ABoVE Science Cloud Introduction and Hands-on Workshop

ABoVE Science Team Meeting
Seattle, 01/25/2018

Computational and Information Sciences and Technology Office (CISTO)
NASA Center for Climate Simulation (NCCS)
Scientific Visualization Studio (SVS)
Goddard Space Flight Center (GSFC)

Agenda

Introduction to the NASA Center for Climate Simulation (NCCS)

Overview of the ABoVE Science Cloud

- Overview of Capabilities
- Data Sets Available
- Success Stories

Scientific Visualization Studio

Logistics

- Gaining access to the system
- Finding out more information

Hands On Demonstration

- Log in to the system
- Finding data
- Transferring data in and out
- Running applications

Team Members

ABoVE

- Elizabeth Hoy
- Peter Griffiths
- Mark Carroll
- Dan Slayback

NCCS

- Jim Shute
- Scott Sinno
- Garrison Vaughan
- Hoot Thompson
- Julien Peters
- Tim Burch
- Laura Carriere
- Ellen Salmon
- Daniel Duffy
- Many others that provide other support, including security, networking, applications, etc.

References/For More Information

NCCS Website

- https://www.nccs.nasa.gov/

ADAPT Specific Information

- https://www.nccs.nasa.gov/services/adapt

ABoVE Science Cloud Specific Information

- https://above.nasa.gov/sciencecloud.html

Who can you contact?

- support@nccs.nasa.gov

NASA Center for Climate Simulation (NCCS)

Provides an integrated high-end computing environment designed to support the specialized requirements of Climate and Weather modeling.

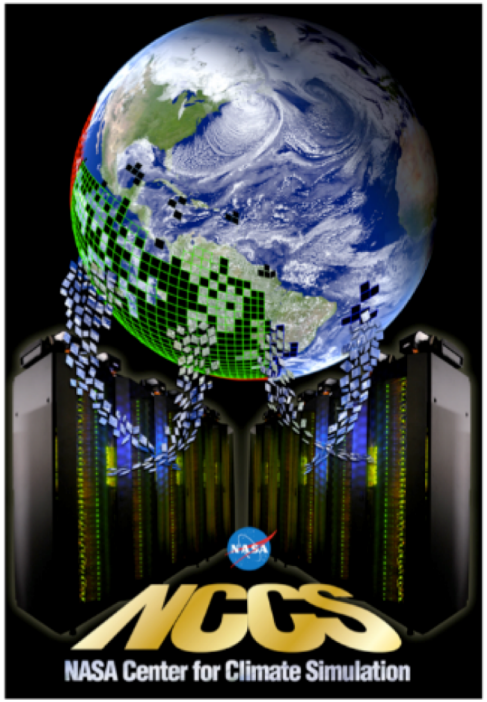
- High-performance computing, cloud computing, data storage, and networking technologies
- High-speed access to petabytes of Earth Science data
- Collaborative data sharing, publication, and analysis services

Primary Customers (NASA Science)

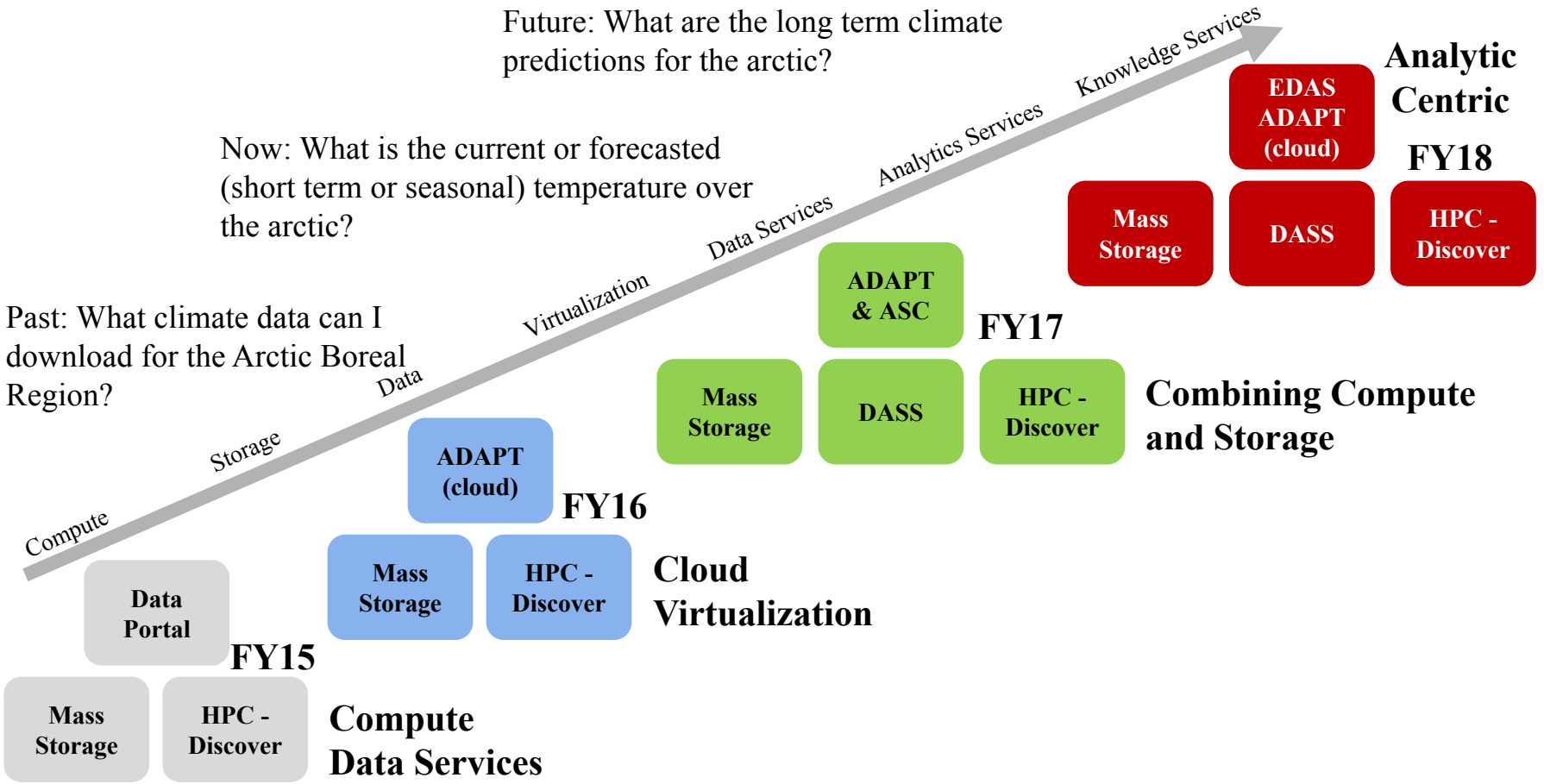
- NASA funded science projects can get access to these resources
- Global Modeling and Assimilation Office (GMAO)
- Land Information Systems (LIS)
- Goddard Institute for Space Studies (GISS)
- Variety of other Research and Development (R&D) and Engineering
  - » ABoVE, HiMAT, CALET, WFIRST

