



# **ABOVE Science Cloud Introduction**

**4/26/2019**

**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
- ArcGIS Services
- Success Stories

### **Scientific Visualization Studio**

#### **Logistics**

- Gaining access to the system
- Finding out more information

#### **Simple Tasks**

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



# Team Members

## **ABoVE**

- Elizabeth Hoy
- Peter Griffith

## **NCCS**

- Jim Shute
- Kerman Bime
- Scott Sinno
- 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](mailto: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.**

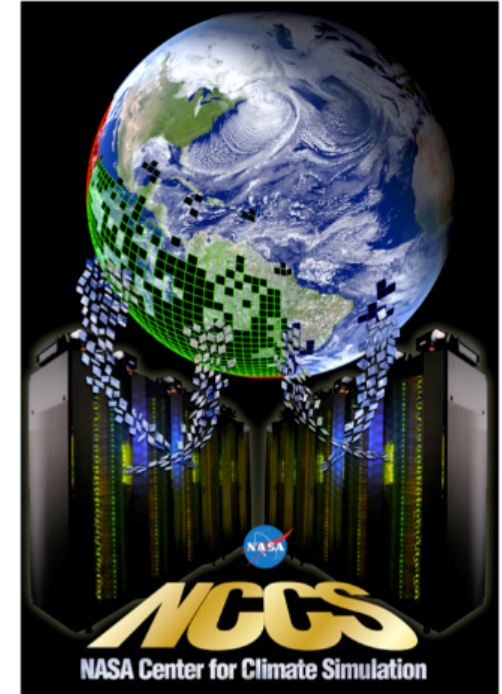
- High-performance computing, cloud computing, data storage, and networking technologies
- High-speed access to petabytes of Earth Science data
- Collaborative data sharing, publication, and analysis services

## **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, IceSAT2

## **High-Performance Science**

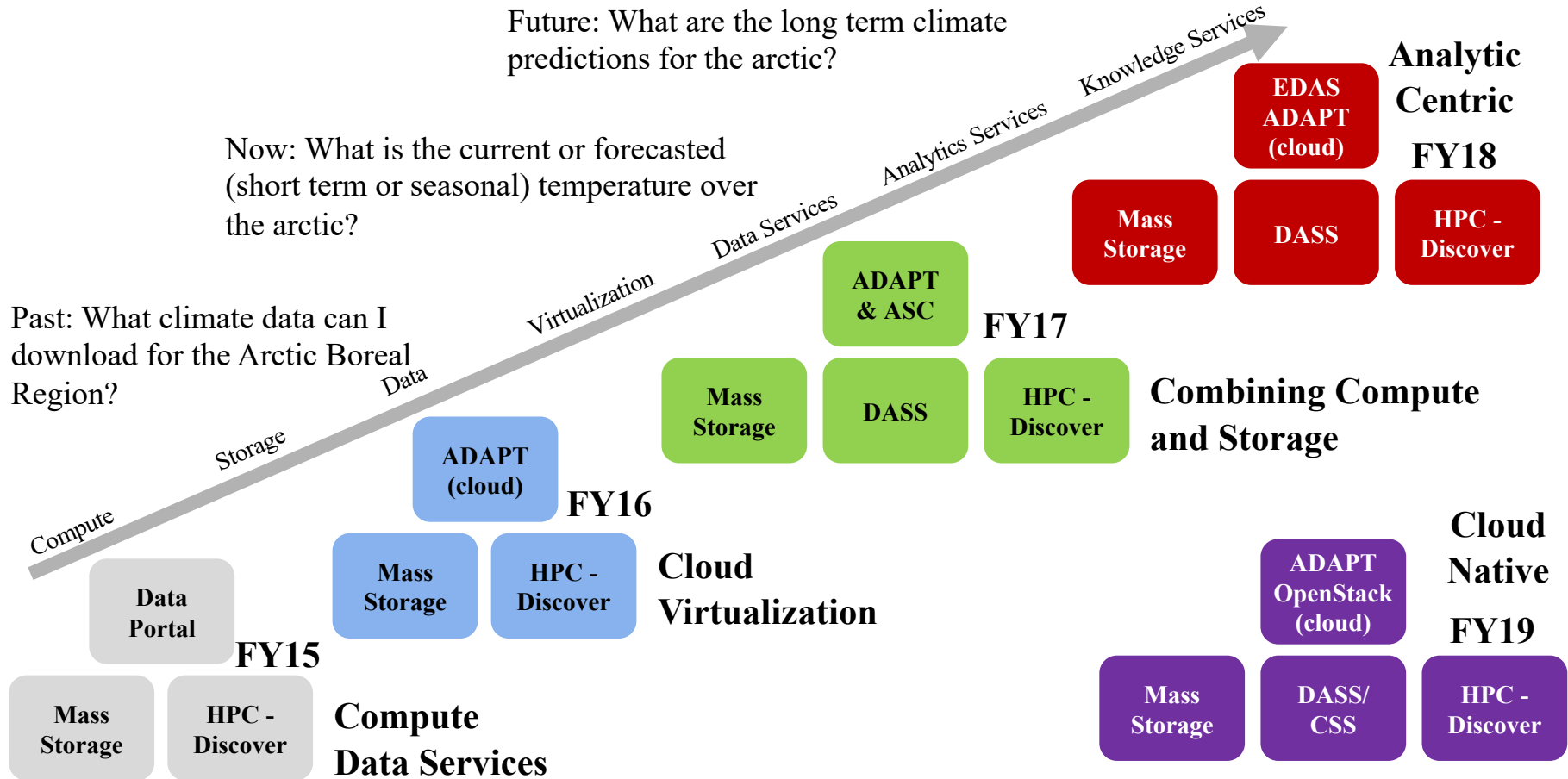
- <http://www.nccs.nasa.gov>
- Funded by the High End Computing (HEC) program under SMD
  - » Dr. Tsengdar Lee, Program Manager
- Code 606.2 at NASA Goddard Space Flight Center in Greenbelt, MD.







