CMAC-11-0008

Parallel Web-Service Climate Model Diagnostic Analyzer

Interim Report of Year 2

PI: Seungwon Lee

April 10, 2014

PAWS-CMDA: Parallel Web-Service Climate Model Diagnostic Analyzer

PI: Seungwon Lee, JPL

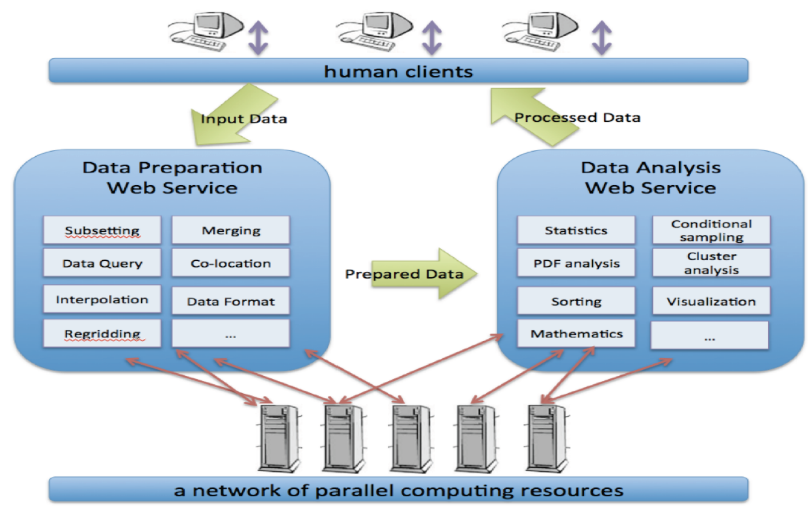
Objective

Develop a parallel, distributed web-service oriented system to enable multi-aspect physics-based and phenomenon-oriented climate model performance evaluation and diagnosis through the comprehensive and synergistic use of multiple observational data, reanalysis data, and model outputs.

Significance

Scientific and technological advances in NASA's climate and weather predictions through multivariate conditional sampling approaches and parallel and distributed web service technologies.

Functional architecture diagram for PAWS-CMDA



Approach

- Develop Python-driven data preparation and analysis system. Wrap the existing data analysis tools in Python for integration into the system.
- Apply Parallel Python to achieve large-scale distributed parallelism. Use Parallel Python to establish a server network, distribute subtasks to and retrieve results from the server network.
- Build web services of the data preparation and analysis system with a dual access mode: web browsers and programs.

Co-Is/Collaborators

Lei Pan, Chengxing Zhai, Jonathan Jiang, Benyang Tang, Joao Teixeira

Task Schedules and Key Milestones

Tasks completed:
- Developed a python-driven data processing/analysis web-service system.
- Demonstrated scientific use of the web-service system.
- Integrated the web-service system into the Amazon Cloud system.

Tasks planned:
- Demonstrate the educational use of the web-service system.
- Integrate the web-service system for multi-step analyses.
- Apply Parallel Python where appropriate (e.g. Co-location).

Milestones Met:
- Implemented the data processing system.
- Implemented the data analysis system.

Milestones Planned:
- Implement integrated data processing/analysis system.

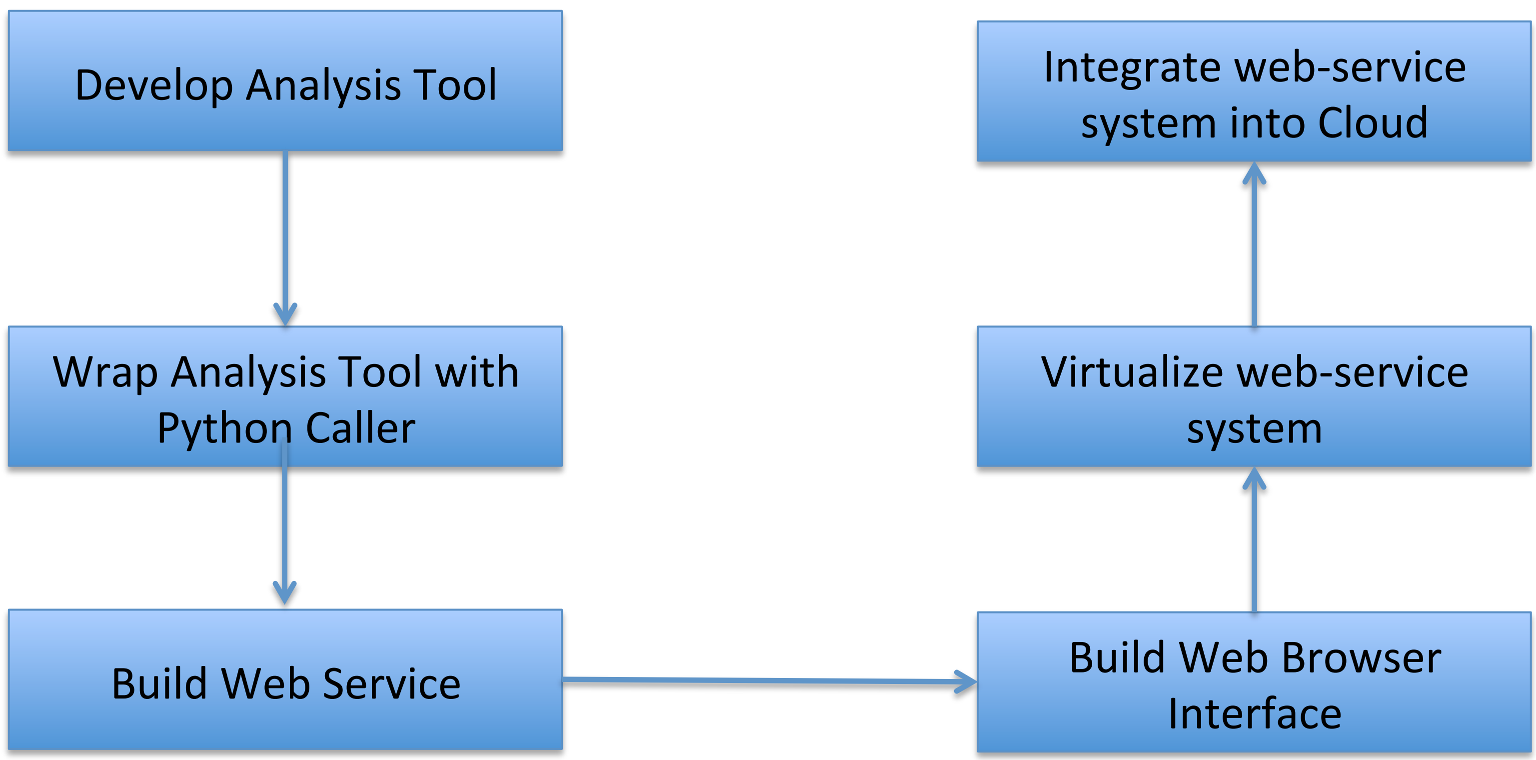
Outline

- **Technical Status**
- Programmatic Status
- Educational and Outreach Status

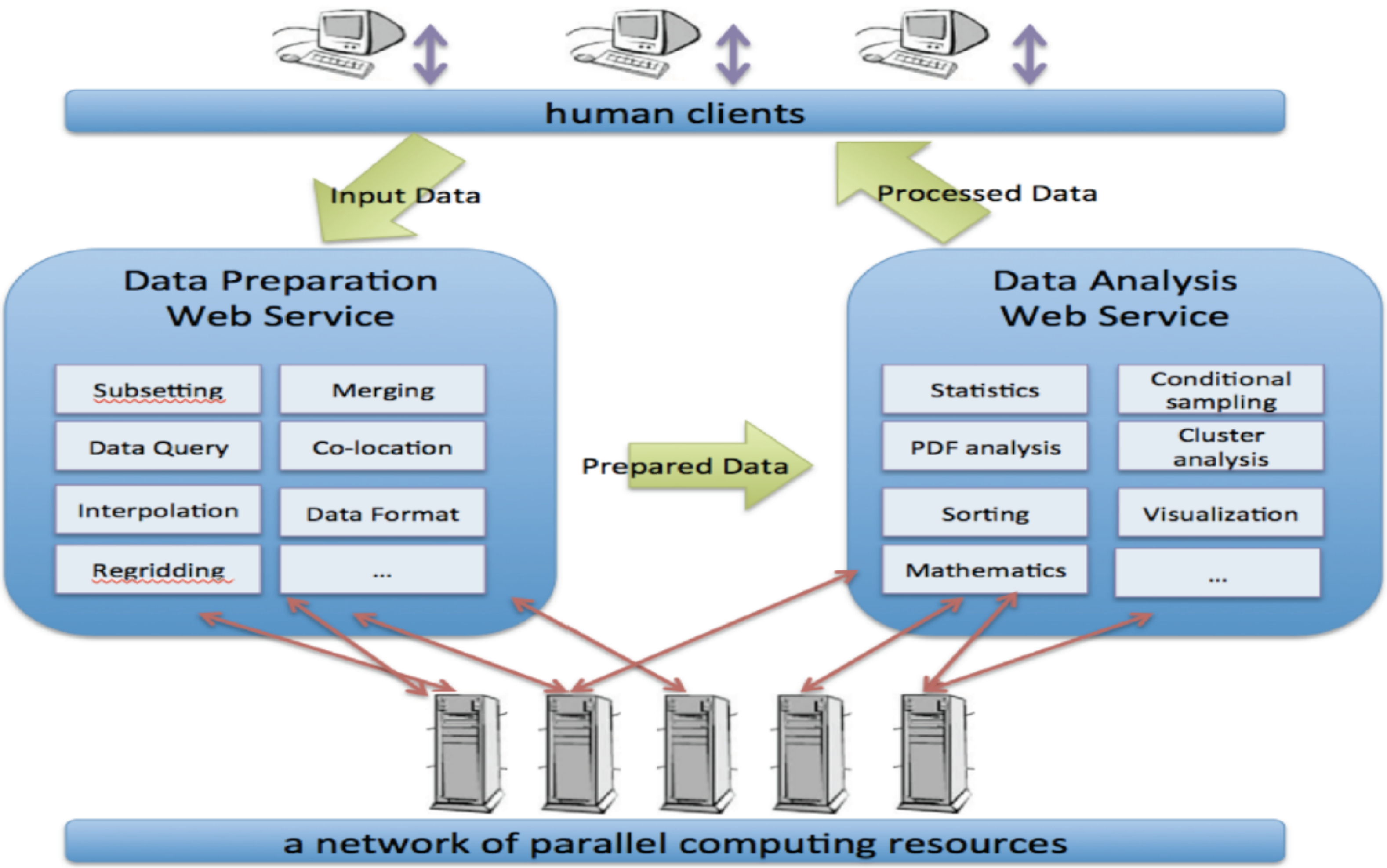
Technical Status Summary

- Purchased a project dedicated machine consisting of four computing servers and one 100-TB RAID storage.
- Developed 10 data analysis tools: 2D/3D averaging and plotting tools, 2 variable correlation, 2 data source difference, 2 variable conditional sampling, 2 dataset collocation in space and time.
- Developed a Python-driven system with the data analysis tools in order to provide uniform interface between the computational code and web services.
- Built web services with the Python-driven web service technologies using Flask, Gunicorn and Tornado.
- Designed and implemented web browser interface for the web services using JavaScript.
- Virtualized web servers for flexible service development and deployment.
- Developed two new web-services since the last report: conditional sampling, co-location.
- Ten web services are available for climate model/satellite data analyses.
- Integrated the web-service system into commercial cloud environment: Amazon.
- Improved existing web services to provide users with more customization capabilities : display options and data downloading.

Approach Flow Chart

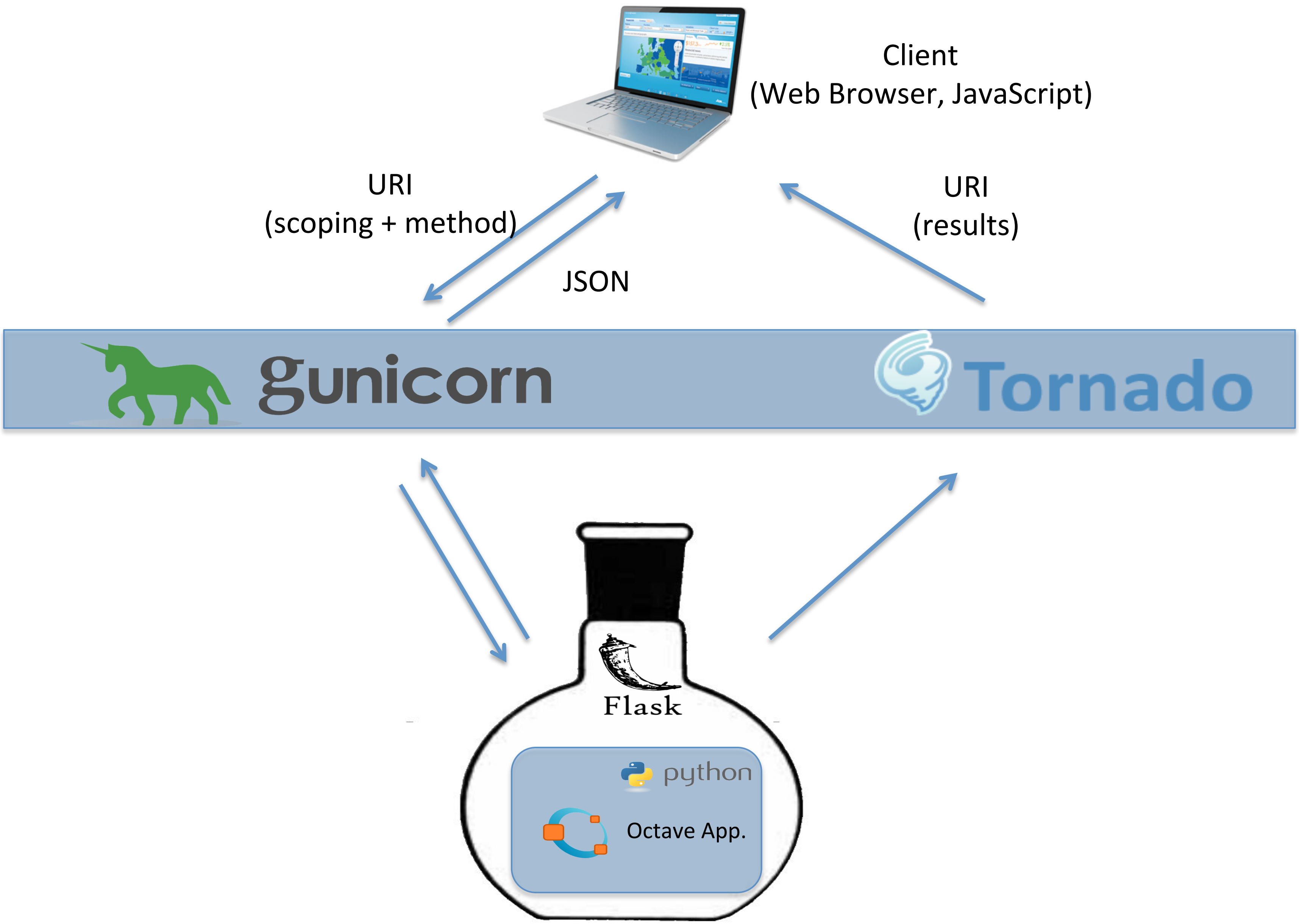


Functional Architecture Diagram



or
Cloud Computing Resources

Service Diagram with Technical Components



Analysis Tools

- 2D Variable Average Map
- 2D Variable Zonal Mean
- 2D Variable Time Series
- 3D Variable 2D Slice Average Map
- 3D Variable Zonal Mean
- 3D Variable Vertical Profile Average
- Scatter Plot and Correlation of Two Variables
- Difference Plot of Two Variables
- Conditional Sampling with Two Variables
- Collocation of Two Datasets in Space and Time

Data Sets

- Model outputs from CMIP5 project
  - Experiments:
    - Historical runs and AMIP runs
  - Physical variables:
    - Surface temperature, Air temperature, Total cloud fraction, Precipitation, Specific humidity, Cloud liquid water content, Cloud ice water content. Sea surface wind, Sea surface height, Radiation fluxes.
  - Models:
    - CCCMA/canesm2, GFDL/esm2g, GISS/e2-h, GISS/e2-r, NCAR/cam5, NCC/noresm, UKMO/hadgem2-es, CCCMA/canam4, CSIRO/mk3.6, GFDL/cm3, IPSL/cm5a-lr, MIROC/miroc5, UKMO/hadgem2-a.
- Observation data from Obs4MIPs
  - AMSR-E surface temperature
  - AIRS and MLS air temperature
  - AIRS and MLS water vapor content
  - MODIS total cloud fraction, leaf area index
  - GPCP and TRMM precipitation
  - AVISO sea surface height
  - CERES radiation fluxes
- Observation data elsewhere
  - Grace equivalent water height
  - NOAA ocean heat content anomaly
  - ARGO ocean temperature and salinity profile

Conditional Sampling

- Conditional sampling displays the physical quantity X according to the values of another physical quantity E, which is related to X via physical processes.
- For example, we may want to display the cloud ice content according to the sea surface temperature. This enables to study the connection between X and E.
- E is a sampling variable and is usually a large scale environment variable such as sea surface temperature and wind field at 500 hPa.
- X is a sampled variable and is usually a physical quantity related to E by physical processes.
- A set of ordered values of E is used to define the boundaries of bins for E. E is sorted according to the bins. X is assigned to a sampling E bin according to its associated E value.

Mathematical Formulation of Conditional Sampling

Given quantities $X(lon,lat,z,t)$ and $E(lon,lat,t)$, where z is an optional dimension, typically the pressure level, and a monotonically increase sequence $\mathrm{B}_m^E, n = 0,1,\cdots,M$, be the boundary values of M bins, into which we want to sort E(lon,lat,t).

