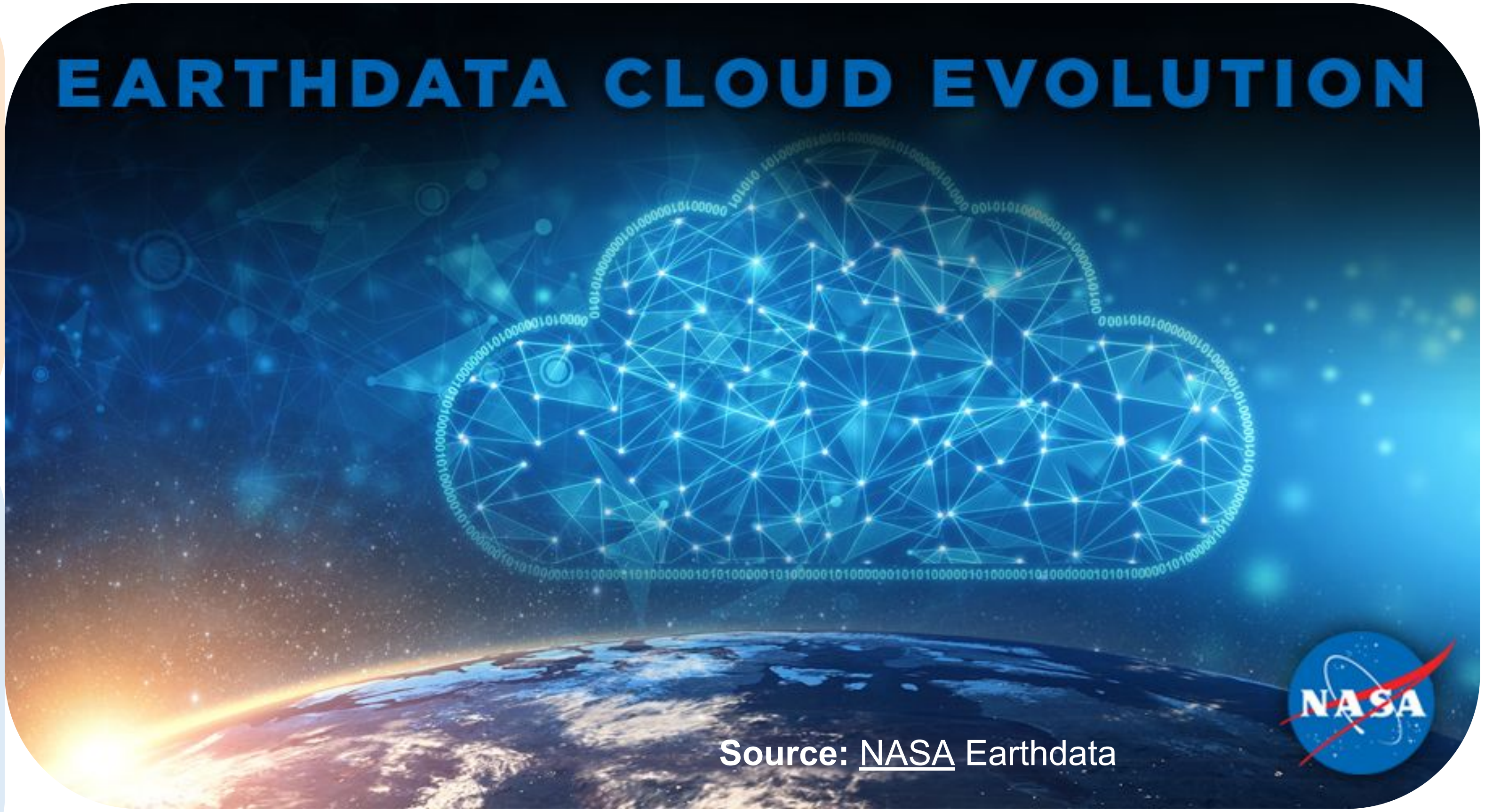
Characterizing Wildfires in Western U.S.:

A Cloud-based Case Study for Interdisciplinary Research using NASA Resources

Xiaohua Pan,
Suhung Shen,
Thomas Hearty,
Jennifer Wei,
Zhong Liu

NASA/GSFC/GES DISC



Acknowledgements

- NASA NCCS Python courses
- GES DISC staff cloud training Christopher Battisto (NASA/GSFC/GES DISC)
- 2021 Cloud Hackthon
- ARSET training - Amy Huff (NOAA/NESDIS)

GESDISC

What are special needs for doing an interdisciplinary research, such as wildfires?

