

Cutting Edge Climate Science in the Cloud with Pangeo

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

Scipy | July 2022 | Julius Busecke

Who am I?

Physical Oceanographer

Studies the role of ocean currents for the climate and ecosystems.

Associate Research Scientist - Columbia University



Core-Maintainer of xgcm

Core-Maintainer of xMIP

Maintainer of the Pangeo CMIP6 data

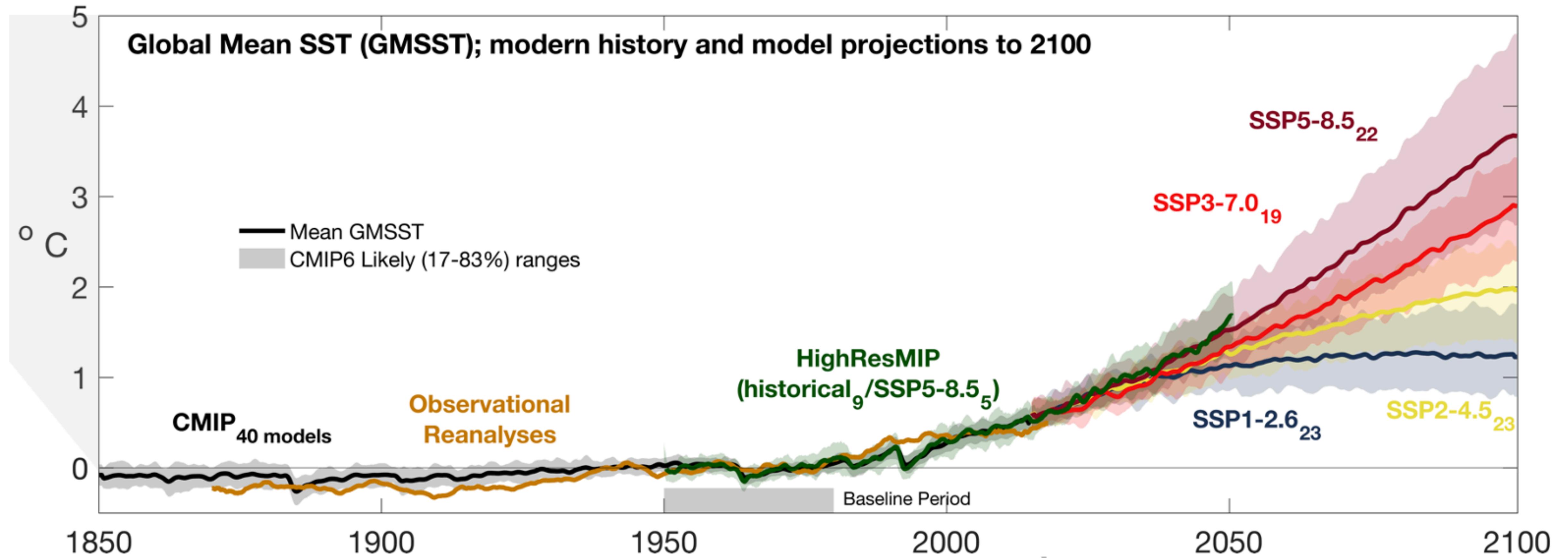
Pangeo Fan, User, and Member

GLOBAL ISSUES

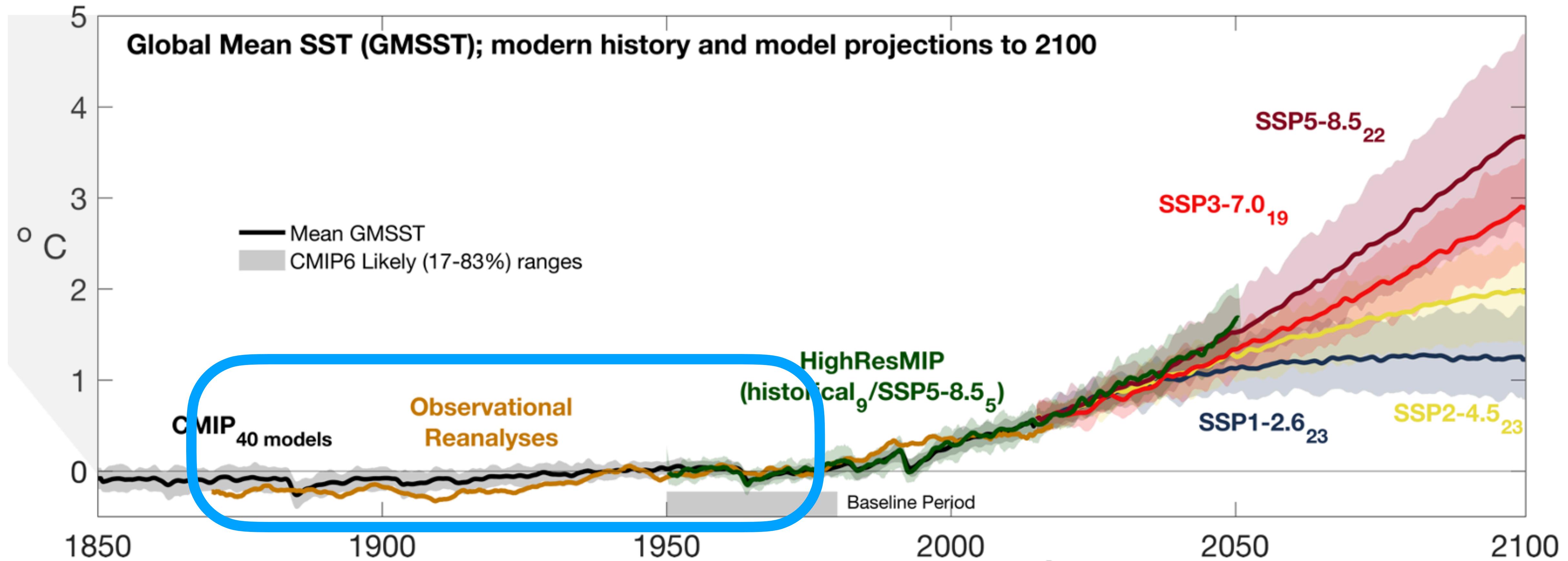
Climate Change

Climate Change is the defining issue of our time and we are at a defining moment. From shifting weather patterns that threaten food production, to rising sea levels that increase the risk of catastrophic flooding, the impacts of climate change are global in scope and unprecedented in scale. Without drastic action today, adapting to these impacts in the future will be more difficult and costly.

