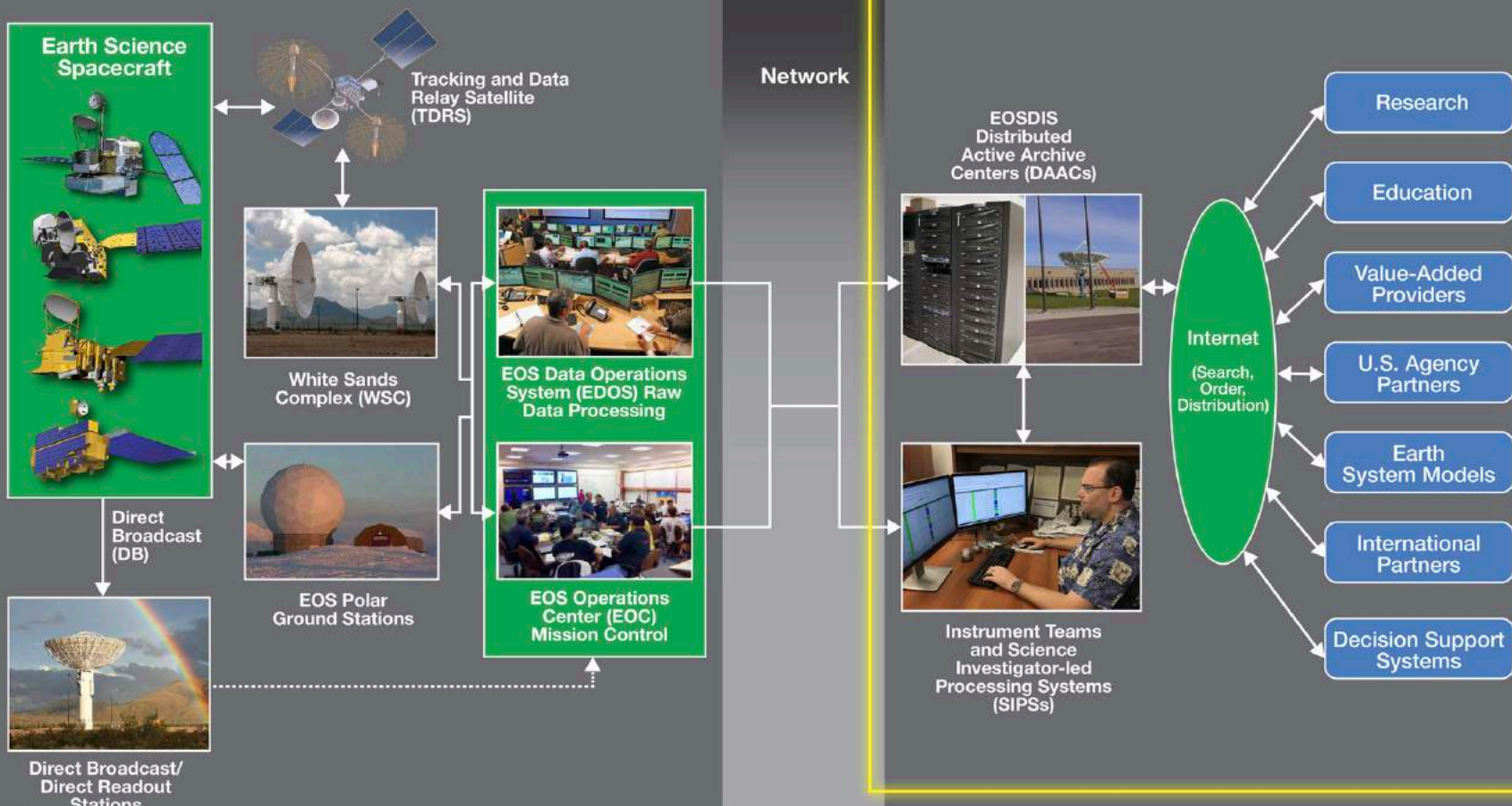
- NASA's Advancing Collaborative Connections for Earth System Science (ACCESS) Program develops and implements technologies to effectively manage, discover, and utilize NASA's archive of Earth observations for scientific research and applications.
- The program complements NASA's Earth Observing System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

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**



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

The screenshot shows the PO.DAAC (Physical Oceanography Distributed Active Archive Center) website. At the top, it features the NASA logo and the Jet Propulsion Laboratory California Institute of Technology name. A navigation bar includes links for JPL HOME, EARTH, SOLAR SYSTEM, STARS & GALAXIES, and SCIENCE & TECHNOLOGY. A red banner at the top states: "PO.DAAC FTP services will be retired on 3 June 2019. For more information and alternate methods of access, please see our latest Announcement." The main header includes the "podaac" logo and the tagline "Physical Oceanography Distributed Active Archive Center". Below this is a navigation menu with options: Home, Dataset Discovery, Data Access, Measurements, Missions, Multimedia, Community, Forum, and About. A central feature is a world map displaying sea surface temperature data with a color scale from blue (cold) to red (warm). To the right of the map is an "Announcements" section with three entries: "JPL GRACE-FO Level 1A/1B ASCII RL04 Dataset Release" (Thursday, May 23, 2019), "NOAA GHRSSST SNPP/M20 VIIRS L2PL3U v2.61 Dataset Released" (Monday, May 20, 2019), and a "REMINDER - PO.DAAC FTP RETIREMENT: Important Information for Users" (Monday, May 6, 2019). Below the announcements are sections for "Spotlight", "Events", and "System Alerts". A sidebar on the left contains links for "Search", "Access", "Visualize", and "Help".

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

The screenshot shows the NSIDC (National Snow and Ice Data Center) website. The header includes the NSIDC logo and the text "National Snow & Ice Data Center". A navigation bar contains links for DATA, RESEARCH, NEWS, and ABOUT, along with a search box and a "Web pages" link. The main content area features a large image of an ice formation with a text overlay that reads: "ICESat-2 has embarked on its mission to measure Earth's ice. Read more ...". Below the image is a section titled "Scientific Data for Research" with six circular icons representing different data categories: SNOW, GLACIERS, ICE SHEETS, SEA ICE, ICE SHELVES, and SOIL MOISTURE. The bottom of the page shows a row of icons for FROZEN GROUND.

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



Estimating the Circulation and Climate of the Ocean (ECCO)

Repositories 10

People 15

Teams 1

Projects 0

Settings

Find a repository...