High-Performance Science

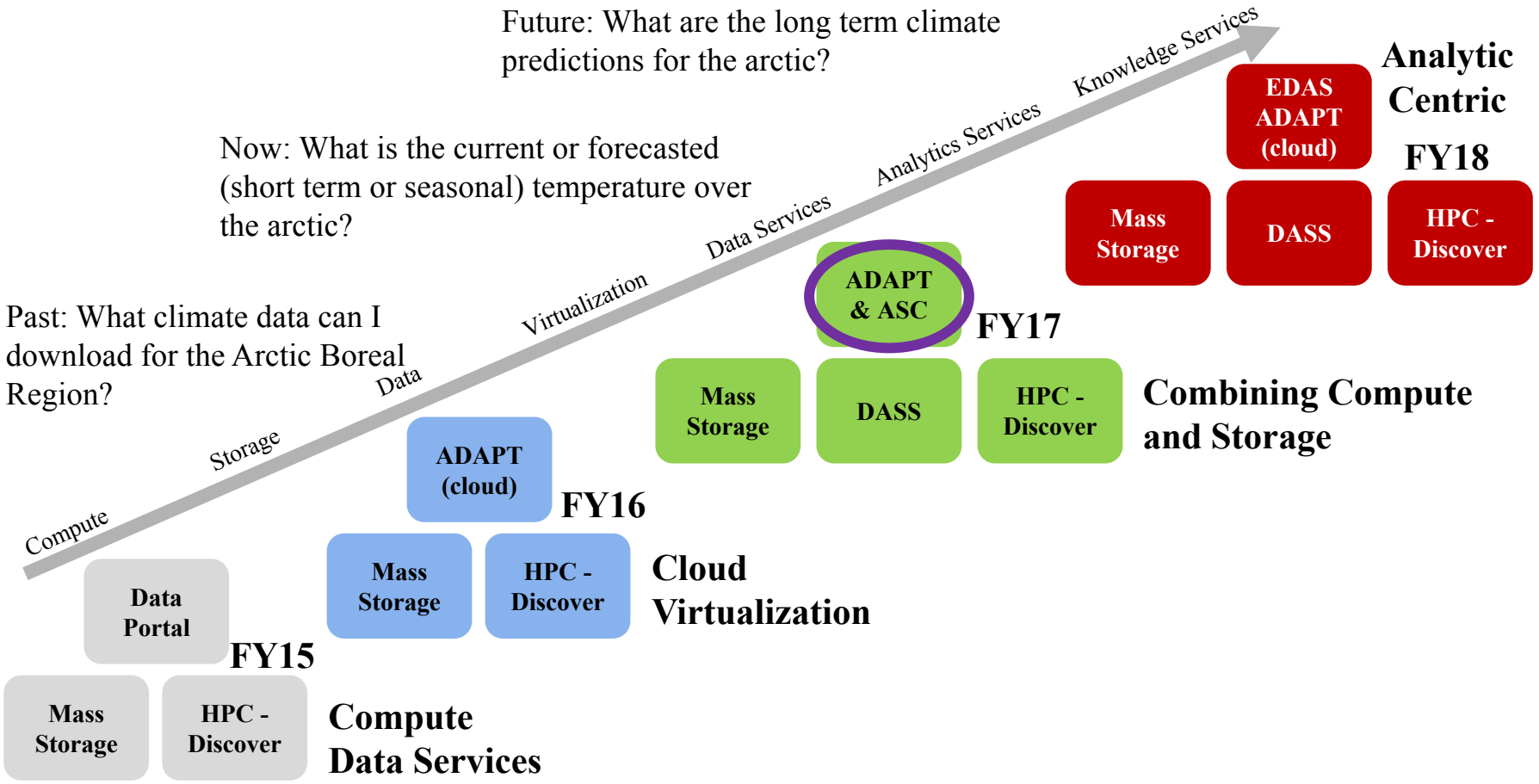
- http://www.nccs.nasa.gov
- Funded by the High End Computing (HEC) program under SMD
  - » Dr. Tsengdar Lee, Program Manager
- Code 606.2 at NASA Goddard Space Flight Center in Greenbelt, MD.



Evolution of Major NCCS Systems



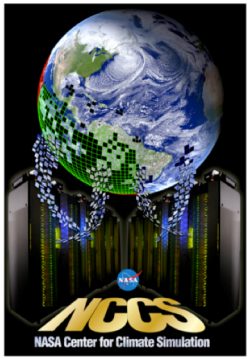
Evolution of Major NCCS Systems



ABoVE Science Cloud (ASC)

- **Partnership between the Carbon Cycle Ecosystems Office, High End Computing Program, and the NASA Center for Climate Simulation**
- **Created within the Advanced Data Analytics Platform (ADAPT) high performance science cloud in the NCCS**
- **Web References**
  - https://above.nasa.gov/
  - https://www.hec.nasa.gov/
  - https://www.nccs.nasa.gov/
- **NCCS User Services**
  - support@nccs.nasa.gov

ABoVE
ARCTIC – BOREAL VULNERABILITY EXPERIMENT



What is cloud computing and why a cloud?

Colocation of Big Data and Compute Resources

- Custom designed virtual machines specific for a user or an application
- To the end user, it looks like a server
- Many virtual machines can be hosted on the same physical device
- Usage is often not consistent – active for short times

Managed Cloud Environment

- End users don't have to worry about security plans and updating operating systems
- Support to install software and manage operating environments
- Support for downloading data, which can save significant amounts of time and effort
- Users don't have to put costs in their proposals for computing and storage

Why not just use Amazon Web Services (AWS)?

- Cost is still quite high in AWS, especially when storing large amounts of data
- Network affinity to large data sets at GSFC enables quicker access than AWS
- Still takes time to set up and manage instances in AWS, taking away from time to do science

Where is this cloud located?

Ease of entry, gradual increase of access to resources

1. **Use ASC Account, run jobs on above101-104**
2. **Submit batch jobs through Slurm to beyond101-106**
3. **Request customized VMs**
4. **Move workload to Discover, the traditional HPC environment**

ADAPT System Components/Configuration
ASC is hosted within ADAPT

Capability and Description	Configuration
 **Persistent Data Services** Virtual machines or containers deployed for web services, examples include ArcGIS, ESGF, GDS, THREDDS, FTP, etc.	Nodes with 128 GB of RAM, 10 GbE, and FDR IB
 **DataBase** High available database nodes with solid state disk.	Nodes with 128 GB of RAM, 3.2 TB of SSD, 10 GbE, and FDR IB
 **Remote Visualization** - planned Enable server side graphical processing and rendering of data.	Nodes with 128 GB of RAM, 10 GbE, FDR IB, and GPUs
 **High Performance Compute** More than 6,000 cores coupled via high speed Infiniband networks for elastic or itinerant computing requirements.	300+ nodes with between 24 and 128 GB of RAM and FDR IB
 **High-Speed/High-Capacity Storage** Petabytes of storage accessible to all the above capabilities over the high speed Infiniband network.	Storage nodes configured with multiple PB's of RAW storage capacity

ASC Software Stack

External License Servers

Open Source Tools
Python, NetCDF, GDAL, R, etc.

Commercial Tools
Intel Compiler (C, C++, Fortran), IDL (4 seats)

Operating Systems
Linux (Debian, CentOS) and Windows

Virtual machines can be set up to reach out to external license servers.

Virtual machines can be customized based on the end user application needs. The NCCS will work with you to create customized VMs specific to meet your needs.

Staged/Common Data Sets in ABoVE Science Cloud

Common datasets "Staged" for ABoVE investigators in ABoVE Science Cloud

- Staged and available for direct use
- Individual investigators don't have to invest time to locate and transfer data into system
- Avoids duplications of large datasets on system
- Additional datasets can be added, including generated data from ABoVE PI
- Data Services Manager to locate data

Example Download Times For 80TB

Speed	Time HH:MM:SS
9.6 Kbps	18518518:31:06
14.4 Kbps	12345679:00:44
28.8 Kbps	6172839:30:22
33.6 Kbps	5291005:17:27
56 Kbps	3174603:10:28
64 Kbps (ISDN)	2777777:46:40
128 Kbps (ISDN-2)	1388888:53:20
256 Kbps	694444:26:40
512 Kbps	347222:13:20
1.024 Mbps	173611:06:40
1.544 Mbps (DS1, T1)	115141:02:52
2.048 Mbps (E1, ISDN-32)	86805:33:20
10 Mbps (10Base-T)	17777:46:40
25.6 Mbps (ATM25)	6944:26:40
34 Mbps (E3)	5228:45:29
45 Mbps (DS3, T3)	3950:37:02
51 Mbps (OC1)	3485:50:19
100 Mbps (100Base-T)	1777:46:40
155 Mbps (OC3)	1146:57:12
622 Mbps (OC12)	285:48:58
1 Gbps (1000Base-T)	177:46:40
2.4 Gbps (OC48)	74:04:26
10 Gbps (OC192)	17:46:40

7.4 Mbps average US internet speed

1 Gbps NASA / Other Gov

@10 Mbps
Days: 741
Weeks: 106
Months: 24

@1 Gbps
Days: 7
Weeks: 1.1
Months: 0.25

Current ABoVE Science Cloud Data Holdings

Large Collections	Amount
Landsat	186 TB
MODIS	MODAPS collection remotely mounted
MERRA & MERRA2	406 TB
NGA/DigitalGlobe Imagery	2.8 PB
Total	> 3 PB

Other Data Sets

- Elevation datasets
  - ArcticDEM
  - CDEM
  - ASTER GDEM
  - Etc.
- Vegetation products
- Land cover products
- Products generated by the science team
- *Others as the team requests…*

Note that the ABoVE Science Cloud is not a permanent repository or the definitive source for this data. Official ABoVE products will be curated by the DAACs.