Source: <https://www.un.org/en/global-issues/climate-change>

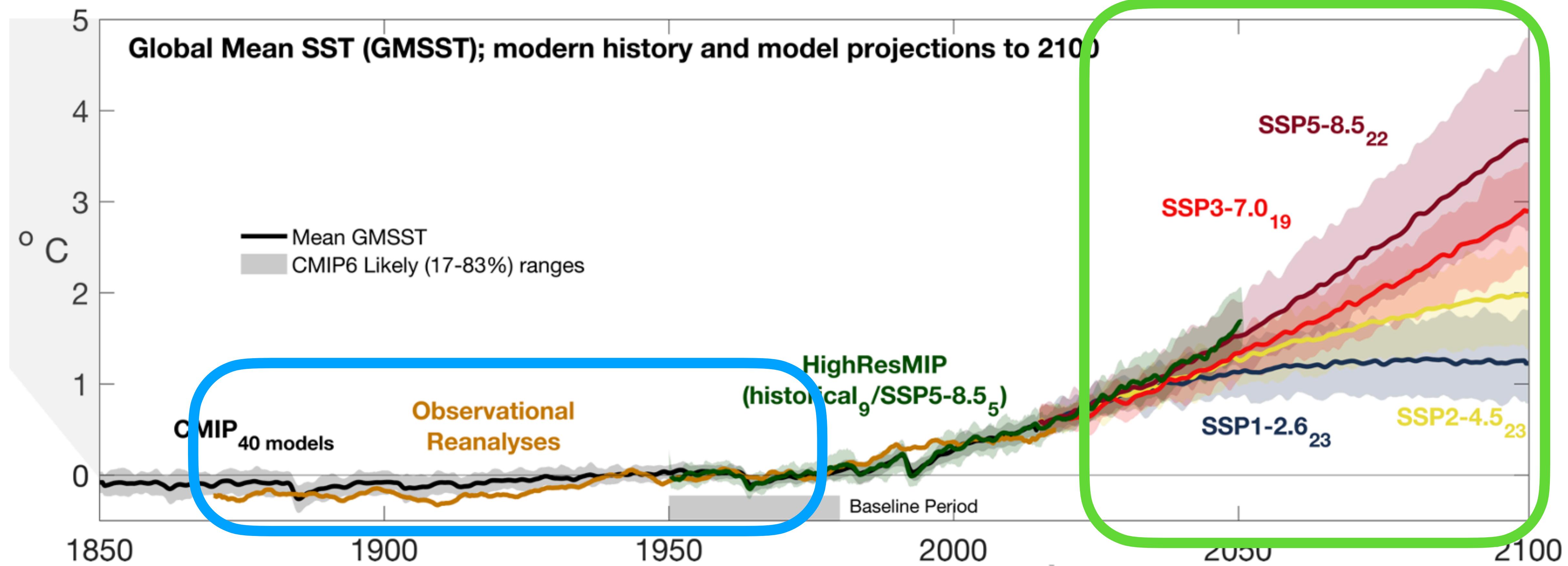


Observations



Observations

Models



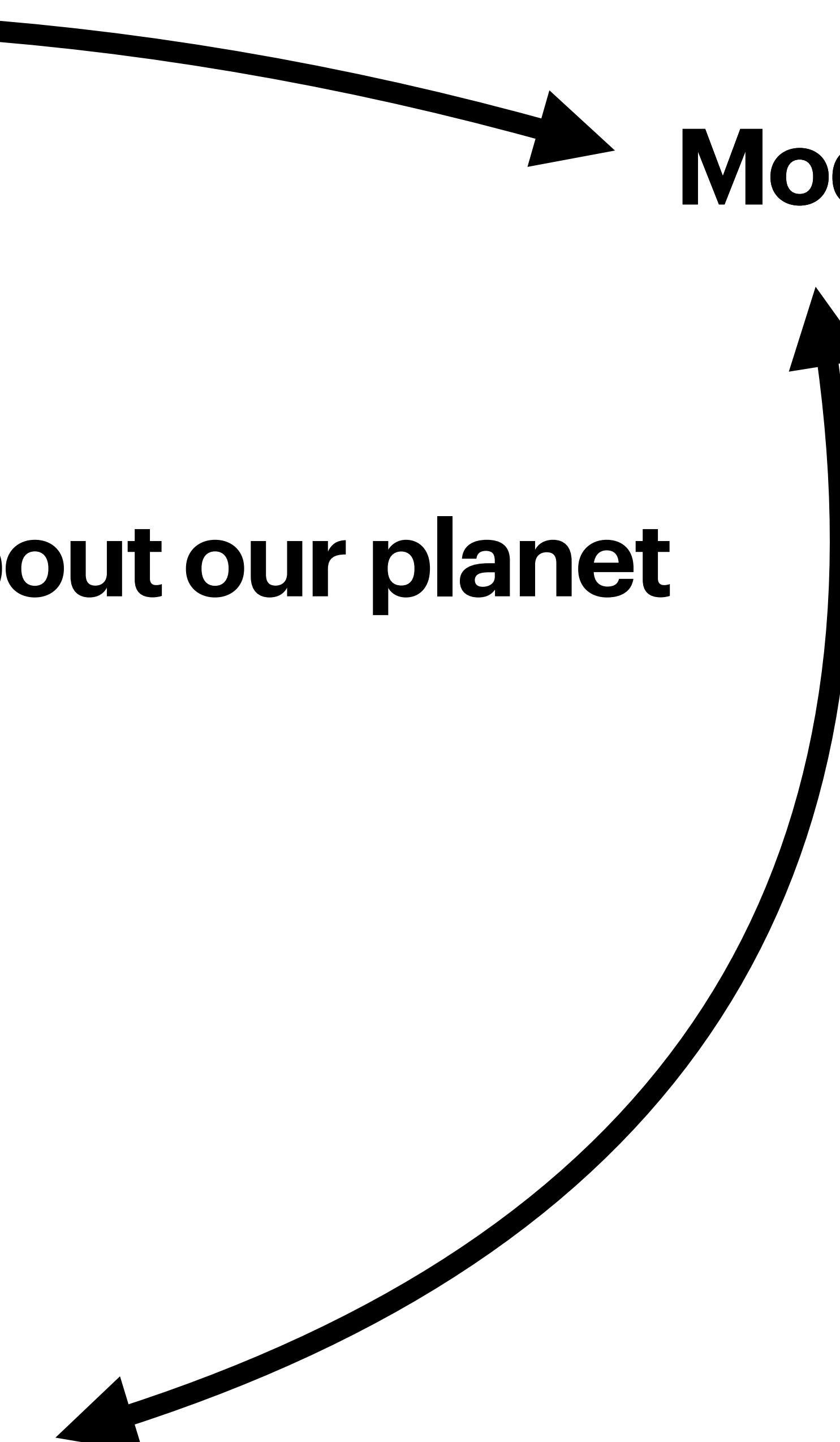
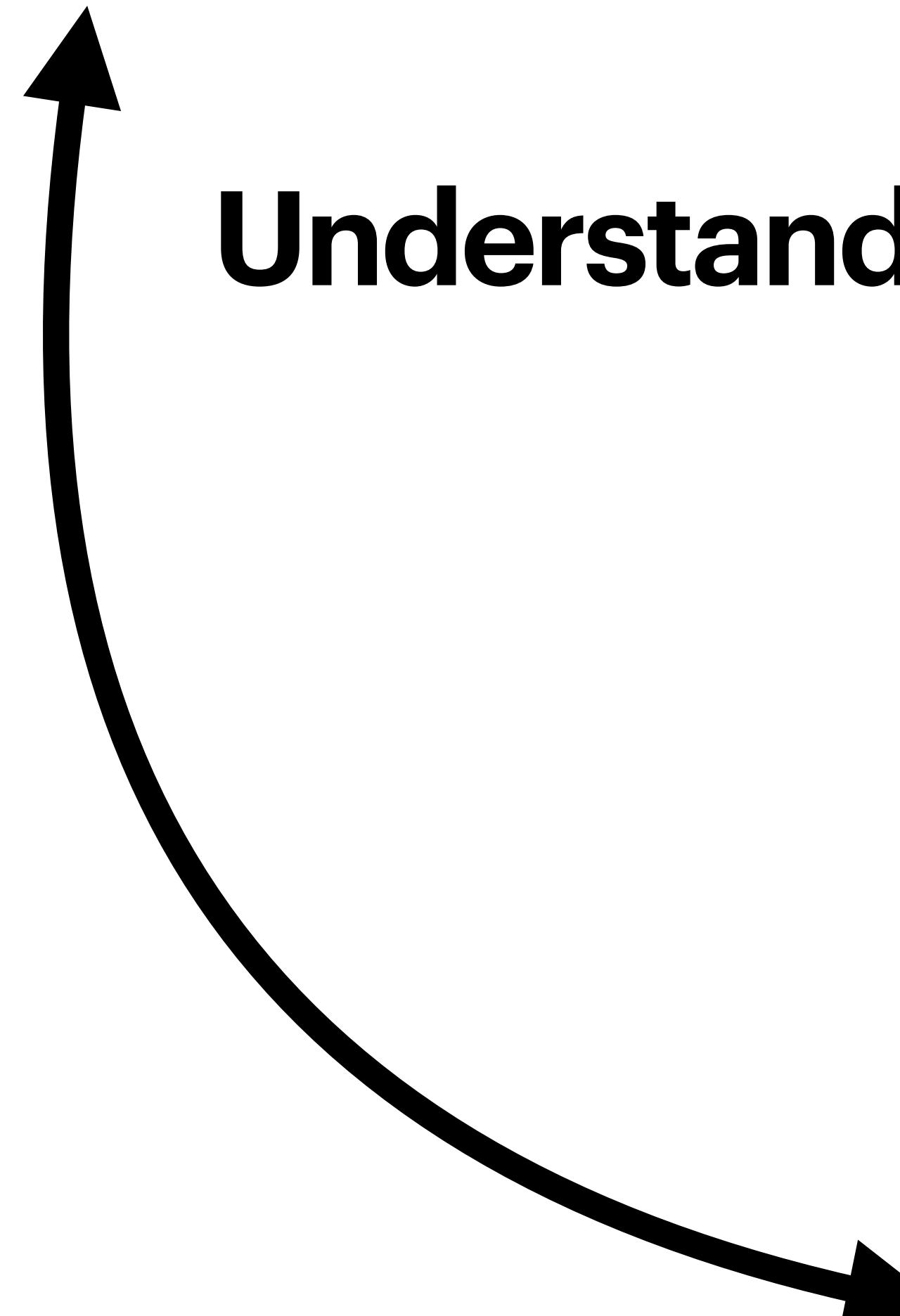
Observations