Need to collect various data

- Characterize the wildfire events:
  - Fire hot spot, burned scar, fire plume height, fire diurnal variation …
- Identify causes:
  - Precipitation, air temperature, drought index, wind, air pressure, sea surface temperature, soil moisture, vegetation growth, spring snowmelt …
- Study the impacts on air quality:
  - $PM_{2.5}$, AOD, CO …

Need to collect data from diverse resources

- Observations (satellites, in-situ, and field campaigns), reanalysis, and models
- Data archived in various places, such as NASA data centers, NOAA, on-premise and in the cloud

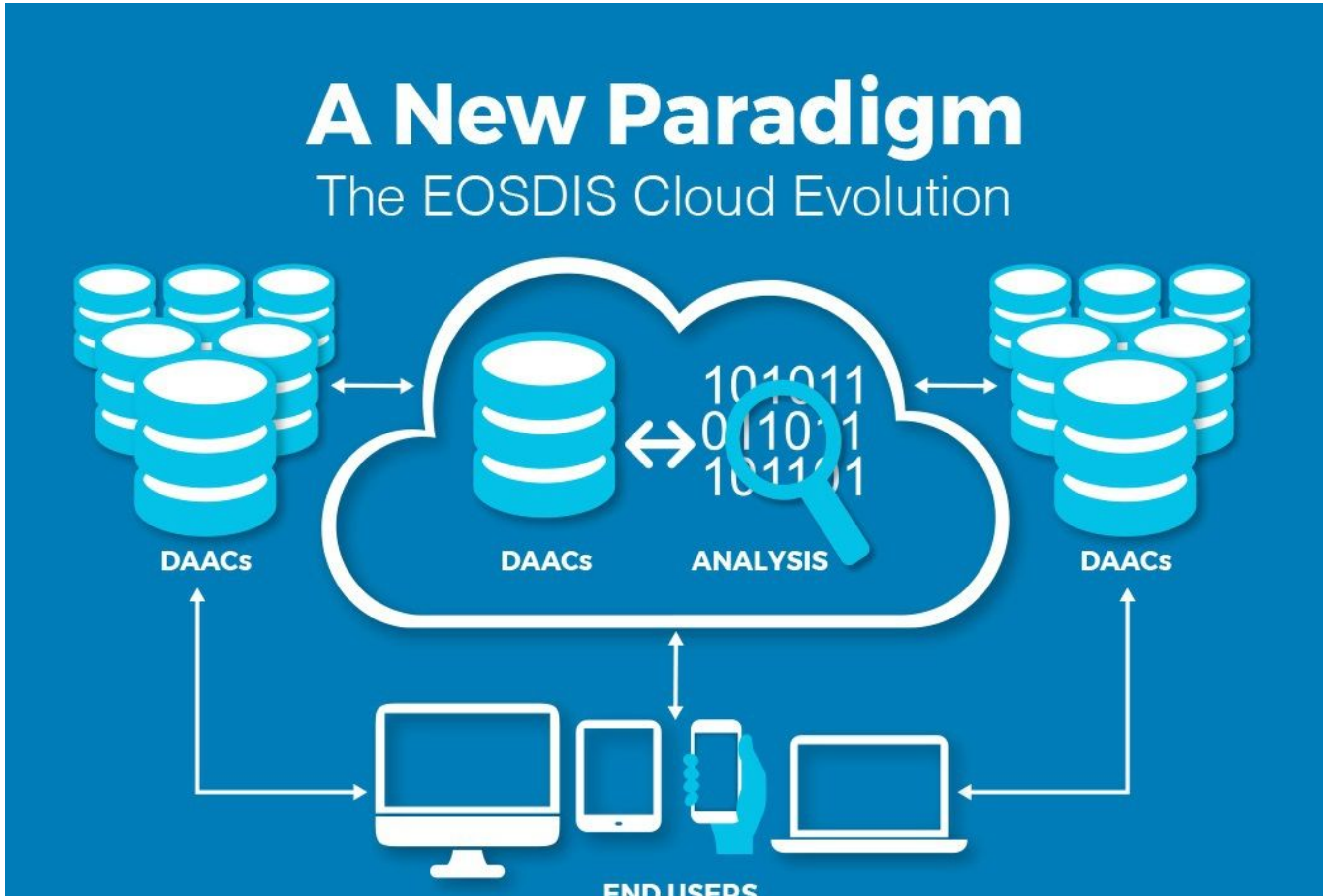
Need long-term climate data, e.g., 10-40 years

- Detect anomalous and extreme events
- Study the trends and interannual variation

GESDISC

$PM_{2.5}$: Fine particulate matter. **AOD**: aerosol optical depth. **CO**: carbon monoxide.

Can we conduct an interdisciplinary study in the cloud?



Source: NASA Earthdata

Benefits of cloud computing:

1. NASA data from various DAACs have been migrated to ONE place - Earthdata Cloud
2. Analysis next to data: directly access data - NO need of downloading

Two-fold objectives of this talk:

1. **As for me**: Test to see if the current NASA cloud data & resources can support an interdisciplinary study
2. **Useful for audiences**:
   a. if you are a scientist, you may expect to experience similar benefits and challenges of cloud computing.
   b. If you play a role in designing cloud data infrastructure, I hope you can better understand user needs and roadblocks.

GESDISC

DAAC: Distributed Active Archive Center.

FAQ on accessing data in the AWS cloud

Analysis Platform

Yes, you CAN!