The set of space time points $\mathrm{S}_m = \left\{(lon,lat,t) \mid B_{m-1}^E \le E(lon,lat,t) \le B_m^E\right\}$,

label elements within this set by $\mathrm{s}_m^i = \left(lon_i,lat_i,t_i\right), i = 1,2,\cdots,N_m$

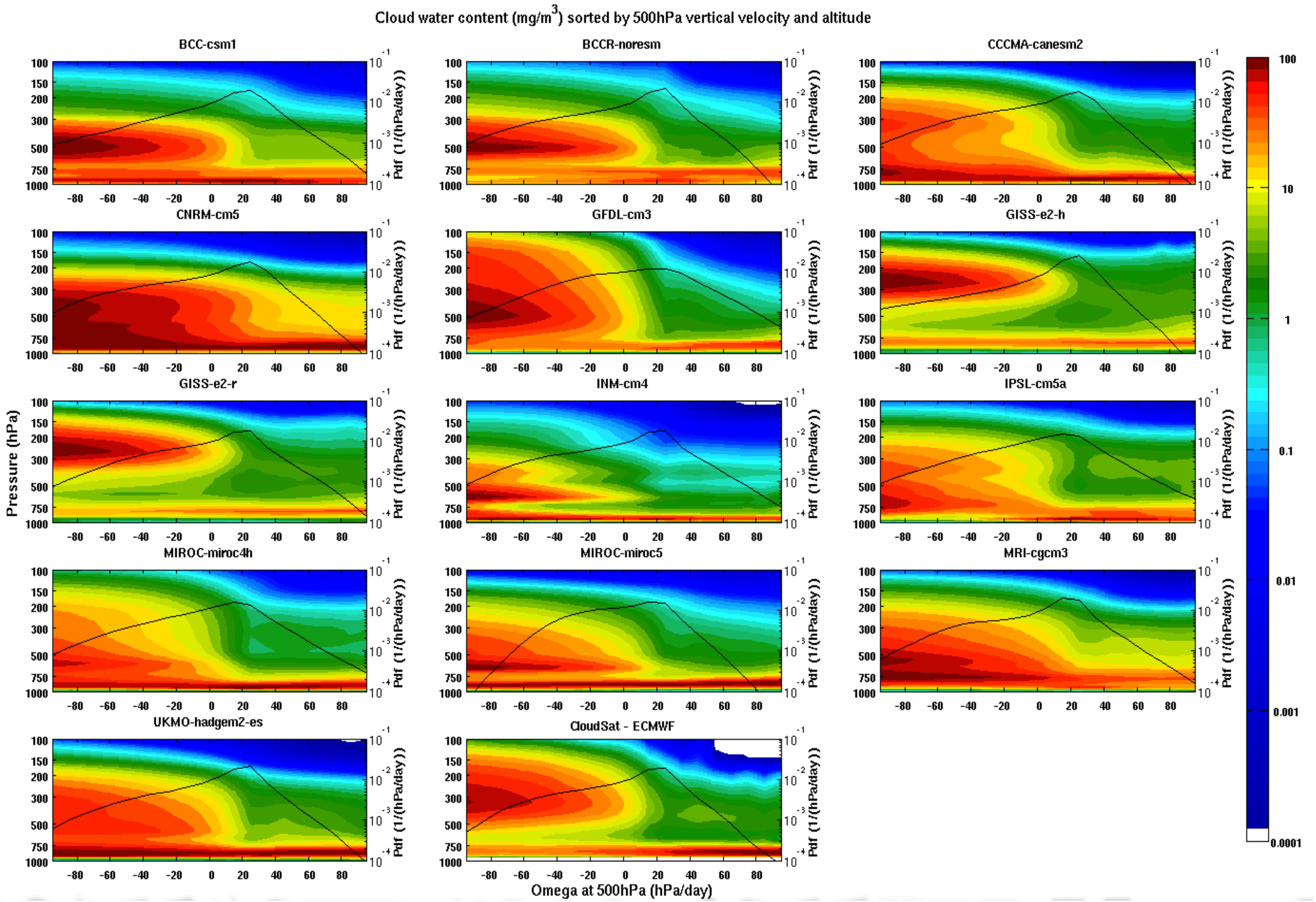
where N_m is the number of samples within m-th bin.

Let $X_m(z,\ i){=}X\left(lon_i,lat_i,z,t_i\right)$ be the value of X correspond to element $\mathrm{s}_m^i = \left(lon_i,lat_i,t_i\right) \in S_m$

The conditional sampling results are expressed as the mean $\bar{\mathrm{X}}_m(z)$ and stadnard deviation $\sigma_m^X(z)$ for the m-th bin:

$$\bar{\mathrm{X}}_m(z) = \frac{1}{N_m}\sum_{i=1}^{N_m} X_m(z,i), \quad \sigma_m^X(z) = \sqrt{\frac{1}{N_m - 1}\sum_{i=1}^{N_m}\left[X_m(z,i) - \bar{\mathrm{X}}_m(z)\right]^2}$$

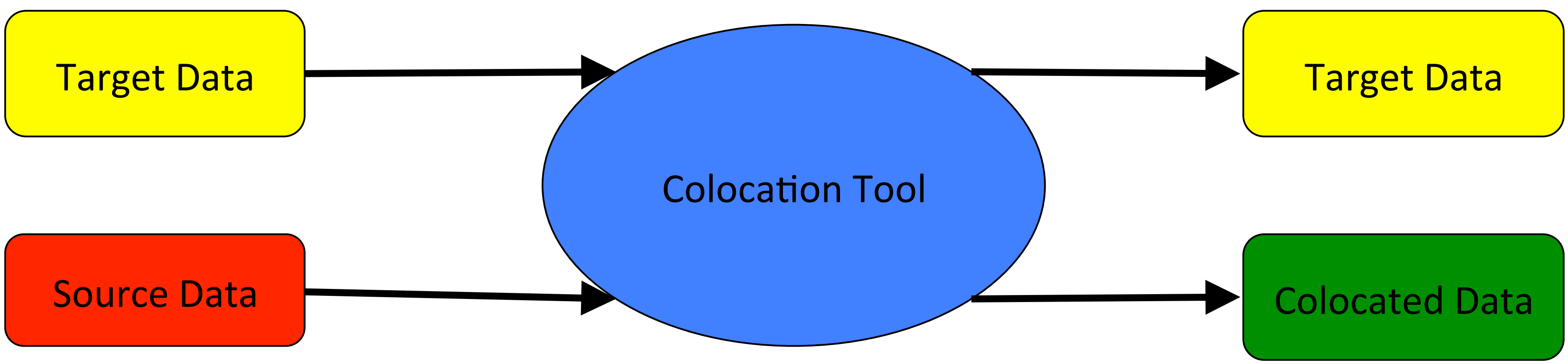
Scientific Use of Conditional Sampling



Cloud water content (cwc) vs vertical velocity (wap) at 500hPa

Co-location of Datasets

- A-train constellation multi-sensor data offer global observational constraints for the Earth's weather and climate models.
- However, each instrument has a different resolution, sensitivity, and characteristics, which make the direction comparison, cross validation, and comprehensive use of the data difficult.
- We developed a tool to temporally and spatially co-locate a source data product with a given target data product.

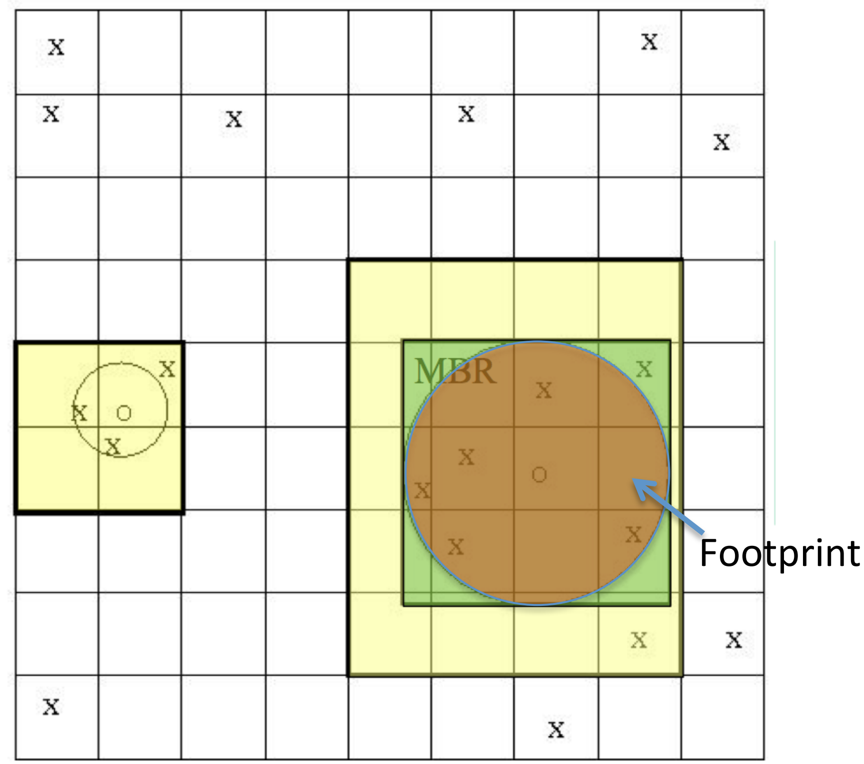


Co-location Algorithm

Nearest neighbor finding algorithm with the complexity of O(m+n), where n is the number of target points and m the number of source points.

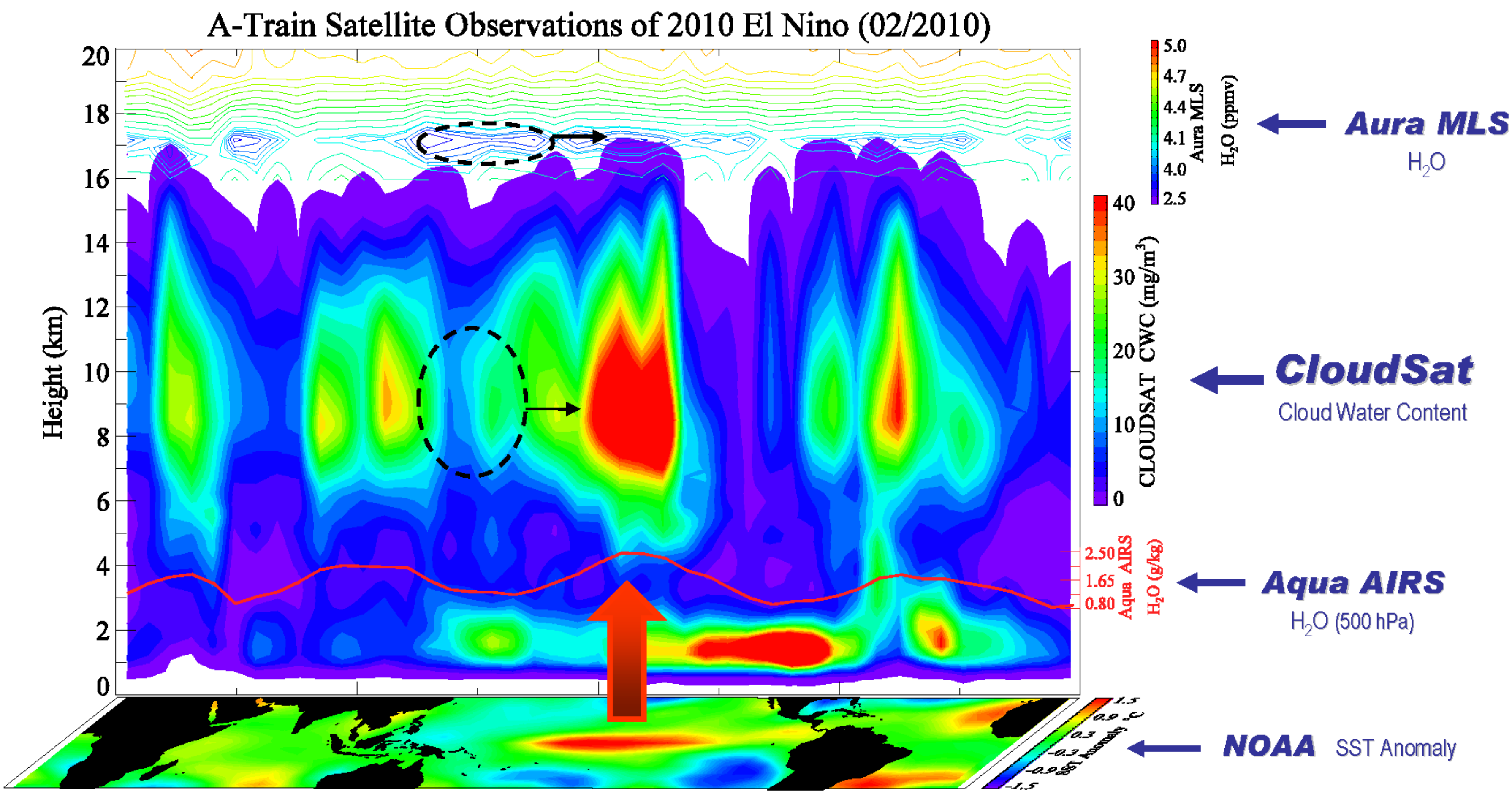
1 Overlay k by k grid on target points where k=sqrt(n) [O(k)].
2 Register each target point to a cell [O(n)].
3 For each source point S [O(m)],
   ① find minimum bounding rectangle (MBR) of its footprint.
   ② find grid cells that cover MBR.
   ③ For each grid cell G [O(G)], store S to T's neighbor list sorted by distance between T and S.
4 For each target point T [O(n)], compute average weighted by the distance among the L nearest neighbor points of T.

X: Target point
O: Source point



C: a constant bounded by max density of target points in MBR, which is not proportional to n.
When L=1, nearest neighbor; when L>1, average of L nearest neighbors

Scientific Use of Co-location



Collocated measurements from Aura MLS, CloudSat, Aqua AIRS and used to study the effect of El Nino event.

Web Services

- 2D Variable Map Service.
- 2D Variable Zonal Mean Service.
- 2D Variable Time Series Service.
- 3D Variable 2D Slice Service.
- 3D Variable Zonal Mean Service.
- 3D Variable Average Vertical Profile Service.
- Scatter Plot of Two Variables Service.
- Difference Plot of Two Variables Service.
- Conditional Sampling Service.
- Co-location Service.

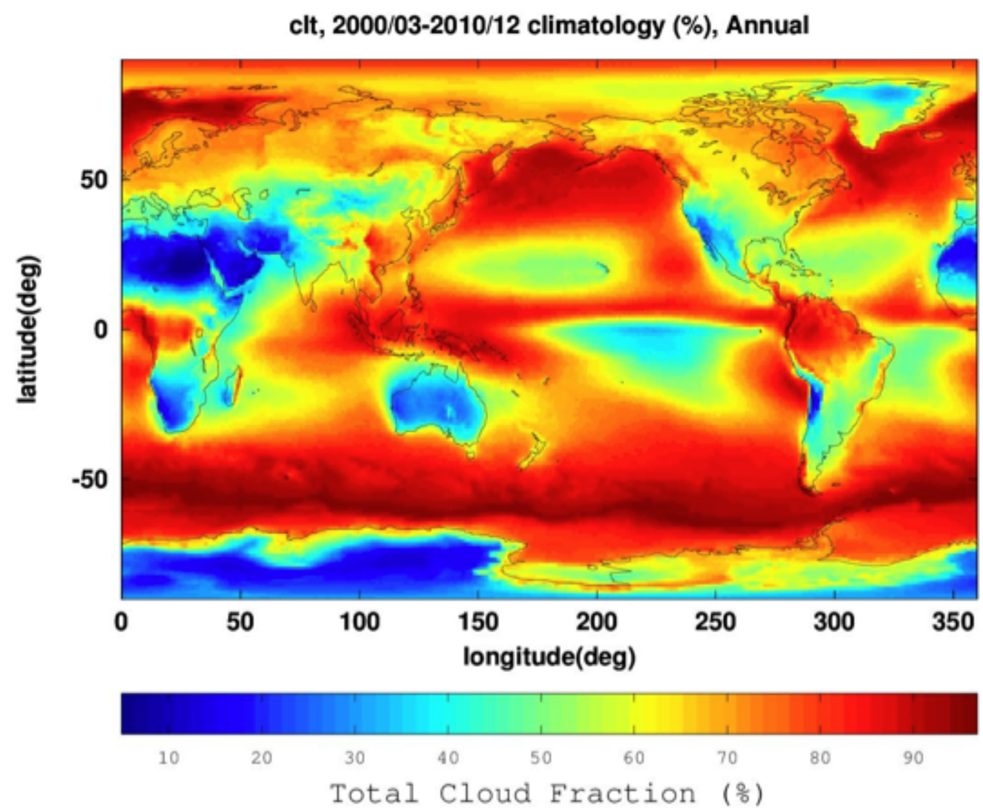
From our web server gate-way site:

http://cmacws4.jpl.nasa.gov:8080/cmac/

Service: 2-D Variable Map

It generates a map of a 2-dimensional variable with time averaging and spatial subsetting.
Select a data source (model or observation), a variable name, a time range, and a spatial range (lat-lon box) below.

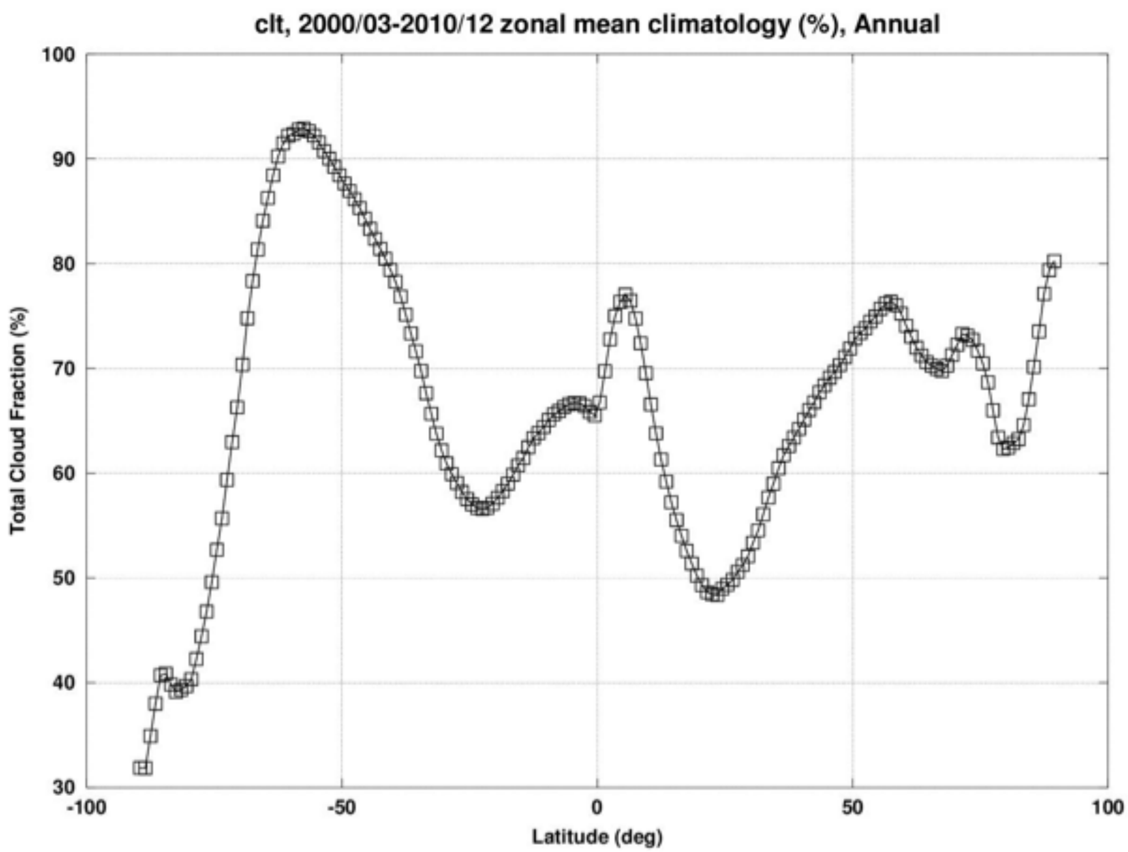
data source:	NASA/MODIS	variable name:	Total Cloud Fraction
start year-month:	2000-01	end year-month:	2010-12
select months:	select all		
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
color scale:	◉ linear ○ logarithmic		
Get Plot		Download Data	



Service: 2-D Variable Zonal Mean

It generates a graph of a 2-dimensional variable's zonal mean with time averaing.
Select a data source (model or observation), a variable name, and a time range below.