Type: All

Language: All

Customize pins

New

ECCOv4-py

Python 8 stars 4 forks MIT Updated 2 days ago



ECCO-v4-Python-Tutorial

Files for the ECCO v4 Python Tutorial

Jupyter Notebook 14 stars 2 forks Updated 3 days ago



ECCO-website

PHP 2 stars Updated 21 days ago



offline-cost

codes to calculate costs offline

Python 1 star Updated on Apr 10



Top languages

Python MATLAB C PHP
Jupyter Notebook

People

15 >



Invite someone

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*

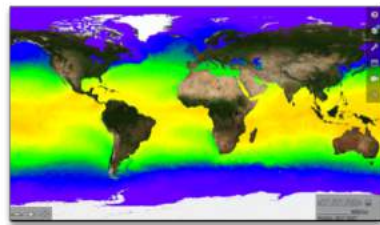
ECCO-CLOUD: Data Distribution System (DDS)

- 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"



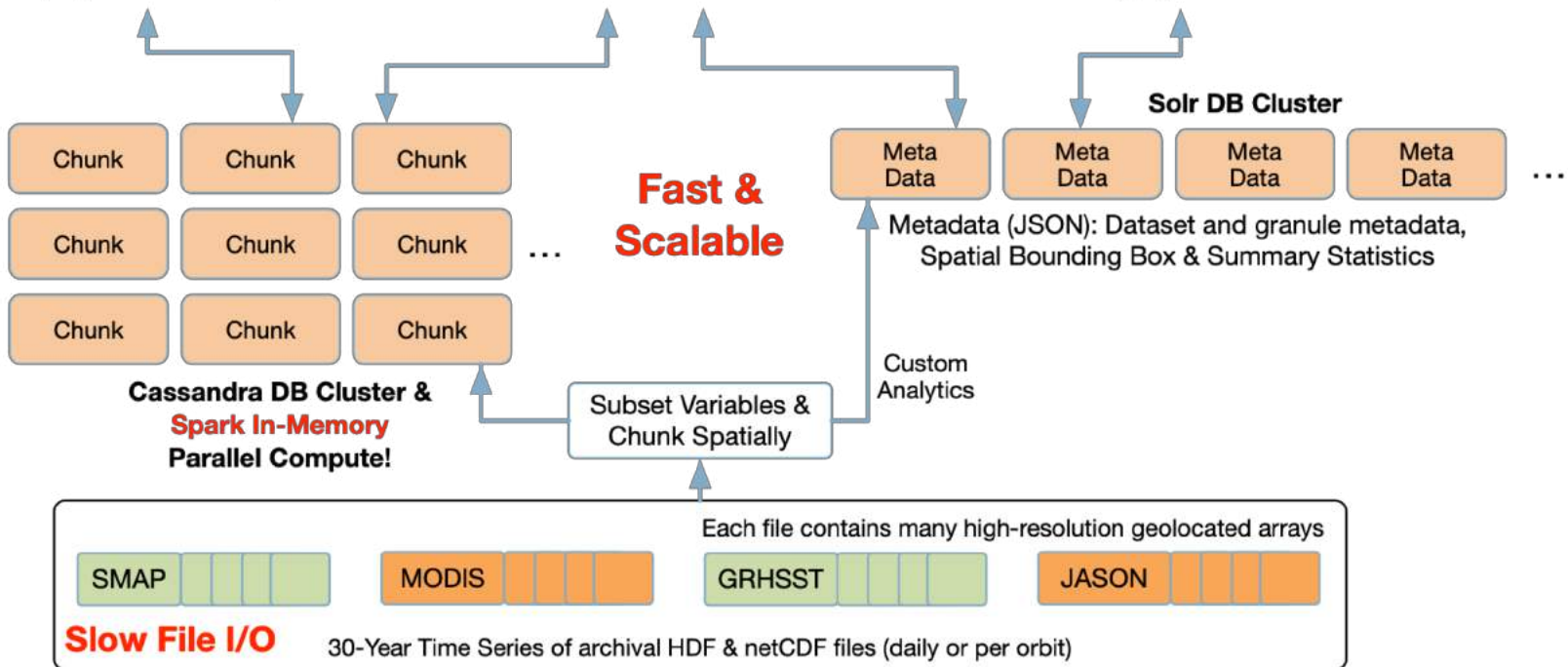
Display Variables on Map



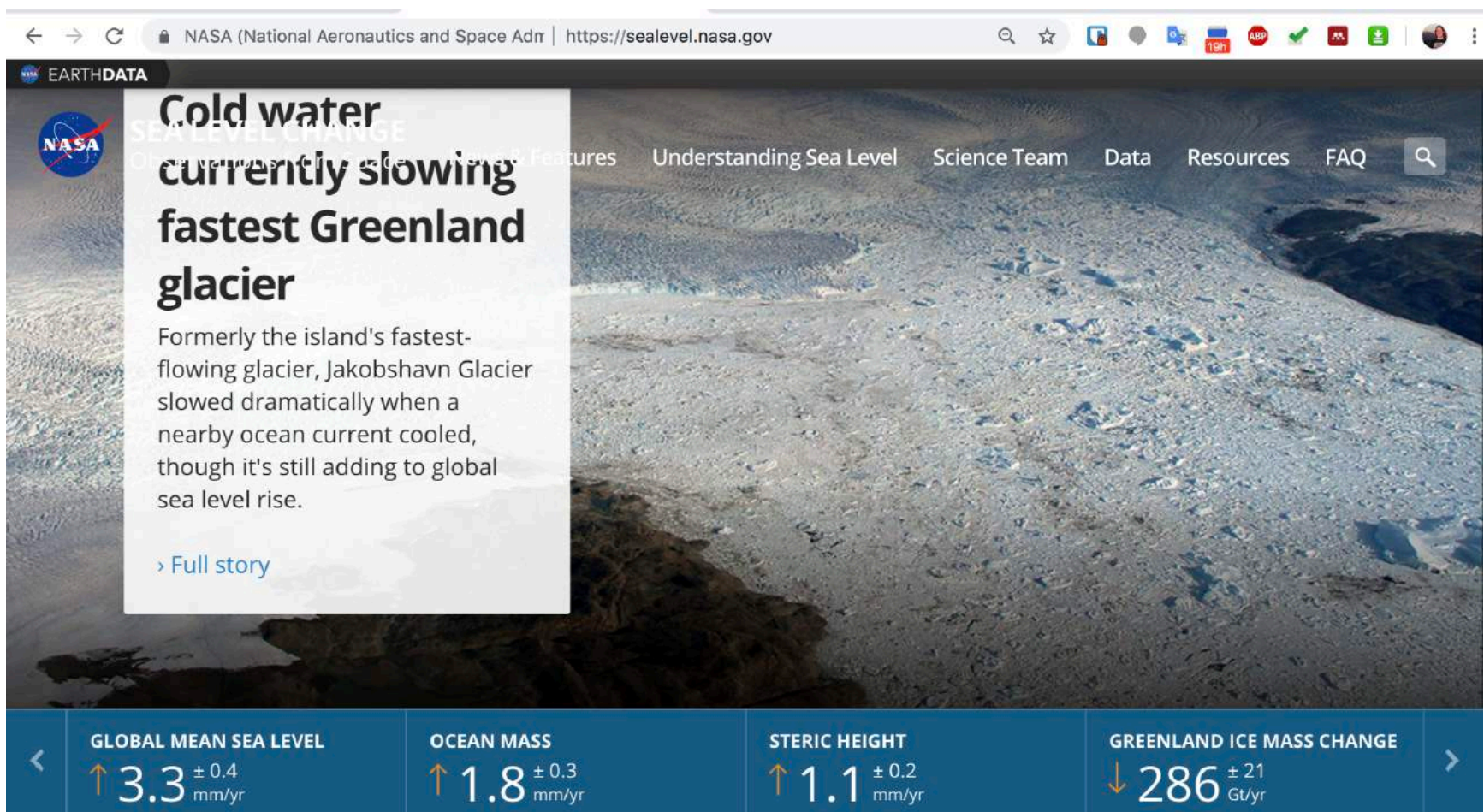
Latitude-Time Hovmoller



Plot Aggregate Statistics



"Data Analysis Tool" on NASA Sea Level Change Portal



Cold water currently slowing fastest Greenland glacier

SEA LEVEL CHANGE

Formerly the island's fastest-flowing glacier, Jakobshavn Glacier slowed dramatically when a nearby ocean current cooled, though it's still adding to global sea level rise.

[Full story](#)

GLOBAL MEAN SEA LEVEL
↑ 3.3 ± 0.4 mm/yr

OCEAN MASS