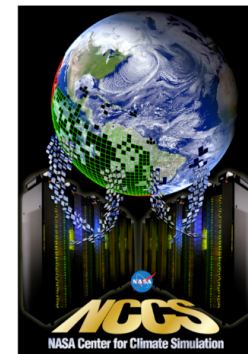
# Evolution of Major NCCS Systems



# 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](mailto:support@nccs.nasa.gov)



# 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



# ASC Compute Options

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 beyond101-125
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, etc.

Nodes with 128 GB of RAM, 10 GbE



#### DataBase

High available database nodes with solid state disk.

Nodes with 128 GB of RAM, 3.2 TB of SSD, and 10 GbE



#### Remote Visualization - planned

Enable server side graphical processing and rendering of data through Jupyterhub.

Nodes with 128 GB of RAM, 10 GbE, FDR IB, and GPUs



#### High Performance Compute

More than 6,000 cores coupled via high speed networks for elastic or itinerant computing requirements.

300+ nodes with between 24 and 128 GB of RAM



#### High-Speed/High-Capacity Storage

Petabytes of storage accessible to all the above capabilities over a high speed network.

Storage nodes configured with multiple PB's of RAW storage capacity

# ASC Software Stack

External License Servers

Virtual machines can be set up to use PI supplied licenses or to reach out to 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 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

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
7.4 Mbps average US internet speed	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 NASA / Other Gov	1 Gbps (1000Base-T) 177:46:40
	2.4 Gbps (OC48) 74:04:26
	10 Gbps (OC192) 17:46:40

**@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
MERRA2	320 TB
NGA/DigitalGlobe Imagery	3.3 PB
Total	> 3.35 PB

## Other Data Sets

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

## Future Data Sets:

- Planet

*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  $\geq 31$  cm satellite imagery for US Federal use, i.e., NSF, NASA and NASA funded projects.

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

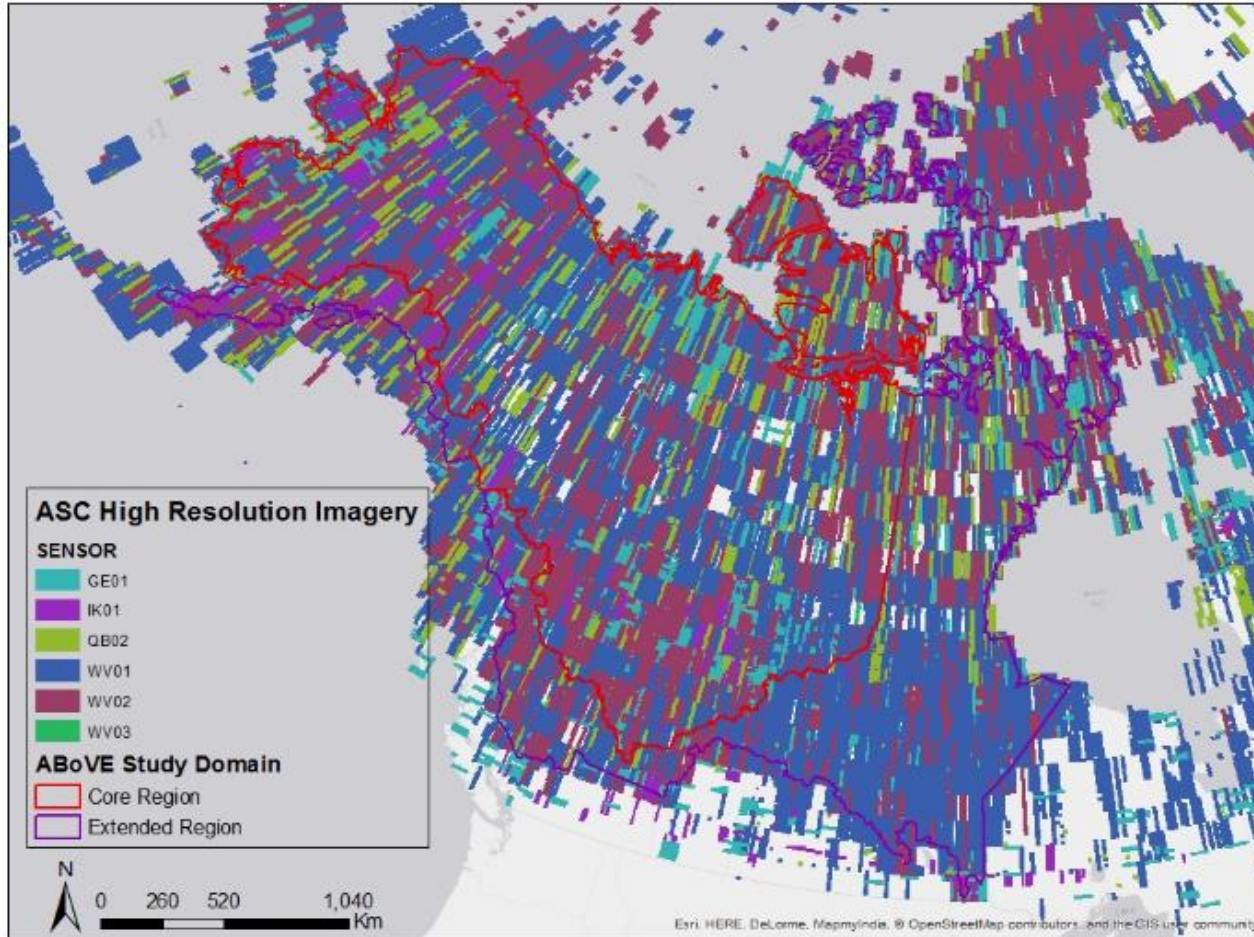
Satellite	Bands	Nadir Panchromatic Resolution (m)	Nadir Multispectral Resolution (m)
Ikonos	Pan, R, G, B, Near IR	0.82	3.2
GeoEye	Pan, R, G, B, Near IR	0.41	1.65
Quickbird	Pan, R, G, B, Near IR	0.55	2.16
WorldView-1	Panchromatic only	0.5	N/A
WorldView-2	Pan, R, G, B, Near IR 1, Near IR 2, Coastal, Red 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