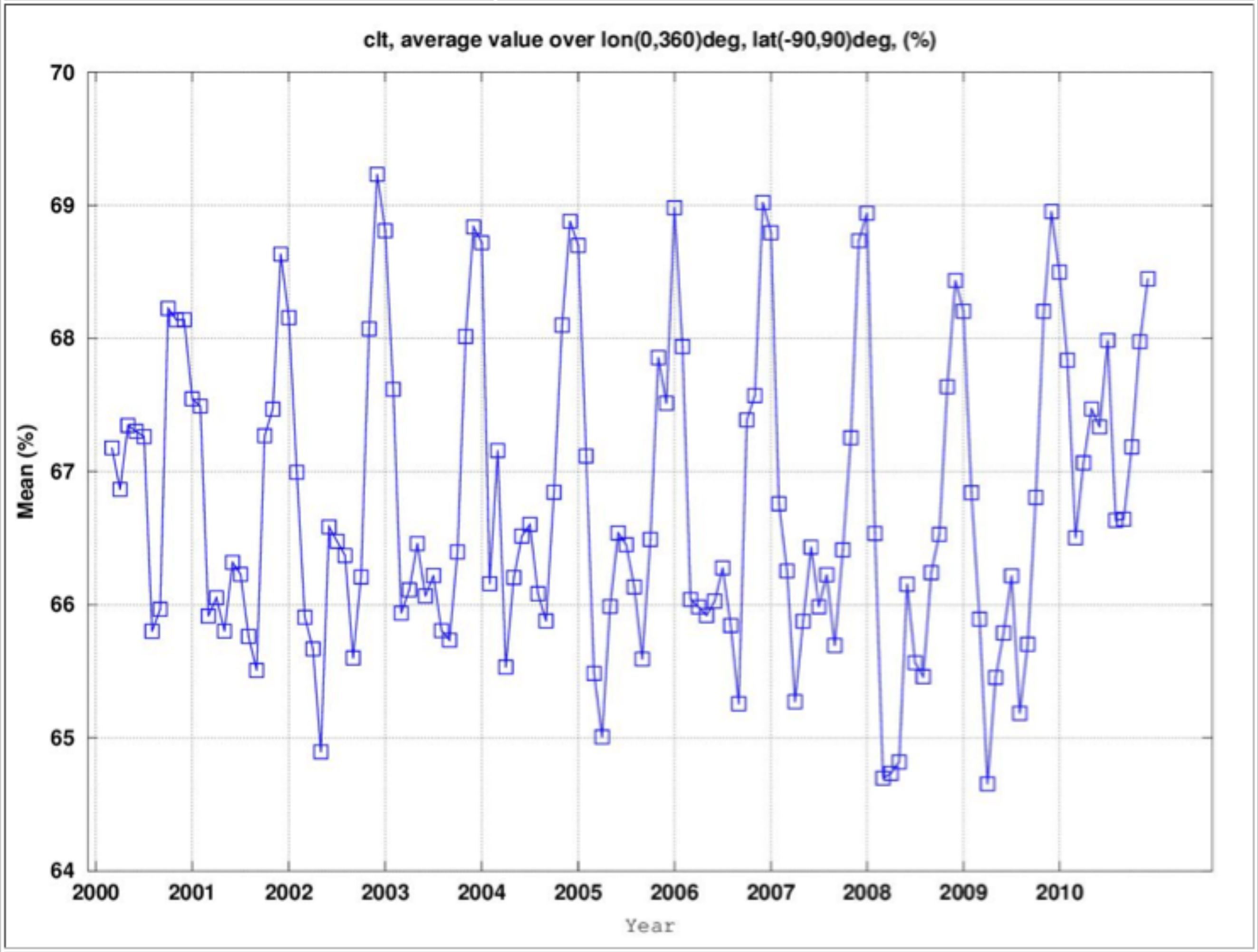
data source:	NASA/MODIS	variable name:	Total Cloud Fraction
start year-month:	2000-01	end year-month:	2010-12
select months:	select all		
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
variable scale:	⊙ linear ○ logarithmic		
Get Plot		Download Data	



Service: 2-D Variable Time Series

It generates a graph of a 2-dimensional variable's time series with monthly averaged values.
Select a data source (model or observation), a variable name, a time range, and a spatial range below.

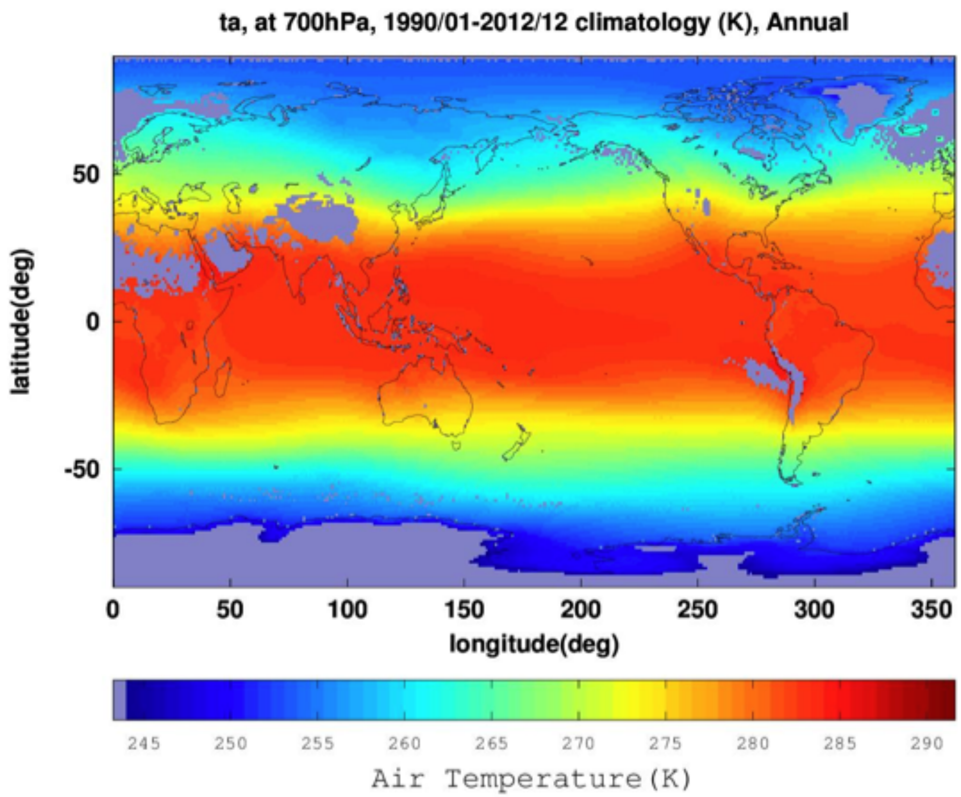
data source:	NASA/MODIS	variable name:	Total Cloud Fraction
start year-month:	2000-01	end year-month:	2010-12
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
variable scale:	linear logarithmic		
Get Plot		Download Data	



Service: 3-D Variable 2-D Slice

It generates a slice map of a 3-dimensional variable at a selected pressure level.
Select a data source, a variable name, a time range, a spatial range, and a pressure level below.

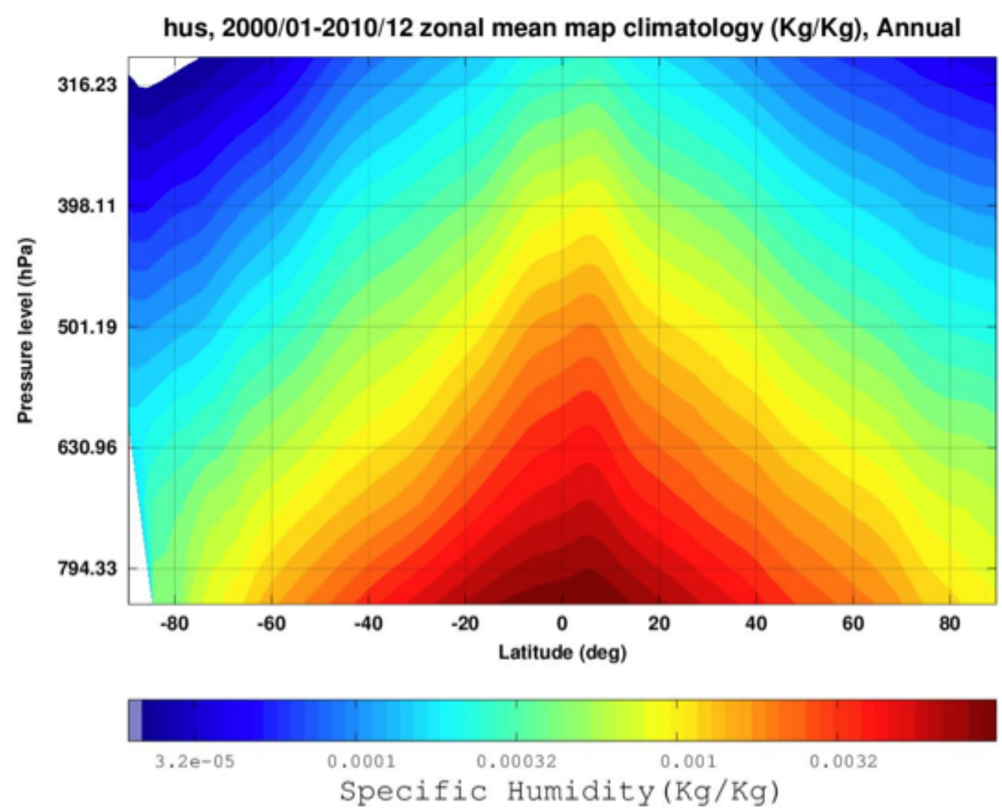
data source:	NASA/AIRS	variable name:	Air Temperature
start year-month:	1990-01	end year-month:	2012-12
select months:	select all	atmosphere pressure level (hPa) or ocean pressure level (dbar)	700
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
color scale:	⦿ linear ○ logarithmic		
Get Plot		Download Data	



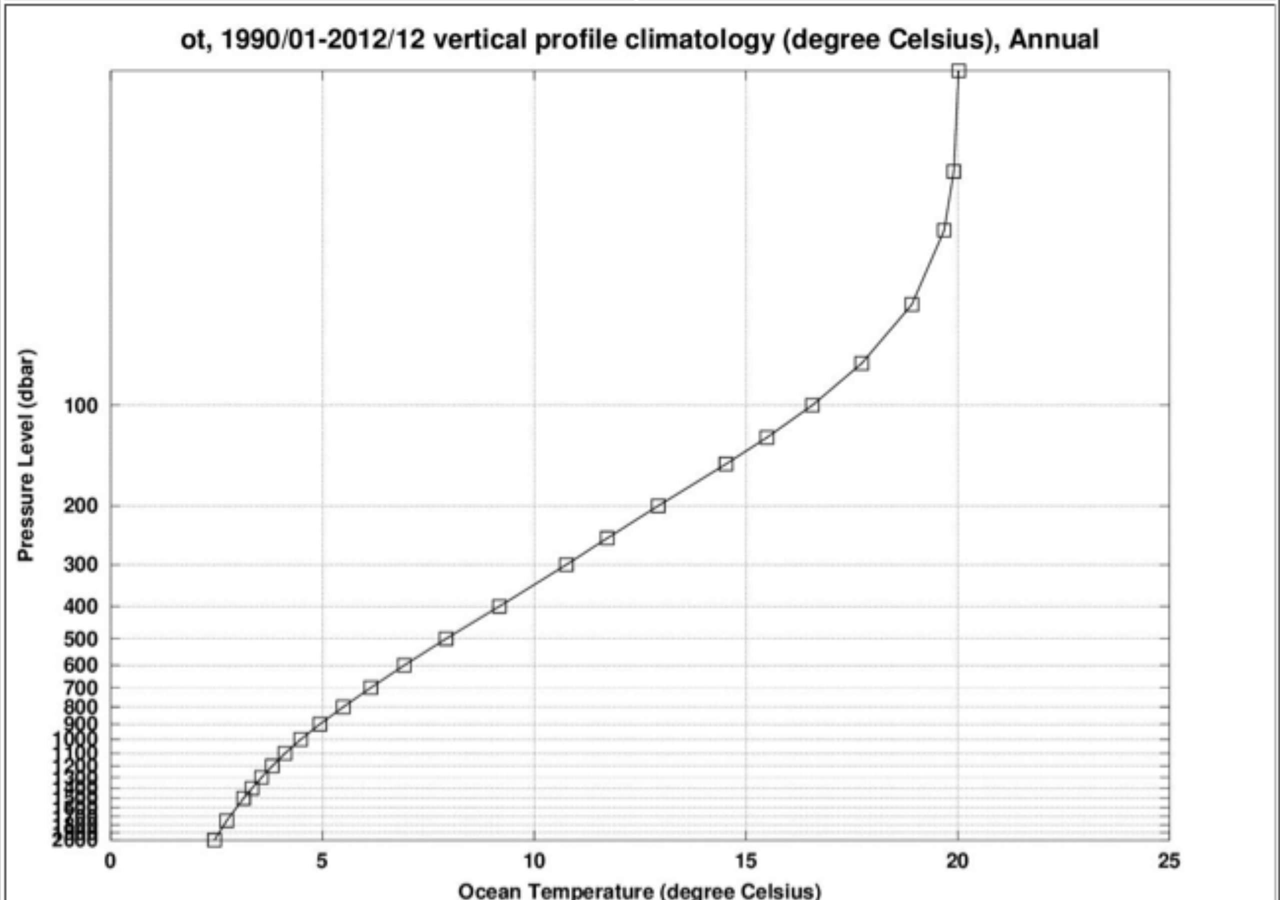
Service: 3-D Variable Zonal Mean

It generates a contour plot of zonal-mean vertical profiles of a 3-dimensional variable.
Select a data source, a variable name, a time range, and a pressure range below.

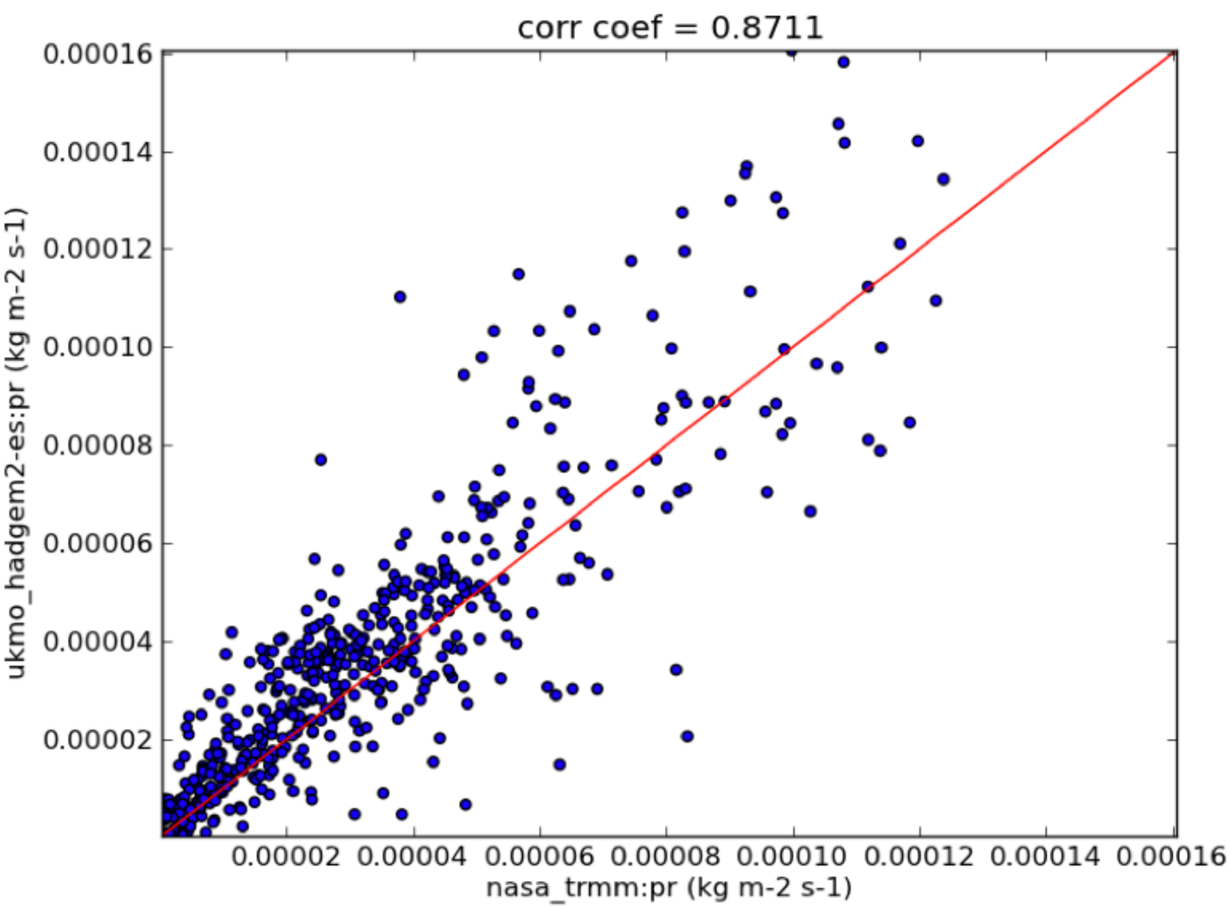
data source:	NASA/AIRS	variable name:	Specific Humidity
start year-month:	2000-01	end year-month:	2010-12
select months:	select all	atmosphere pressure range (hPa) or ocean pressure range (dbar):	900,200
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
pressure level scale:	○ linear ⦿ logarithmic	color scale:	○ linear ⦿ logarithmic
Get Plot		Download Data	



Service: 3-D Variable Average Vertical Profile			
data source:	ARGO/ARGO	variable name:	Ocean Temperature
start year-month:	1990-01	end year-month:	2012-12
select months:	select all		
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
pressure level scale:	○ linear ⦿ logarithmic	variable scale:	⦿ linear ○ logarithmic
Get Plot		Download Data	

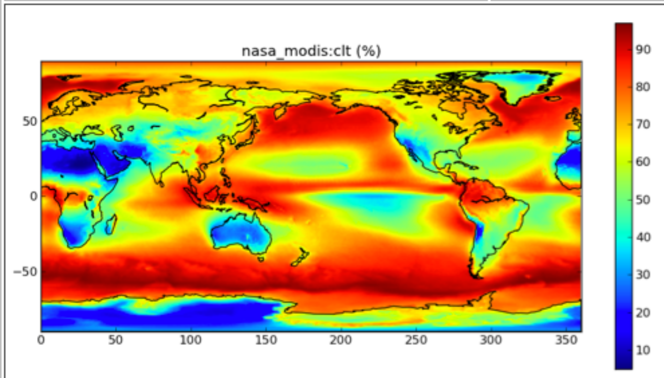


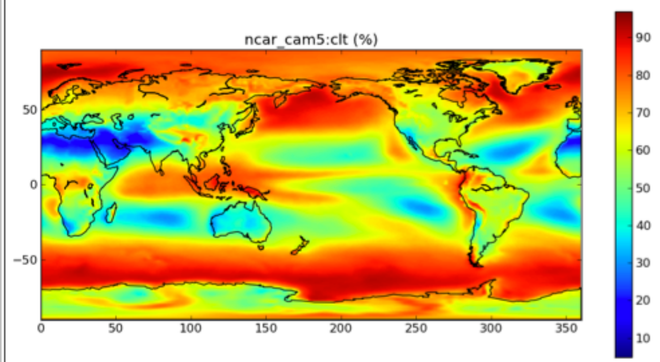
Service: Scatter Plot of Two Time Averaged Variables			
data 1		data 2	
source:	NASA/TRMM	source:	UKMO/HadGEM2-ES
variable name:	Precipitation Flux	variable name:	Precipitation Flux
atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A	atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A
start year-month:	2000-01	end year-month:	2012-12
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
Get Plot		Download Data	