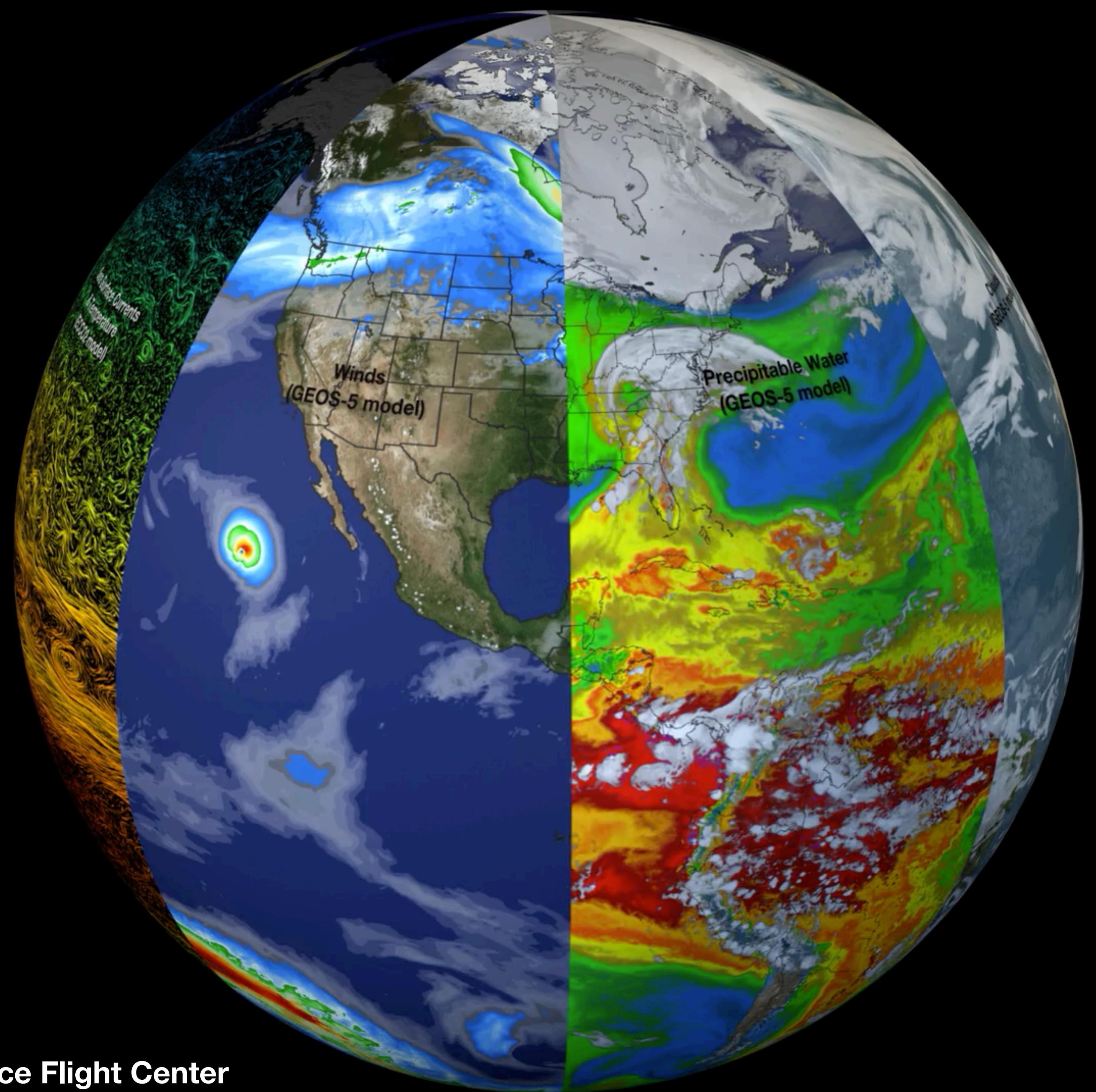
Models

Understanding more about our planet

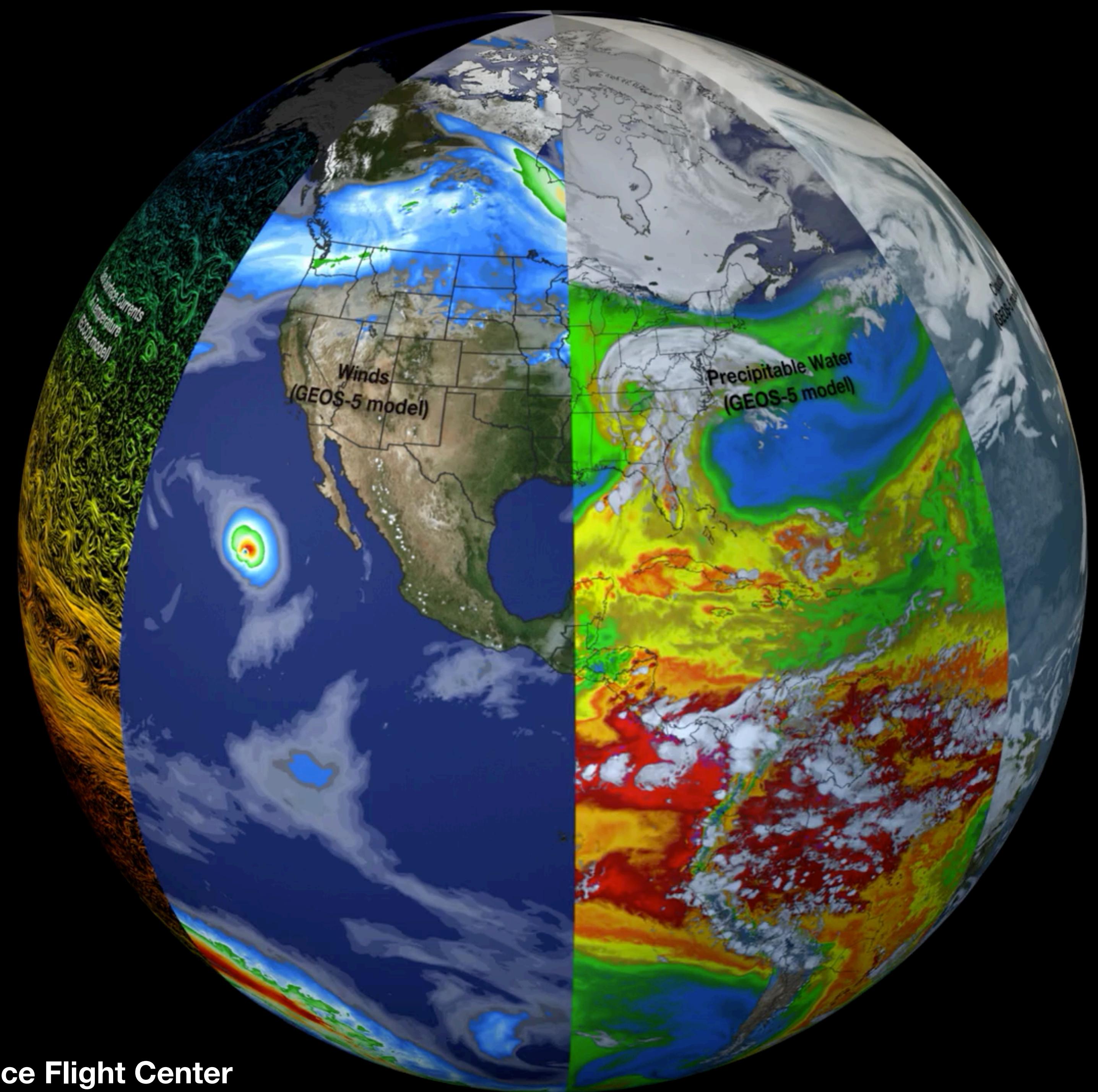