NGA/DigitalGlobe High Resolution Commercial Satellite Imagery

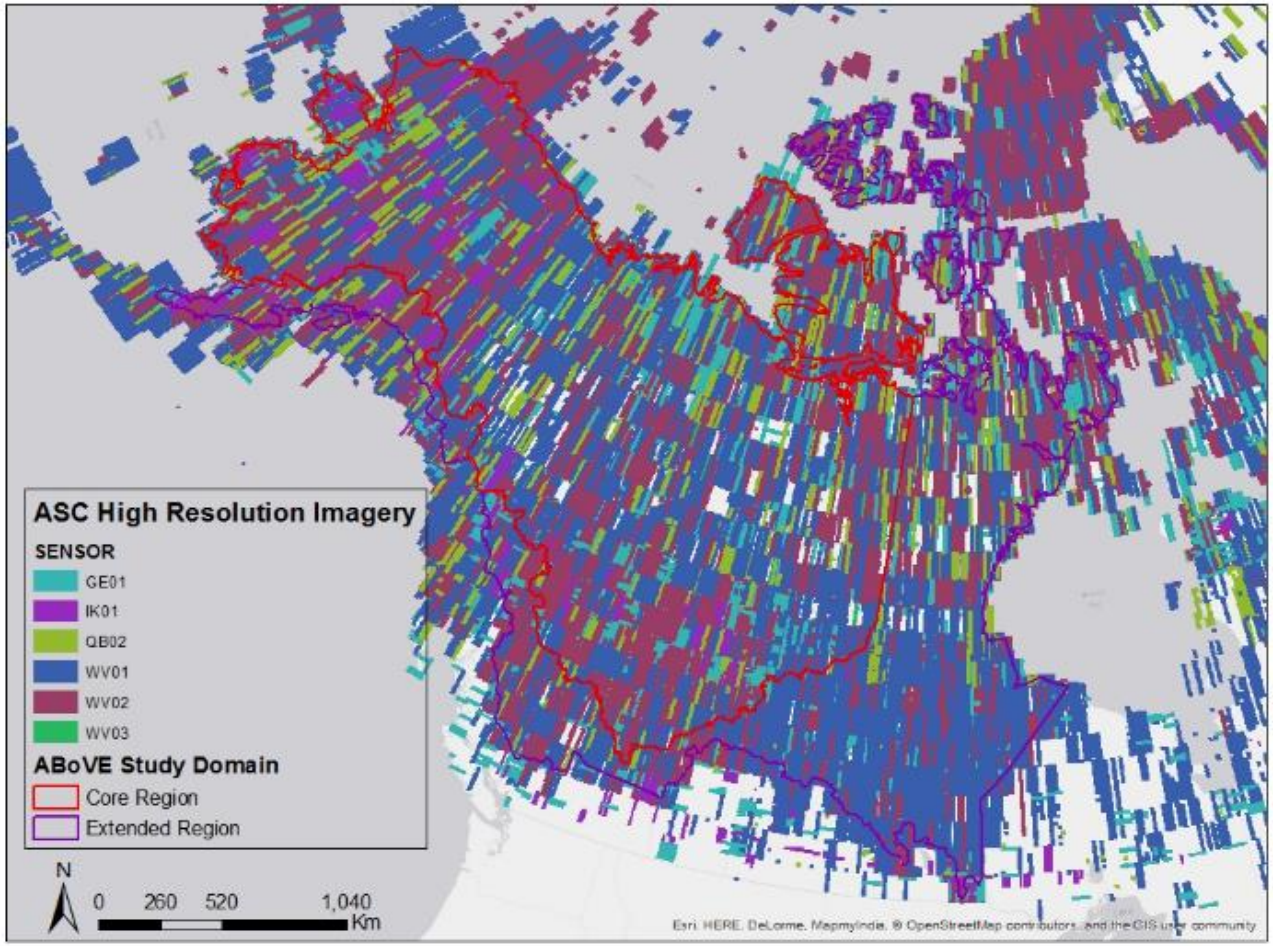
National Geospatial Intelligence Agency (NGA) has licensed all DigitalGlobe ≥ 31 cm satellite imagery for US Federal use, i.e., NSF, NASA and NASA funded projects.

- Archive of >4.2 billion km^2 of data from 2000 to present
- Data from six different satellites: Worldview-1, 2 and 3; Ikonos; Quickbird; and Geoeye-1

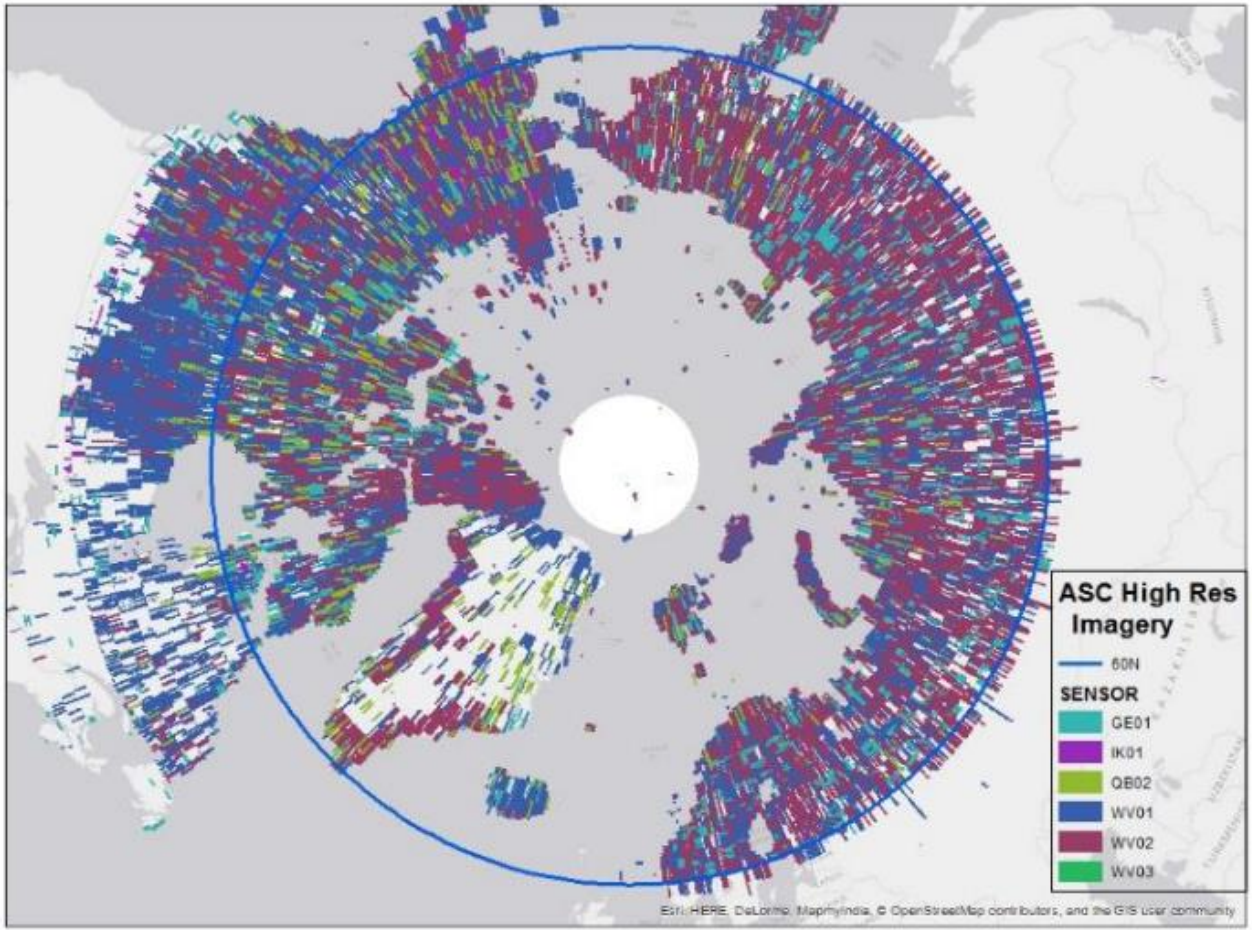
Satellite	Bands	Nadir Panchromatic Resolution (m)	Nadir Multispectral Resolution (m)
Ikonos	Pan, R, G, B, Near IR	0.82	3.2
GeoEye	Pan, R, G, B, Near IR	0.41	1.65
Quickbird	Pan, R, G, B, Near IR	0.55	2.16
WorldView-1	Panchromatic only	0.5	N/A
WorldView-2	Pan, R, G, B, Near IR 1, Near IR 2, Coastal, Red Edge, Yellow	0.46	1.85
WorldView-3	Same as WV-2 plus 8 SWIR bands and 12 CAVIS bands	0.31	1.24

DigitalGlobe Satellite Fleet

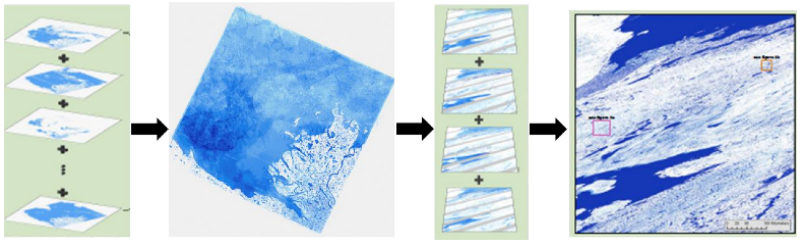
ABoVE Science Cloud DigitalGlobe Imagery: Study Domain