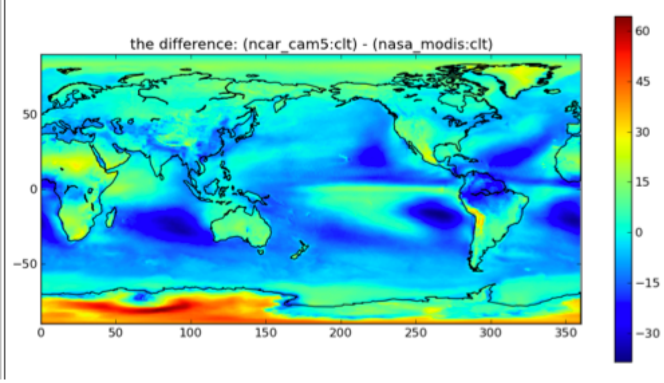


Service: Difference Plot of Two Time Averaged Variables

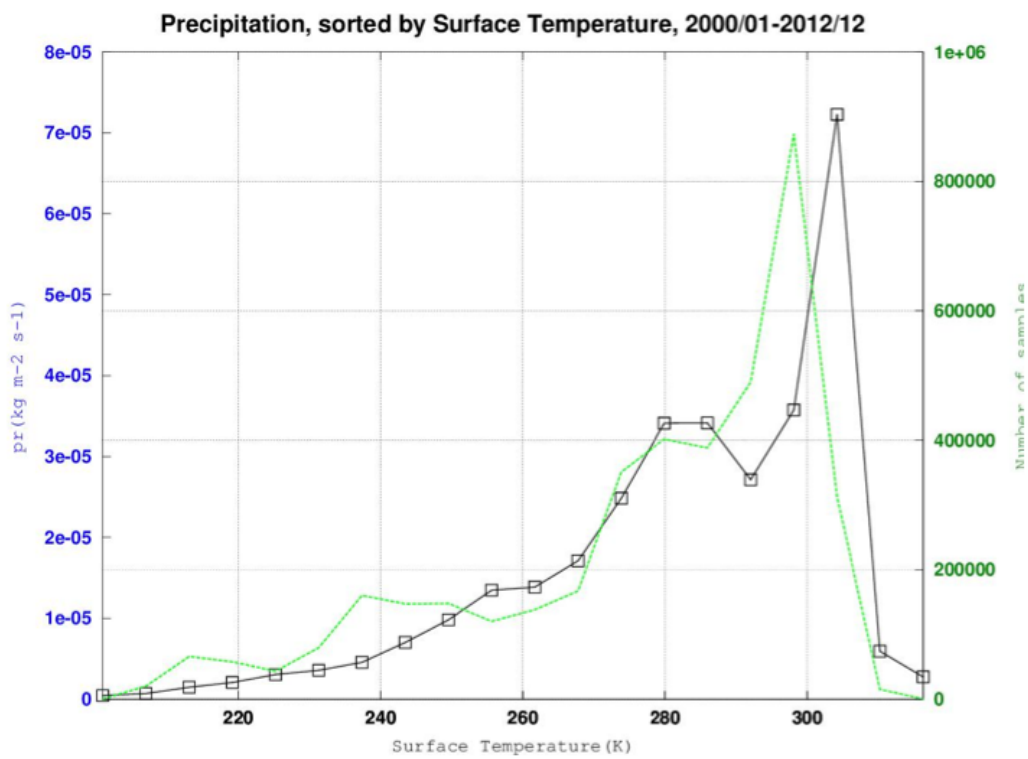
data 1		data 2	
source:	NASA/MODIS	source:	NCAR/CAM5
variable name:	Total Cloud Fraction	variable name:	Total Cloud Fraction
atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A	atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A
start year-month:	2000-01	end year-month:	2012-12
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
Get Plot		Download Data	



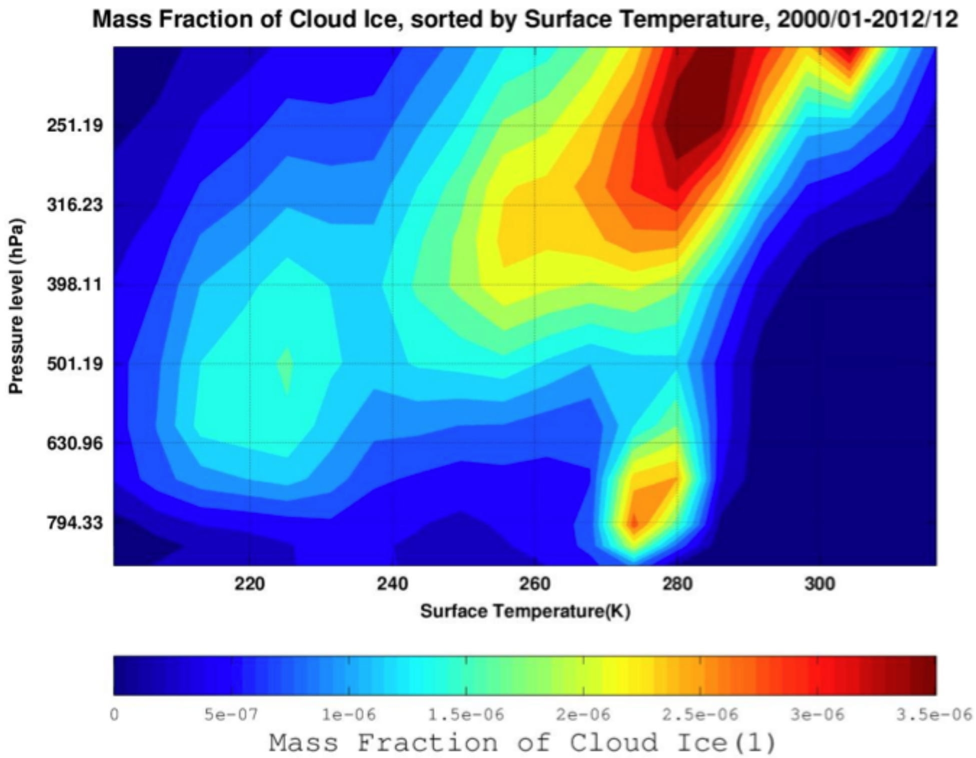


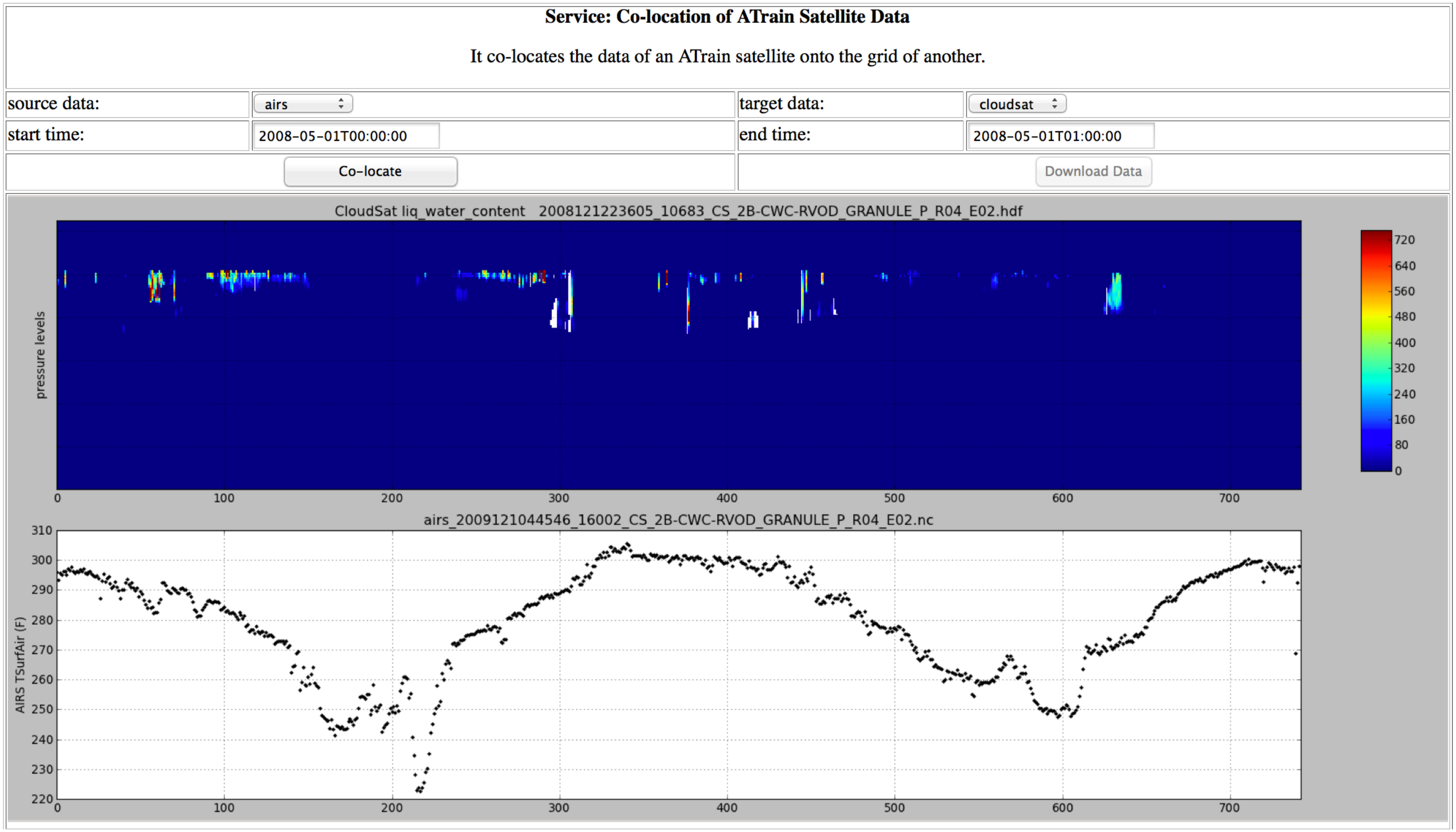


Service: Conditional Sampling			
Physical Variable (sampled variable)		Environmental Variable (sampling variable)	
source:	NCAR/CAM5	source:	NCAR/CAM5
variable name:	Precipitation Flux	variable name:	Surface Temperature
atmosphere pressure range (hPa) or ocean pressure range (dbar):	N/A	atmosphere pressure level (hPa) or ocean pressure level (dbar):	N/A
start year-month:	2000-01	end year-month:	2012-12
select months:	select all		
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
display option:	X-axis (sampling variable) scale: ⦿ linear ○ logarithmic		
	Y-axis (sampled variable or pressure) scale: ⦿ linear ○ logarithmic		
	Z-axis (color) scale: ⦿ linear ○ logarithmic		
sampling variable binning specification:	default: ⦿		
	customized: ○ Min: N/A Max: N/A number of bins: N/A		
Get Plot		Download Data	



Service: Conditional Sampling			
Physical Variable (sampled variable)		Environmental Variable (sampling variable)	
source:	NCAR/CAM5	source:	NCAR/CAM5
variable name:	Cloud Ice Water Content	variable name:	Surface Temperature
atmosphere pressure range (hPa) or ocean pressure range (dbar):	200, 900	atmosphere pressure level (hPa) or ocean pressure level (dbar):	N/A
start year-month:	2000-01	end year-month:	2012-12
select months:	select all		
☑ Jan ☑ Feb ☑ Mar	☑ Apr ☑ May ☑ Jun	☑ Jul ☑ Aug ☑ Sep	☑ Oct ☑ Nov ☑ Dec
start lat (deg):	-90	end lat (deg):	90
start lon (deg):	0	end lon (deg):	360
display option:	X-axis (sampling variable) scale: ⦿ linear ○ logarithmic		
	Y-axis (sampled variable or pressure) scale: ○ linear ⦿ logarithmic		
	Z-axis (color) scale: ⦿ linear ○ logarithmic		
sampling variable binning specification:	default: ⦿		
	customized: ○ Min: N/A	Max: N/A	number of bins: N/A
Get Plot		Download Data	

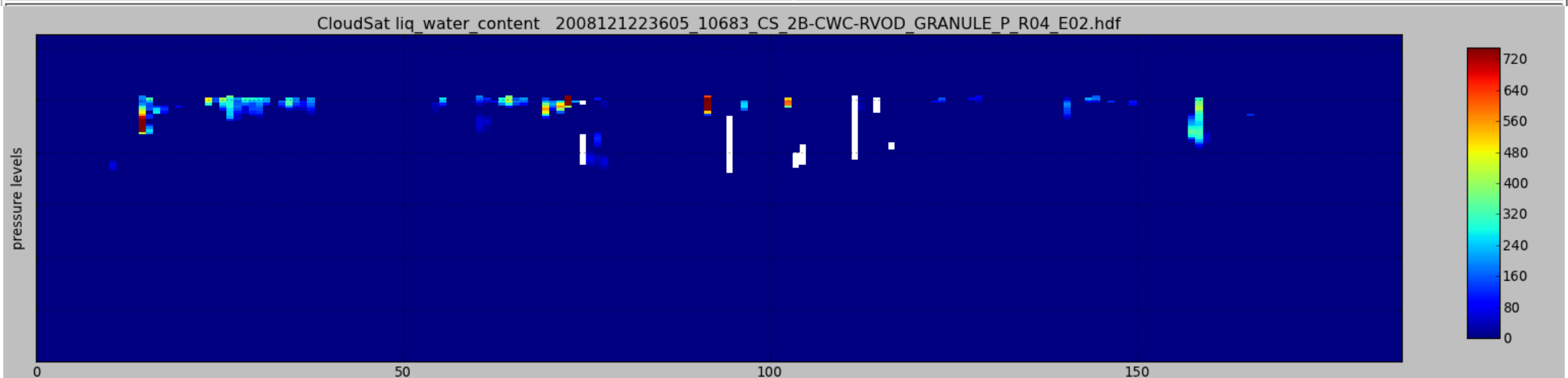


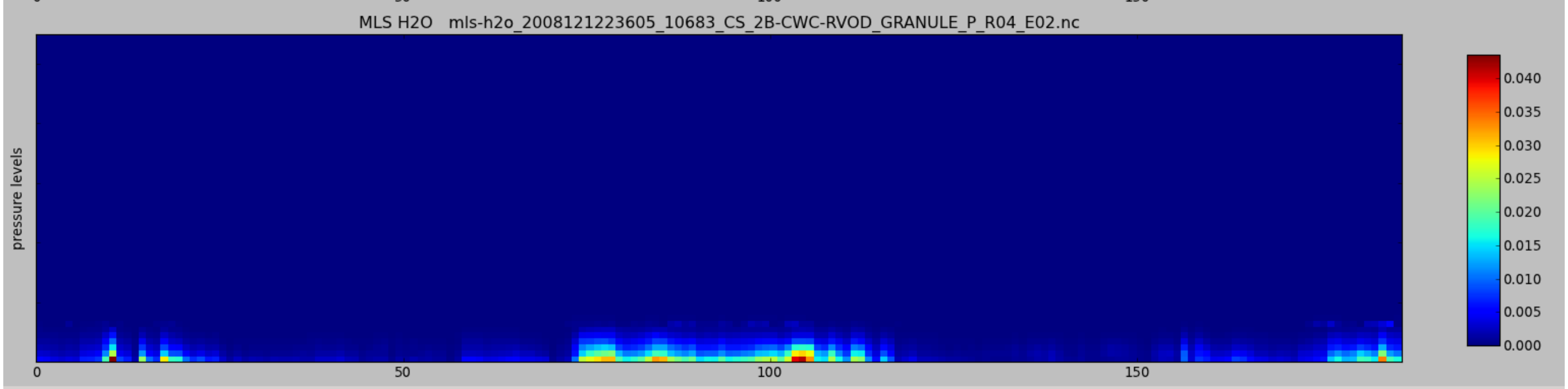

Service: Co-location of ATrain Satellite Data
It co-locates the data of an ATrain satellite onto the grid of another.
source data:
airs
target data:
cloudsat
start time:
2008-05-01T00:00:00
end time:
2008-05-01T01:00:00
Co-locate
Download Data
CloudSat liq_water_content 2008121223605_10683_CS_2B-CWC-RVOD_GRANULE_P_R04_E02.hdf
pressure levels
airs_2009121044546_16002_CS_2B-CWC-RVOD_GRANULE_P_R04_E02.nc
AIRS TSurfAir (F)

Service: Co-location of ATrain Satellite Data

It co-locates the data of an ATrain satellite onto the grid of another.

source data:	mls-h2o	target data:	cloudsat
start time:	2008-05-01T00:00:00	end time:	2008-05-01T01:00:00
Co-locate		Download Data	

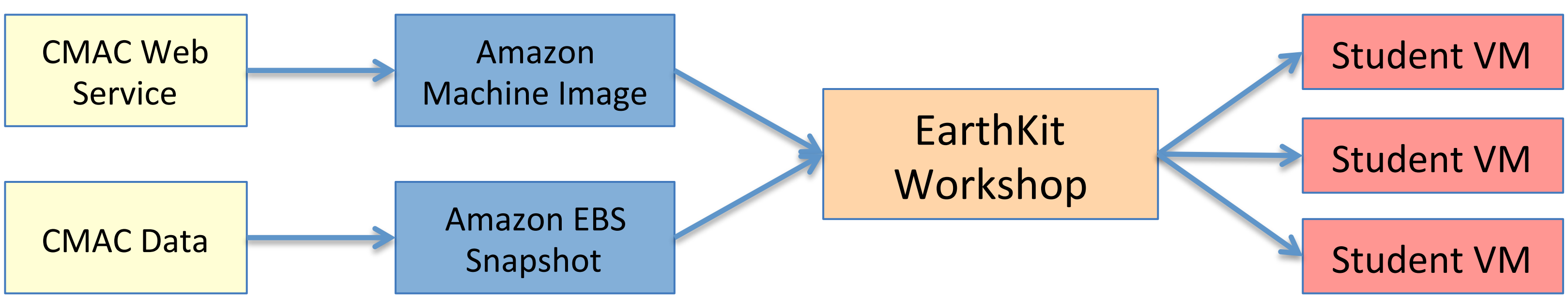




Cloud Computing

- Integrated the web service system into Amazon cloud environment using EarthKit.
  - Adapted existing EarthKit workshop platform.
  - Simplied UI by removing in-browser tutorial and SSH console.
  - Streamlined platform for dynamically launching virtual machines (VMs) for students.
  - Integrated CMAC web service system into Earthkit workshop platform.
  - Deployed CMAC web service into Amazon EC2 image.
  - Created CMAC data snapshot for use during the workshop.