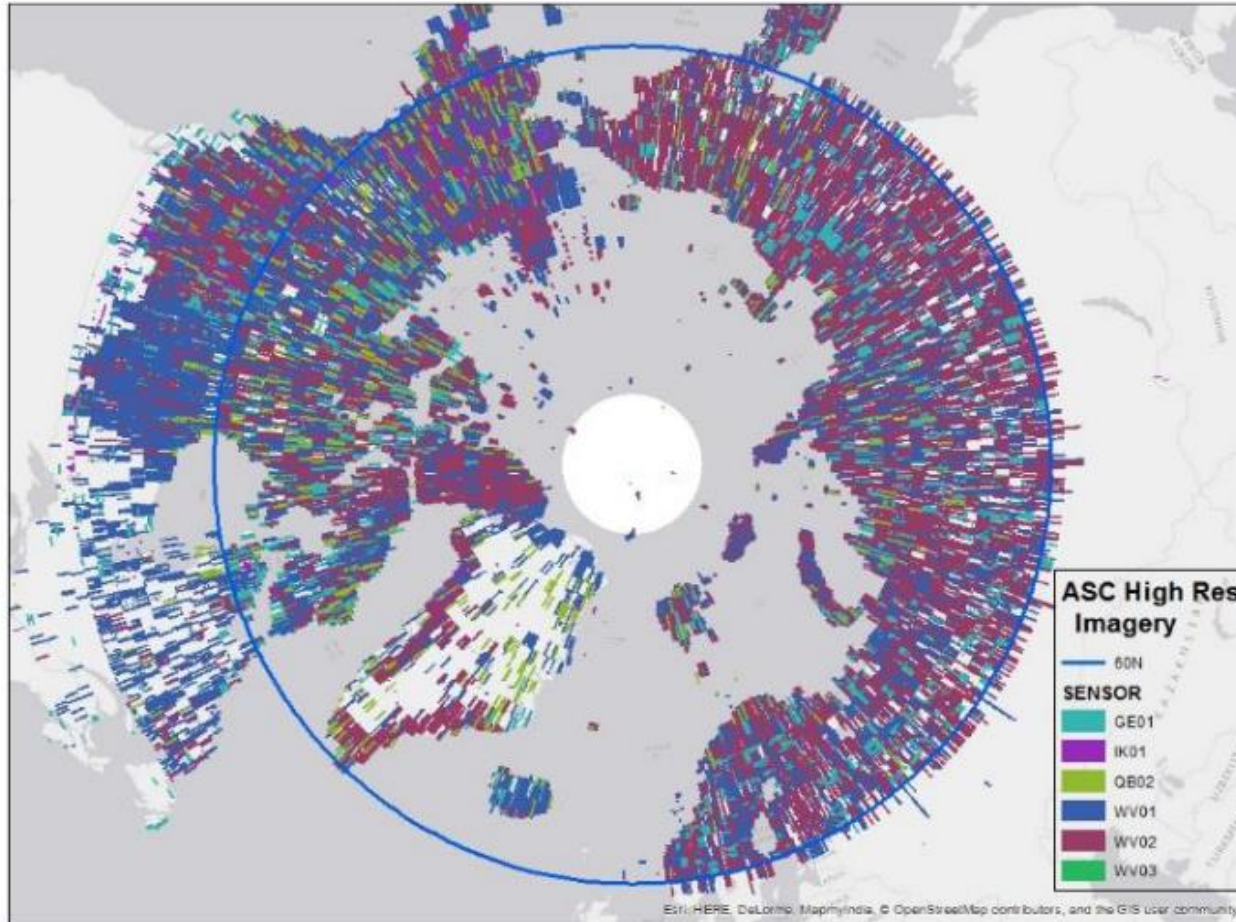
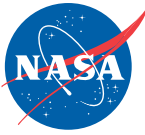