ABoVE Science Cloud DigitalGlobe Imagery: Circumpolar

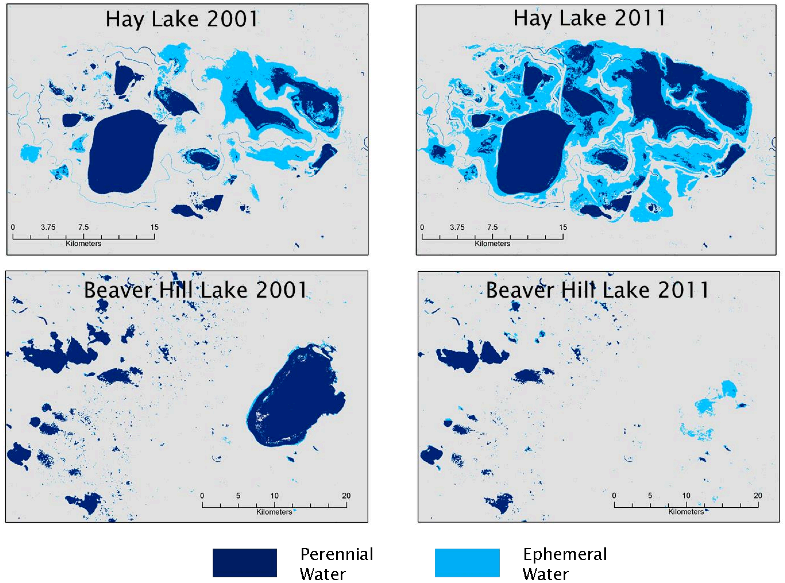


ABoVE Water Maps: 30 meter spatial resolution surface water 1991-2011," M.L. Carroll, et. al



Processing work flow for the generation of the ABoVE water maps from Landsat scenes to ABoVE tiles.

100,000 LandSat Scenes
20 TB of Data



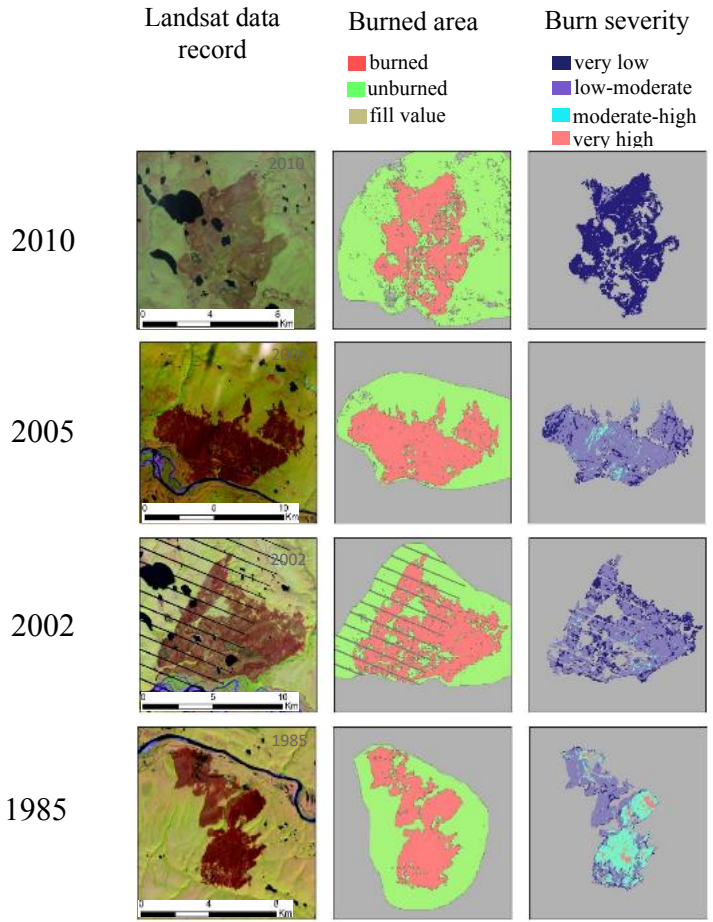
AWM for 2001 and 2011 for Hay Lake and Beaver Hill Lake in Canada. Hay Lake has clearly expanded over this time frame while Beaver Hill Lake has diminished.

- Takes in large amounts of input and creates small output
- Using large amounts of observation or model data
- Python code of 100's of lines
- Easily run in parallel across multiple virtual machines

Taken from "AboVE Water Maps: 30 meter spatial resolution surface water 1991-2011," M.L. Carroll, et. al, http://above.nasa.gov/pdfs/ABoVE_water_maps_user_guide_05102016.pdf

Fire History for ABoVE – T. Loboda & M. Miller

- Fire history across the ABoVE study region is compiled from available and new (Miller et al. in prep) data products and enhanced
- Multiple VMs on the ASC are used to process Landsat and MODIS data to develop the burn severity characterization



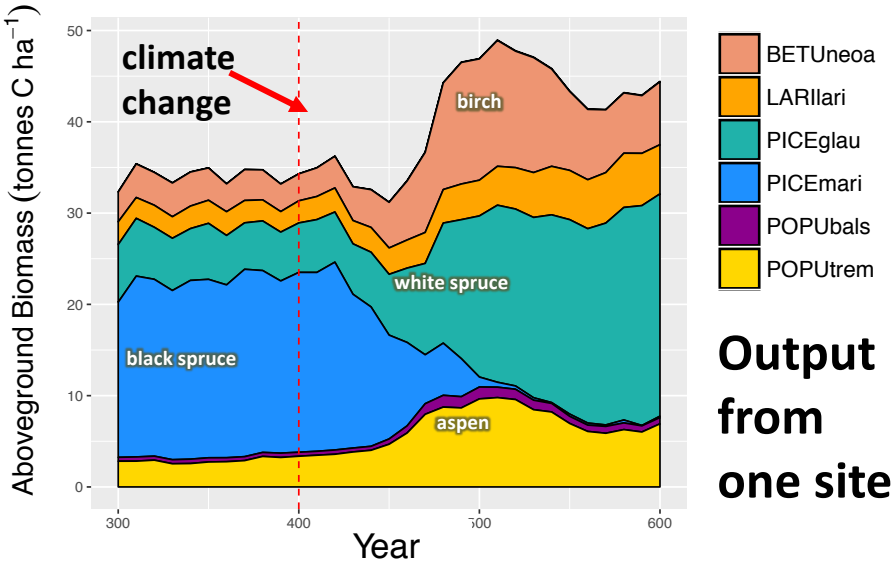
Large-scale simulation results with climate change, Tanana Valley – A. Foster

University of Virginia Forest Model Enhanced (UVAFME) – individual tree based model that simulates tree growth and response to external factors & tree-tree competition

For each site/grid cell:
Run 200 plots (Monte-Carlo simulation) for at least 500 years

Can take up to 1 minute per site
1 minute **x** 131,000 sites = **~90 days** for only one simulation