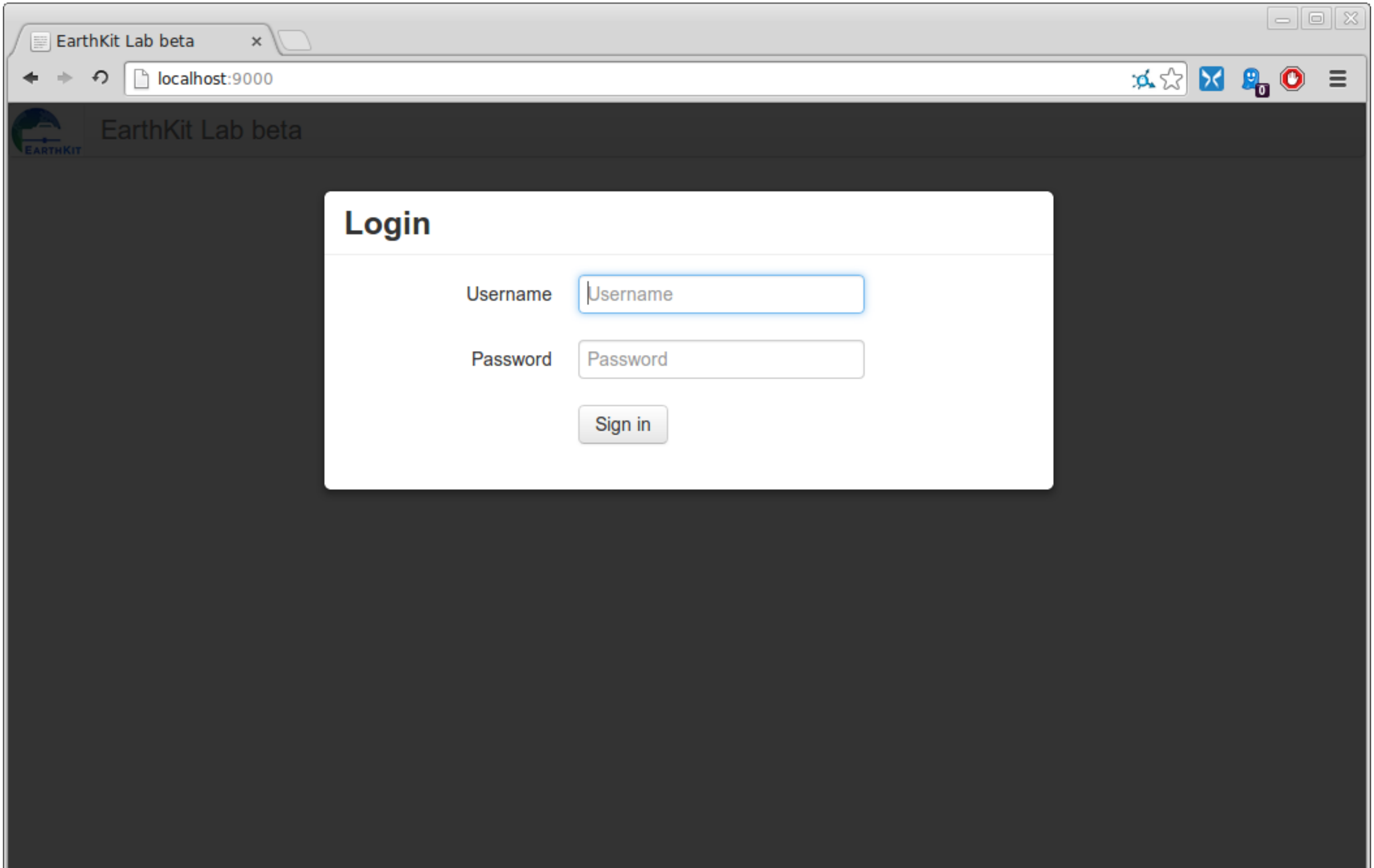
Overview of Architecture



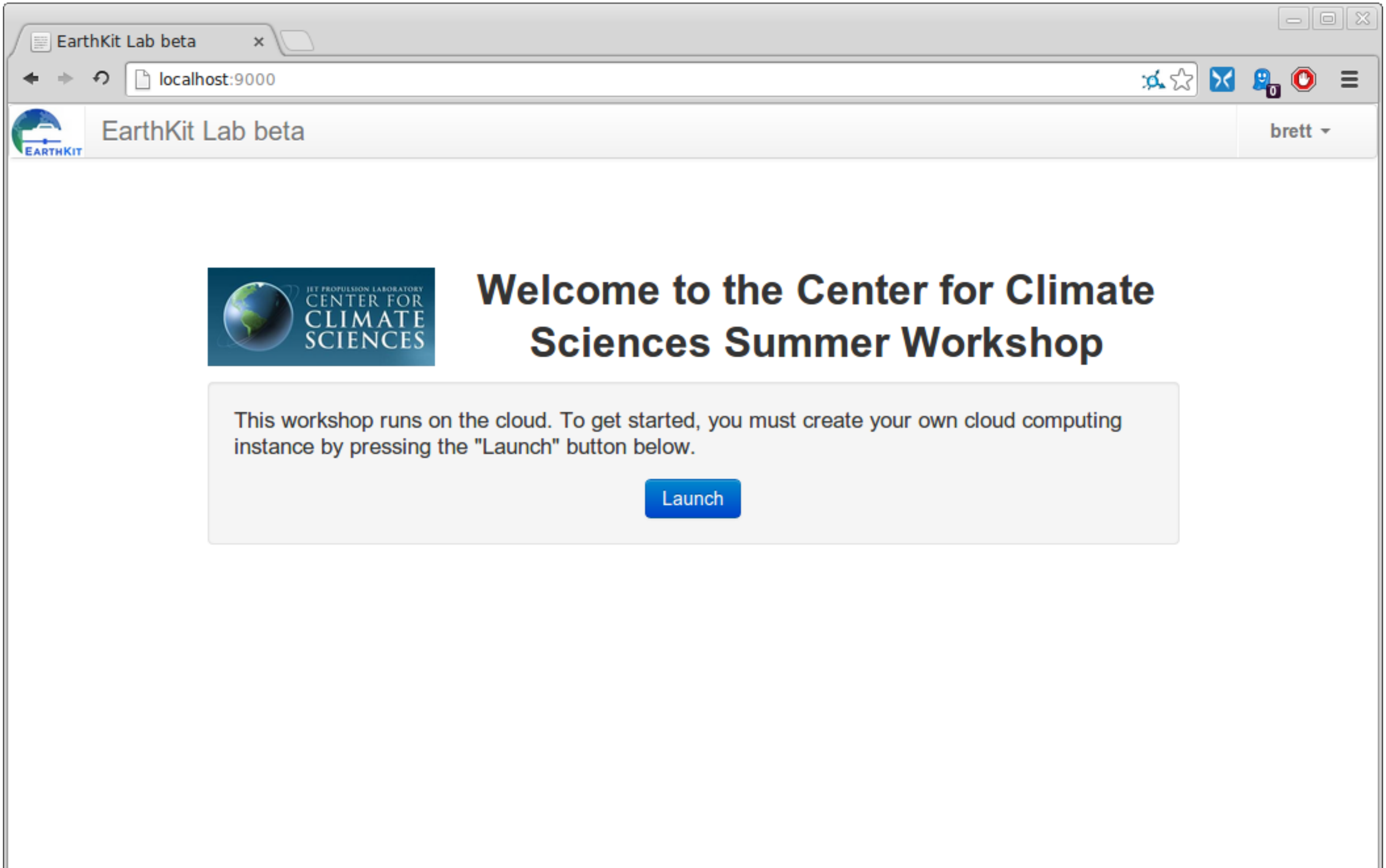
Cloud Computing Demo

http://climate.earthkit.jpl.net

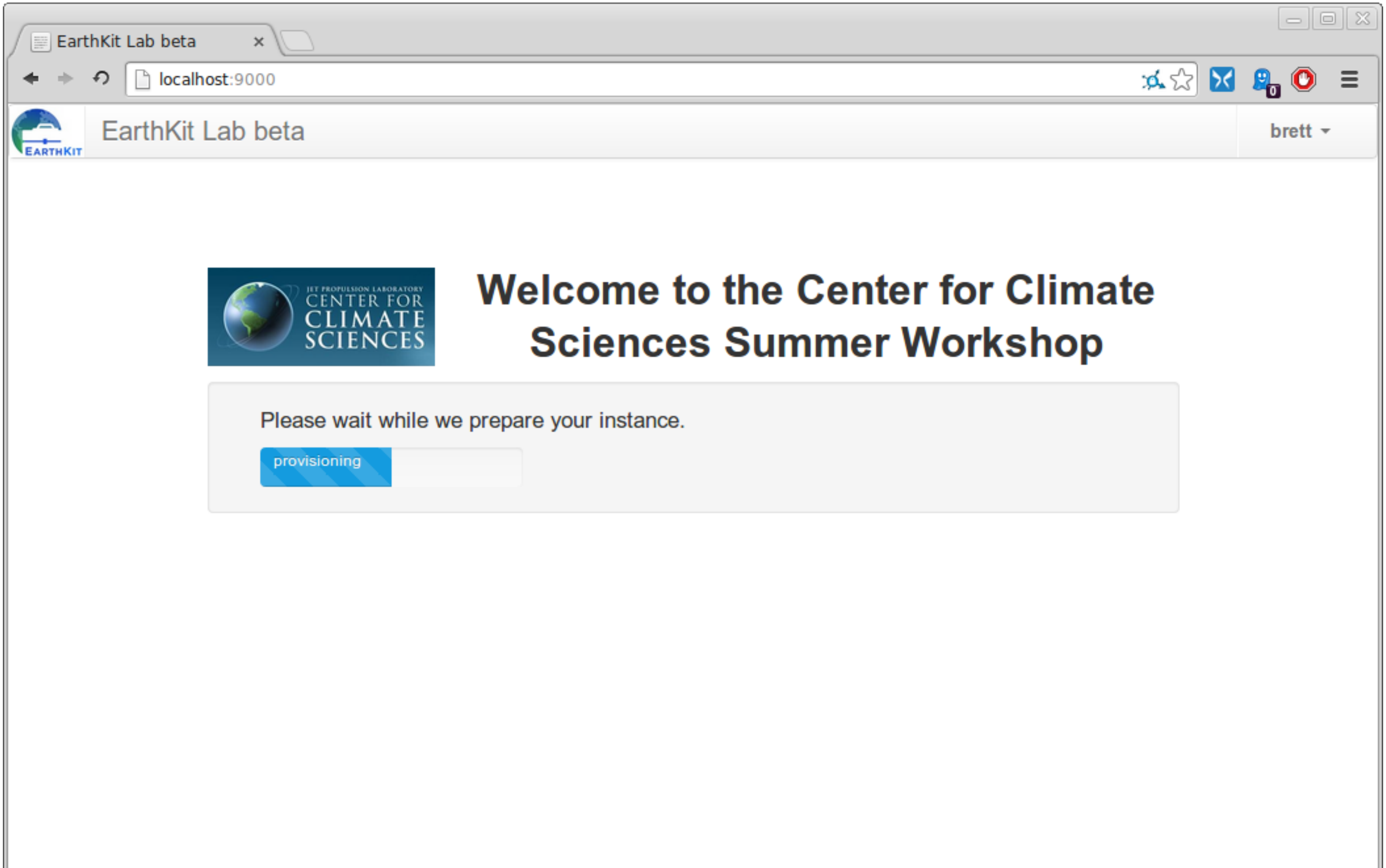
1. Login
2. Welcome screen with cloud launcher
3. Launching: provisioning stage
4. Launching: booting stage
5. Cloud computing ready to start
6. Landing page for climate summer school tool
7. Select analysis web service
8. Select parameters for the analysis
9. View analysis result
10. Cloud computing shutting down


EarthKit Lab beta
localhost:9000
EARTHKIT
EarthKit Lab beta
Login
Username
Username
Password
Password
Sign in