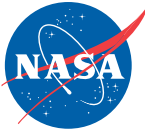


# ABOVE Science Cloud DigitalGlobe Imagery: Study Domain



# ABOVE Science Cloud DigitalGlobe Imagery: Circumpolar



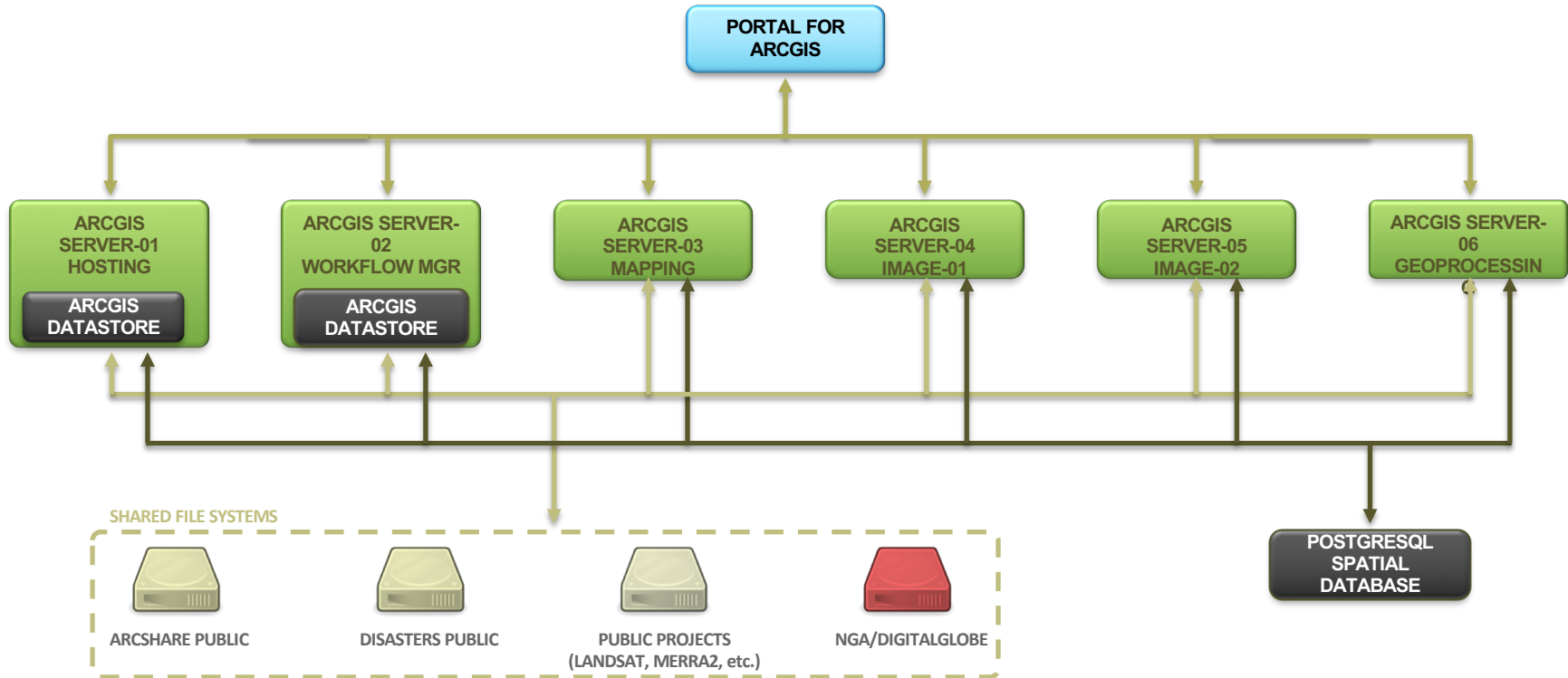


# NGA Footprints

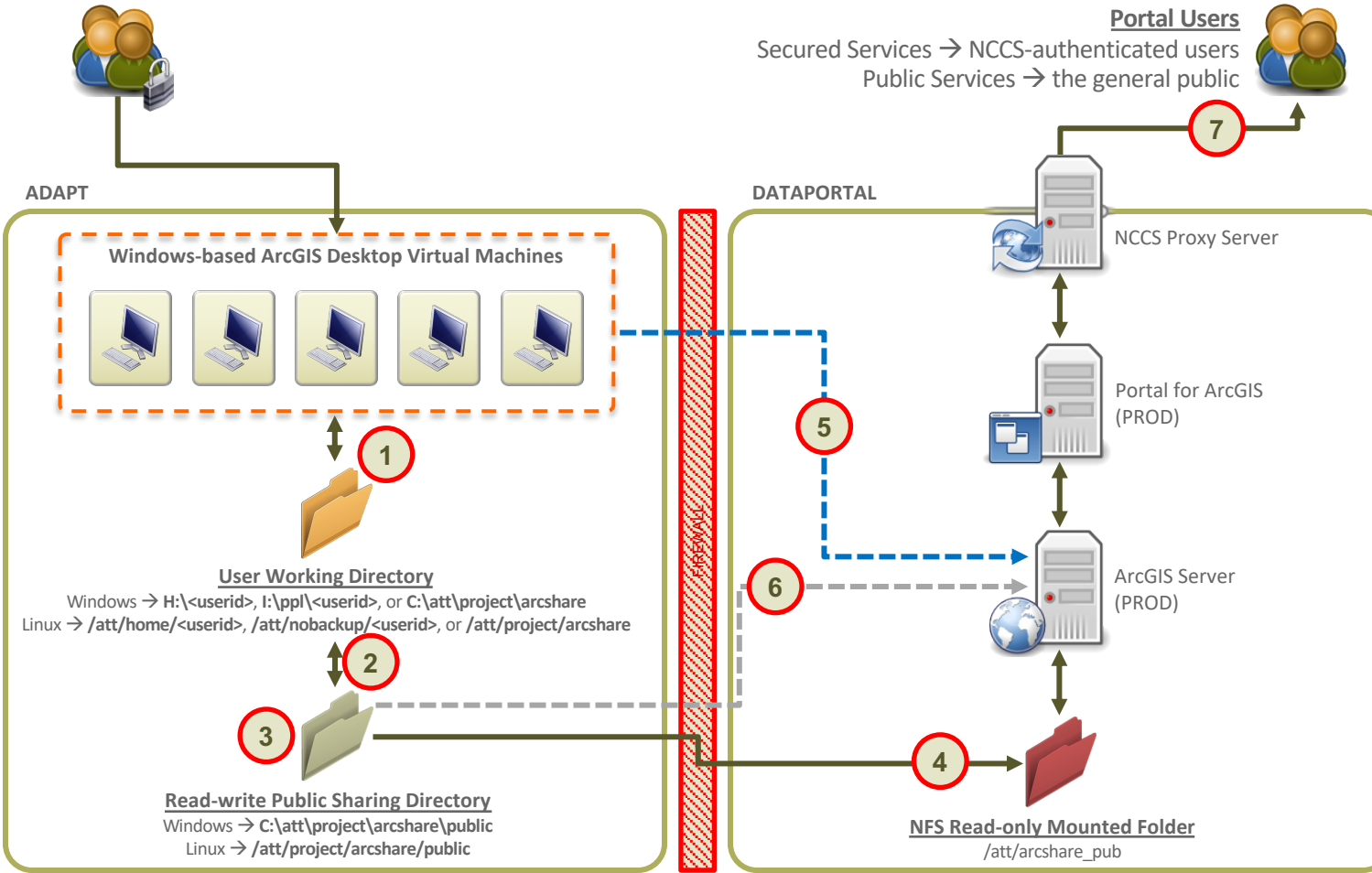
- Path to Footprints (shape file and geodatabase) for data discovery:
  - /att/pubrepo/NGA/INDEX/Footprints/current/newest
  - Older dates are available
  - Shapefiles are available by Sensor, Product Level, and year.
- Qgis, ArcMap can read the Footprints files
- Bi-monthly updates to the Footprints file
- Data is added to the system weekly