Data

Access

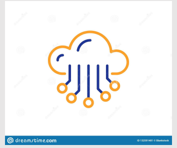
CASE 1:

Access cloud-based data archived by two or more different NASA DAACs

CASE 2:

Access on-premise data

CASE 3:

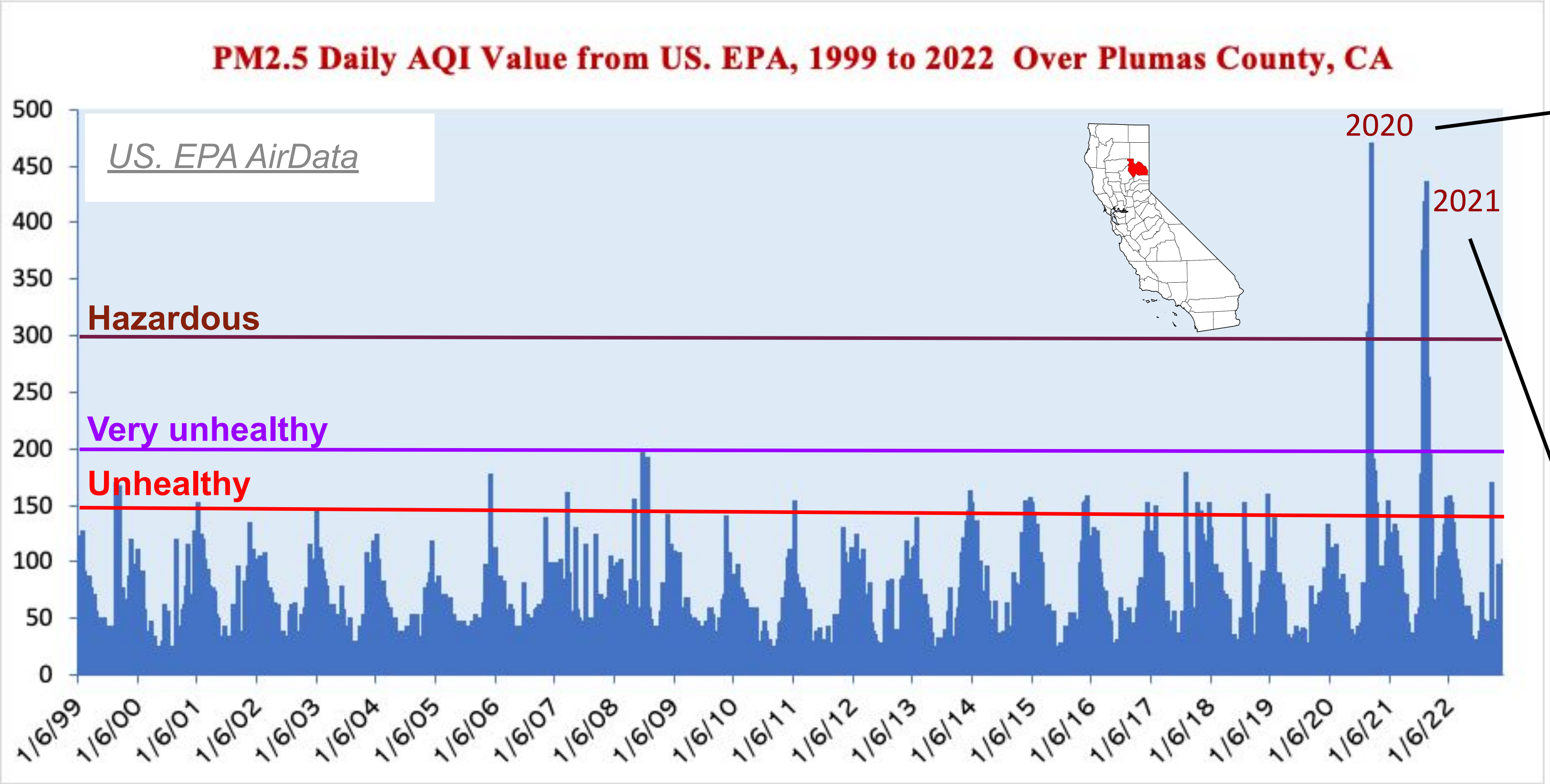


Access cloud-based data archived by NOAA in different AWS region

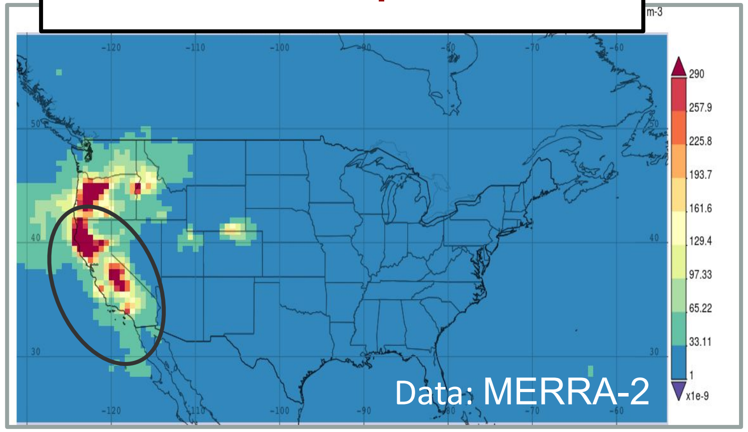
GESDISC

AWS: Amazon Web Services. **DAAC**: Distributed Active Archive Center.