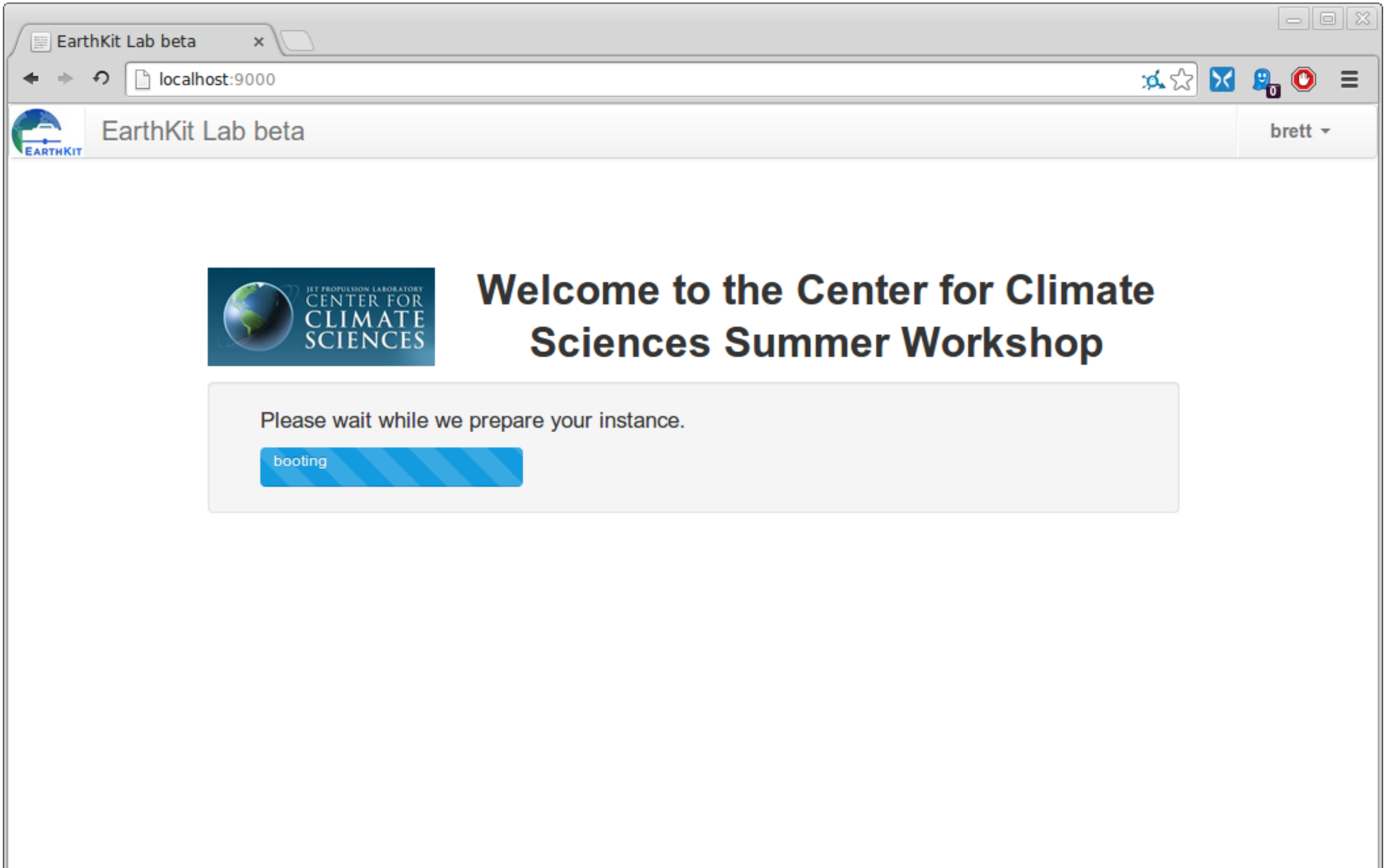
Login



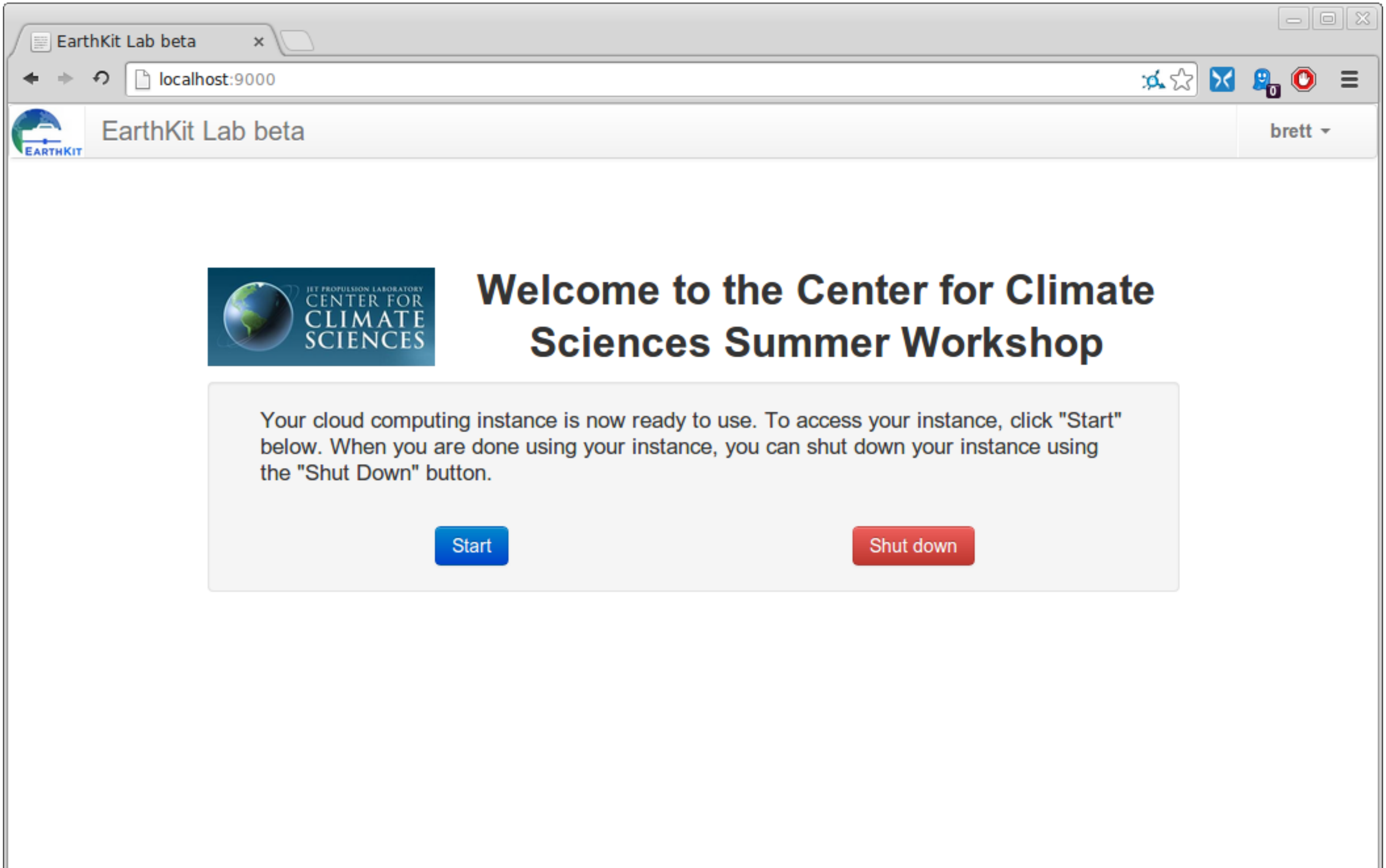
Welcome screen with cloud launcher



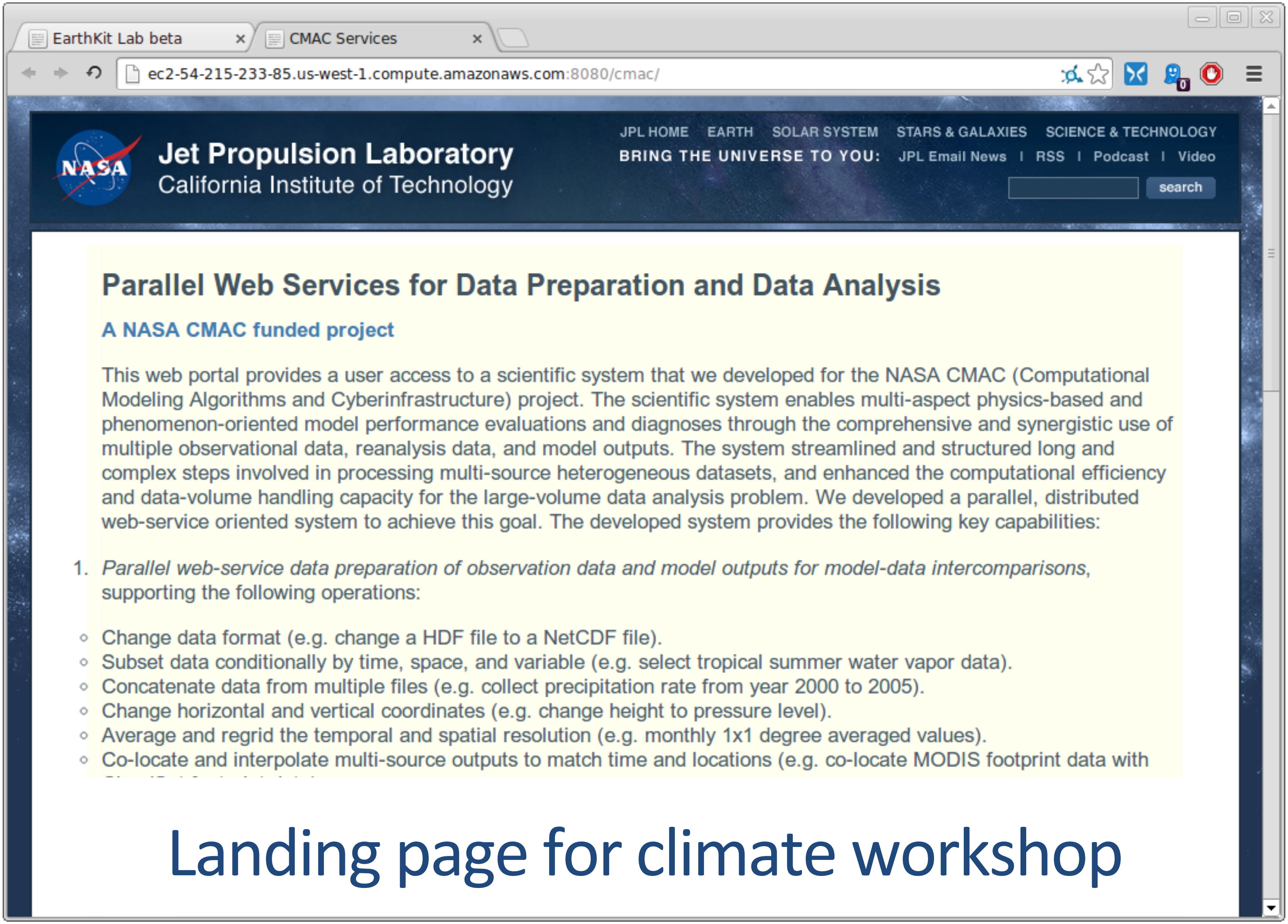
Launching: provisioning stage



Launching: booting stage

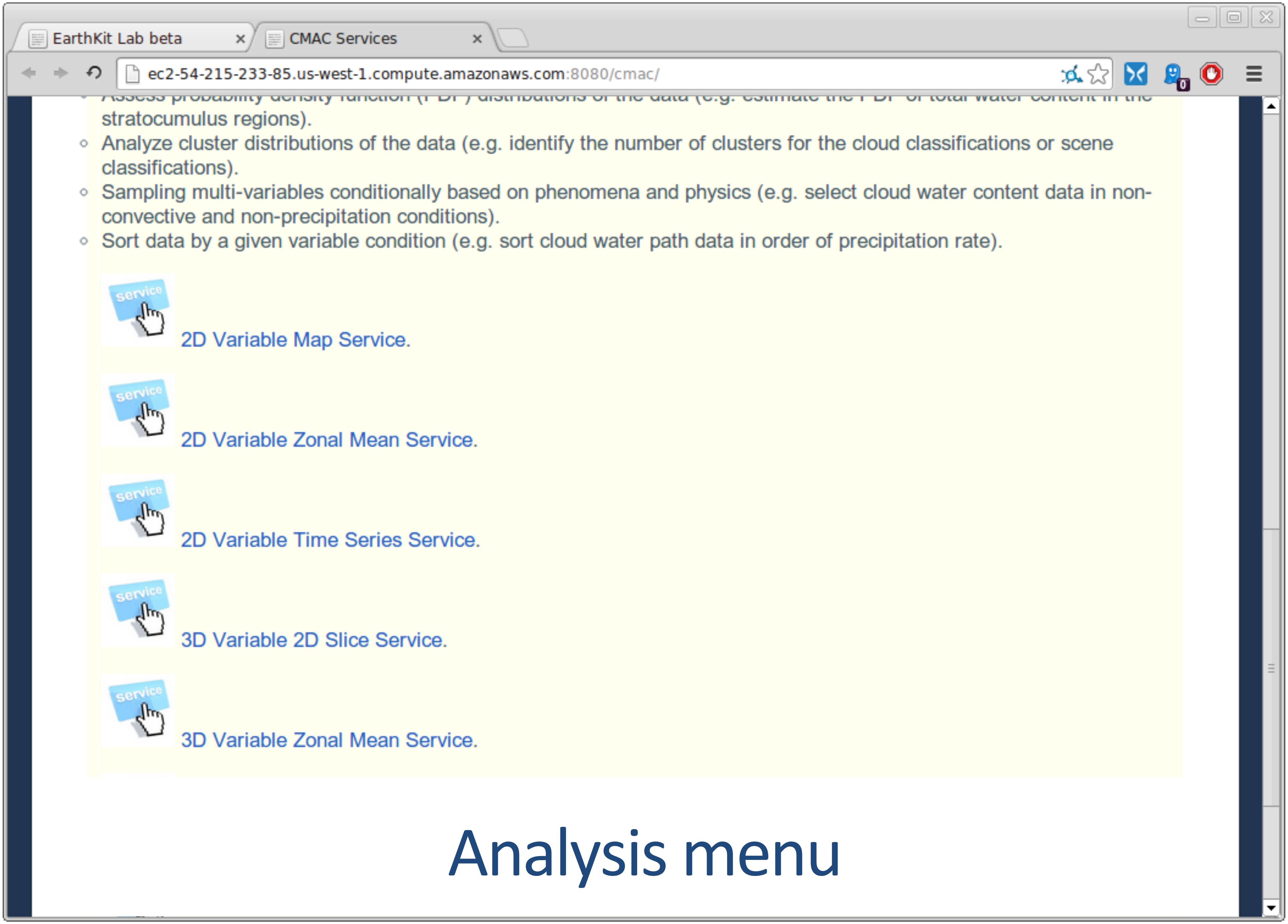


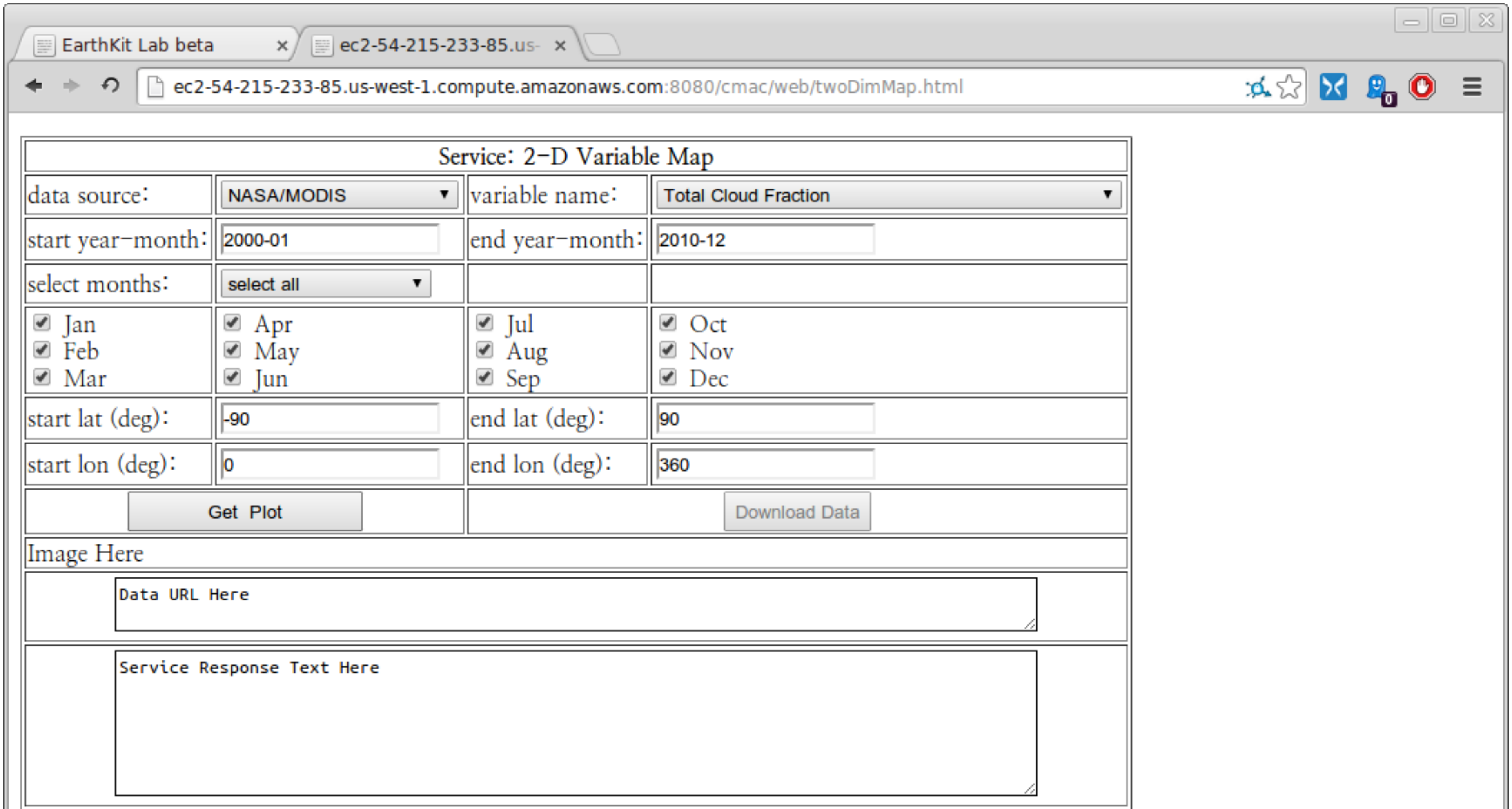
Cloud computer ready to start



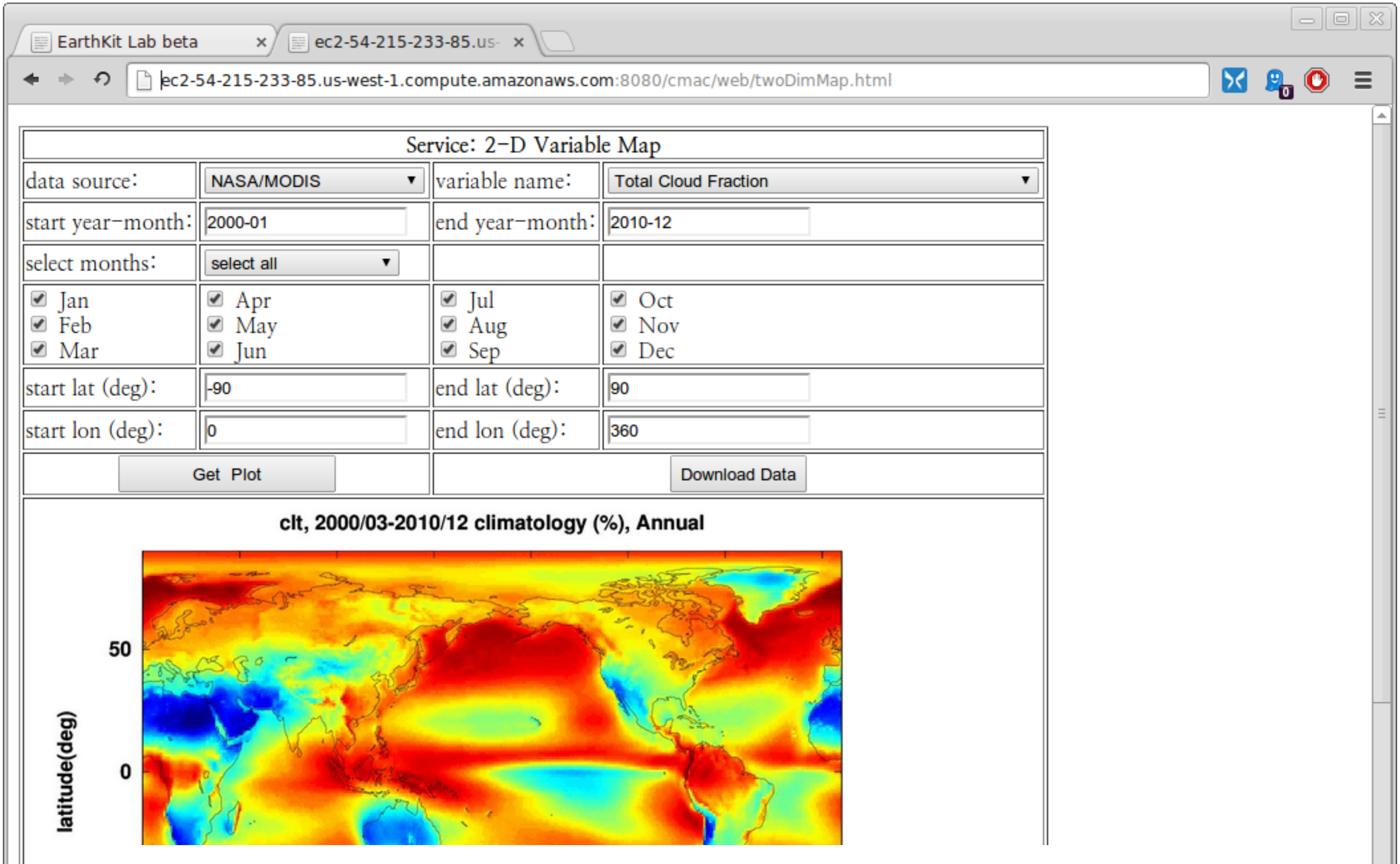
Landing page for climate workshop

Analysis menu

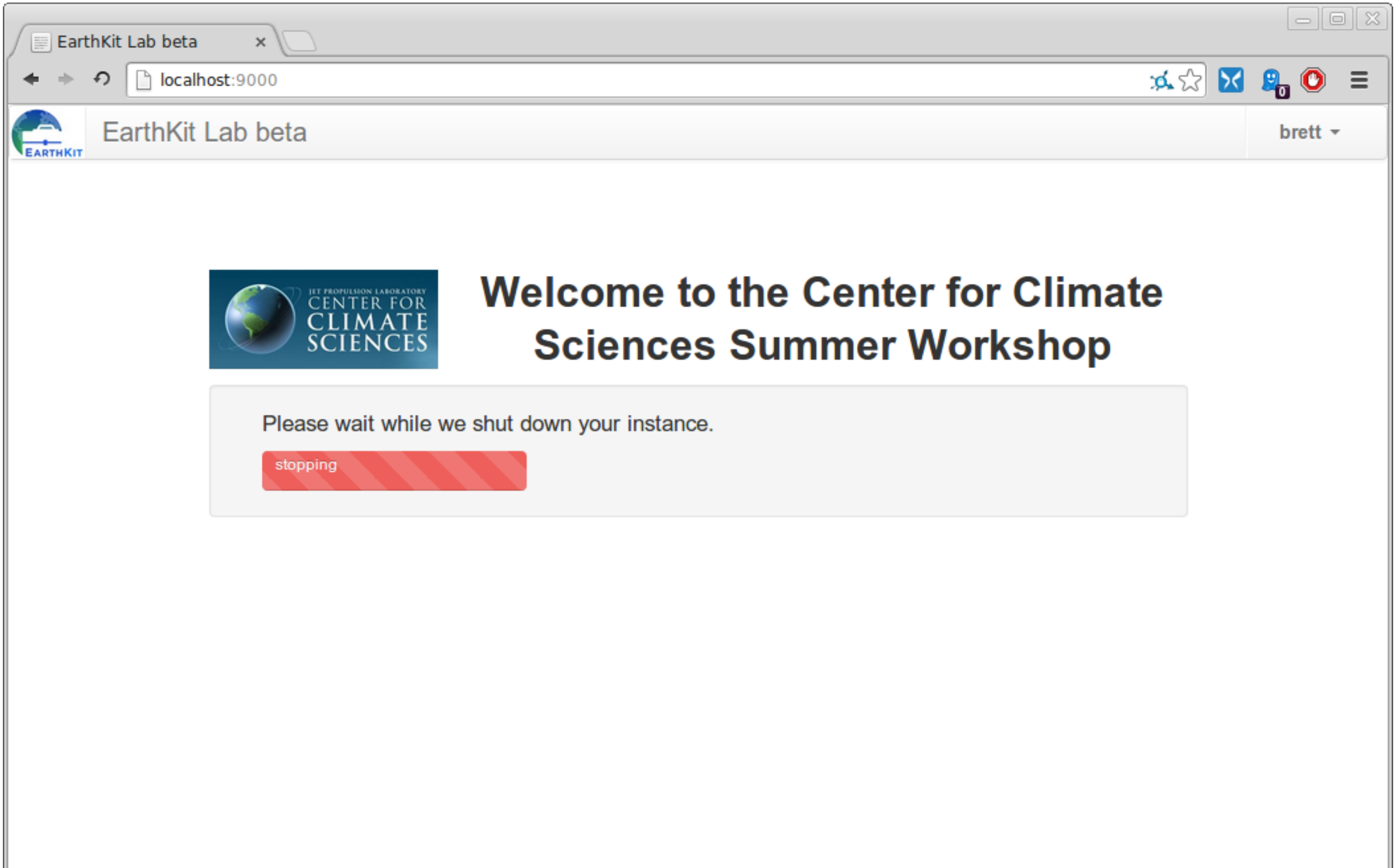




Parameter page



Analysis visualization



Cloud computer shutting down

Planned Technical Work

- Improve the functionality and dataset support of co-location web service.
- Develop a regridding web service.
- Apply Parallel Python where appropriate.
- Integrate the web-service system for multi-step analyses (e.g. regridding and conditional sampling, co-location and averaging).
- Deploy the latest version of the web-service system and dataset into Amazon EC2 image.