# ArcGIS: SPATIAL PLATFORM ARCHITECTURE

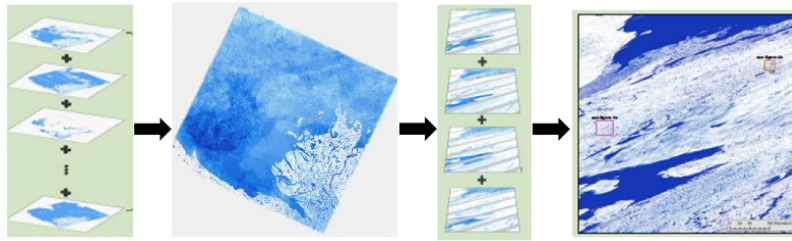


# PUBLISHING TO THE NCCS MAPPING PORTALS FROM THE FILE SYSTEM



- Publishing Steps**
1. Using the Windows VM environment, user creates and manages data in their working directory (or shared ArcGIS folder)
  2. When ready to publish, data is copied to the /arcshare/public folder
  3. User sets file ownership and permissions
  4. (System) NFS read-only mounted folder automatically updates with data changes
  5. Using ArcGIS Desktop tools from the Windows VM environment, user publishes service(s) to the Production Portal. Note: Map documents, etc. should point to the /arcshare/public folder
  6. (System) Folders are registered with ArcGIS Server and replace /arcshare/public data pointers with the NFS read-only mounted folder location
  7. End user accesses services that reference the NFS read-only version of the data (/att/arcshare\_pub)

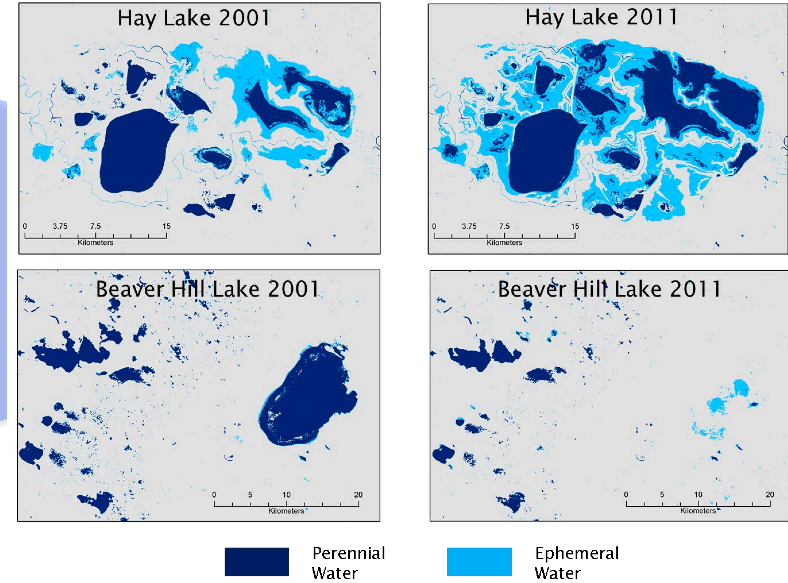
# ABoVE Water Maps: 30 meter spatial resolution surface water 1991-2011,” M.L. Carroll, et. al