↑ 1.8 ± 0.3 mm/yr

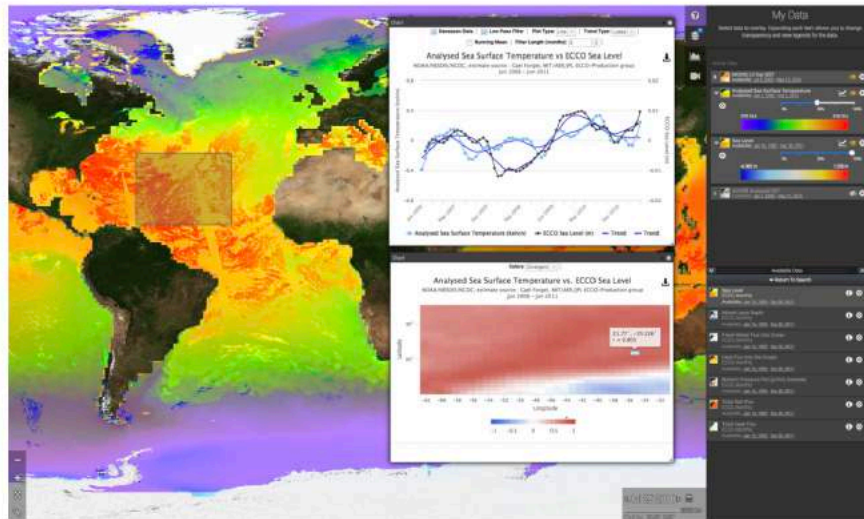
STERIC HEIGHT
↑ 1.1 ± 0.2 mm/yr

GREENLAND ICE MASS CHANGE
↓ 286 ± 21 Gt/yr



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

My Data

Select data layers to add to the map in the panel below. Expand each item to set the layer transparency, and view the legend for the data. You can drag layers in the list to change the order.

North Pole Global South Pole

Active Layers

No Data has been added. Add data from the product catalog.

Available Data

Parameter	Mission/Project
Data Search...	
Freshwater Flux	3 >
Heat Flux	6 >
Ice	7 >
Mass Flux	5 >
Mixed Layer Depth	3 >
Ocean Temperature	2 >
Paleo-topography	122 >
Salinity	4 >
Sea Surface Topography	5 >

North: 90.0 West: -180.0 East: 180.0 South: 90.0

< May 2019 >

Position: 68.14°, -20.84° 2000 km

"Data Analysis Tool" on NASA Sea Level Change Portal

My Data

Select data layers to add to the map in the panel below. Expand each item to set the layer transparency, and view the legend for the data. You can drag layers in the list to change the order.

North Pole Global South Pole

Active Layers

No Data has been added. Add data from the product catalog.