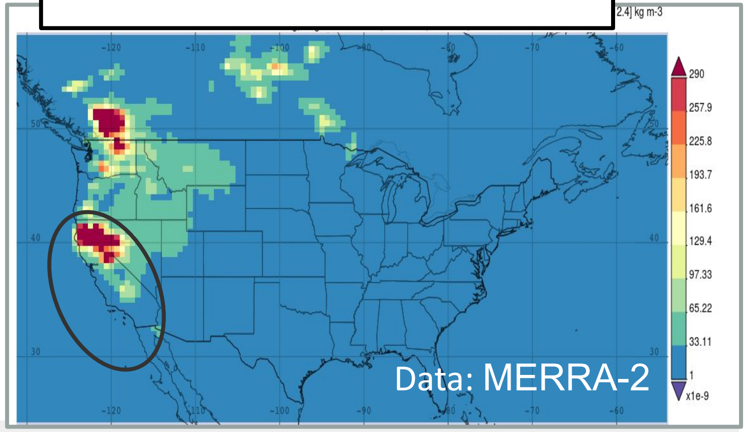
Extreme high $PM_{2.5}$ events in northern California



$PM_{2.5}$ in Sep. 2020



$PM_{2.5}$ in Aug. 2021



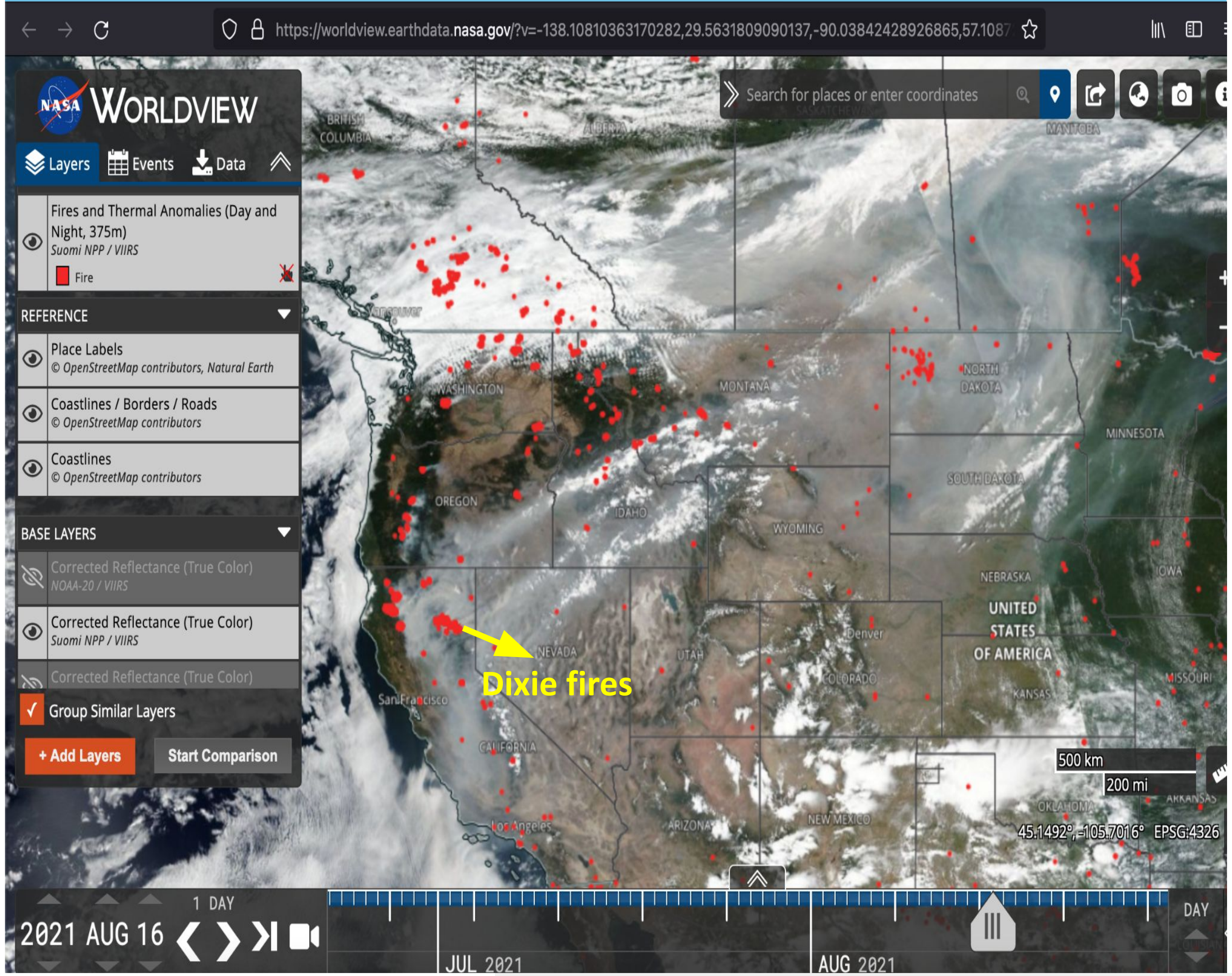
PM2.5 peaked during each winter.
However, two extreme events occurred around Sep. 2020 and Aug. 2021, due to wildfires.

GESDISC

PM2.5: Fine particulate matter. **AQI:** Air Quality Index. **EPA**: U.S. Environmental Protection Agency.
MERRA-2: Modern-Era Retrospective analysis for Research and Application Version 2.

Interdisciplinary Case Study: Characterize Wildfires in Western U.S.

