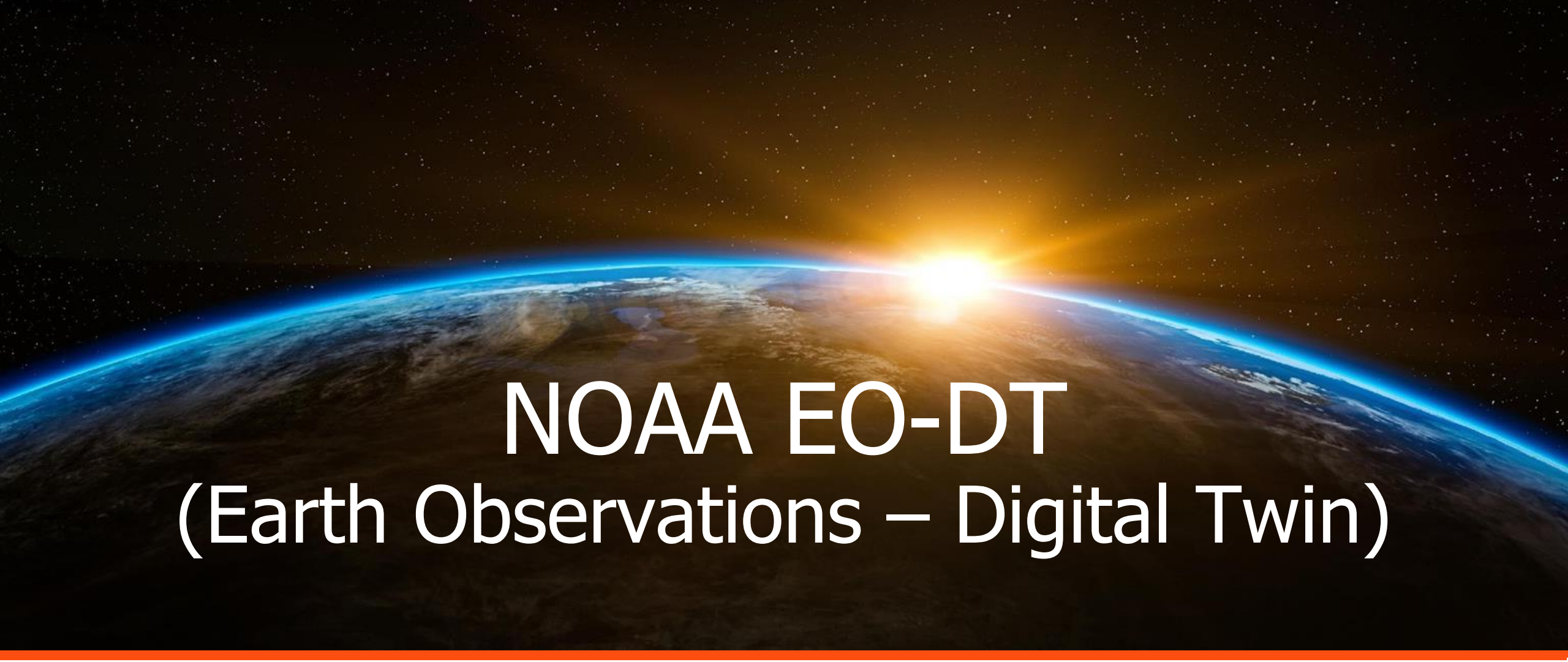


NOAA EO-DT
(Earth Observations – Digital Twin)