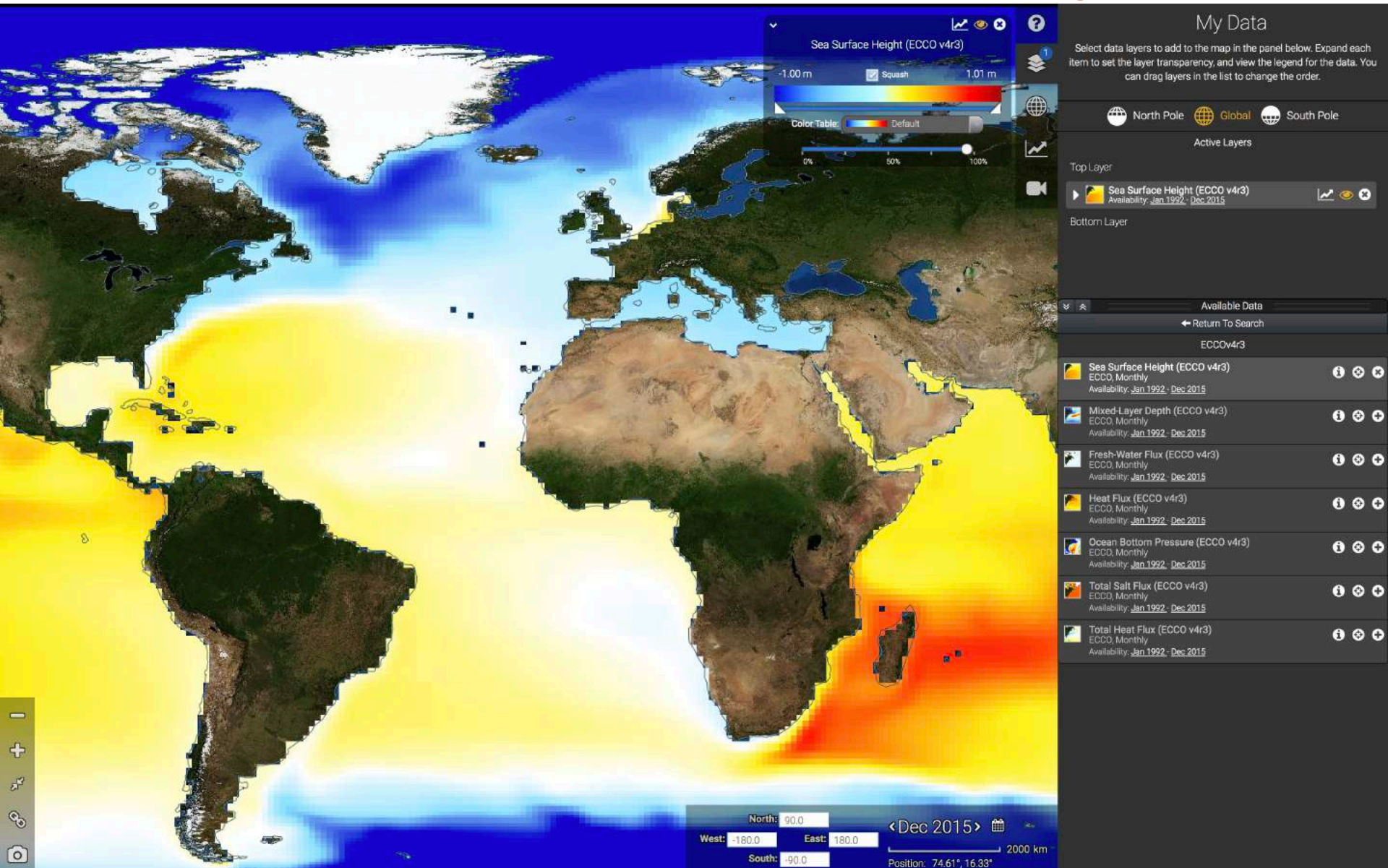
Available Data

Parameter	Mission/Project
Data Search...	
Altimetry	2 >
ARGO	2 >
ECCOv4r1	7 >
ECCOv4r2	7 >
ECCOv4r3	7 >
GHRSSST	1 >
Grace	2 >
Ice Shelf Mass Balance	4 >
ICE-6G	122 >
ISSM	3 >

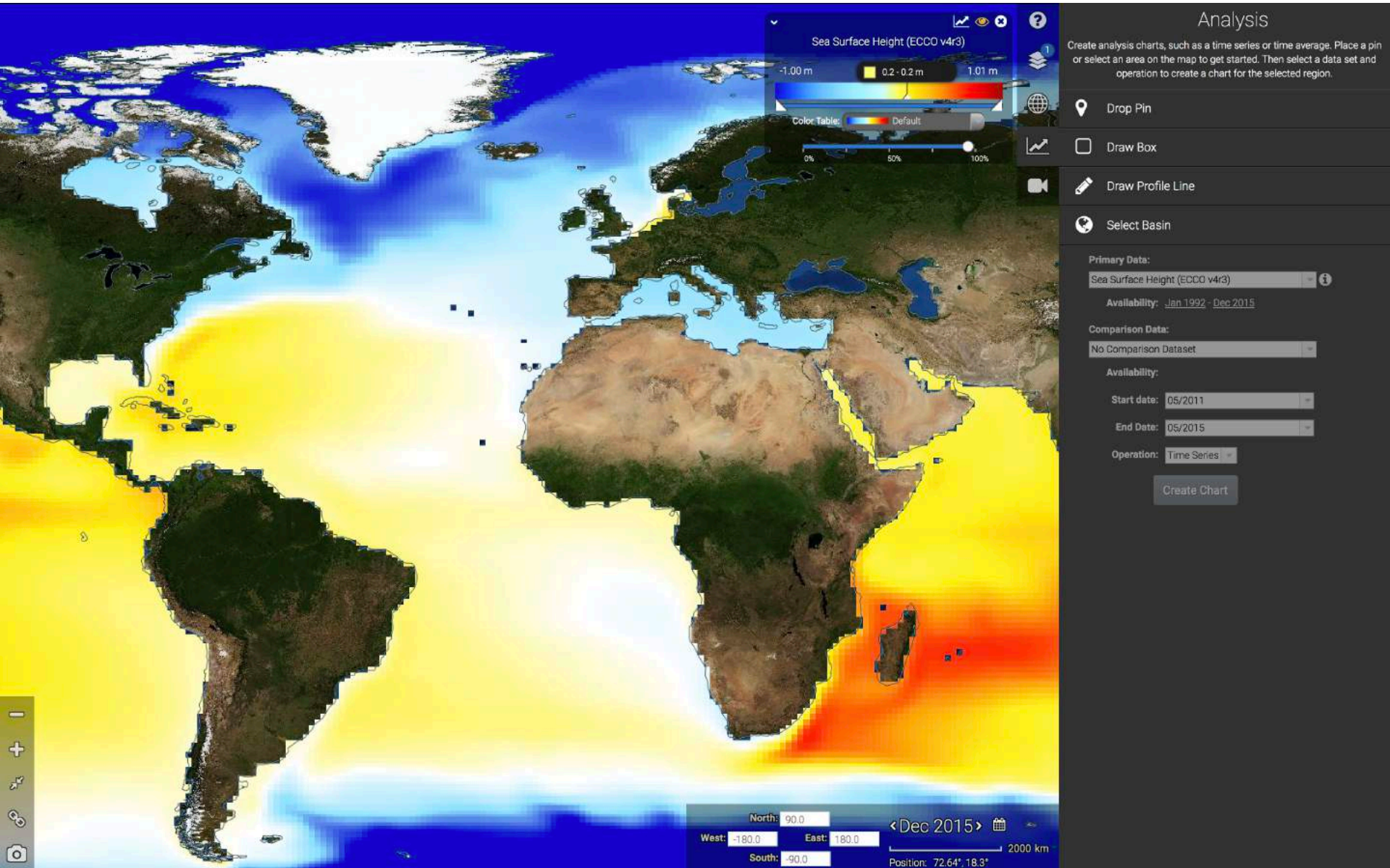
North: 90.0
West: -180.0 East: 180.0
South: 90.0