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

- 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



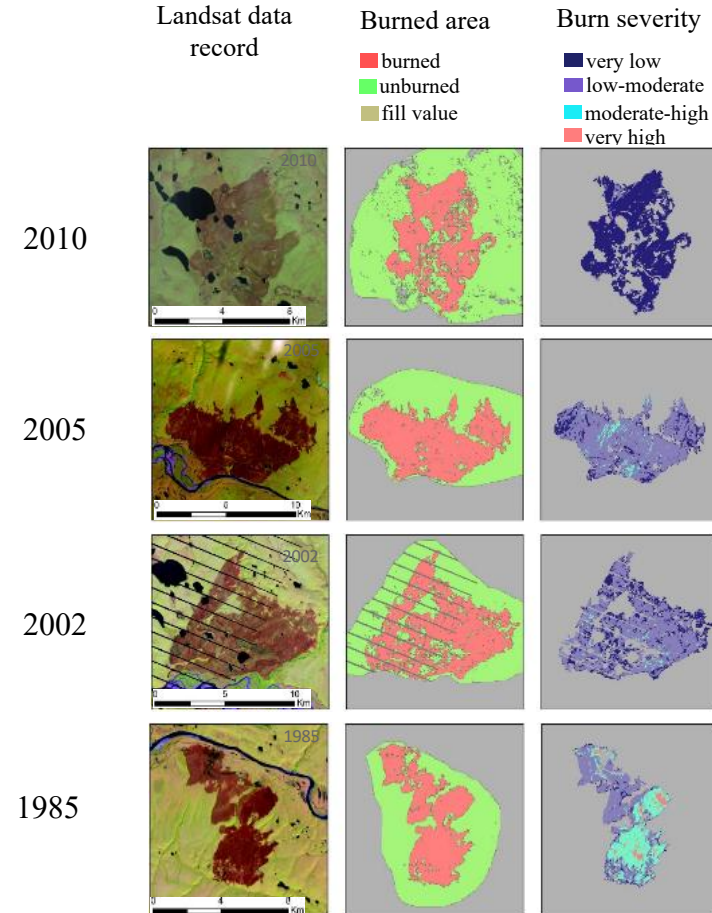
AWM for 2001 and 2011 for Hay Lake and Beaver Hill Lake in Canada. Hay Lake has clearly expanded over this time frame while Beaver Hill Lake has diminished.

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](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

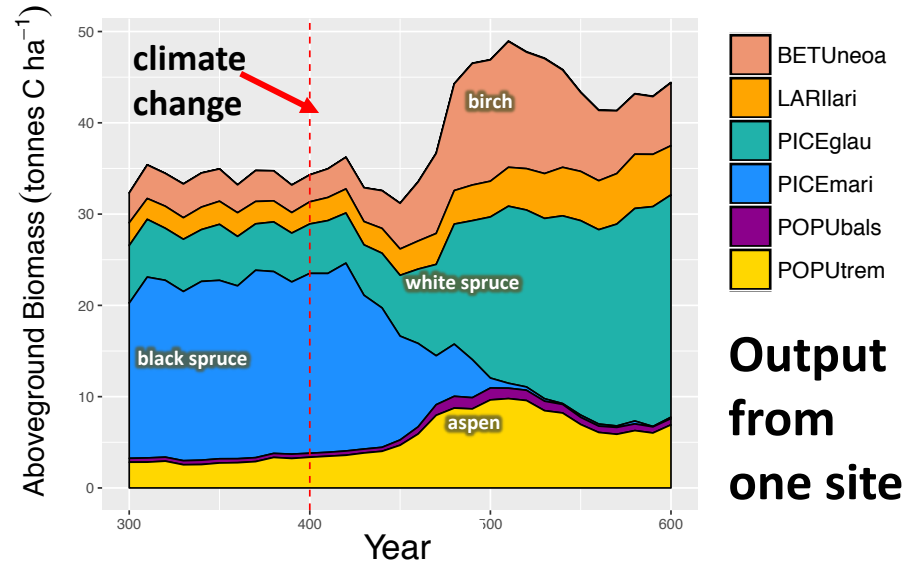
For each site/grid cell:

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**

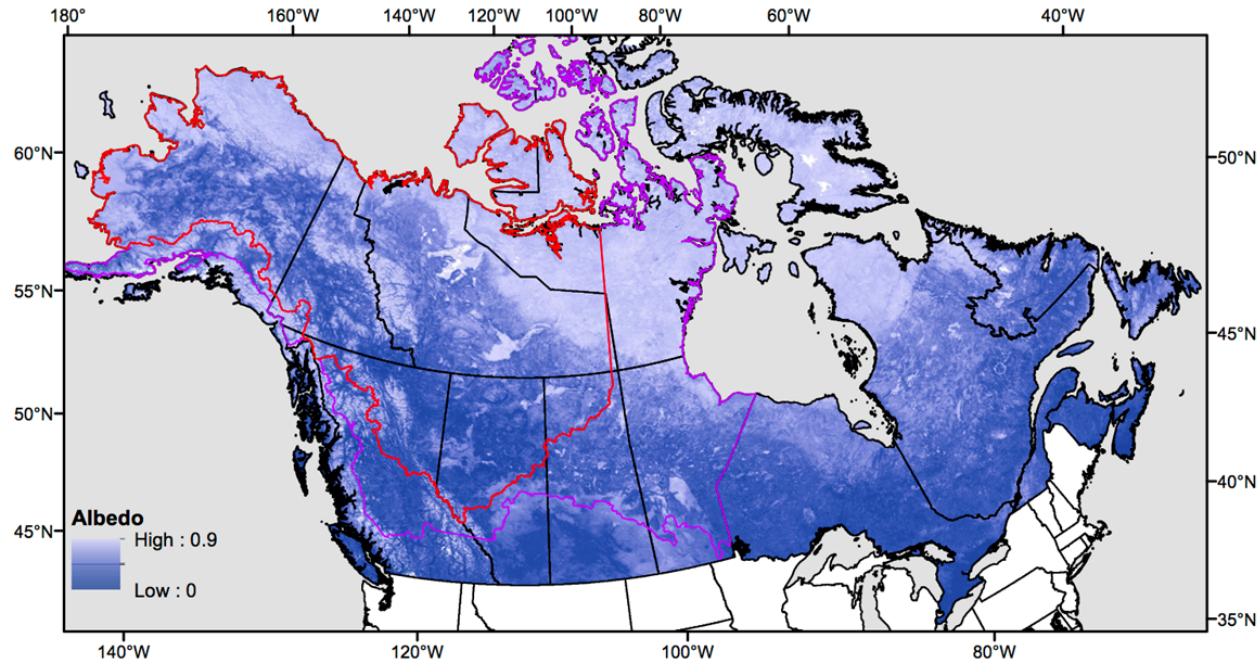
About to use 4 cores/node to see if that can be reduced to less than half a day