S-NPP RGB image and fire hotspots on Aug.16, 2021



- **Cloud Platform-AWS EC2 instance:**

  Openscapes [2i2C JupyterHub](#) in AWS openscapes gion

- **Data tool:**

  Jupyter notebook in Python

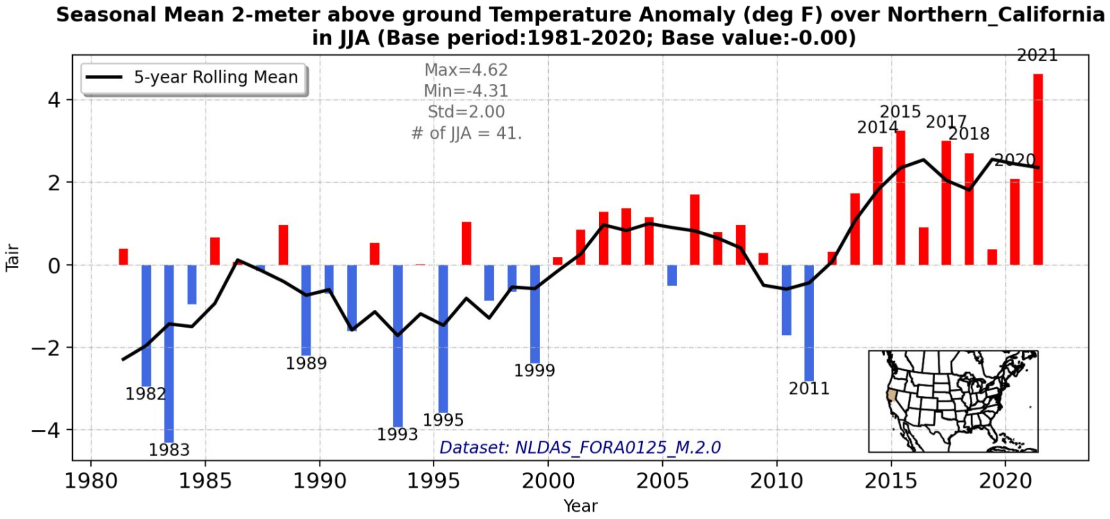
- **Data from two DAACs:**

  AWS cloud S3 buckets in the us-west-2 region from two NASA data centers, GES DISC and PO.DAAC.

GESDISC

EC2: Elastic Compute Cloud. **S3:** Simple Storage Service. **GES DISC:** NASA Goddard Earth Sciences Data and Information Services Center. **PODAAC:** NASA Physical Oceanography DAAC.

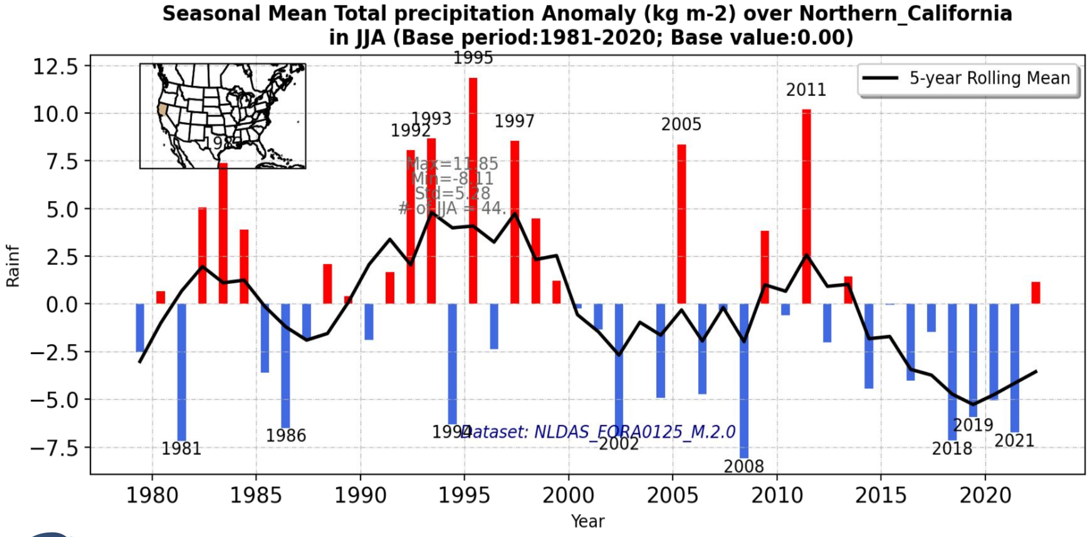
Meteorological conditions of wildfires in northern California



Data from GES DISC AWS S3 buckets:

40-year monthly data from NLDAS_FORA0125

⇐ 2-meter air temperature (upper panel)