< May 2019 > 2000 km
Position: 70.53°, 15.63°

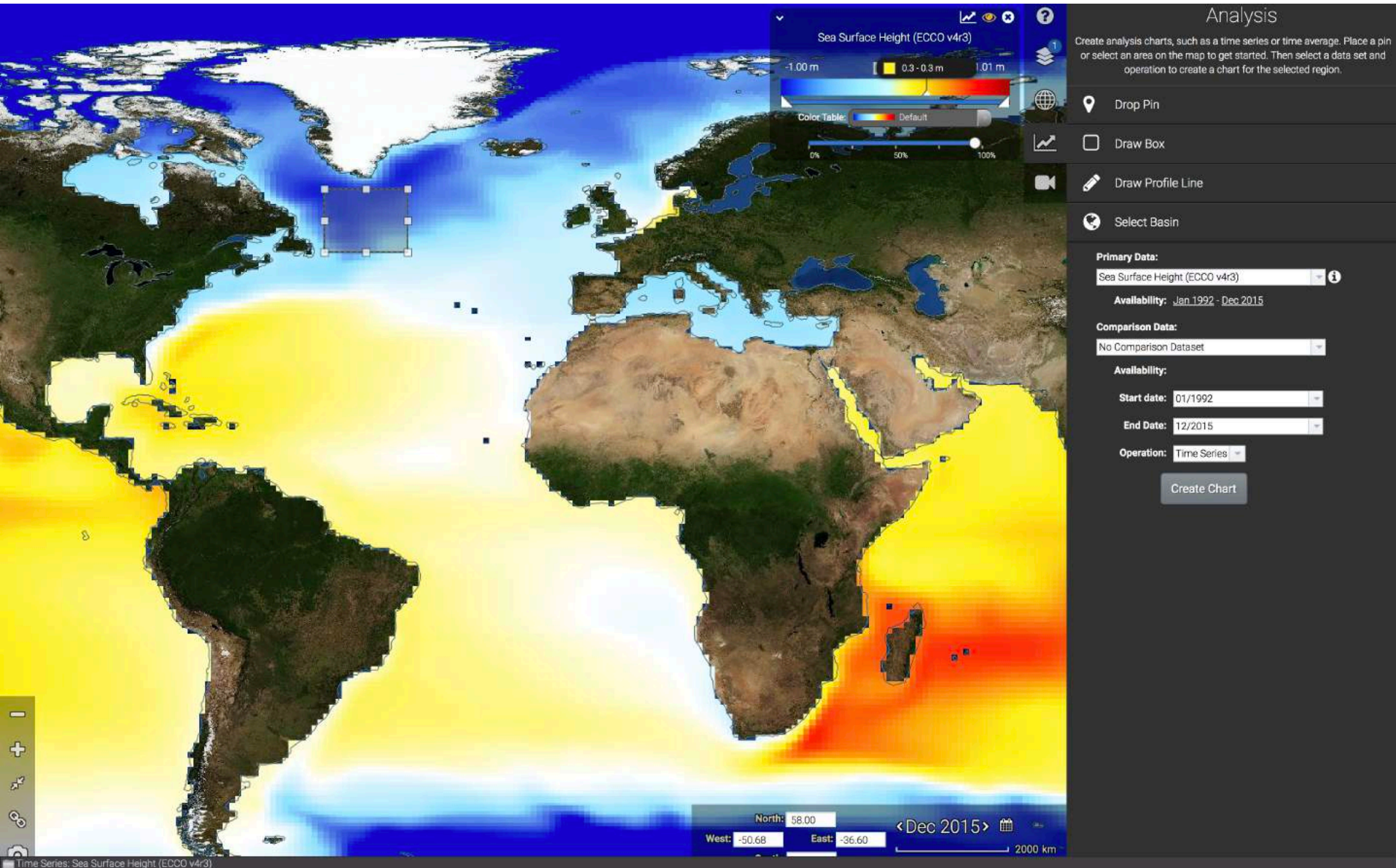
"Data Analysis Tool" on NASA Sea Level Change Portal



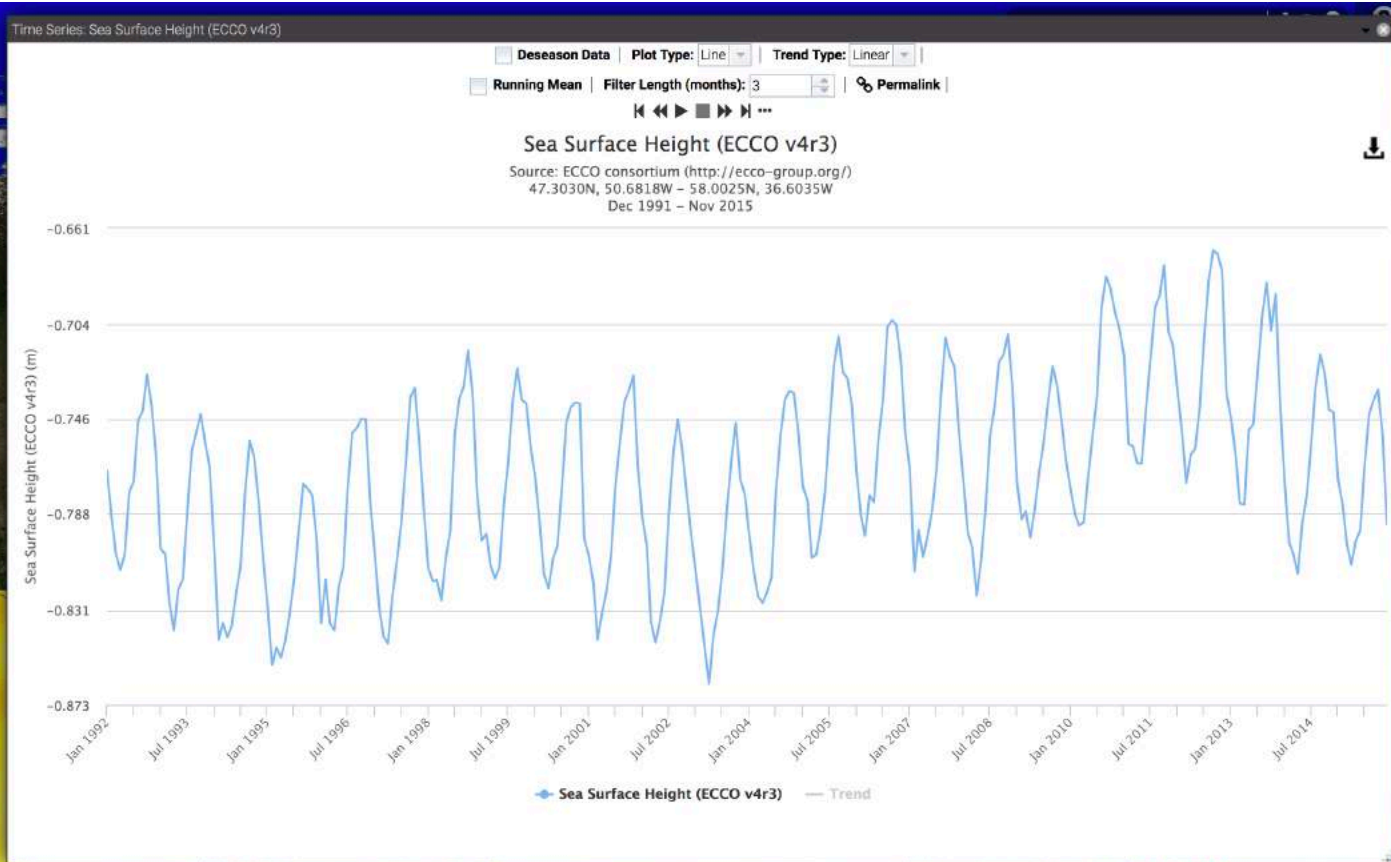
"Data Analysis Tool" on NASA Sea Level Change Portal