Theory

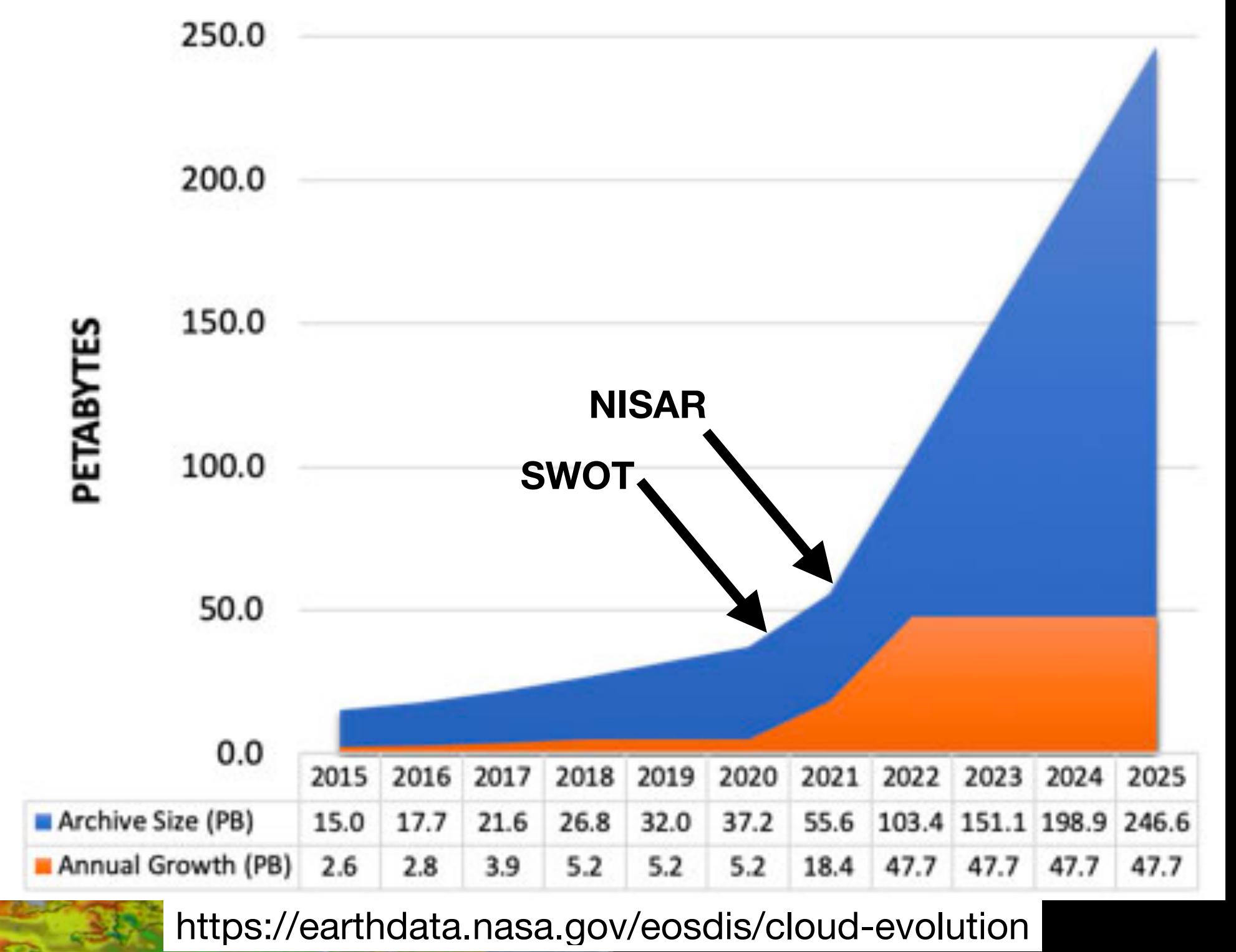
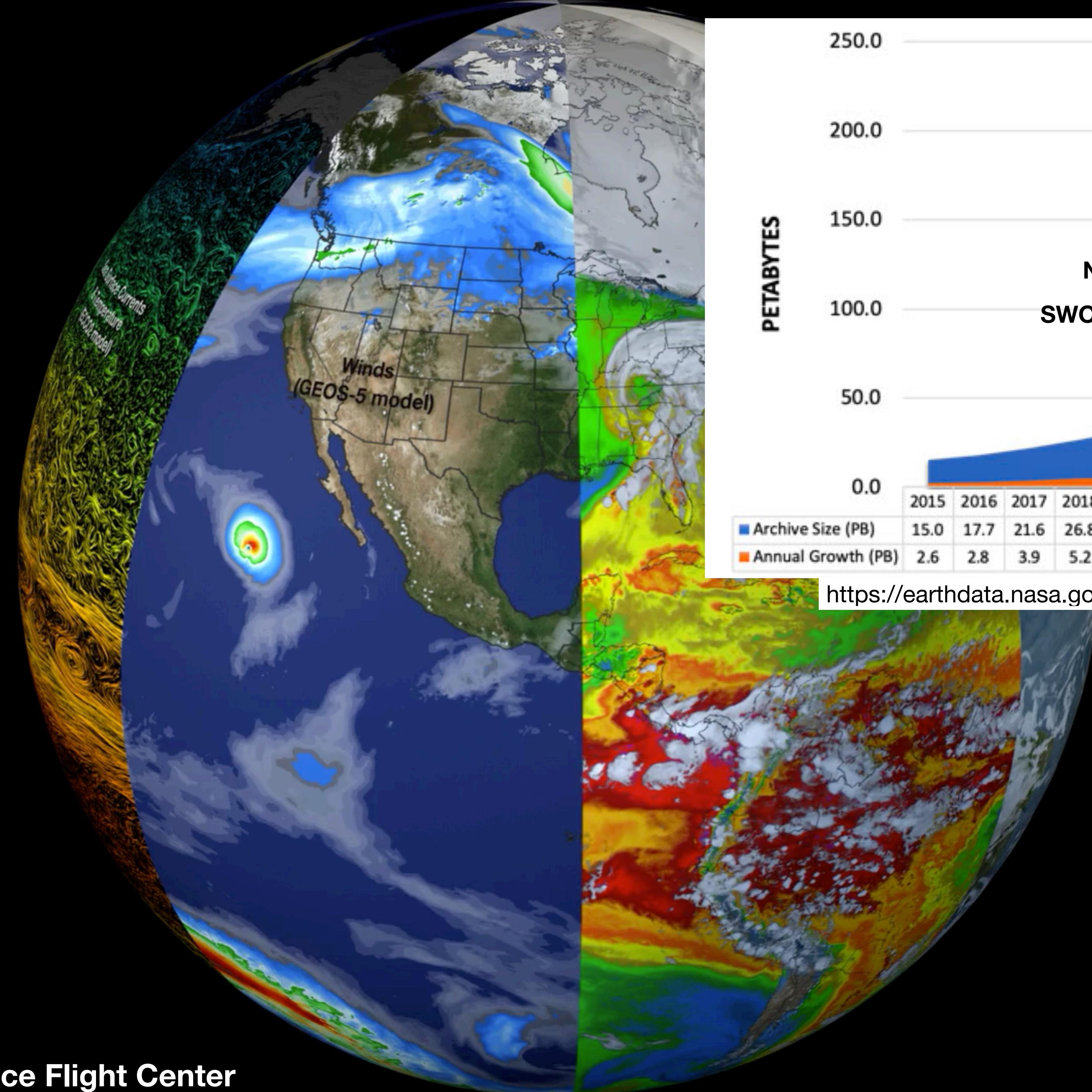