Ran ~35,000 sites – spread out across ADAPT nodes --> took about **1.5 days**

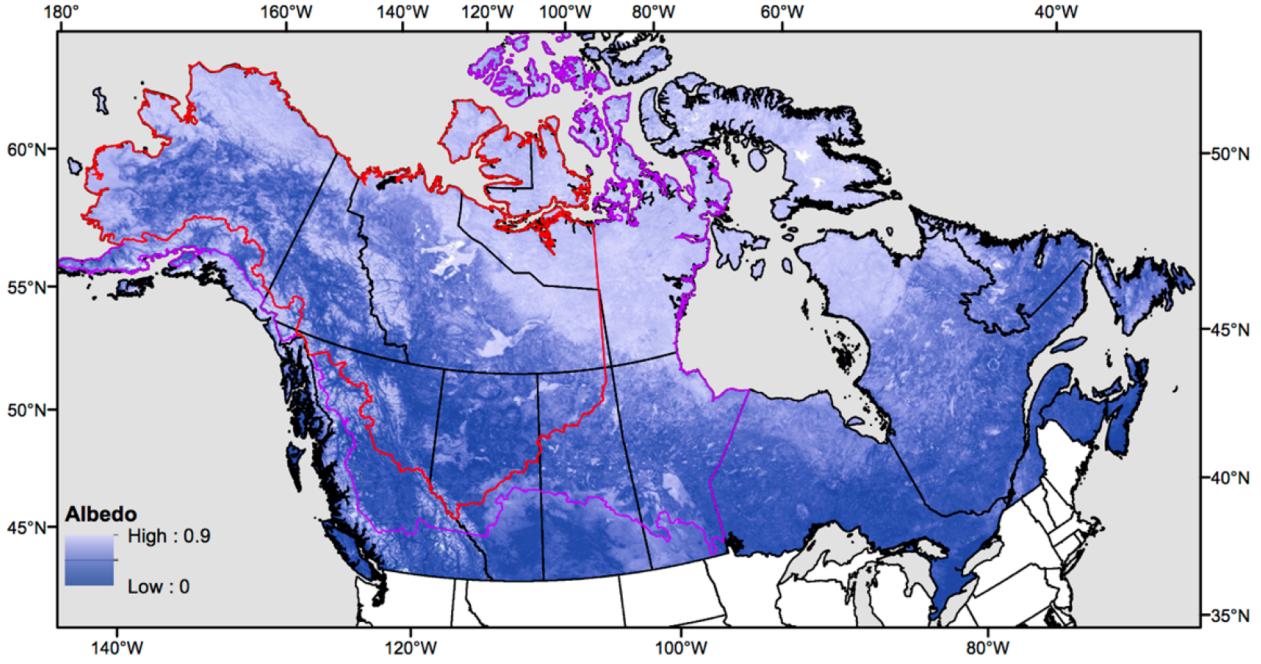


Output from one site

About to use 4 cores/node to see if that can be reduced to less than half a day

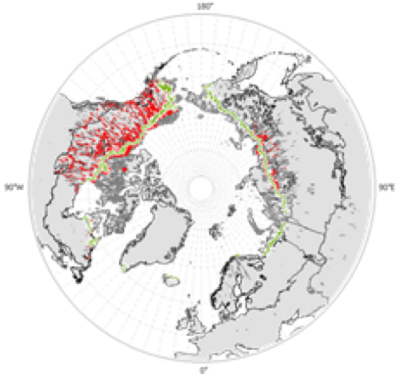
Understanding the Causes and Implications of Enhanced Seasonal CO_2 Exchange in Boreal and Arctic Ecosystems – B. Rogers

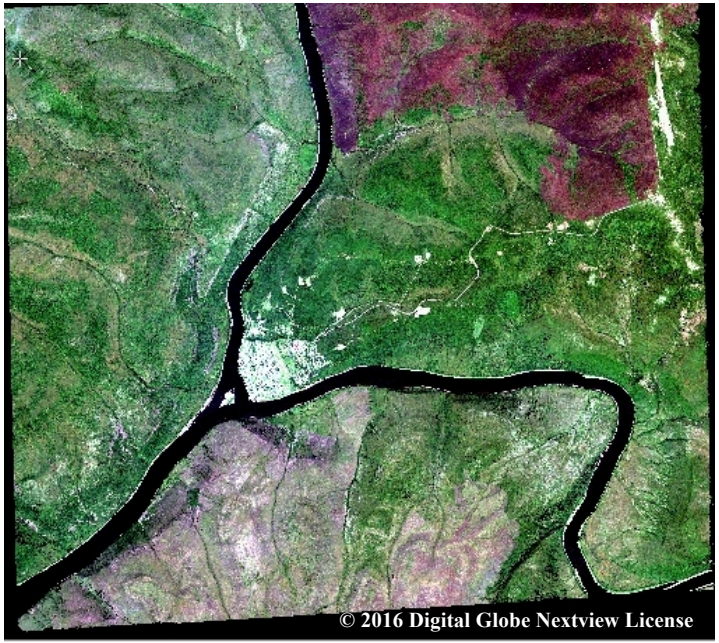
- Modeling driving factors of post-fire albedo trajectories
- Creation of mean albedo maps
- Fire combustion mapping



Forest Canopy Surface Elevations – C. Neigh & P. Montesano

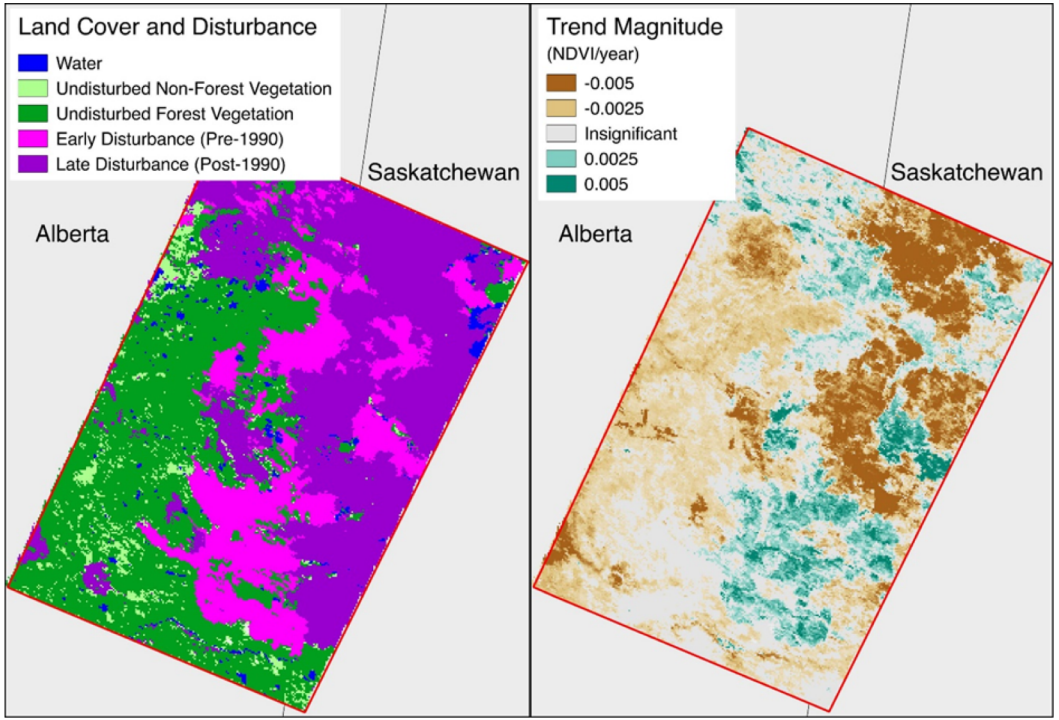
- Understanding forest patterns using DigitalGlobe high-resolution satellite imagery
- Using multiple VMs and Ames Stereo Pipeline (ASP) on the ASC to process Digital Elevation Models





Landscape-Scale Histories of Disturbance, Seasonality and Greenness Trends - C. Woodcock & D. Sulla-Menashe

- 30+ year historical record and ongoing characterization of disturbance events and phenology across the ABoVE study domain
- Using multiple VMs to move Landsat data into the ABoVE grid and then develop the landscape histories



NASA Scientific Visualization Studio

A Project unique in the federal government

- Create compelling visual content from data to communicate NASA Science to the public through all media venues
- Provide a free-to-use public archive of all visual content
- Empowered to seek the best in NASA Science across the entire agency
- Trusted connections to the research community and the media

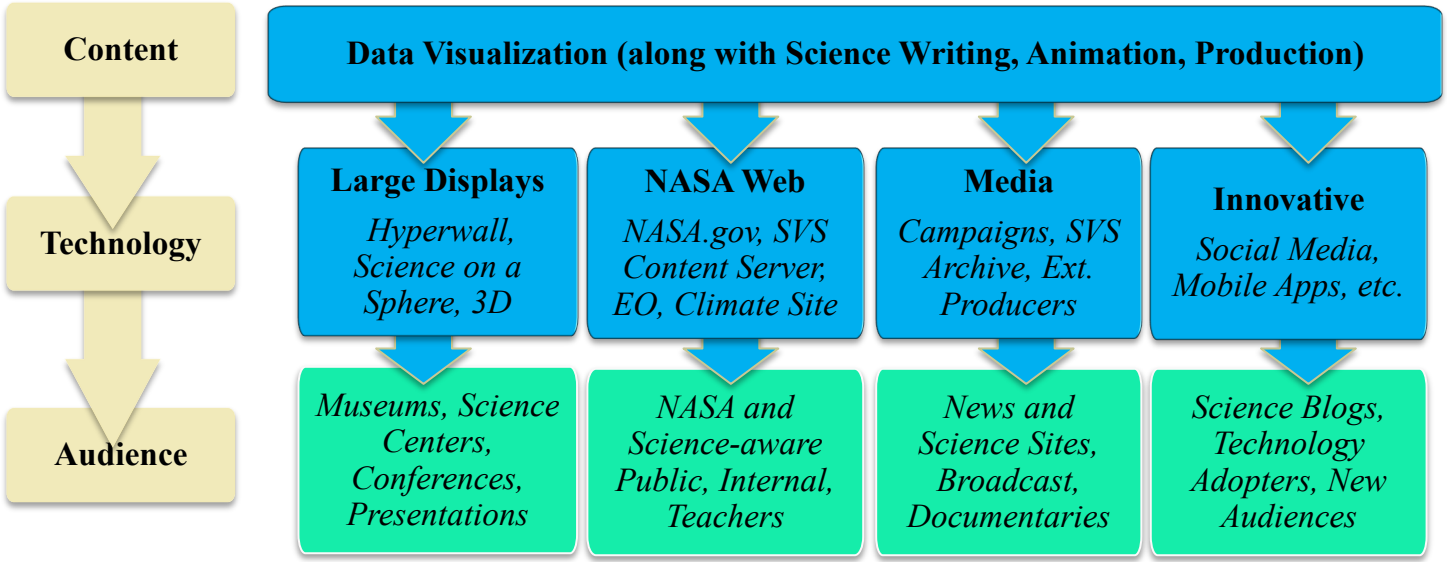
Available to support the ABoVE project

- Will require some funding to create award winning quality visualizations
- Visualizations made available through the SVS website, youtube, hyperwall, and more