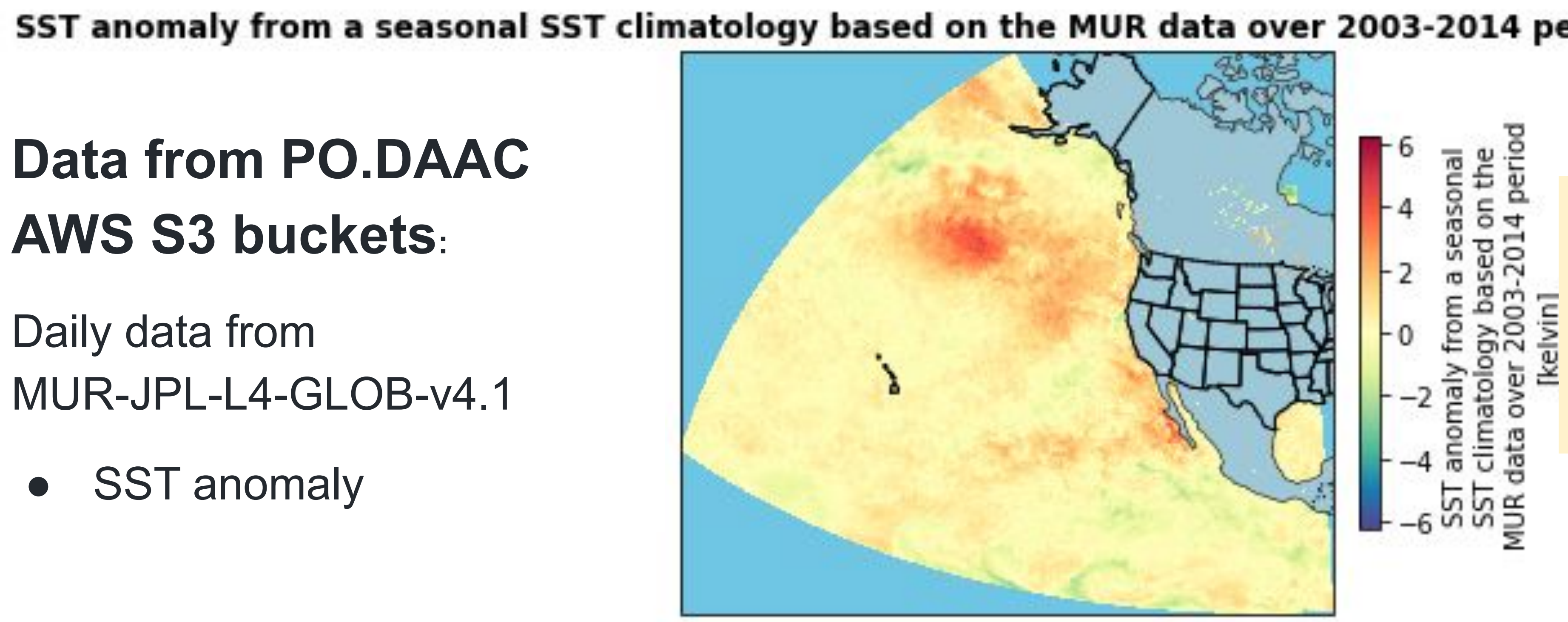
⇐ Total precipitation (lower panel)

The weather in summer (JJA) has been persistently anomalous hot and dry in northern California since 2014, especially in **2021**. This led to **drought conditions**, which are conducive to occurrence of wildfires by drying out vegetation that fuels wildfires.

GESDISC

S3: Simple Storage Service. **NLDAS:** North American Land Data Assimilation System

Meteorological conditions of wildfires in Northern California (cont.)



Data from PO.DAAC AWS S3 buckets:

Daily data from MUR-JPL-L4-GLOB-v4.1

- SST anomaly

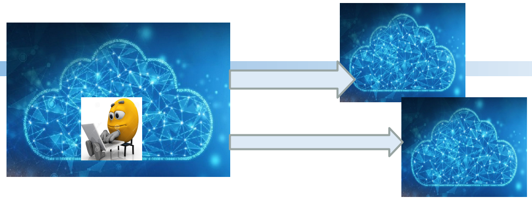
A warm SST anomaly occurred in the northern Pacific ocean during wildfire season in August 2022.

With the same Python notebook, we can S3 access two DAACs' data and study the relationships between them.

GESDISC

SST: Sea Surface Temperature. **MUR**: Multi-scale Ultra-high Resolution

Accessing cloud-based data in GES DISC vs. PO.DAAC



DAAC	Collection	Direct S3 Cloud Access
NASA GES DISC	**NLDAS_FORA0125_M.2.0** - L4 in NetCDF format - 0.125 x 0.125 deg - Monthly	1) **s3_credential_endpoint_gesdisc**:"https://data.gesdisc.earthdata.nasa.gov/s3credentials" 2) **gesdisc_collection_s3_url**: e.g.,"s3://gesdisc-cumulus-prod-protected/NLDAS/NLDAS_FORA0125_M.2.0/" 3) **Python data access modules:** s3fs and xarray
NASA PO.DAAC	**MUR-JPL-L4-GLOB-v4.1** - L4 in NetCDF format - 0.01 x 0.01 deg - Daily	1) **s3_credential_endpoint_podaac**: "https://archive.podaac.earthdata.nasa.gov/s3credentials" 2) **podaac_collection_s3_url**: e.g.,"s3://podaac-ops-cumulus-protected/MUR-JPL-L4-GLOB-v4.1/" 3) **Python data access modules:** s3fs and xarray

Similarities: Both data are in AWS us-west-2 region.

Differences: s3_credential_endpoint and collection_s3_url are structured slightly different in each DAAC.

