

Visualizing Climate Data

Duncan Watson-Parris, SIO/HDSI

DSC 106: 1st April 2026

Overview

- Introduce three example datasets:
 - CMIP Climate model simulations
 - Geostationary Operational Environmental Satellite (GOES)
 - NASA MODIS Satellite Imagery
- Example data processing scripts for each

CMIP6 Overview

- CMIP6 (Coupled Model Intercomparison Project Phase 6) is a global collaboration to advance understanding of the climate system through coordinated simulations.
- Provides the core datasets underpinning IPCC AR6 and UNFCCC assessments.
- Enables systematic comparison and evaluation of Earth System Models (ESMs) developed by research centers worldwide.



How It's Organized

- Participating centers: >50 modeling groups contribute ESMs with unique representations of the climate system.
- Experimental framework:
 - DECK (Diagnostic, Evaluation and Characterization of Klima): standard baseline experiments (e.g. preindustrial control, abrupt CO₂).
 - Historical simulations: reproduce observed 1850–present climate.
 - MIPs (Model Intercomparison Projects): focused experiments addressing distinct scientific questions, e.g.:
 - ScenarioMIP – Future socioeconomic pathways and emission scenarios.
 - DAMIP – Detection and attribution of historical climate change.
 - AerChemMIP – Aerosol–chemistry interactions.
 - LS3MIP / C4MIP / HighResMIP, etc., targeting land, carbon, and high-resolution processes.
- Multiple realizations per experiment quantify internal variability

What CMIP6 Doesn't Provide

- CMIP6 is not designed to provide deterministic forecasts or predictions.
- Relatedly, it does not capture short-term weather variability.
- Represents plausible *futures* conditioned on assumptions, not guaranteed outcomes.

CMIP6 for Data Science

- It consists of a 14PB of neatly (mostly!) organized gridded data
- Why It's Useful:
 - Time series analysis across scenarios and models.
 - Spatial analysis for extremes, teleconnections, and climate indices.
 - Benchmarking machine learning emulators, uncertainty quantification, explainable AI.
- Common Data Formats:
 - NetCDF, consistent metadata using CF conventions, often accessed through OPeNDAP or cloud-optimized Zarr.
- Tools & Platforms:
 - Ideal for use with xarray + dask for out-of-core and parallel workflows.
 - Data also accessible via Pangeo, Google Cloud, AWS Registry of Open Data, NCAR Glade, etc.

Example Visualizations and Applications

- IPCC Atlas: <https://interactive-atlas.ipcc.ch/regional-information>
- CMIP Visualization tool: <https://cmip6.science.unimelb.edu.au/>
- This is the data we trained the ClimateBench emulators on. You can play with a ClimateBench emulator here: [Streamlit App](#)

GOES Imagery – What is GOES?

- GOES: The latest generation of Geostationary Operational Environmental Satellites, developed by NOAA and NASA.
- Includes GOES-16 (East), GOES-17 (West), GOES-18, and GOES-U (future).
- Positioned in geostationary orbit (about 36,000 km altitude) — always watching the same part of Earth.
- Provides continuous, near-real-time imagery of weather, atmosphere, oceans, and the Sun.

GOES Imagery – What is GOES?



Why is GOES Important

- Critical for severe weather forecasting (e.g. hurricanes, thunderstorms, wildfires).
- Supports aviation safety, air quality monitoring, solar/space weather, and disaster response.
- Up to 60x faster and 4x higher resolution than previous GOES generations.
- Enabled tools like lightning mapping, volcanic ash detection, and fog identification.

Key Instruments on GOES

- ABI (Advanced Baseline Imager): 16-channel multispectral imaging (visible to infrared) — most widely used data stream.
- GLM (Geostationary Lightning Mapper): First-ever lightning mapper in geostationary orbit.
- SUVI (Solar Ultraviolet Imager) and EXIS: Monitor solar flares and space weather.
- Additional sensors: For particle flux, magnetometers, X-ray flux, etc.

GOES-R Data for Data Science

- Spatial resolution: Up to 0.5 km (visible bands).
- Temporal resolution: Full-disk scans every 10 minutes; targeted regions every 30 seconds.
- Provides raw and derived products: cloud-top heights, fire hotspots, precipitation estimates, etc.
- Full list of available products is here: <https://github.com/NOAA-Big-Data-Program/nodd-data-docs/tree/main/GOES>
- Data is open access through platforms such as NOAA CLASS, Google Cloud, and Amazon AWS Open Data Registry.