Credit: NASA's Goddard Space Flight Center

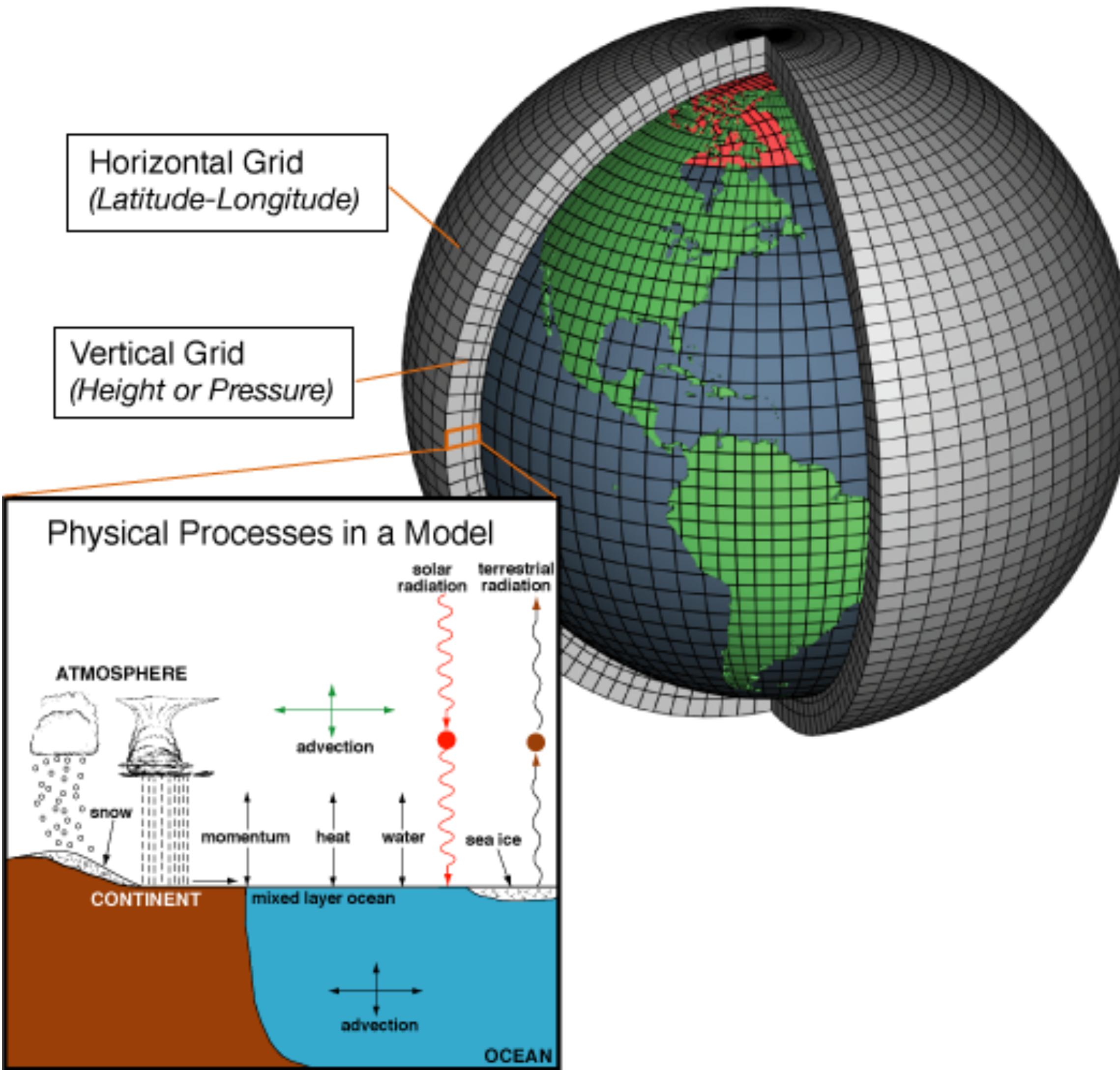


Credit: NASA's Goddard Space Flight Center

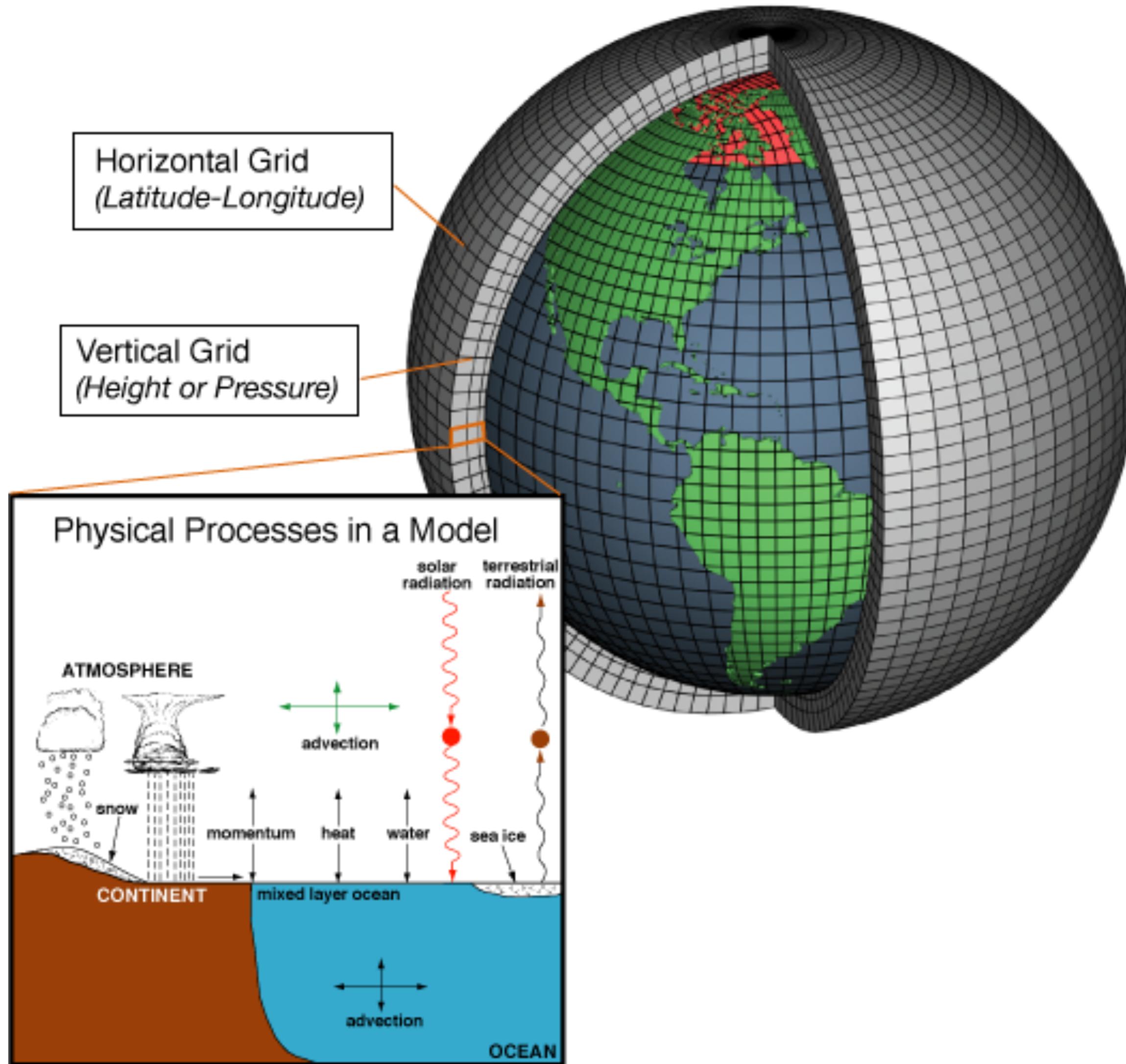


<https://earthdata.nasa.gov/eosdis/cloud-evolution>

Climate Models

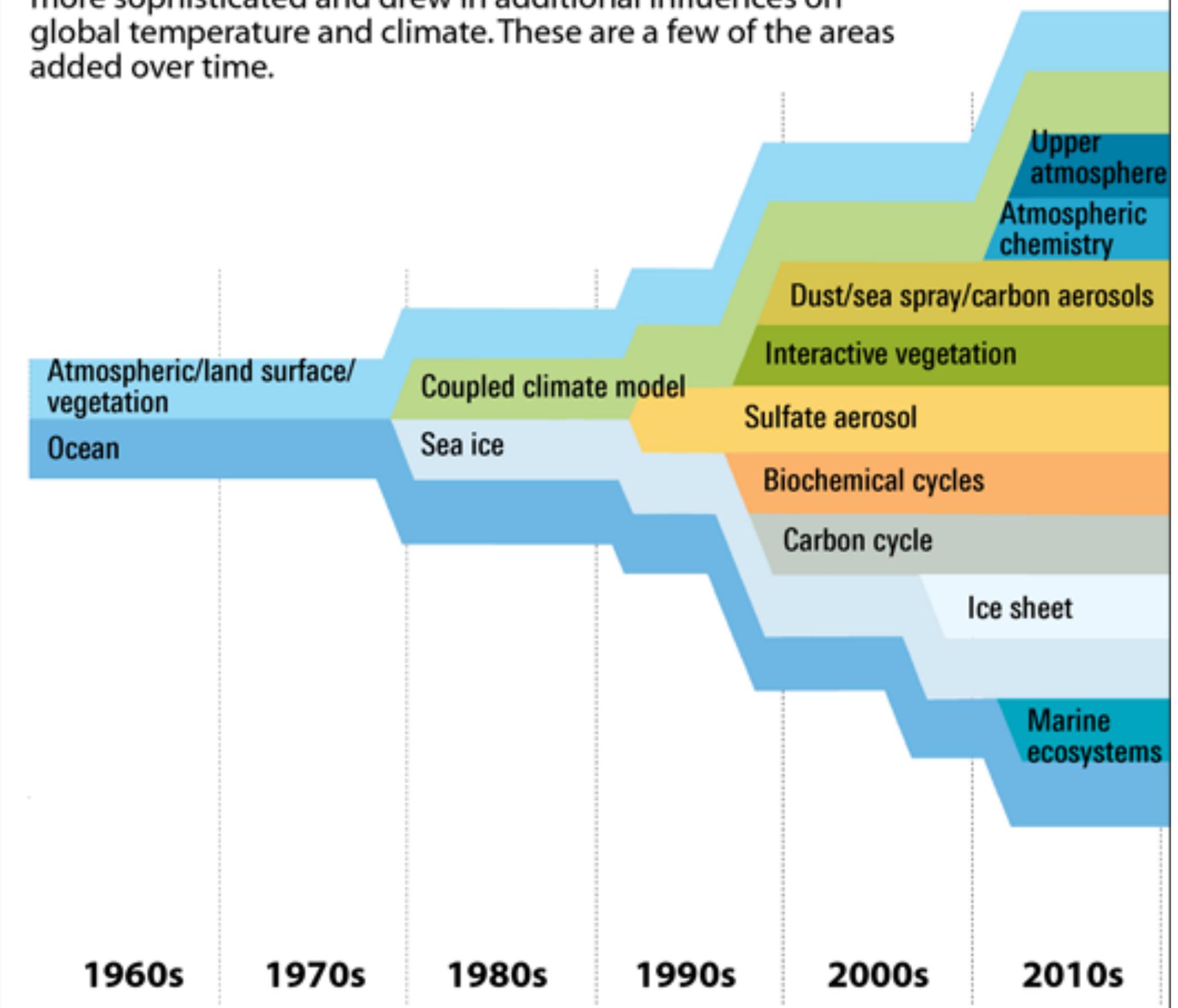


Increasing Resolution and Complexity



Growth of Climate Modeling