Outline

- Technical Status
- **Programmatic Status**
- Educational and Outreach Status

Programmatic Status Summary

- This CMAC project started on October 1st, 2012.
- This project has 9 tasks, 4 deliverables, and 3 milestones.
- The schedule/order of the tasks is adjusted to improve the development cycle and support the new educational opportunity.
- 6 tasks are completed, 3 deliverables are completed, and 2 milestones are met.
- This project total budget is $616K with $303K for year 1 and $313K for year 2.
- $165K (53% of year 2 budget) is spent as of April 3, 2014.
- The actual cost is on track with the planned budget.

Task Status

- Completed:
  - Developed a python-drive data processing system.
  - Developed a web service for python driven data processing system.
  - Developed a python-drive data analysis system.
  - Developed a web service for python driven data analysis system.
  - Demonstrated the scientific use of the data processing system.
  - Demonstrated the scientific use of the data analysis system.
  - Integrated the web-service system into the Amazon Cloud system.

- Planned:
  - Demonstrate the educational use of the web-service system.
  - Integrate the web-service system for multi-step analyses.
  - Apply Parallel Python where appropriate (e.g. Co-location).

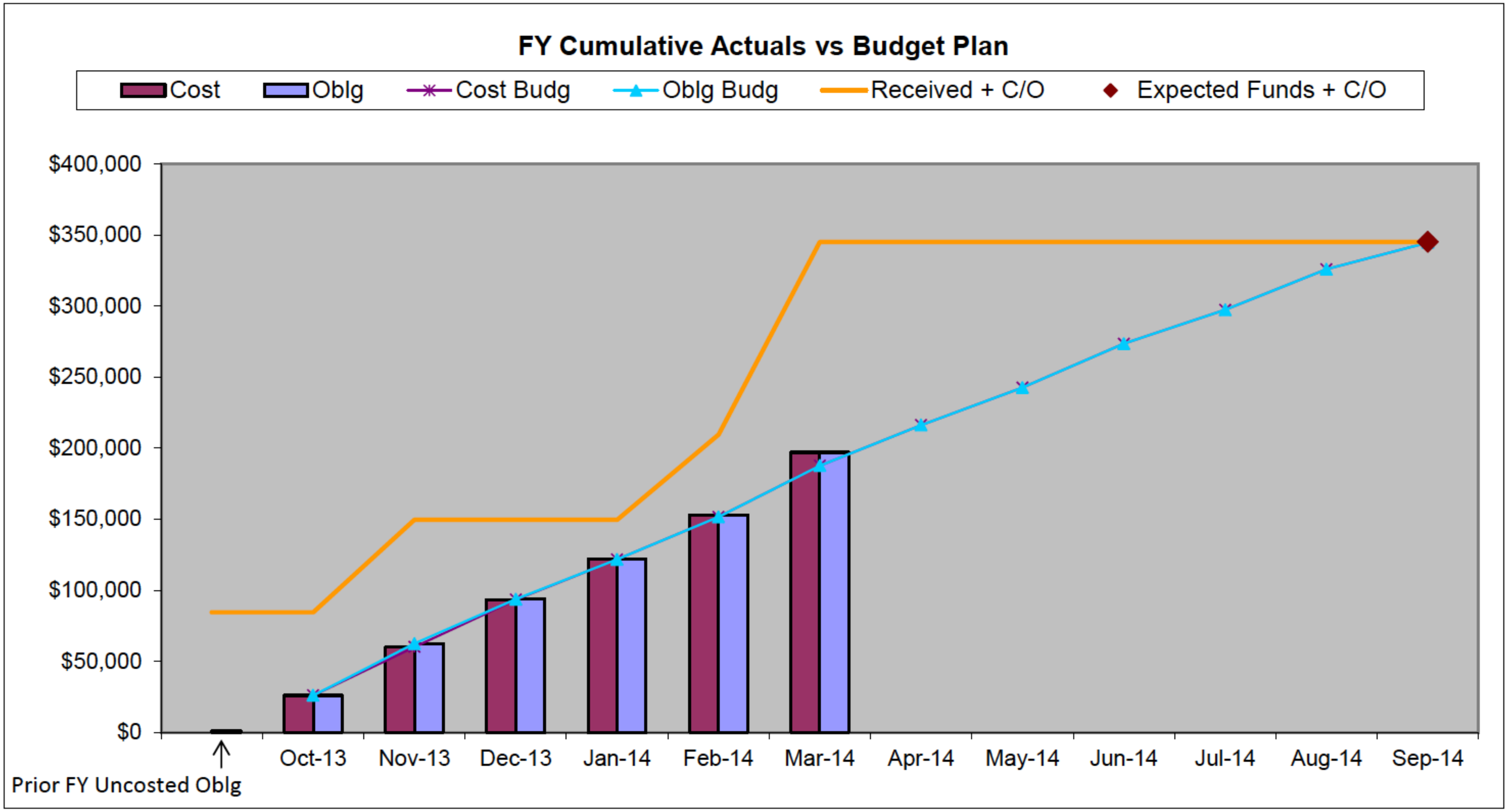
Deliverable/Milestone Status

- Completed:
  - Delivered an Interim Project Report in Year 1 (D1).
  - Delivered an Annual Project Report in Year 1 (D2).
  - Delivered an Interim Project Report in Year 2 (D3).
  - Implemented the data processing system (MS 1).
  - Implemented the data analysis system (MS 2).

- Planned:
  - Deliver a Final Project Report (D4).
  - Implement integrated data processing/analysis system (MS 3).

Project Cost Status

- Total project budget for Year 2 is $313K.
- As of the end of March 2014, the project spent $165K (53%) and is on track with the planned budget.



Outline

- Technical Status
- Programmatic Status
- **Educational and Outreach Status**

Educational and Outreach Status

- The Center for Climate Sciences (CCS) plans to host a summer school in 2014.
- CCS is working on obtaining the JPL and NASA approval for the summer school in 2014.
- The summer school will bring together the next generation of climate scientists – about 30 graduate students and postdocs from around the world – to engage with premier climate scientists from JPL and elsewhere.
  http://climatesciences.jpl.nasa.gov/events/summer-school

- This CMAC project demonstrated the capability of the tool to CCS in April, 2013.
- CCS will use the PAWS-CMDA tool for its summer school in 2014.
- We designed requirements for the education tool: the user interface should be web-browser based; the web server or web system should support >30 simultaneous web-service requests.
- Several web-services to support model-data intercomparisons are developed: analysis and visualization of 2D/3D variable multi-annual means using Coupled Model Intercomparison Project Phase 5 (CMIP5) model outputs and Obs4MIPs datasets.
- Multi-variable analysis web-services are developed: correlation, difference, conditional sampling.
- Collocation of observational Level 2 datasets is developed as a web service.
- New datasets are introduced and integrated into the web service system.
- The web service system is imported and integrated into the Amazon cloud environment.

CCS Summer School Tool & Curriculum

- The CCS summer school education tool will be a web-browser based tool.
- The tool will support the following functionalities:
  - 2D/3D variable map of annual means with spatial and temporal selection
  - 2D/3D variable zonal means with temporal selection
  - 2D variable time series with spatial and temporal selection
  - 3D variable vertical profile with spatial and temporal selection
  - Two 2D/3D variable scatter plot and correlation analyses with spatial, temporal and vertical level selection.
  - Two 2D/3D variable difference plot with spatial, temporal and vertical level selection.
  - Physical variable conditionally sampled by another physical variable
  - Co-location of A-Train satellite data sets.
- The tool will be used to support the following curriculum:
  - Learn the aspects of both model outputs and NASA satellite observations.
  - Compare the model outputs with the observation datasets.
  - Consider the effect of the observation uncertainty and limitation on the model comparison.
  - Report model systematic biases if exist.
  - Examine the potential causes of the model biases.

Backup Slides

Project Machine Purchase and Installation

- We purchased a machine dedicated for this project at the cost of $30K in January 2013.
- The project machine consists of
  - Two servers, each having 12 cores, 64 GB physical memory, and 4 TB disk space.
  - One RAID storage server with 100 TB usable space, 12 cores, and 64 GB physical memory.
  - One Gigabit switch to connect the three servers.
- Machine installation
  - CentOS Linux
  - RAID server made head node (JPL LAN IP)
  - Three servers made compute nodes (local IP: 192.xxx.xxx.xxx)
  - RAID cross mounted to all nodes
  - /home shared by all nodes
  - VMWare player (free) installed for virtualization

Adapting Existing Analysis Tools

- Existing Matlab code is adapted to this project.
  - The adaptation only requires to develop interface routines that facilitate data passing between user and core functions.
- To meet the requirement of large scalability of PAWS-CMDA, we use GNU Octave, which is a license-free version of Matlab.
- A minimal adjustment of the existing Matlab code is needed to make it executable in Octave environment.
  - Data format difference in reading netcdf files
  - A few capabilities from some matlab toolboxes, e.g. coastal line for map, juliandate
- Current capabilities of the adapted Octave code:
  - 2D variable map
  - 2D variable time series
  - 2D zonal mean
  - 3D variable zonal mean
  - 3D vertical profiles
  - 3D pressure-level slice map

Developing New Tools

- Chose Python for new tool development since the rest of the system is Python driven.
- Computational performance and data access performance are sufficient with Python code.
- Developed new Python-driven data analysis tools
  - Two variable scatter plot and correlation analysis
  - Two variable difference plot

Developing Python-Driven System

- Wrap legacy data analysis tools for integration into the system.
  - Application code in GNU Octave (a free Matlab)
  - Use python caller code to spawn off the App. as a child process (using python subprocess)
  - The python caller prepares all input arguments for the child process
  - The python caller defines where the child process puts its results (e.g., images) using input argument
  - The stdout and stderr of child process are captured by the python caller at joining point (the way for the App. to communicate back to the python caller)
  - The legacy App. now looks like a python application

Building Web Services

- Use Flask, an open source microframework for python apps.
  - Provide entry point code
    - Parse input arguments from client
    - Call aforementioned python app., passing input arguments
    - Get return values and pass them to client in Json format
    - REST-ful style: scoping info. placed in URI and method info. conveyed in HTTP method
    - Example: http://cmacws.jpl.nasa.gov:8089/twoDimMap/display?model=ukmo_hadgem2-a&var=ts&start_time=199001&end_time=199512&lon1=0&lon2=100&lat1=-29&lat2=29&months=1,2,3
  - Define port number and location of static data
- Separate app. traffic from static HTTP traffic
  - Use Gunicorn to provide Web Server Gateway Interface (WSGI) service (app.)
  - Use Tornado to provide web service (static HTTP)

Building Web Browser Interface

- Designed web browser interface for our web services using JavaScript.
  - Provide an intuitive and easy access to the web services.

- Implemented the web browser interface for 8 web services.
  - 2D variable map display
  - 2D variable time series
  - 2D zonal mean plots.
  - 3D variable zonal mean
  - 3D vertical profiles
  - 3D pressure-level slice map
  - Two variable scatter plot and correlation analysis
  - Two variable difference plot

- Users can display plots and download the output result in NetCDF file.

Virtualizing Service

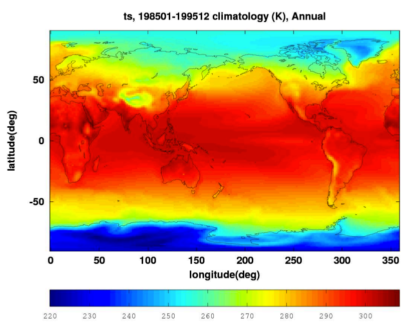
- Use VMWare (VMPlayer)
  - Host OS: CentOS (Linux)
  - Guest OS: Ubuntu (Linux)
  - Open source software, python packages, etc. installed on guest OS
  - Services deployed to guest OS
  - Developers' user accounts are on guest OS
  - RAID disk, with 100TB capacity, is mapped from host OS to guest OS
  - Host ports need to be open and forwarded to guest OS
  - Benefits: guest OS runs on any host OS, so moving is quick; no need to reinstall all the dependent packages again

Public Access to the web servers

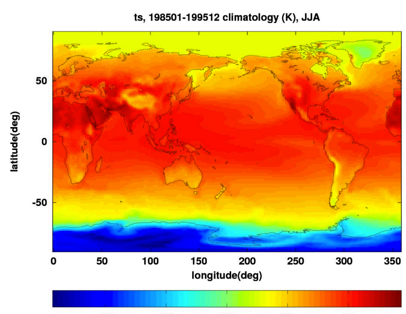
- Perimeter access approval for our web servers is necessary for summer school support since the school will be held at the Caltech campus, which is outside JPL network.
- JPL approved the content of the web page for export control compliance (June 2013).
- JPL approved inbound perimeter access to our web servers (July 2013).
- Three web servers have inbound access approved from outside JPL network.
  - http://cmacws.jpl.nasa.gov:8080/cmac/
  - http://cmacws2.jpl.nasa.gov:8080/cmac/
  - http://cmacws4.jpl.nasa.gov:8080/cmac/
- One web server is kept inside the JPL network for development.
  - http://cmacws3.jpl.nasa.gov:8080/cmac/

Analysis Capabilities (1)

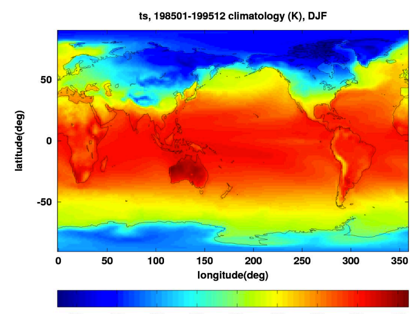
- Display a two dimensional physical variable on a map as function of longitude and latitude
  - Temporal average is computed for a specified range
  - User may choose a subset of months in a year to study seasonal variation
  - Displayed region may be customized for examining details



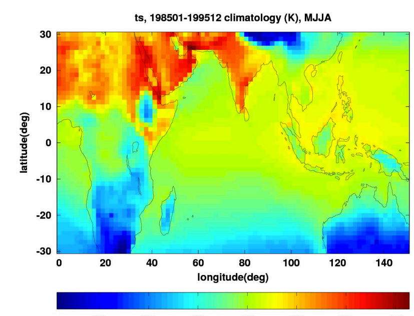
Surface temperature, annual



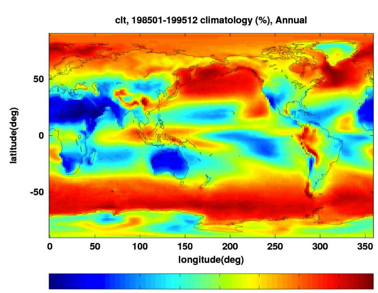
summer



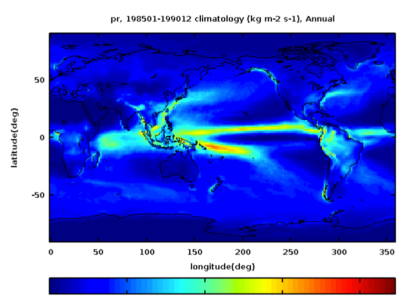
winter



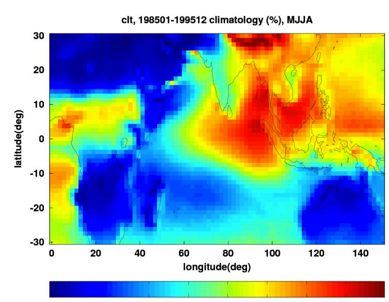
summer, region



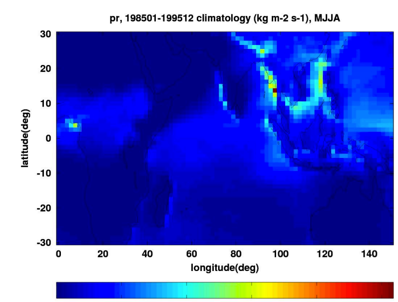
cloud fraction



Precipitation



cloud fraction, region

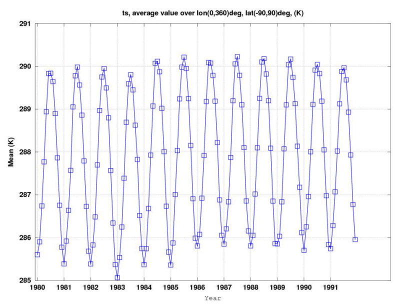


Precipitation, region

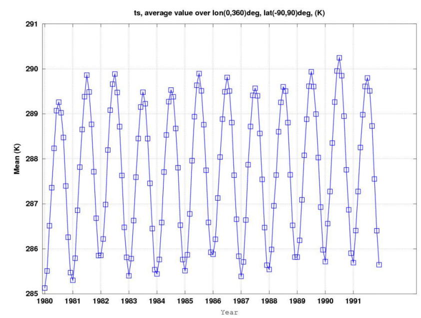
Analysis Capabilities (2)

- Plot the time series of a 2D physical variable averaged over a specified region
- Plot the zonal mean of a 2D physical variable averaged over a temporal period
  - User may select months in a year for study seasonal variation