# Understanding the Causes and Implications of Enhanced Seasonal CO<sub>2</sub> Exchange in Boreal and Arctic Ecosystems – B. Rogers



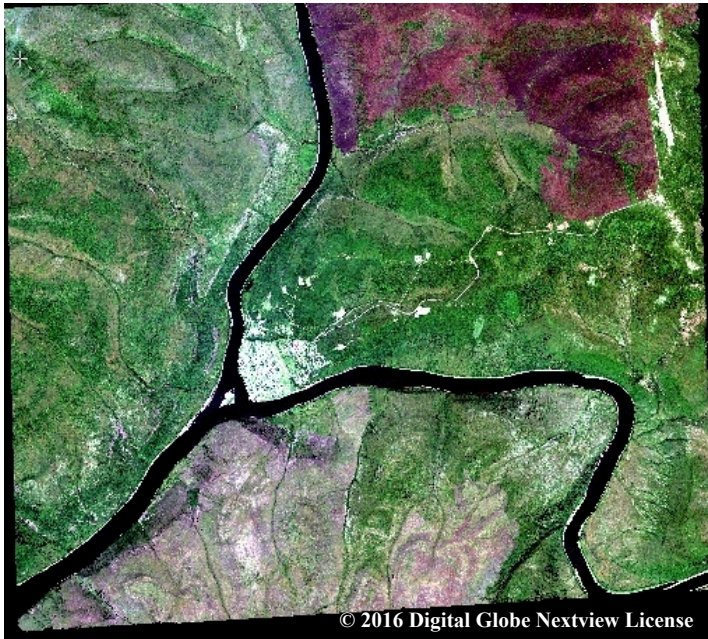
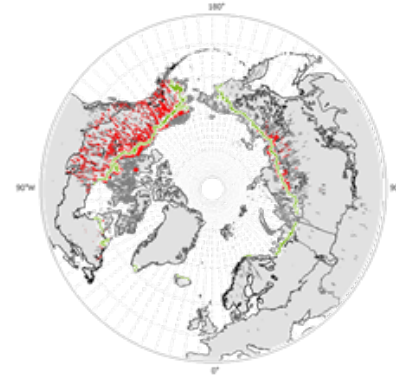
- Modeling driving factors of post-fire albedo trajectories
- Creation of mean albedo maps
- Fire combustion mapping