For more information

- https://svs.gsfc.nasa.gov/
- Horace Mitchell at horace.g.mitchell@nasa.gov

NASA Scientific Visualization Studio



Logistics

Gaining Access – ABoVE Website:

- Instructions under “Data”, “ABoVE Science Cloud”, link to Science Cloud Setup Instructions
- Direct url: https://above.nasa.gov/sciencecloud_setup.html
- Need NASA identity, IT Security training, RSA Token, process takes a while
- Optional: signed NGA paperwork, new version in process
- Links to monthly webinars, other instructional videos
- POC – Liz Hoy, support@nccs.nasa.gov, Laura Carriere

More Information:

- NCCS Website: Look under “Services”, “ADAPT”
  - How to login, data locations, Windows FAQ, ABoVE FAQ, including orthorectification instructions
- ODISEA – Access from login node using firefox or NCCS website

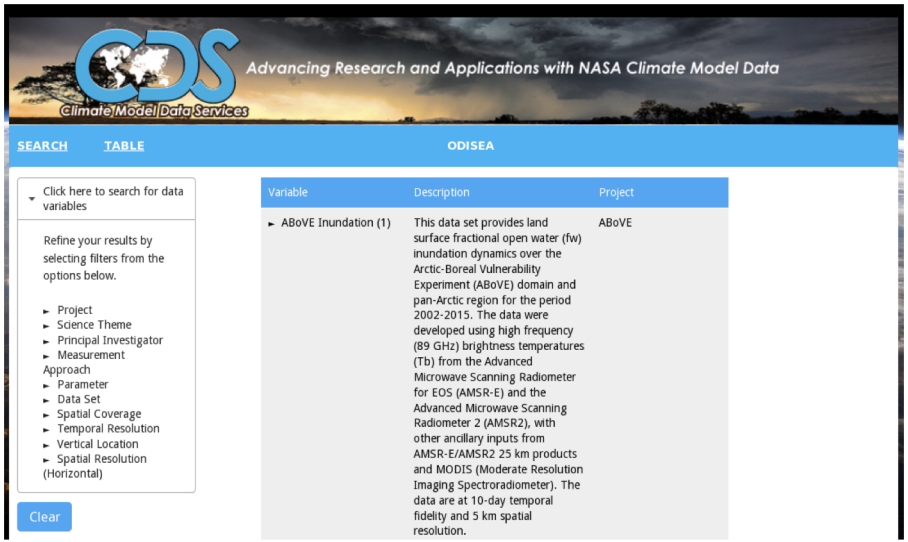
Hands On Demonstration

Log in to the system (Windows - MobaXterm, MacOS - terminal window + Xquartz)

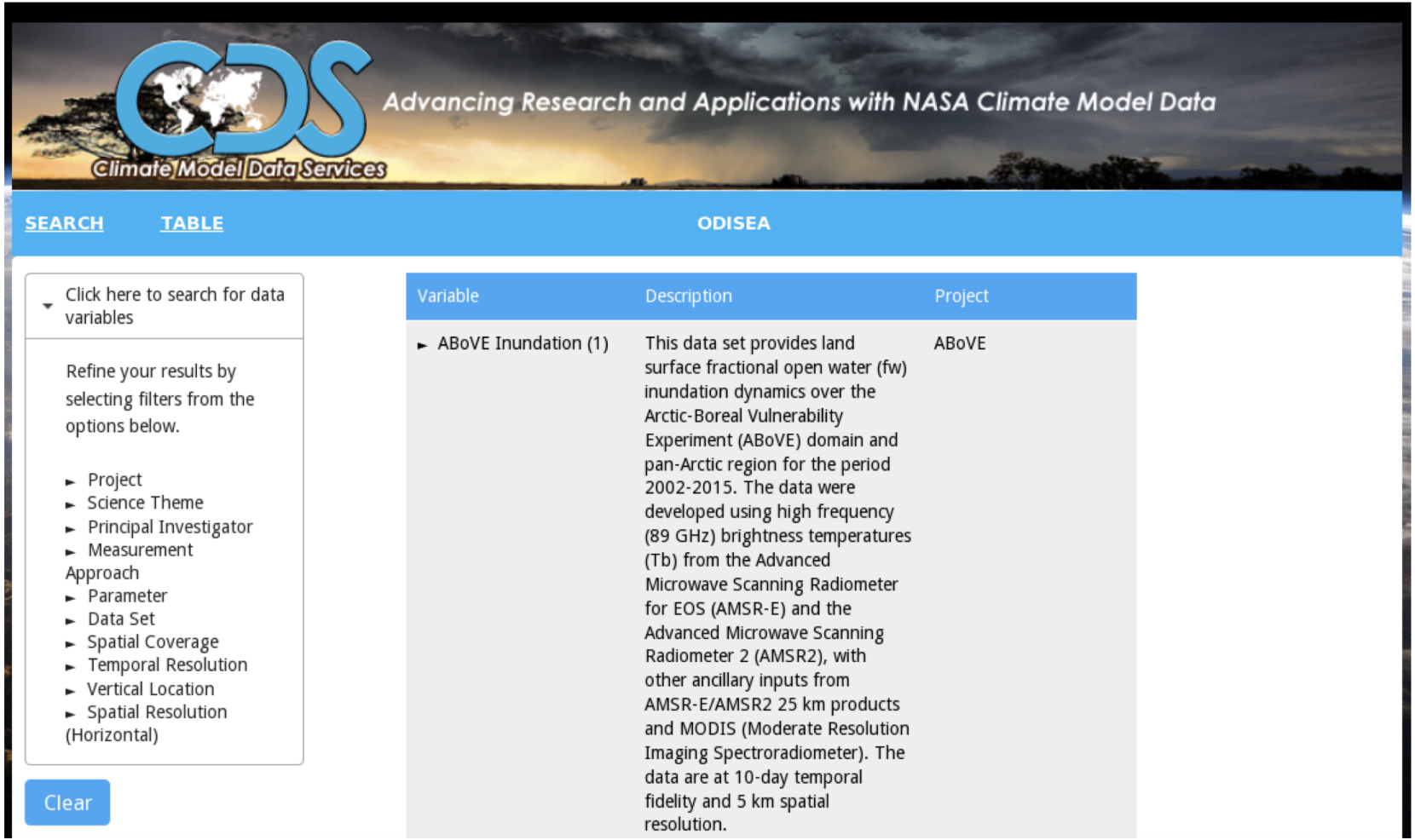
- ssh -X dsclogin.nccs.nasa.gov
- ssh above101.nccs.nasa.gov – above101-104
- If no NGA paperwork – ssh foyer101.nccs.nasa.gov – foyer101-102

Finding data

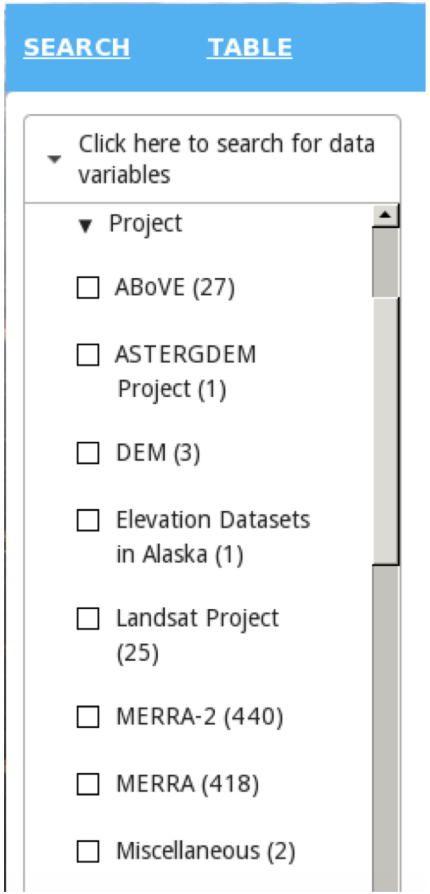
- /att/nobackup/*username*
- /etc/motd
- /att/pubrepo
- ODISEA – firefox – search, table, video, website