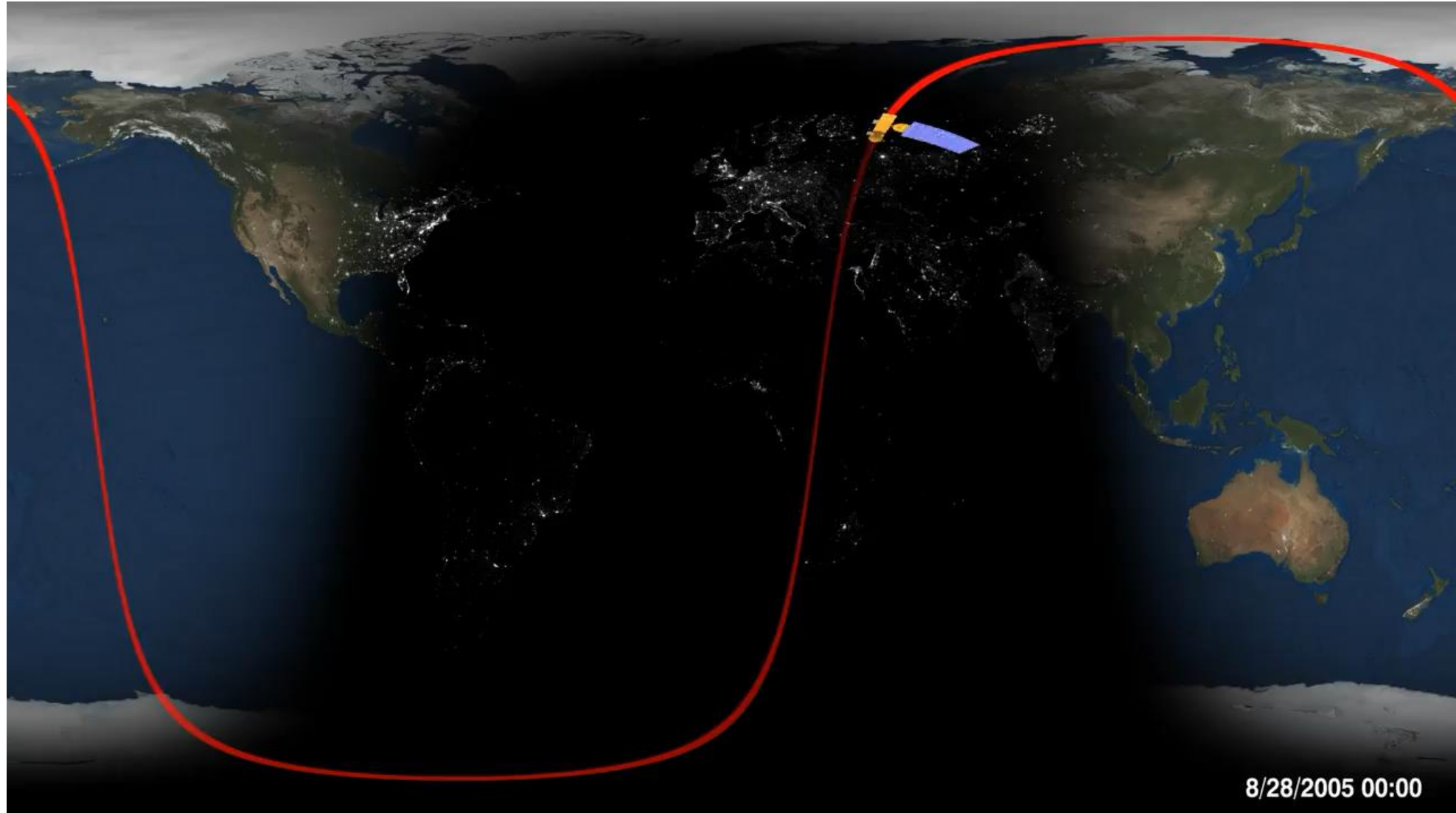
Example Visualizations

- GOES Viewer: <https://www.star.nesdis.noaa.gov/goes/index.php>
- Example recent snapshot:
https://www.linkedin.com/posts/simon-proud_hurricane-melissa-activity-7388980160961617920-ikhC?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAag6MFYBDjKQIQqmDGsBNmcyMmAKZnUQONo

What is MODIS?

- MODIS: Moderate Resolution Imaging Spectroradiometer.
- Onboard NASA's Terra (1999) and Aqua (2002) satellites — part of the Earth Observing System (EOS).
- Provides global coverage every 1–2 days.
- Continuously collects data across 36 spectral bands (from visible to thermal infrared).

Polar orbit



Why is MODIS Important?

- A foundational source for long-term, global environmental monitoring.
- Tracks land, ocean, and atmospheric properties: vegetation, sea surface temps, aerosols, snow cover, fires, and more.
- Enabled thousands of scientific studies since launch — central to climate, ecology, oceanography, and disaster science.
- Data continuity for over two decades has made MODIS a “gold standard” reference dataset.