GESDISC

S3: Simple Storage Service. **GES DISC:** NASA Goddard Earth Sciences Data and Information Services Center. **PO.DAAC:** NASA Physical Oceanography DAAC.

Cloud resources and tutorials available at NASA

Besides data, what resources and techniques do you need for the cloud computing?

- ❑ Login a AWS EC2 instance, e.g., Openscape 2i2c Jupyterhub
- ❑ Have the NASA Earthdata account (create one if you don't have it) and save your account information to a .netrc file in your cloud home directory
- ❑ Use the S3 Credential endpoint (varying with DAAC) to gain an NASA Earthdata login token to access S3 bucket in that DAAC.
- ❑ Know data's S3 urls (find it in the Earthdata data information page)
- ❑ Know how to read data (follow the tutorials which are mostly in Python)
- ❑ Conduct data analysis: the same as you have done on a local machine

Cloud Resources

- ❑ NASA GES DISC Cloud Dashboard: (link)
- ❑ NASA Earthdata Cloud Cookbook (link)

Tutorials

- ❑ Python courses by NASA: NASA NCCS Python courses
- ❑ Cloud tutorial by OPeNDAP and NASA: 2021 Cloud Hackthon
- ❑ ARSET training - Accessing and Analyzing Air Quality Data from Geostationary Satellites

Take-home messages about cloud computing based on this case study

- **Benefits**
  1. No need of downloading data
  2. Easily access data from various NASA data centers in one place (EARTHDATA cloud) and from NOAA

- **Roadblocks**
  1. A sharp learning curve (e.g., S3 access method and cloud concepts)
  2. Limited **data** in the cloud so far
  3. Limited **data services** available in the cloud (e.g., subset or cloud optimized data format)
  4. Cost for computing and storage (accessing NASA data is FREE)

- **Recommendations to cloud users**
  1. Be courageous and curious
  2. Take the Python and cloud tutorials
  3. Contact us
     - GES DISC Help Desk: gsfc-dl-help-disc@mail.nasa.gov
     - Earthdata Forum: forum.earthdata.nasa.gov

Additional slides

FAQ on cloud

Q1: Can I download the cloud-based data to my local machine as before?

Yes, you can. Downloading is still free to you, but do your analysis in the cloud next to the data is strongly recommended for many reasons!

Q2: What is the status of cloud migration?

- 161 out of 1817 collections in GES DISC have been migrated to the AWS cloud US West 2 region As of December 5 (data link)
- 1739 out of 9110 collections in EARTHDATA have been migrated to the AWS cloud US West 2 region (data link)

Dataset citations

NLDAS project (2022), NLDAS Primary Forcing Data L4 Monthly 0.125 x 0.125 degree V2.0, Edited by David M. Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [*2022-12–03*], 10.5067/2DPKB5B5N14O

JPL MUR MEaSUREs Project. 2015. GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis. Ver. 4.1. PO.DAAC, CA, USA. Dataset accessed [2022-12-03] at https://doi.org/10.5067/GHGMR-4FJ04

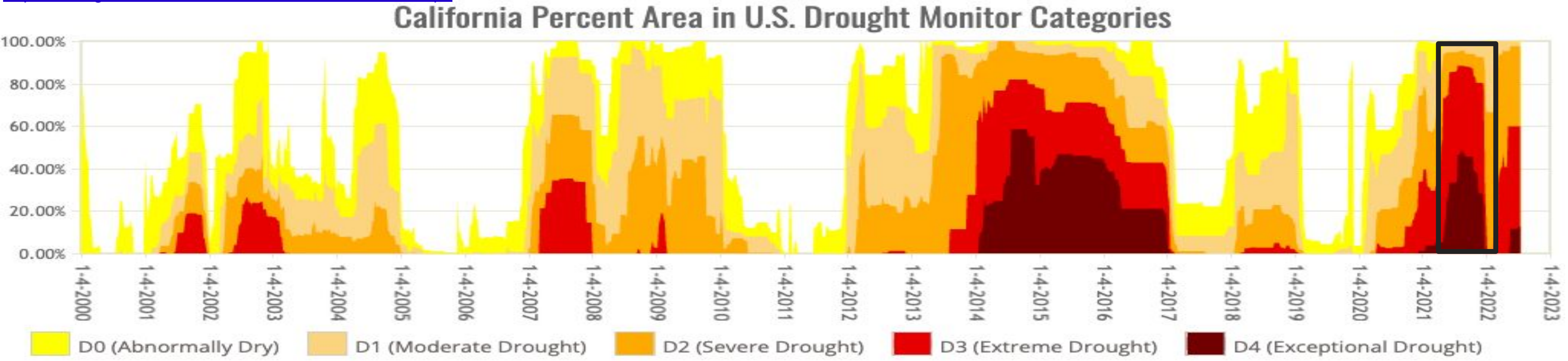
References

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J. Geophys. Res., 117, D03109, doi:10.1029/2011JD016048

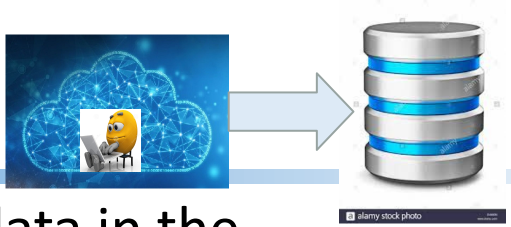
Chin, T.M, J. Vazquez-Cuervo, and E.M. Armstrong. 2017. A multi-scale high-resolution analysis of global sea surface temperature, Remote Sensing of Environment , 200 . https://doi.org/10.1016/j.rse.2017.07.029