As computing power expanded, climate modeling became more sophisticated and drew in additional influences on global temperature and climate. These are a few of the areas added over time.



SOURCE: University Corporation for Atmospheric Research (UCAR)

InsideClimate News

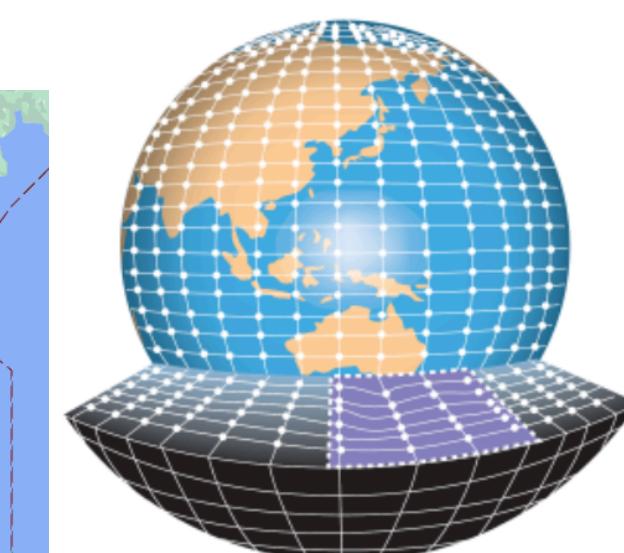
Climate Science: A global distributed effort



BOM 2003.



BOM 2003.



BOM 2003.



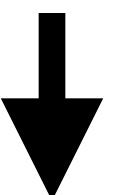
BOM 2003.