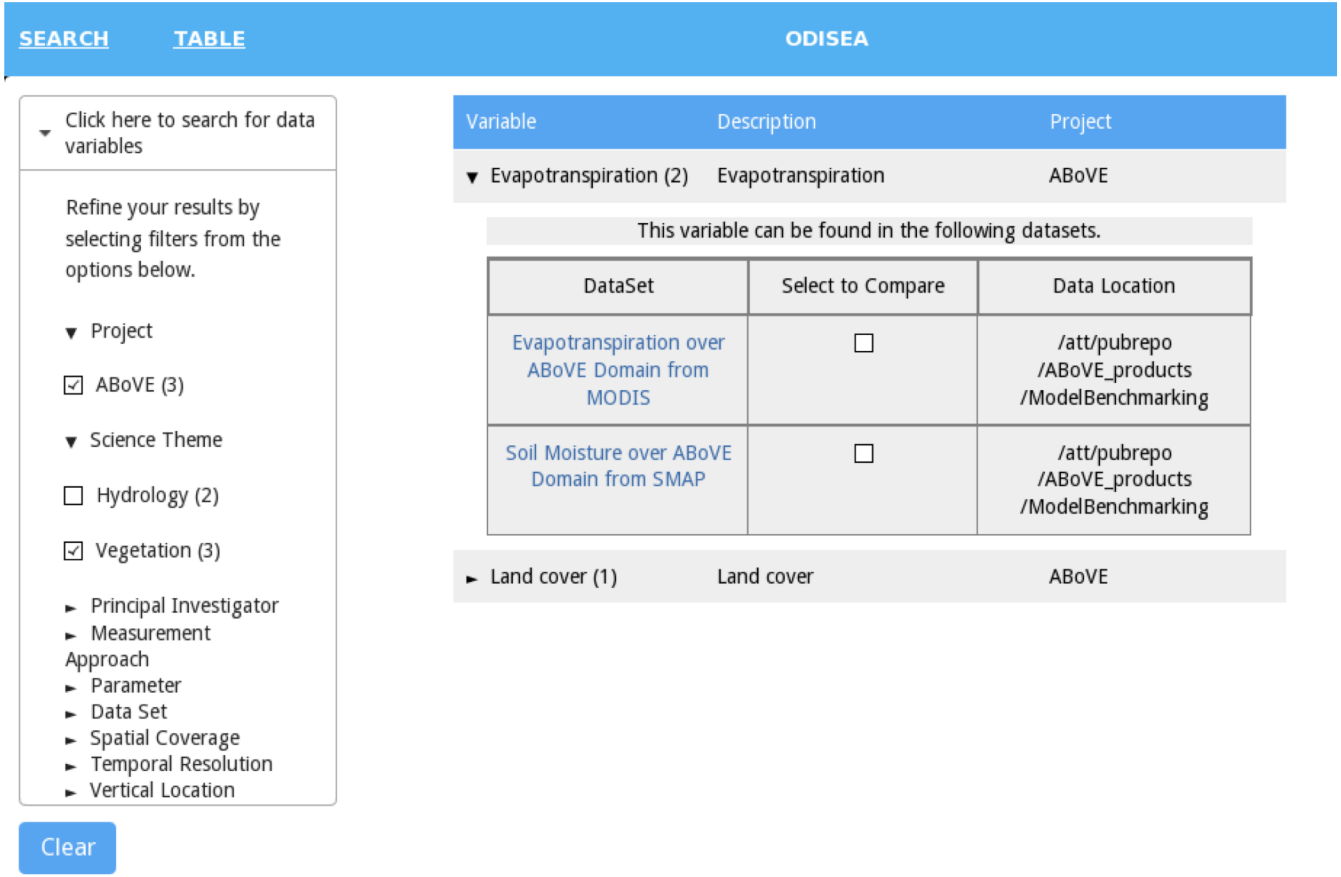


ODISEA


Climate Model Data Services
Advancing Research and Applications with NASA Climate Model Data
SEARCH
TABLE
ODISEA
Click here to search for data variables
Refine your results by selecting filters from the options below.
► Project
► Science Theme
► Principal Investigator
► Measurement Approach
► Parameter
► Data Set
► Spatial Coverage
► Temporal Resolution
► Vertical Location
► Spatial Resolution (Horizontal)
Clear
Variable
Description
Project
► ABoVE Inundation (1)
This data set provides land surface fractional open water (fw) inundation dynamics over the Arctic-Boreal Vulnerability Experiment (ABoVE) domain and pan-Arctic region for the period 2002-2015. The data were developed using high frequency (89 GHz) brightness temperatures (Tb) from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) and the Advanced Microwave Scanning Radiometer 2 (AMSR2), with other ancillary inputs from AMSR-E/AMSR2 25 km products and MODIS (Moderate Resolution Imaging Spectroradiometer). The data are at 10-day temporal fidelity and 5 km spatial resolution.
ABoVE

ODISEA


SEARCH
TABLE
Click here to search for data variables
Project
ABoVE (27)
ASTERGDEM Project (1)
DEM (3)
Elevation Datasets in Alaska (1)
Landsat Project (25)
MERRA-2 (440)
MERRA (418)
Miscellaneous (2)


SEARCH
TABLE
ODISEA
Click here to search for data variables
Refine your results by selecting filters from the options below.
Project
ABoVE (3)
Science Theme
Hydrology (2)
Vegetation (3)
Principal Investigator
Measurement Approach
Parameter
Data Set
Spatial Coverage
Temporal Resolution
Vertical Location
Clear
Variable
Description
Project
Evapotranspiration (2)
Evapotranspiration
ABoVE
This variable can be found in the following datasets.
DataSet
Select to Compare
Data Location
Evapotranspiration over ABoVE Domain from MODIS
/att/pubrepo /ABoVE_products /ModelBenchmarking
Soil Moisture over ABoVE Domain from SMAP
/att/pubrepo /ABoVE_products /ModelBenchmarking
Land cover (1)
Land cover
ABoVE

Hands On Demonstration

Transferring data in and out

- In/out – Filezilla to login nodes, nobackup
  - Configure dsclogin in Site Manager, select “Logon Type” to be “Interactive” for RSA token login

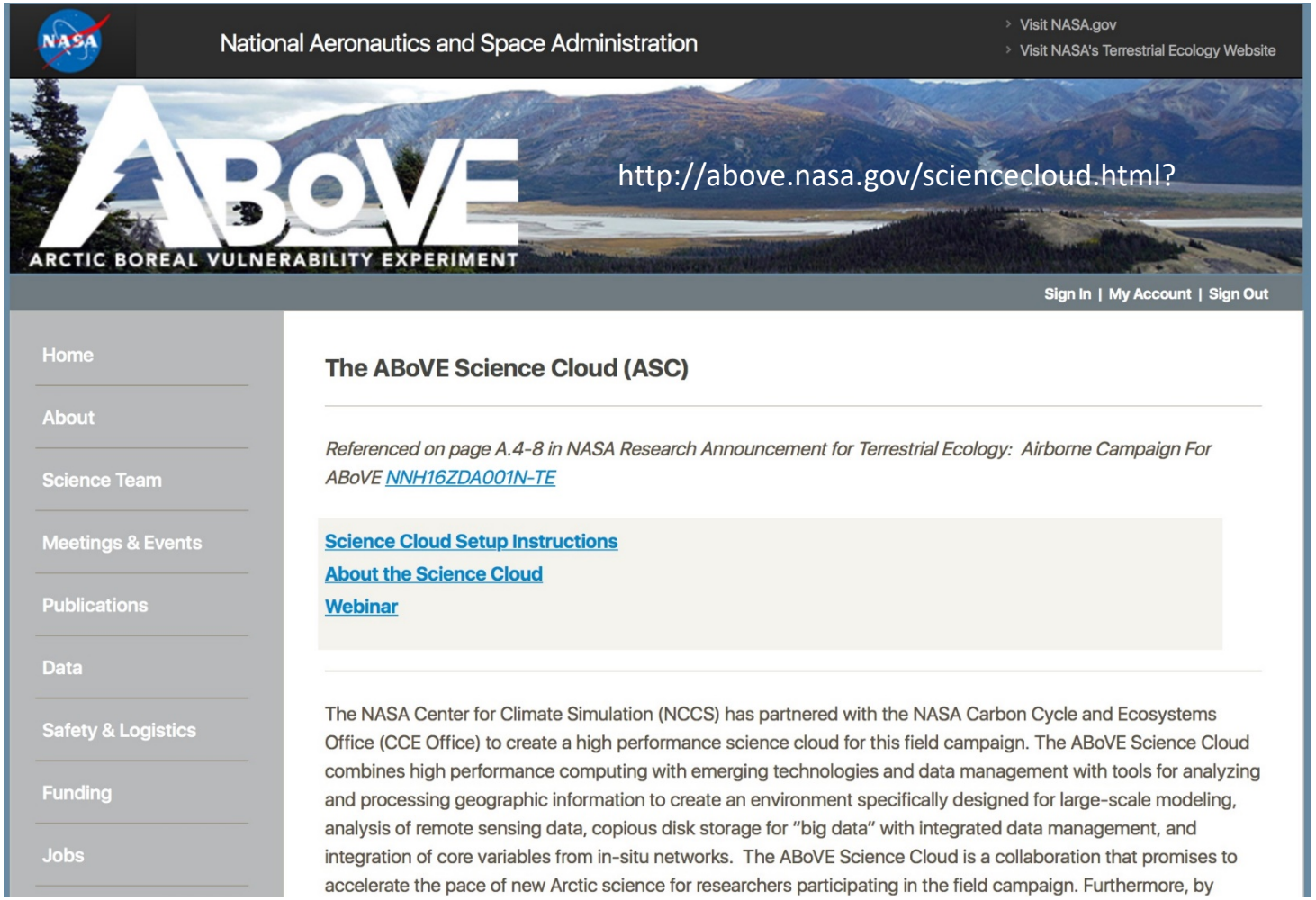
NGA Out – need written permission from Jim Tucker (email), use ngaaccess