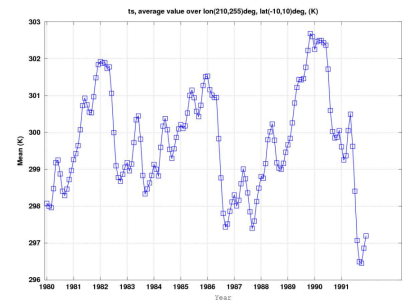
Surface Temperature Time series



UKMO_hadgem2-es

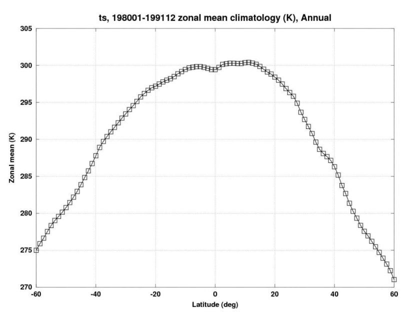


NCAR-cam5

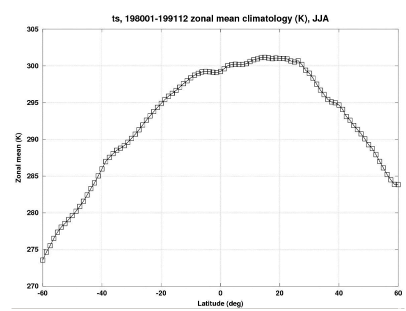


NCAR-cam5, Nino 3

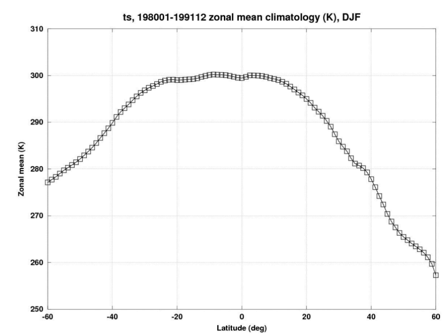
Surface Temperature zonal means



annual



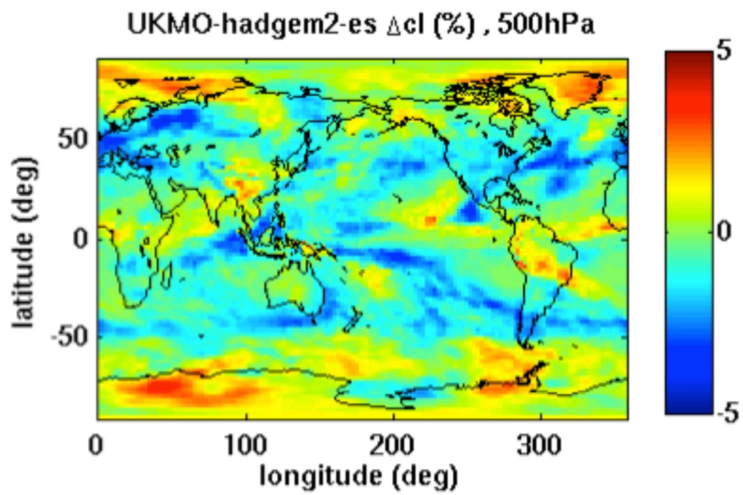
summer, JJA



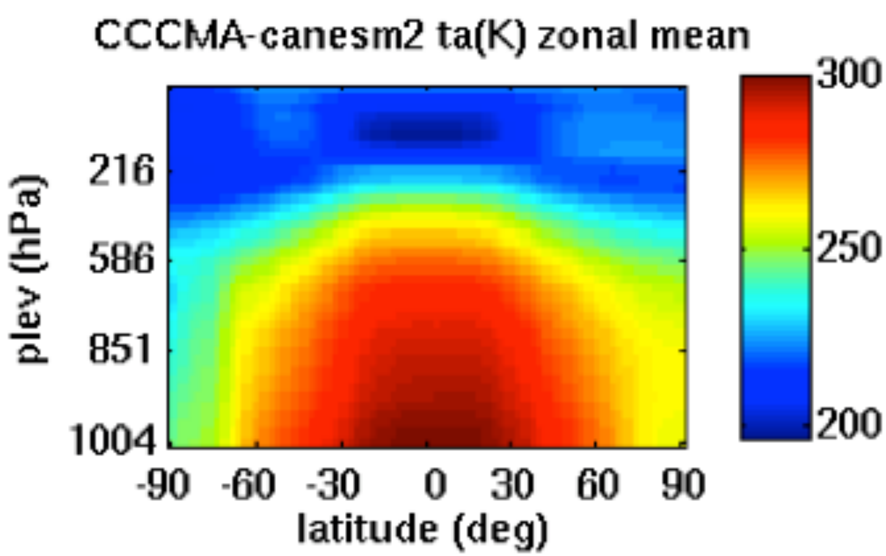
winter, DJF

Analysis Capabilities (3)

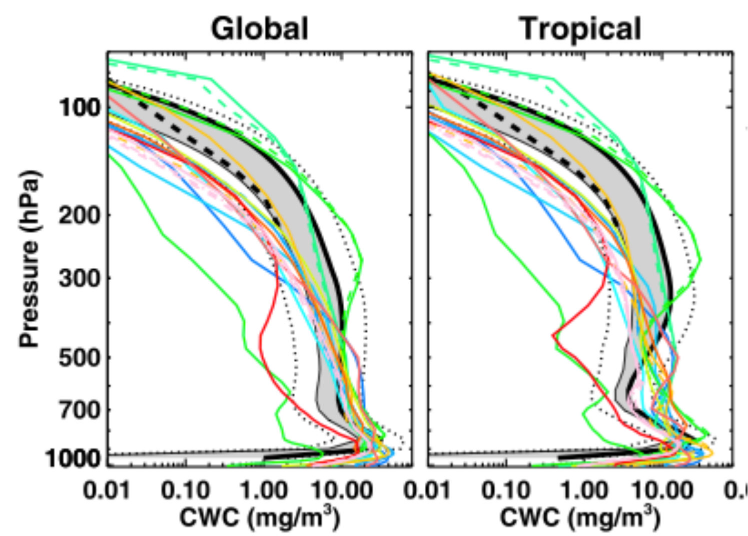
- Plot the pressure-level slice of a 3D physical variable averaged over a specified region and temporal period.
- Plot the zonal mean of a 3D physical variable averaged over a temporal period.
- Plot the vertical profile of a 3D physical variable averaged over a temporal period.



Cloud fraction change at 500hPa



Zonal mean of air temperature

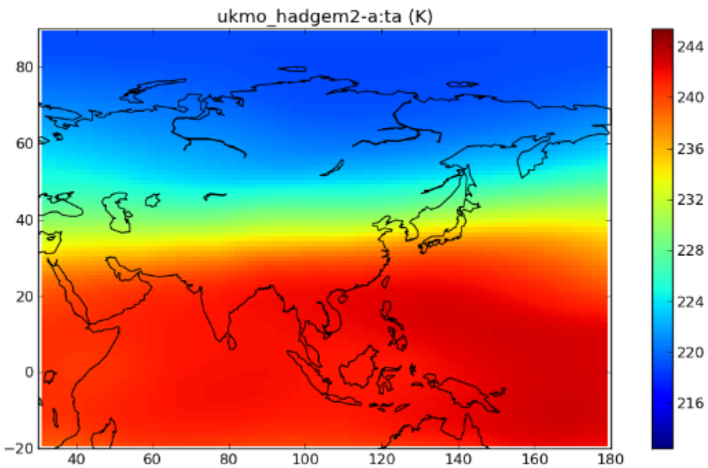


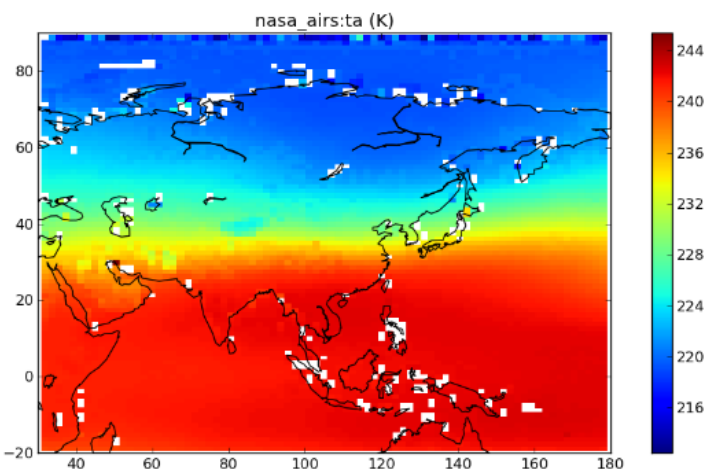
Vertical profile of cloud water content

Analysis Capabilities (4)

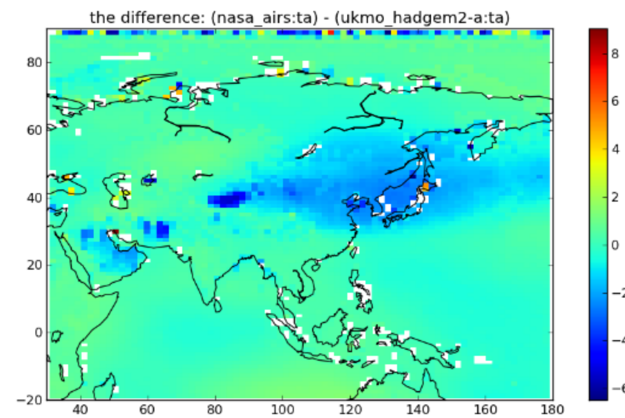
- Plot the difference between two variables over a specified region and averaged over a temporal period.
- Plot the scatter of two variables and calculate the correlation between the two.

Air temperature at 300 hPa

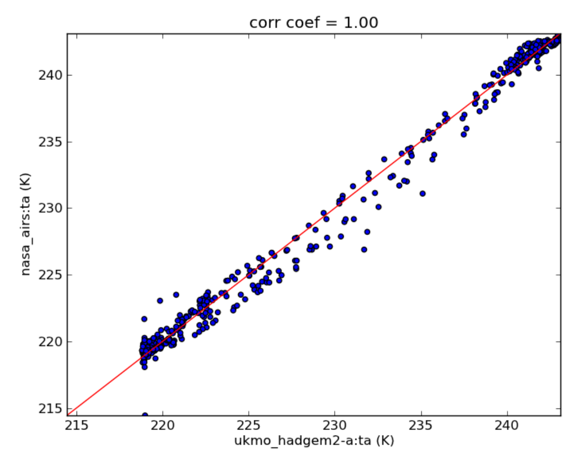




Two variable difference

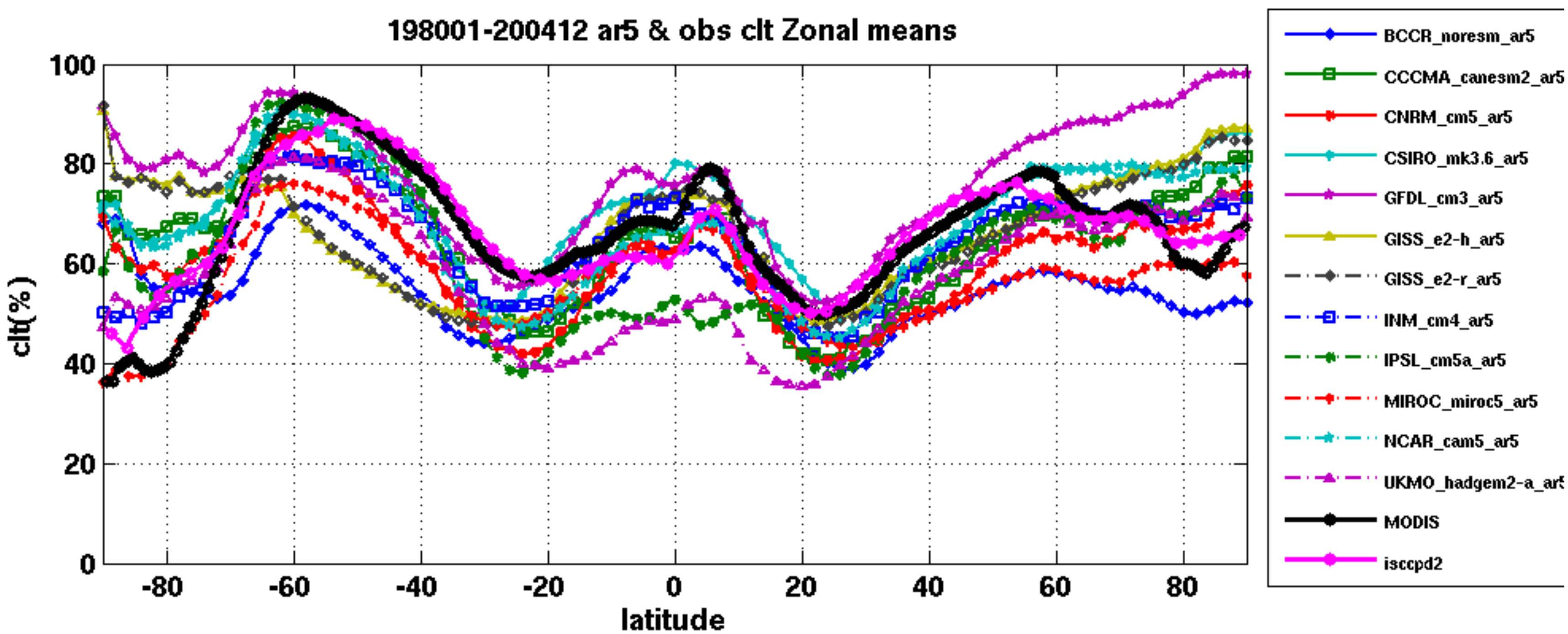


Two variable scatter plot and correlation



Example of planned summer school exercises (1)

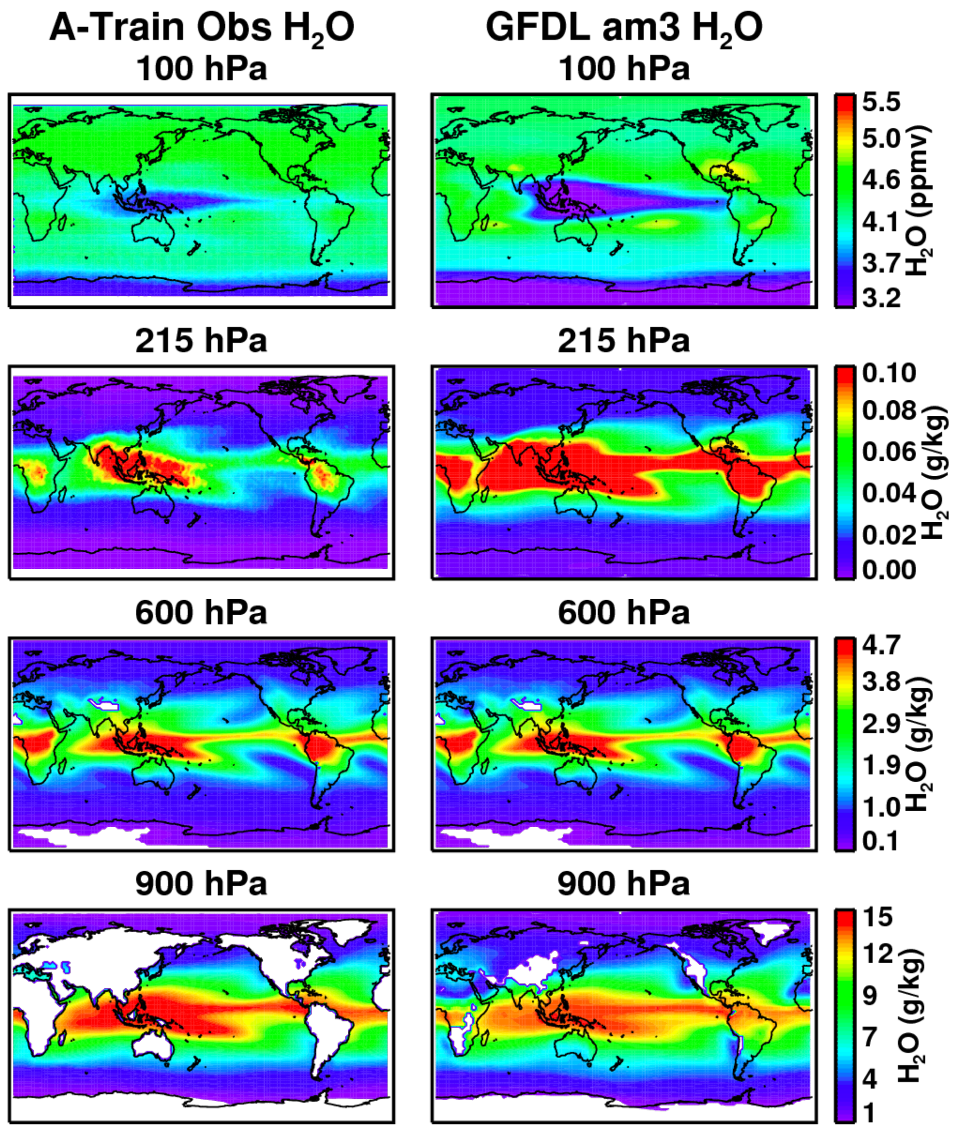
The students will be asked to use the monthly mean CMIP5 model outputs and Obs4MIPs datasets to compute and plot the zonal mean total cloud fractions as shown below and report model systematic biases if exist. The students will be also asked to consider the observation uncertainty and limitation and their effect on the comparison with model outputs.



Example of planned summer school exercises (2)

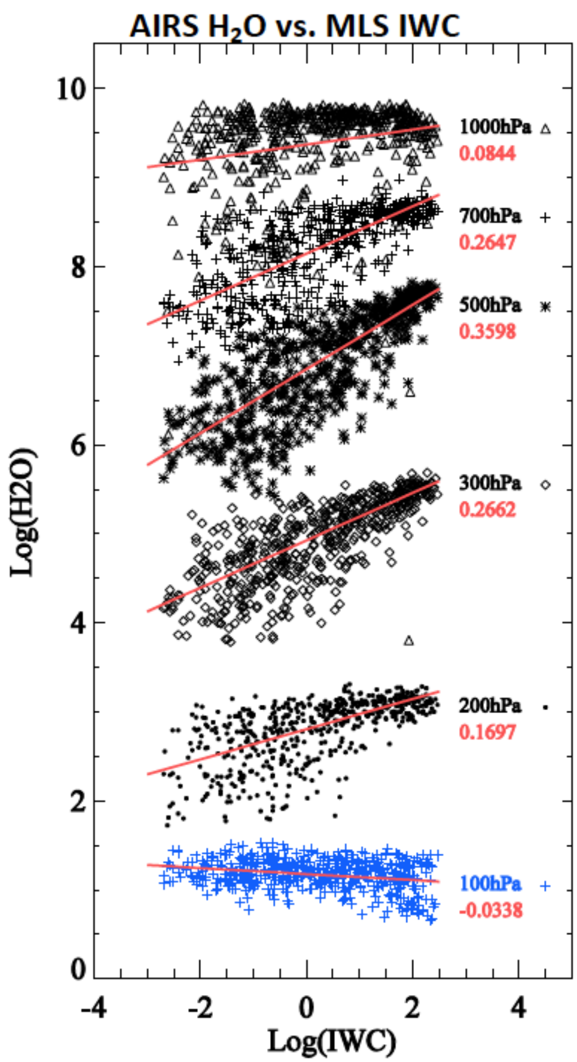
The students will be asked to plot a 3d model parameter (e.g. water vapor H_2O, air temperature) at several different pressure levels (e.g. 100hPa, 215hPa, 600hPa, 900hPa) and compare the model results with Obs4MIPs observational data and report model biases if exist.

The students will be also asked to analyze the potential causes of the systematic model biases.

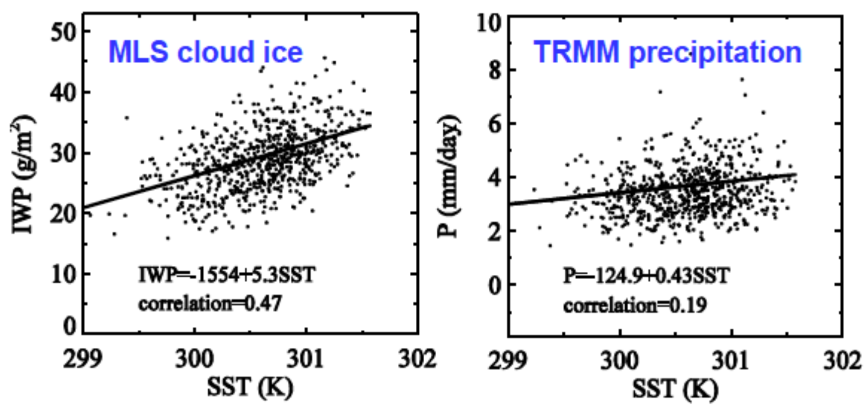


Example of planned summer school exercises (3)

Study correlation/interplay between cloud, water vapor, precipitation and SST.



Left: AIRS water vapor (H_2O) at five different levels (1000 hPa to 100 hPa) and MLS H_2O at 100 hPa scattered against MLS IWC at 215 hPa (used as a convective index, *Jiang et al. 2010*). AIRS H_2O product above 200 hPa altitude is not recommended for use so MLS H_2O data at 100 hPa are used as complements.



Above: Daily tropical mean upper tropospheric ice water path increase with SST, faster than the increase of daily mean cloudy-area precipitation with SST. (*Su and Jiang et al. 2008*)