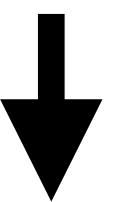
ESGF
Earth System Grid Federation

 CF MetaData

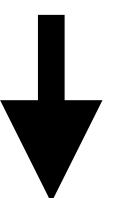
Download



Clean and Combine



Crunch the data

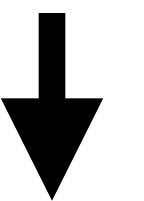


Interpret Results

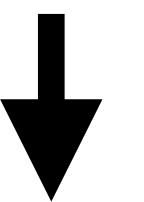
CLIMATE DATA IS SO HOT RIGHT NOW

**I AM ALMOST CONSIDERING
SPENDING A MONTH DOWNLOADING IT**

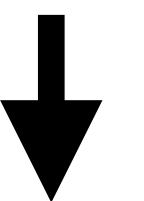
Download



Clean and Combine

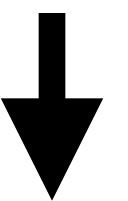


Crunch the data

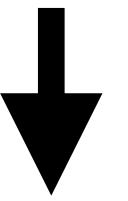


Interpret Results

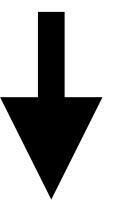
Download



Clean and Combine

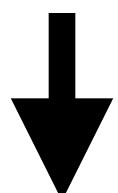


Crunch the data

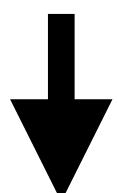


Interpret Results

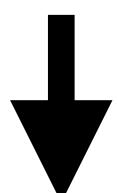
Download



Clean and Combine

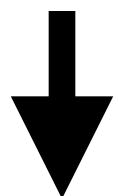


Crunch the data

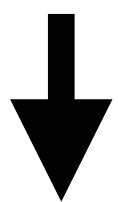


Interpret Results

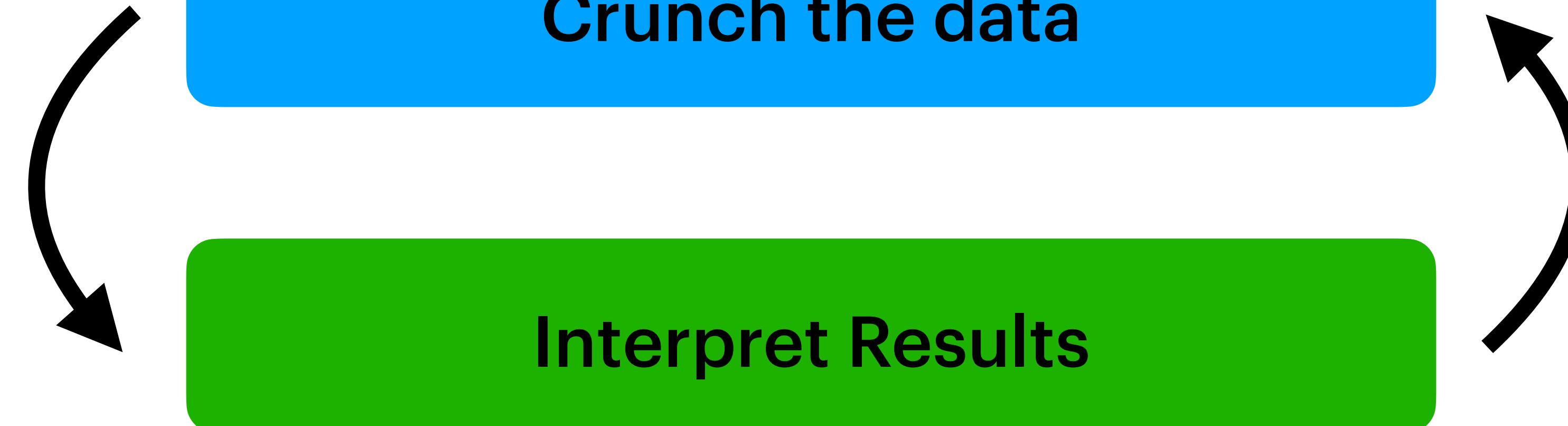
Download



Clean and Combine

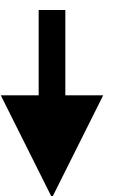


Crunch the data

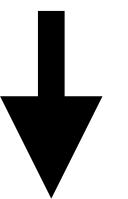


Interpret Results

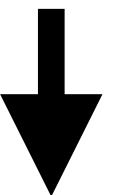
Download



Clean and Combine



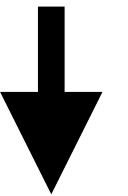
Crunch the data



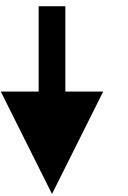
Interpret Results

~~Download~~

Clean and Combine



Crunch the data

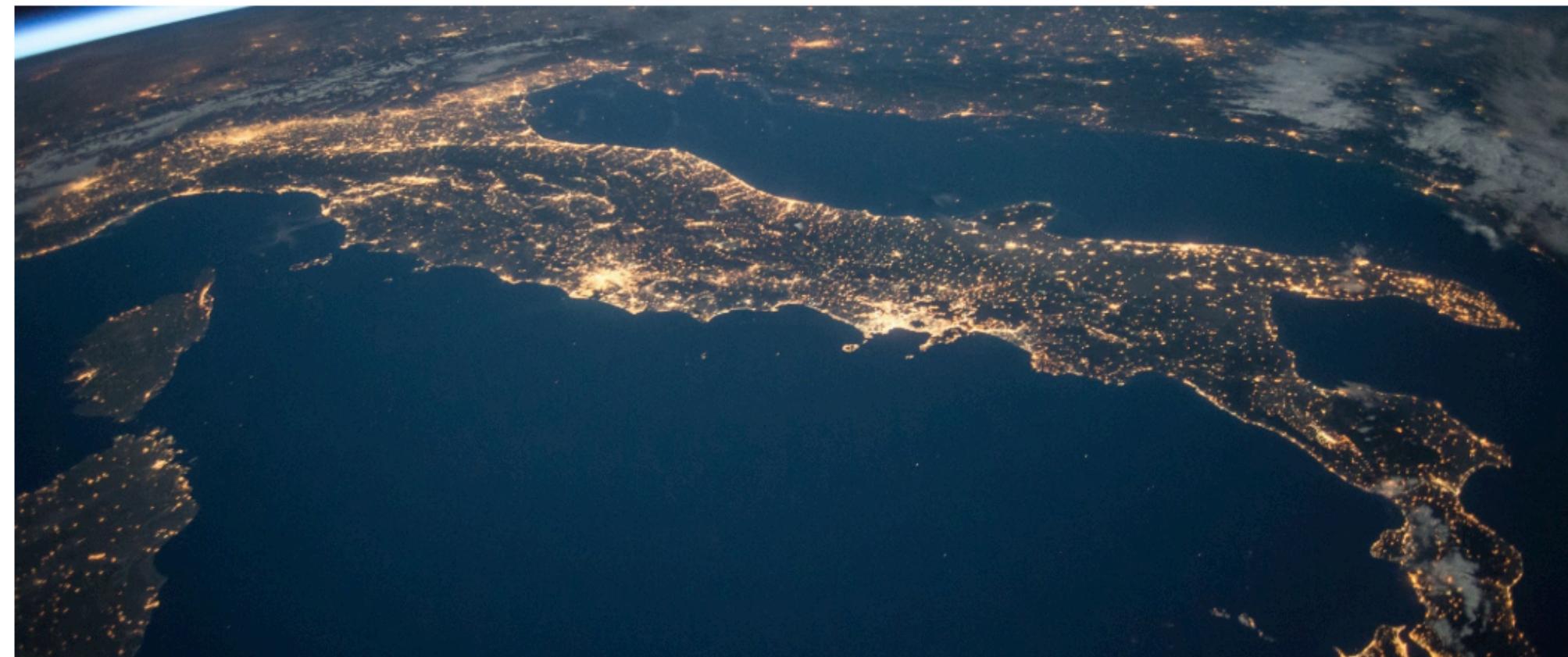


Interpret Results

CMIP6 Cloud Dataset

DATA ANALYTICS

New climate model data now in Google Public Datasets

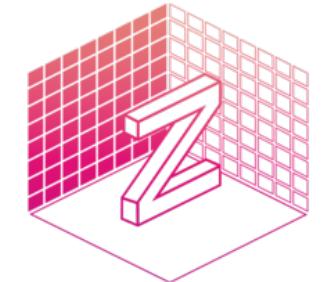


Shane Glass
Program Manager, Google
Cloud Public Dataset Program
December 9, 2019

Exploring [public datasets](#) is an important aspect of modern data analytics, and all this gathered data can help us understand our world. At Google Cloud, we maintain a collection of public datasets, and we're pleased to collaborate with the [Lamont-Doherty Earth Observatory](#) (LDEO) of Columbia University and the Pangeo Project [to host the latest climate simulation data in the cloud](#).