Running applications

- Python, Fortran, C
- Script to run slurm --->
- Run with:
  - srun -n 8 ~/coin
- More options available (example)
- squeue –u *username*
- *scancel*

```
lcarrier@above101:~$ cat coin.j
#!/bin/bash
#SBATCH --job-name=coin
#SBATCH --time=01:35:00
#SBATCH --nodes=1
#SBATCH --ntasks=8
```

Backup Slides


National Aeronautics and Space Administration
Visit NASA.gov
Visit NASA's Terrestrial Ecology Website
ABoVE
ARCTIC BOREAL VULNERABILITY EXPERIMENT
http://above.nasa.gov/sciencecloud.html?
Sign In | My Account | Sign Out
Home
About
Science Team
Meetings & Events
Publications
Data
Safety & Logistics
Funding
Jobs
The ABoVE Science Cloud (ASC)
Referenced on page A.4-8 in NASA Research Announcement for Terrestrial Ecology: Airborne Campaign For ABoVE NNH16ZDA001N-TE
Science Cloud Setup Instructions
About the Science Cloud
Webinar
The NASA Center for Climate Simulation (NCCS) has partnered with the NASA Carbon Cycle and Ecosystems Office (CCE Office) to create a high performance science cloud for this field campaign. The ABoVE Science Cloud combines high performance computing with emerging technologies and data management with tools for analyzing and processing geographic information to create an environment specifically designed for large-scale modeling, analysis of remote sensing data, copious disk storage for "big data" with integrated data management, and integration of core variables from in-situ networks. The ABoVE Science Cloud is a collaboration that promises to accelerate the pace of new Arctic science for researchers participating in the field campaign. Furthermore, by

Scientific Visualization Studio Web Site
342 TB of free content

Scientific Visualization Studio Visualizations

Conceptual Image Lab Animations

Science Support Office Presentations

svs.gsfc.nasa.gov
342 TB online
384 TB delivered in 2016

Multimedia Videos, Interviews, Social Media Content

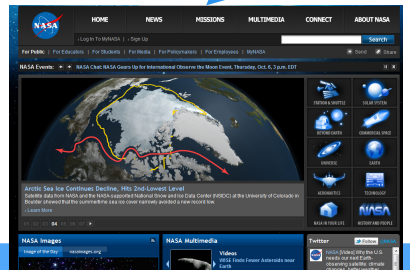
SMD Presentation HyperwallShows

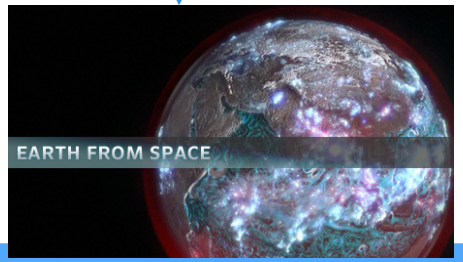
NASA Visualization Explorer Stories

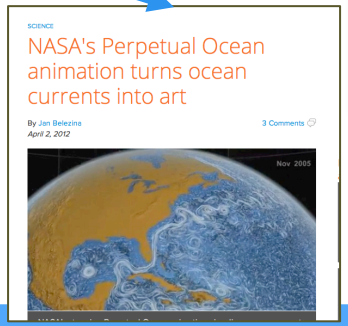
SMD Presentation Hyperwall Content Feed

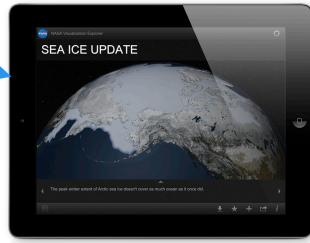
NASA Visualization Explorer Story Feed

Science and Theme-based Communications Content Server



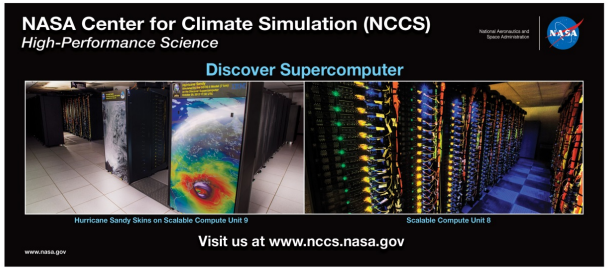


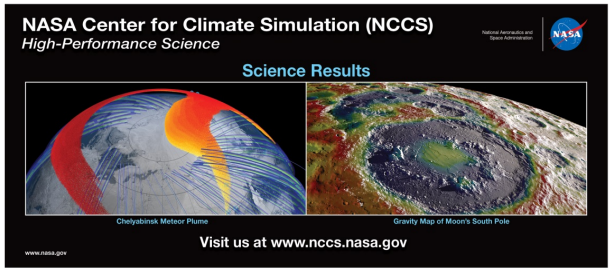




NCCS Fun Facts: Did You Know?

- It would take the world's population tapping on calculators for more than 145 hours to equal what the NCCS Discover supercomputer can calculate in one second!
- The NCCS holds enough science data to produce a music playlist 260,000 years long!
- If the NCCS printed out all of its data, the stack of paper would reach from the Earth out to more than ten times the distance to the Moon!
- Making the paper to print all that data would require about 4.8 billion trees!
- Using the NCCS network, you could download an HD movie in about 6 seconds, or 560 HD movies every hour!
- The NCCS currently has 2,000 kilowatts of power-the equivalent of approximately 200 U.S. households!

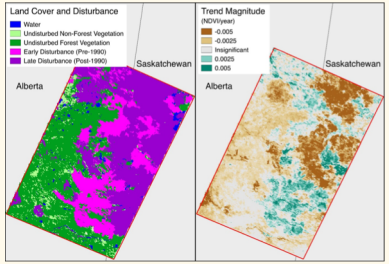




ABoVE Science Cloud on ADAPT

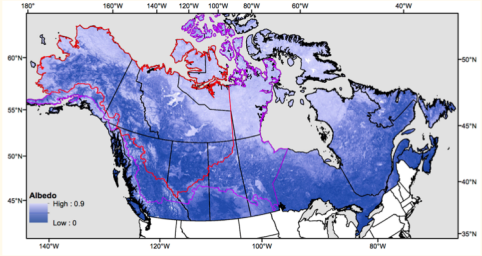
ABoVE is a large-scale NASA-led study of environmental change in Arctic and boreal regions and the implications for ecological systems and society. The ABoVE Science Cloud on the NCCS Advanced Data Analytics Platform (ADAPT) hosts large datasets and compute space for researchers. Over 120 users from 45 projects are using ADAPT to generate data products across many Earth science disciplines. Examples include:

Landsat (>200 terabytes)



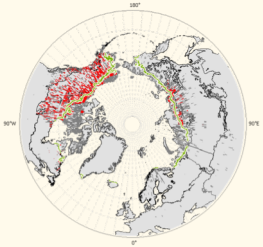
- **C. Woodcock** (above): Domain-wide mapping of disturbance, growing season length, and greening/ browning
- **Others**: Domain-wide mapping of water, fire disturbance, vegetation characterization, and tree canopy cover

MODIS (>60 terabytes)



- **B. Rogers** (above): Mean albedo maps, modeling driving factors of post-fire albedo, and fire combustion mapping
- **Others**: Domain-wide mapping of active fire detection, NDVI, and cloud climatology

DigitalGlobe (>2 petabytes)



- **C. Neigh** (above): Tundra-taiga ecotone mapping and digital elevation models
- **Others**: Land cover/land use classification, digital surface models, calibration/validation of model data, site selection, and logistics