"Data Analysis Tool" on NASA Sea Level Change Portal



"Data Analysis Tool" on NASA Sea Level Change Portal



Analysis

Create analysis charts, such as a time series or time average. Place a pin or select an area on the map to get started. Then select a data set and operation to create a chart for the selected region.

Drop Pin

Draw Box

Draw Profile Line

Select Basin

Primary Data:

Sea Surface Height (ECCO v4r3)

Availability: Jan 1992 - Dec 2015

Comparison Data:

No Comparison Dataset

Availability:

Start date: 01/1992

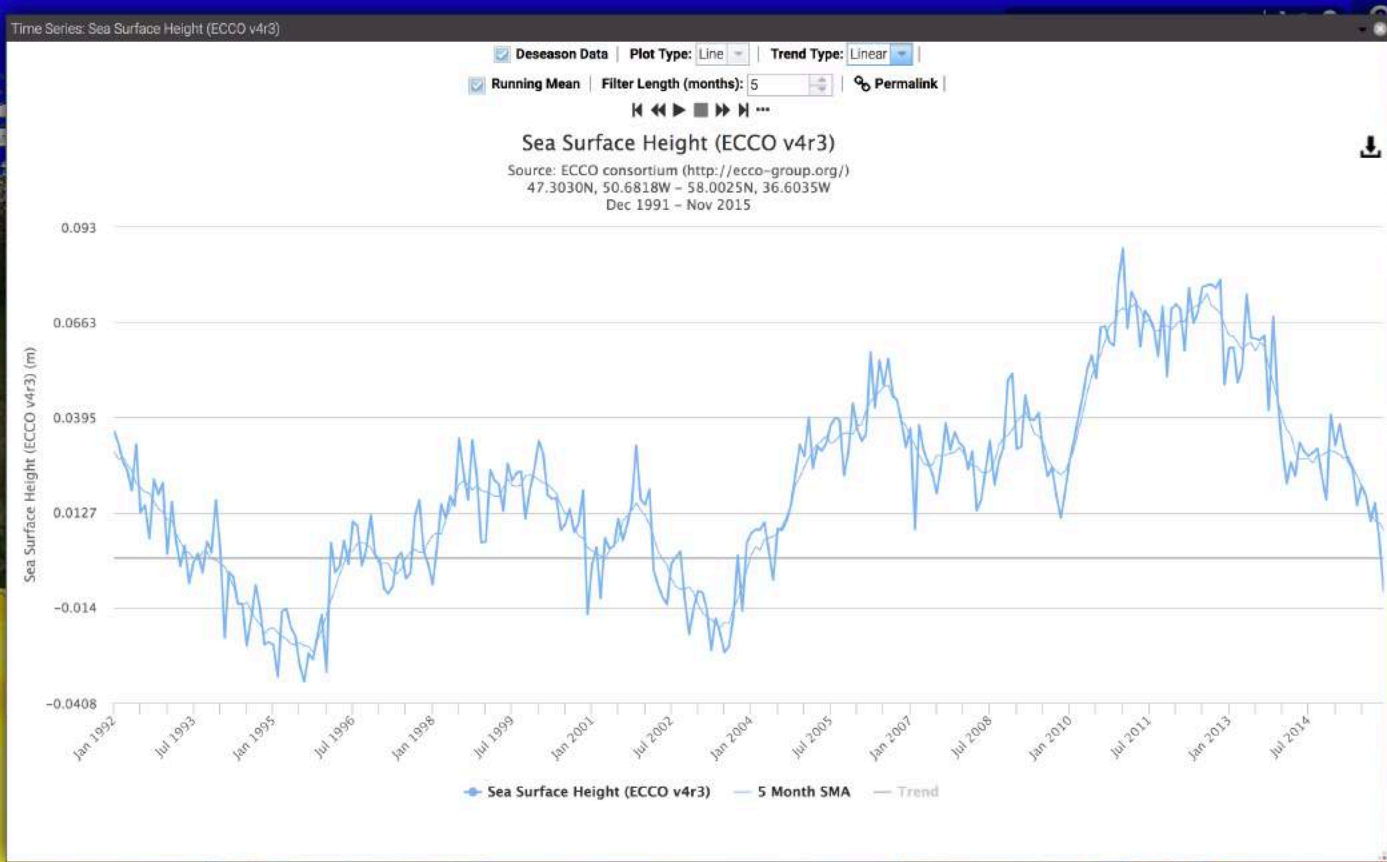
End Date: 12/2015

Operation: Time Series

Create Chart

North: 58.00
West: -50.68 East: -36.60
South: 47.30
Position: 73.07° 0.7°
< Dec 2015 >
2000 km

"Data Analysis Tool" on NASA Sea Level Change Portal



Analysis

Create analysis charts, such as a time series or time average. Place a pin or select an area on the map to get started. Then select a data set and operation to create a chart for the selected region.