Drought and heat wave in California

https://droughtmonitor.unl.edu/DmData/TimeSeries.aspx



Data analysis from local vs. in the cloud



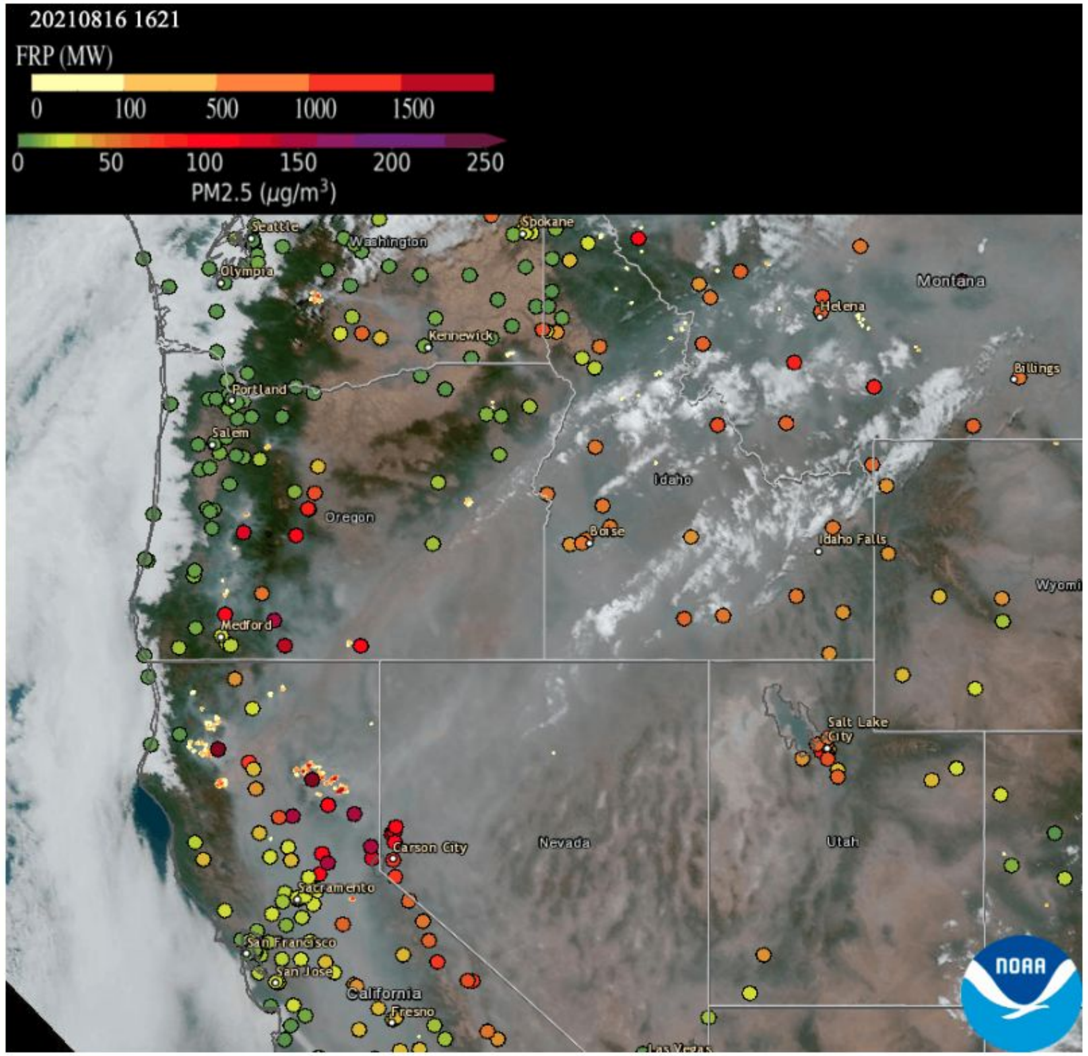
Preliminary result: it takes less time(~⅓) to access and analyze the cloud data in the cloud than otherwise

Table: Comparison of running a notebook to read 14-day hourly MERRA-2 data M2T2NXSLV and plot the spatial map from the local machine vs. the AWS cloud

Platform	Data	Setup	Machine configuration	Time for reading data	Time for plotting the weekly map	Time in toal
Local machine	on-premises OPeNDAP	.netrc, modules	2.5 GHz Dual-Core Intel Core i7, 16G RAM	18s - 1min 50s	2min 5s - 3min 53s	~ 3-6 min
The AWS cloud	On-premises OPeNDAP	Same as above	2 CPU, ~ 8G RAM	25s - 2min 4s	24s - 38s	~ 1-3 min
The AWS cloud	in the cloud S3	Same as bove	2 CPU, ~ 8G RAM	40s -43s	5s - 7s	~ 1 min

GESDISC

Animation of Dixie fire on Aug. 16, 2021 from GEOS 17



Click the link below for animation:

Credit: [NOAA AerosolWatch]