# 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



© 2016 Digital Globe Nextview License



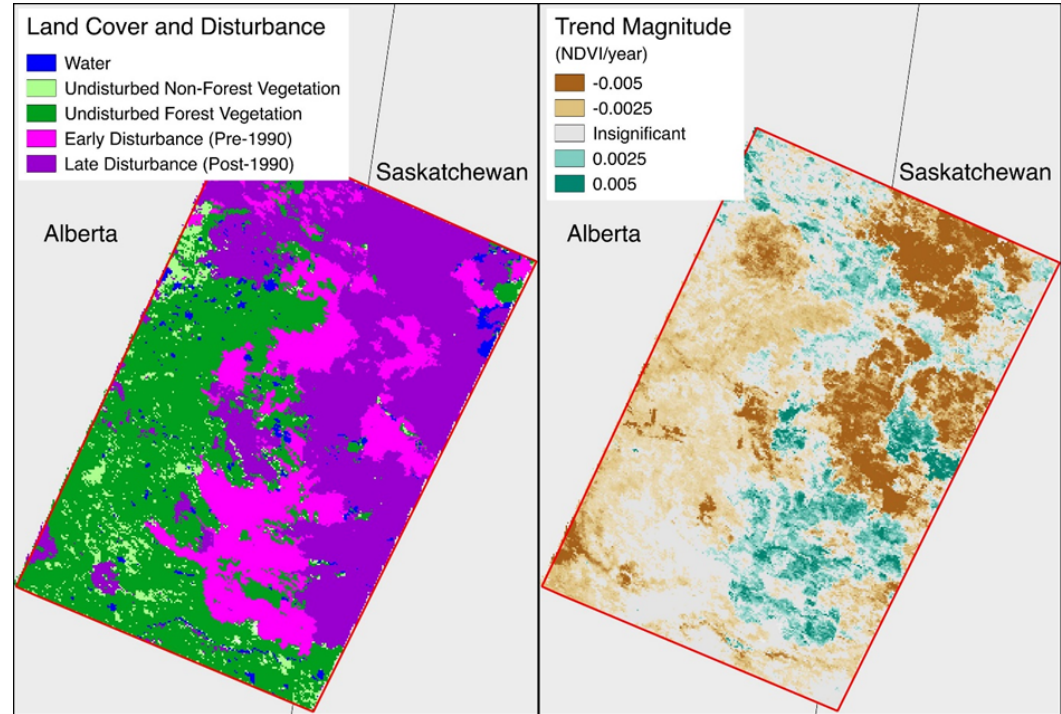
© 2016 Digital Globe Nextview License



# 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



Melaas, E. K., Sulla-Menashe, D., Friedl, M. A. **2018**. Multidecadal Changes and Interannual Variation in Springtime Phenology of North American Temperate and Boreal Deciduous Forests. *Geophysical Research Letters*. 45(6), 2679-2687. DOI: [10.1002/2017GL076933](https://doi.org/10.1002/2017GL076933)

Sulla-Menashe, D., Woodcock, C. E., Friedl, M. A. **2018**. Canadian boreal forest greening and browning trends: an analysis of biogeographic patterns and the relative roles of disturbance versus climate drivers. *Environmental Research Letters*. 13(1), 014007. DOI: [10.1088/1748-9326/aa9b88](https://doi.org/10.1088/1748-9326/aa9b88)

# 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

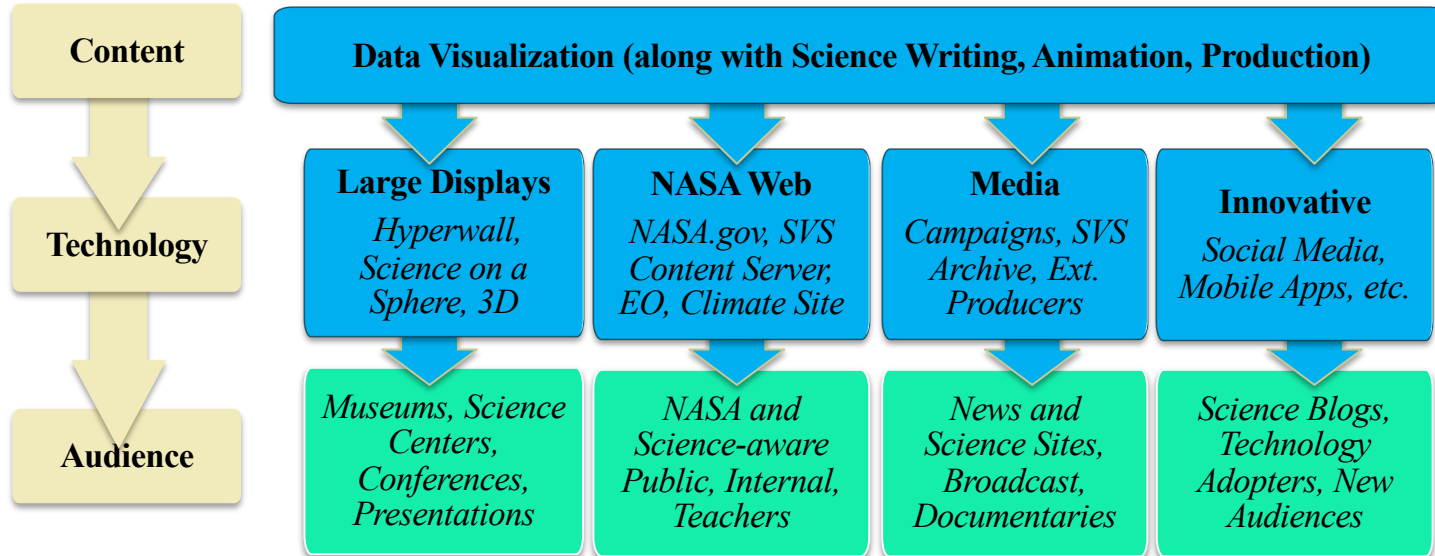
## **Available to support the ABoVE project**

- 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](mailto: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](https://above.nasa.gov/sciencecloud_setup.html)
- Need NASA Identity, IT Security training, RSA Token, process takes a while
- Optional: signed NGA paperwork
- Links to monthly webinars, other instructional videos
- POC – Liz Hoy, [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov), Laura Carriere

## More Information:

- NCCS Website: See <https://www.nccs.nasa.gov/nccs-users/instructional/adapt-instructional>
  - How to login, data locations, Windows FAQ, ABoVE FAQ, including orthorectification instructions
  - More information will be added in early May





# New User Account Notes

- To streamline the user account process, complete the following before submitting a request to the NCCS:
  - FY2019 SATERN
  - Applied for the "Agency RSA SecurID Token" through NAMS, needed for (hard or soft) token access
- Foreign national (FN) users, with either new or existing NASA Identities, the user's affiliation sponsor will need to work with the GSFC International Visit Coordinator (IVC) to have the "GSFC-Code 600-NCCS" Access Control Plan (ACP) added to the FN's NASA Identity.



# Simple Tasks

## Log in with:

- Linux: `ssh adaptlogin.nccs.nasa.gov`
- Windows: <https://guac.nccs.nasa.gov>

## Transferring data in and out

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

NGA Out – check the appropriate box on the NGA paperwork, use `ngaaccess`

## Running applications

- Python, Fortran, C
- Script to run slurm --->
- Run with: `srun -n 8 ~/coin`
- `queue -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
```

# Questions?



## 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](mailto:support@nccs.nasa.gov)