Key MODIS Characteristics

- Spectral coverage: 0.4 μm to 14.4 μm across 36 bands.
- Spatial resolution: 250 m (bands 1–2), 500 m (bands 3–7), and 1 km (remaining bands).
- Temporal coverage: Near-daily global revisit with Terra and Aqua combined.
- Products: >40 standard products, including:
 - LAI/FPAR (Leaf Area Index / Photosynthesis)
 - NDVI (Vegetation index)
 - LST (Land Surface Temperature)
 - Aerosol Optical Depth
 - Burned area, snow cover, clouds, etc.

MODIS Data for Data Science

- Accessible via NASA's LAADS DAAC, Google Earth Engine, AWS Open Data, and more.
- Rich source for data science projects involving time series analysis, classification, anomaly detection, and spatial modeling.
- Often used for trend detection (e.g. greening, drought), hazard mapping, and AI training data (e.g. for cloud masking or land cover models).
- This is the data we used for the ship-track detection work discussed in the last lecture

Example Visualizations

- NASA Worldview: <https://worldview.earthdata.nasa.gov>
- Image of the day: <https://modis.gsfc.nasa.gov/gallery>
- Other idea: Make a geolocator game?

Working with the datasets

The xarray library

- xarray is a Python library for working with labeled, multi-dimensional arrays, built on NumPy.
- Designed for efficient handling of gridded climate and Earth science data (e.g. NetCDF, GRIB, Zarr).
- Works seamlessly with data formats from CMIP6, GOES, MODIS, etc.
- Deep integration with:
 - pandas: shares the same indexing and alignment concepts.
 - dask: supports parallel and out-of-core computation for big datasets.

xarray vs pandas

xarray brings **pandas-like usability** to **n-dimensional arrays**, enabling intuitive slicing, grouping, and computation based on **named dimensions** (e.g. "time", "lat", "lon").

Feature	pandas (DataFrame)	xarray (DataArray / Dataset)
Data Type	2D tabular data	N-dimensional labeled arrays (e.g. time, lat, lon)
Indexing	1 index (rows), columns	Labeled axes (dims) and coordinates
Use Case	Flat, relational data	Geospatial, time series, climate model output
Example Shape	Rows × Columns	Time × Latitude × Longitude × Level (4D)
Data I/O	CSV, HDF5, SQL	NetCDF, GRIB, Zarr, OPeNDAP, etc.

Navigating CMIP6 Data

- CMIP6 data files are highly structured to ensure global consistency and reproducibility.
- Each file name and directory encodes key metadata describing who, what, and how the simulation was run:

```
/activity_id/institution_id/source_id/experiment_id/  
member_id/table_id/variable_id/
```

- Note, that not all institutions (or models) ran all experiments, or provide all the same variables

Navigating CMIP6 Data: terminology

Term	Meaning	Example
activity_id	The <i>MIP</i> or <i>project</i> organizing the experiment	ScenarioMIP, DAMIP, C4MIP
institution_id	Modeling center or institution	NCAR, MOHC, IPSL
source_id	Specific model name/version	CESM2, UKESM1-0-LL
experiment_id	Experiment being run	historical, ssp585, piControl
member_id	Realization or ensemble member	r1i1p1f1 (run/init/physics/forcing)
table_id	Variable frequency & domain table	Amon (monthly atmosphere), Omon (monthly ocean), day, etc.
variable_id	Actual physical variable stored	tas (surface temperature), pr (precipitation), clt (cloud fraction)

More details can be found [here](#)

MODIS API

- There are extensive docs covering the various MODIS APIs
- A simple starting point is the Tile Web Map Service endpoint though.
- You can ask for specific geographic regions based on a pre-defined tile grid. E.g.:
 - https://gibs.earthdata.nasa.gov/twms/epsg4326/best/twms.cgi?request=GetMap&layers=MODIS_Terra_CorrectedReflectance_TrueColor&srs=EPSG:4326&format=image/jpeg&styles=&time=2012-06-08&width=1024&height=1024&bbox=-118.1250,32.6250,-117.0000,33.7500

What are some interesting questions?

Dataset	Example questions you can answer
CMIP	<ul style="list-style-type: none">- How many days over 100F might Los Angeles experience now vs. 50 years from now?- How different is that if we do nothing (SSP2.45) or aggressively mitigate (SSP1.26)
GOES	<ul style="list-style-type: none">- Can you predict how much sunlight a given location will experience 2 hours from now?- What does the diurnal (daily) cycle of clouds look like in San Diego? Seattle?
MODIS	<ul style="list-style-type: none">- How has the amount of clouds changed over the last 20 years in different locations?- Maybe a geolocation game?