Drop Pin

Draw Box

Draw Profile Line

Select Basin

Primary Data:

Sea Surface Height (ECCO v4r3)

Availability: Jan 1992 - Dec 2015

Comparison Data:

No Comparison Dataset

Availability:

Start date: 01/1992

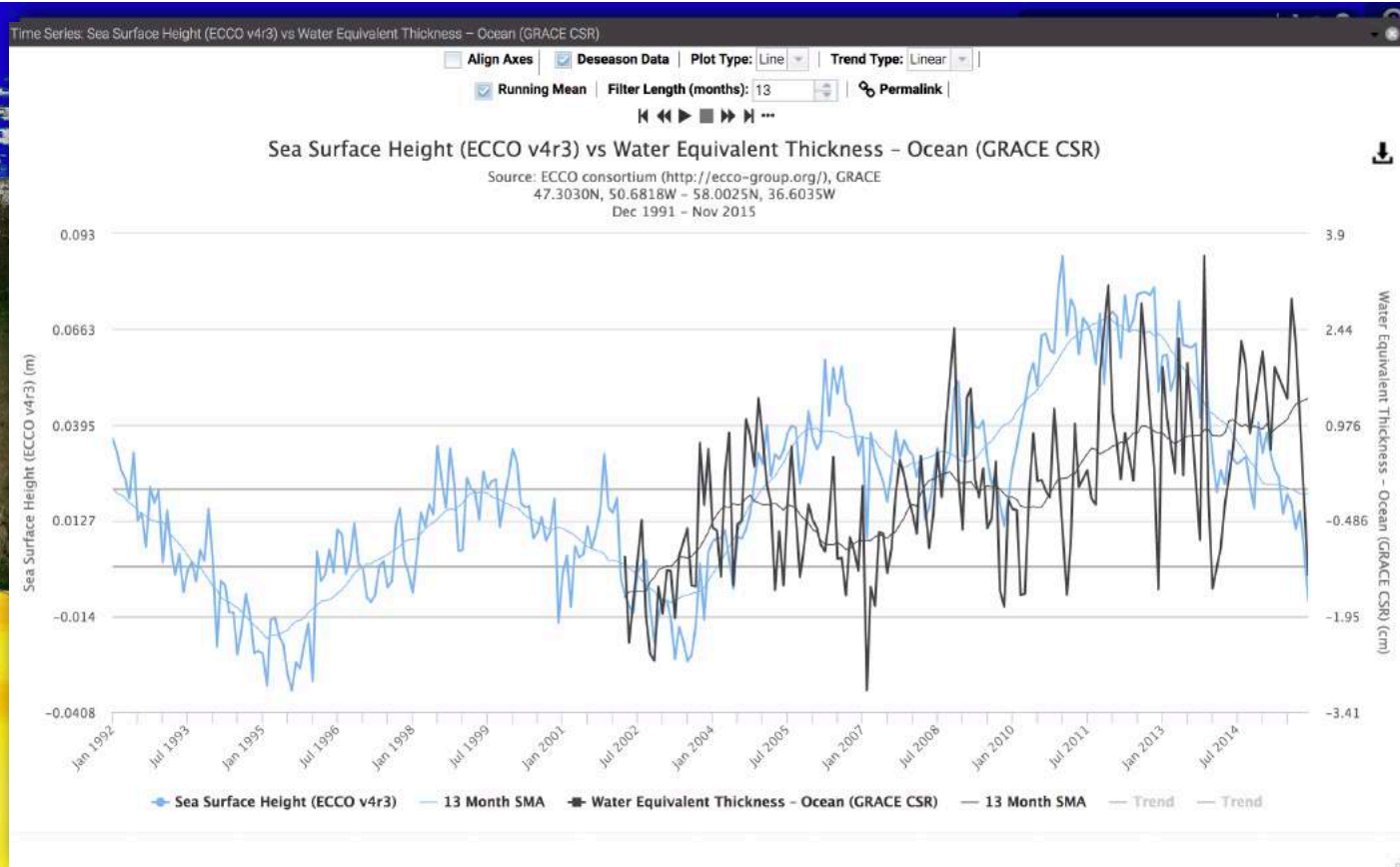
End Date: 12/2015

Operation: Time Series

Create Chart



"Data Analysis Tool" on NASA Sea Level Change Portal



Analysis

Create analysis charts, such as a time series or time average. Place a pin or select an area on the map to get started. Then select a data set and operation to create a chart for the selected region.

Drop Pin

Draw Box

Draw Profile Line

Select Basin

Primary Data:

Sea Surface Height (ECCO v4r3) ⓘ

Availability: Jan 1992 - Dec 2015

Comparison Data:

No Comparison Dataset

Availability:

Start date: 01/1992

End Date: 12/2015

Operation: Time Series

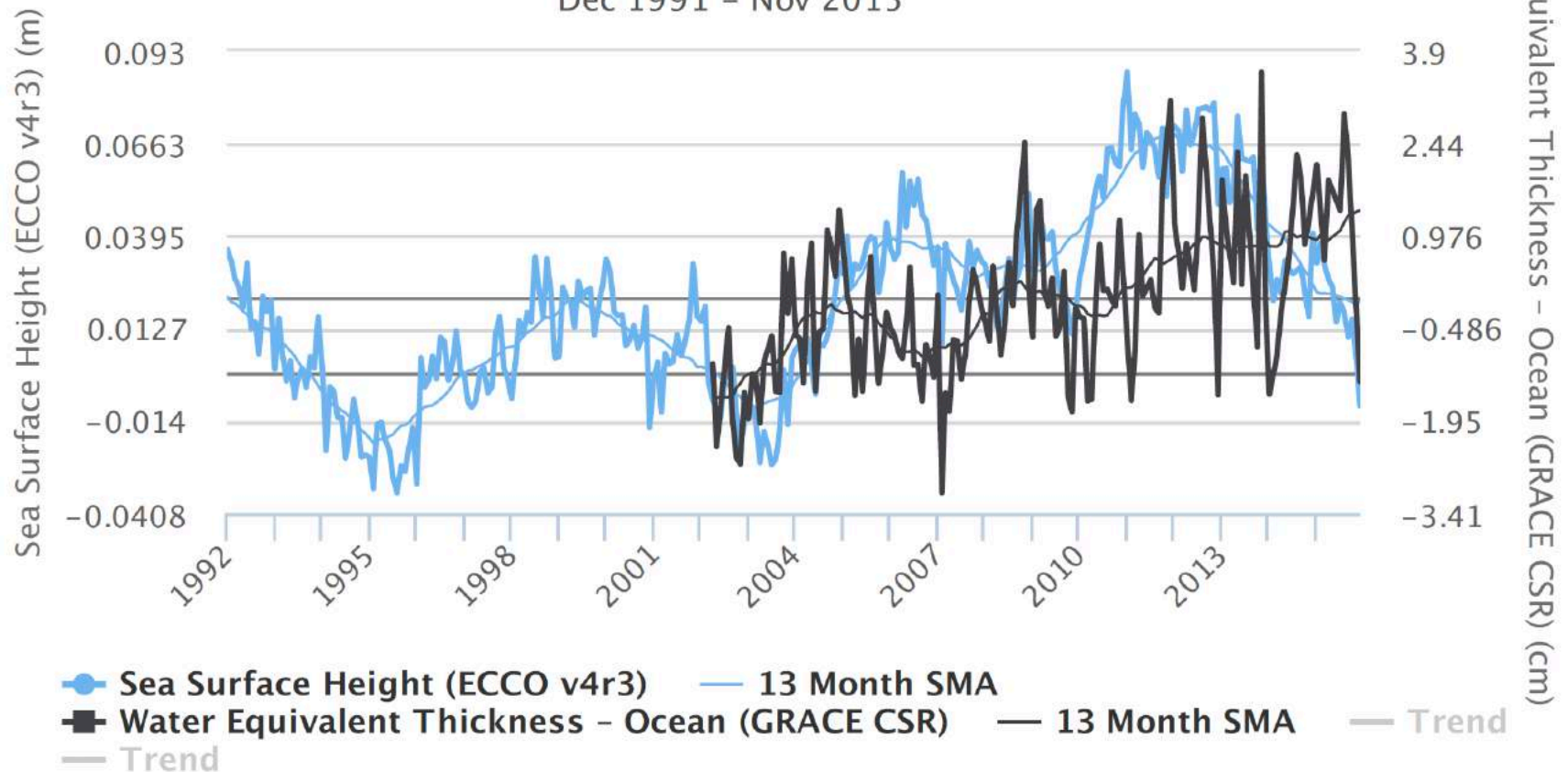
Create Chart



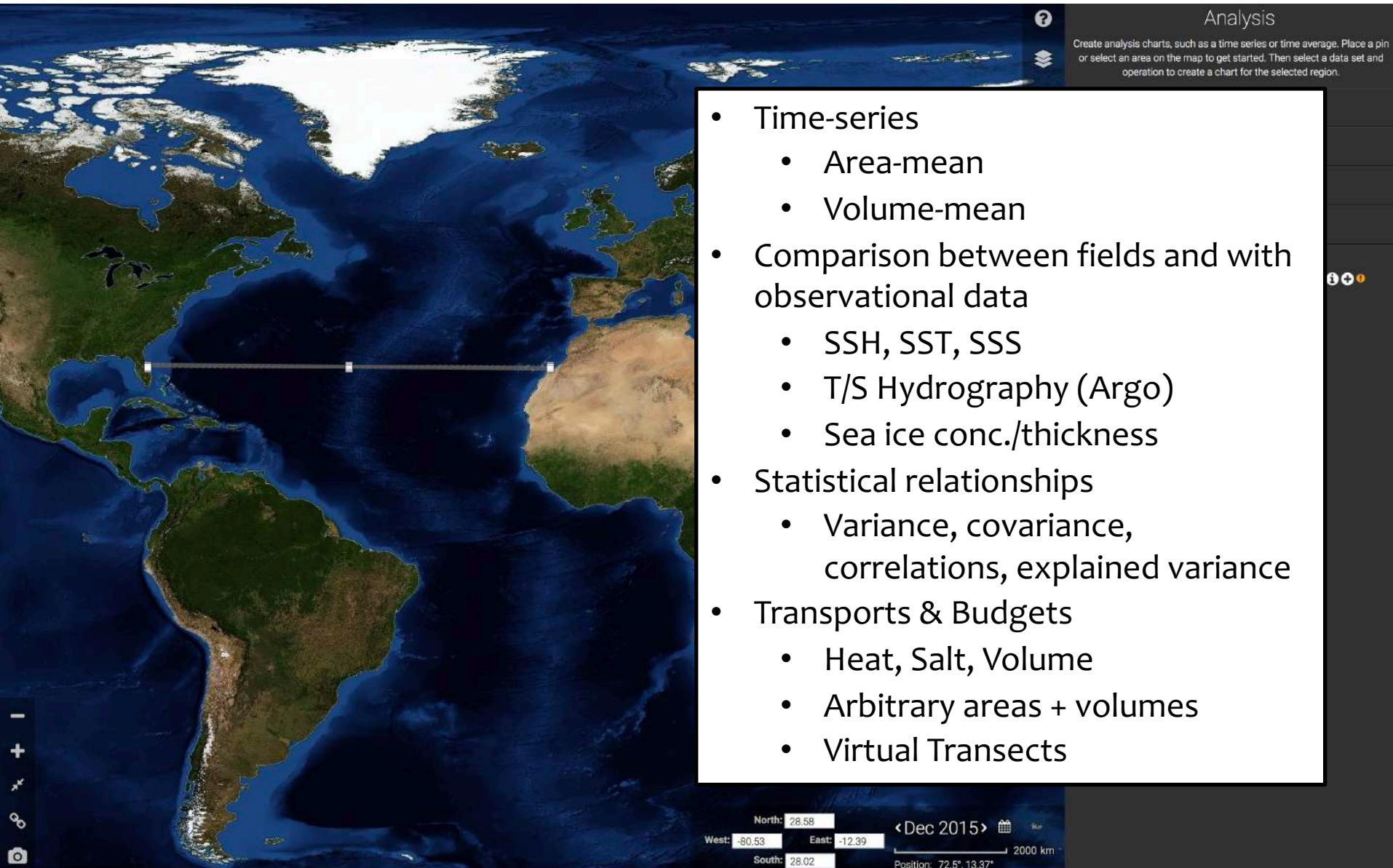
“Data Analysis Tool” on NASA Sea Level Change Portal

Sea Surface Height (ECCO v4r3) vs Water Equivalent Thickness - Ocean (GRACE CSR)

Source: ECCO consortium (<http://ecco-group.org/>), GRACE
47.3030N, 50.6818W - 58.0025N, 36.6035W
Dec 1991 - Nov 2015



"Data Analysis Tool" on NASA Sea Level Change Portal



Analysis

Create analysis charts, such as a time series or time average. Place a pin or select an area on the map to get started. Then select a data set and operation to create a chart for the selected region.

- 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

North: 28.58
West: -80.53 East: -12.39
South: 28.02
< Dec 2015 > 2000 km
Position: 72.5°, 13.37°

"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_GHRSSST_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_GHRSSST_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' RESTful 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 be moved to local machine
 - Remotely → users submit chains of operations

$$V = \sum_{i,j,k} r A(i, j) dr F(k) hFacC(i, j, k)$$
$$\langle T(t) \rangle = \frac{1}{V} \sum_{i,j,k} T(i, j, k, t) dr F(k) r A(i, j) hFacC(i, j, k)$$

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 Vinogradova¹, Mark Shiffer¹, Gael Forget², and Chris Hill²

¹Cambridge Climate Institute, MA

²Massachusetts 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* `cfnccluster` 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