- Pangeo partnered with ESGF and Google Cloud to provide a new public dataset
- > 1 PB and counting
- Data stored in Zarr format
- Google provides free hosting in GCS
- Mirrored on AWS

<https://pangeo-data.github.io/pangeo-cmip6-cloud/>



Zarr

Efficient and fast access to arrays in object store



Easy parallelization -> fast iteration on analysis



Flexible scientific data structure



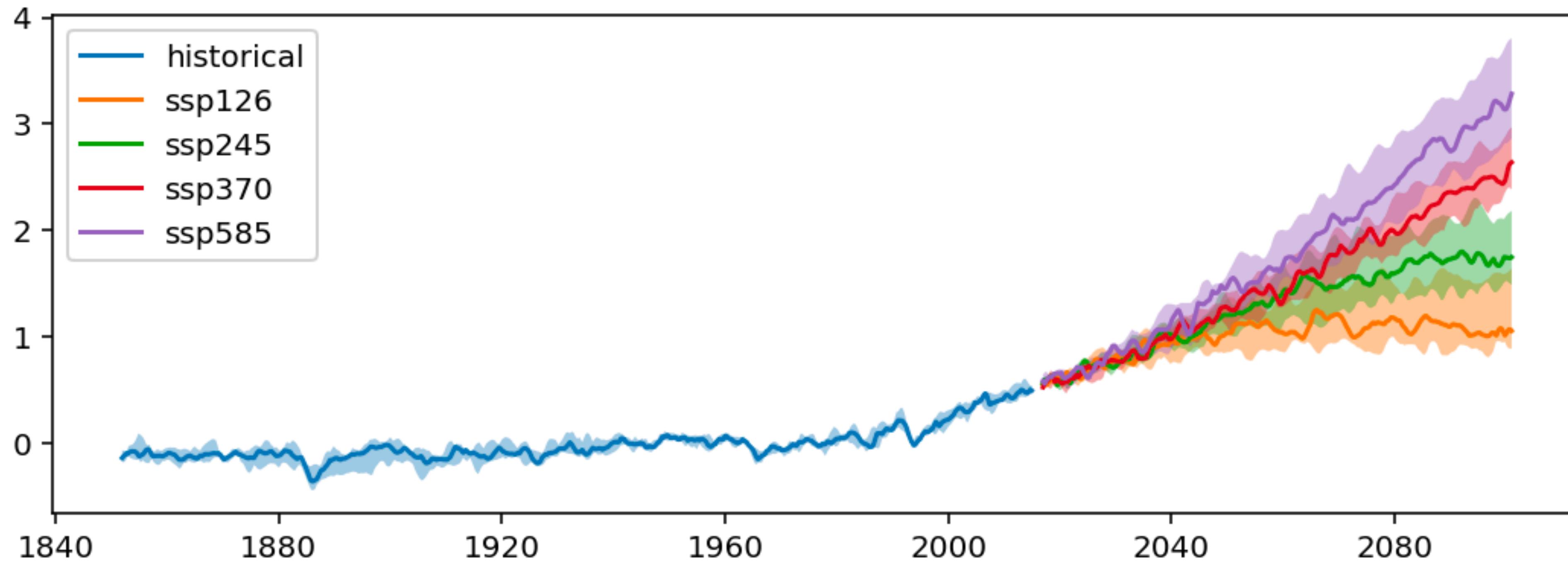
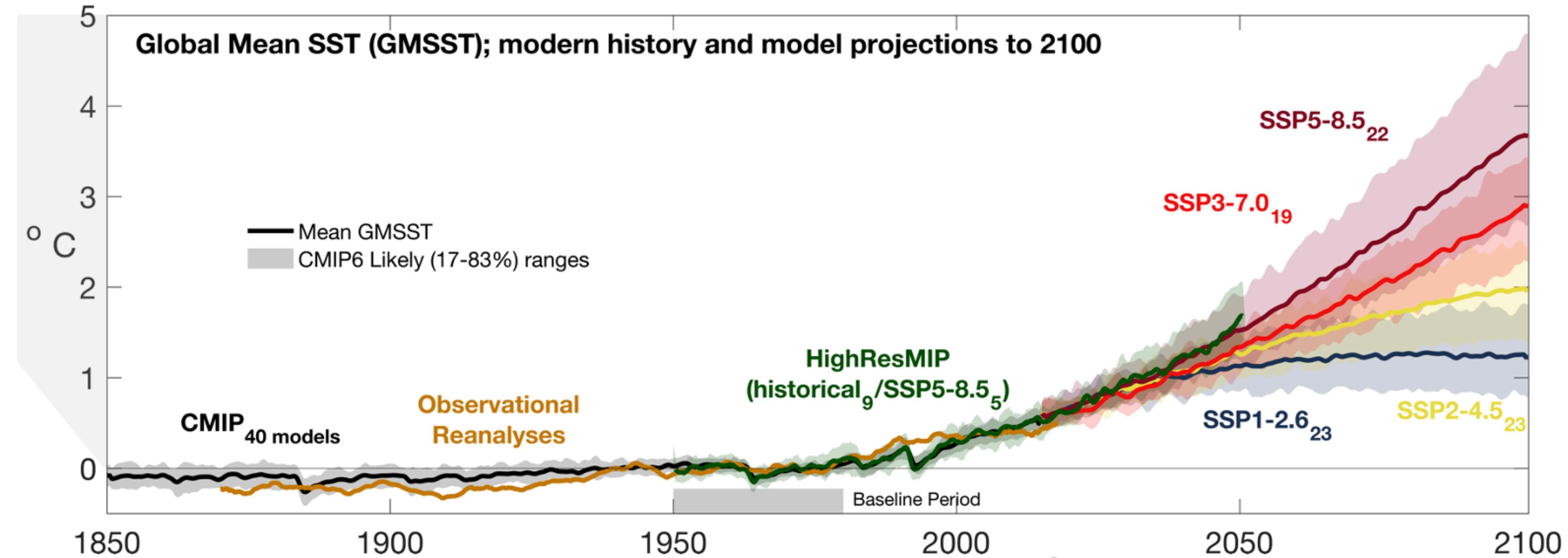
Crowd Sourced Cleaning + Combining



Your favorite code to analyze/interpret

Demo

[**https://github.com/jbusecke/presentation_scipy_2022_cmip**](https://github.com/jbusecke/presentation_scipy_2022_cmip)



Start exploring CMIP6 data now!

If you run into issues with CMIP data please check out xMIP!

The screenshot shows the GitHub repository page for `jbusecke/xMIP`. The repository is public and has 119 stars, 28 forks, and 5 watchers. It contains 125 branches and 57 tags. The repository description is: "Analysis ready CMIP6 data in python the easy way with pangeo tools." Key tags include `cmip6-preprocessing`, `readthedocs.io`, `preprocessing`, `cmip6`, `climate-models`, `climate-analysis`, `pangeo`, `analysis-ready-data`, `xgcm`, and `cmip6-data`. The repository has 540 commits, with the most recent commit being "Unpin sparse and remove numba (#247)" by `jbusecke` on 9ca1606 2 days ago. Other commits listed include "Switch cloud data tests to opt-in (#255)", "Unpin sparse and remove numba (#247)", "batch rename the second (#251)", and "Delete _version.py". The repository also includes files like `.github`, `ci`, `docs`, `notebooks`, `tests`, `xmip`, `.gitattributes`, `.gitignore`, `.pre-commit-config.yaml`, `BLM.png`, `LICENSE.txt`, and `MANIFEST.in`.

jbusecke / xMIP Public

Code Issues 40 Pull requests 4 Discussions Actions Projects 2 Wiki Security Insights Settings

main 125 branches 57 tags Go to file Add file Code

jbusecke Unpin sparse and remove numba (#247) ... 9ca1606 2 days ago 540 commits

.github Switch cloud data tests to opt-in (#255) 2 days ago

ci Unpin sparse and remove numba (#247) 2 days ago

docs batch rename the second (#251) 5 days ago

notebooks batch rename the second (#251) 5 days ago

tests batch rename the second (#251) 5 days ago

xmip Delete _version.py 5 days ago

.gitattributes batch rename the second (#251) 5 days ago

.gitignore batch rename the second (#251) 5 days ago

.pre-commit-config.yaml batch rename the second (#251) 5 days ago

BLM.png Added banner and message 2 years ago

LICENSE.txt Rename LICENSE.TXT to LICENSE.txt 2 years ago

MANIFEST.in batch rename the second (#251) 5 days ago

About

Analysis ready CMIP6 data in python the easy way with pangeo tools.

cmip6-preprocessing.readthedocs.io/...

preprocessing cmip6 climate-models
climate-analysis pangeo
analysis-ready-data xgcm cmip6-data

Readme MIT license 119 stars 5 watching 28 forks

Releases 31

v0.6.1rc0 Latest 5 days ago + 30 releases