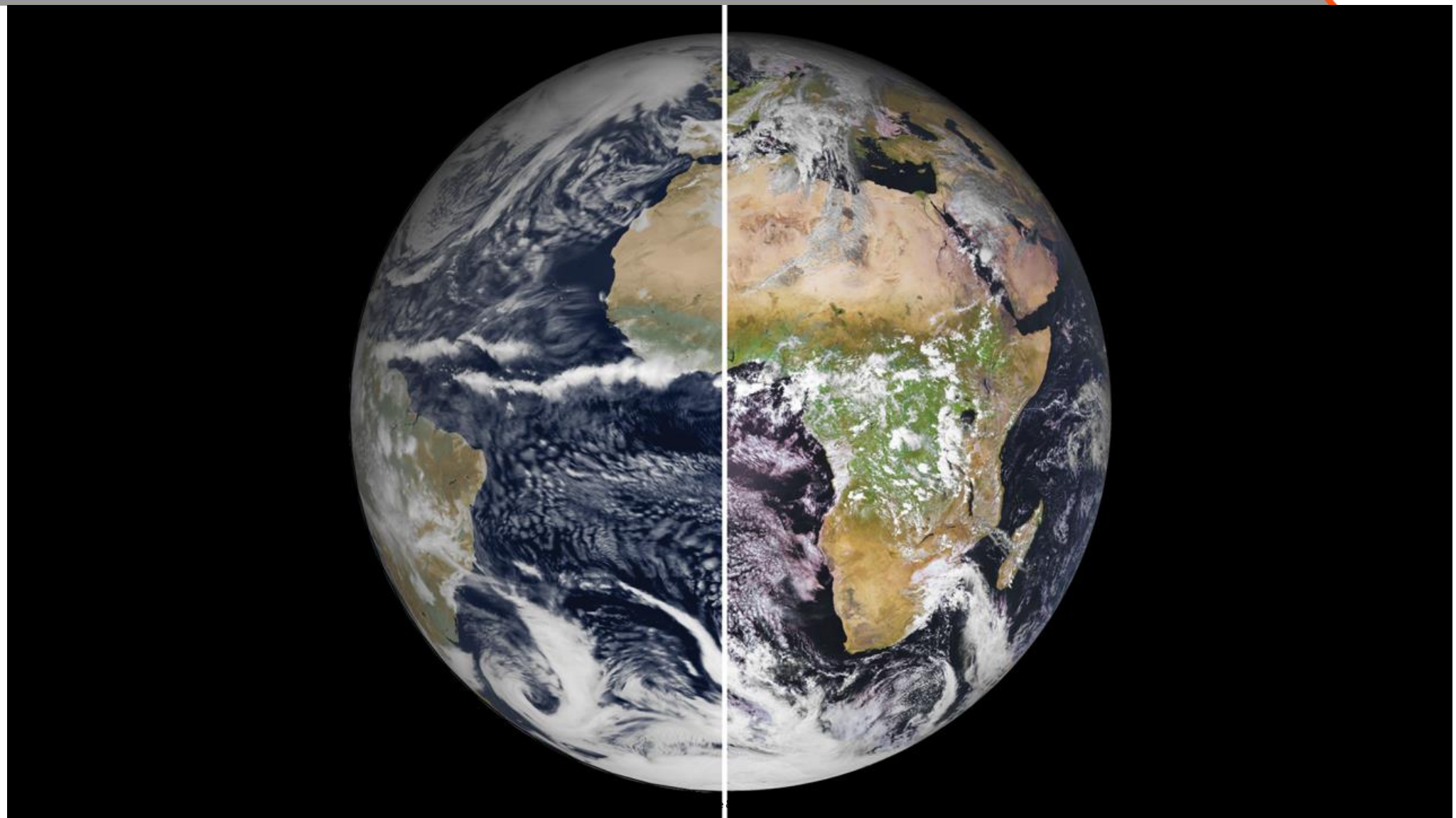
orion SPACE SOLUTIONS
An Arcfield Company

PIs: Patrick McBride, Jeff Steward
Contributors: John Furlong, Ryan Kelly, Rachel Stutz, Jeremy Highley, Russell Cox, Ryan Nguyen
Program Manager: Jeff Campbell, Katie Nyland
Portfolio Lead: Junk Wilson

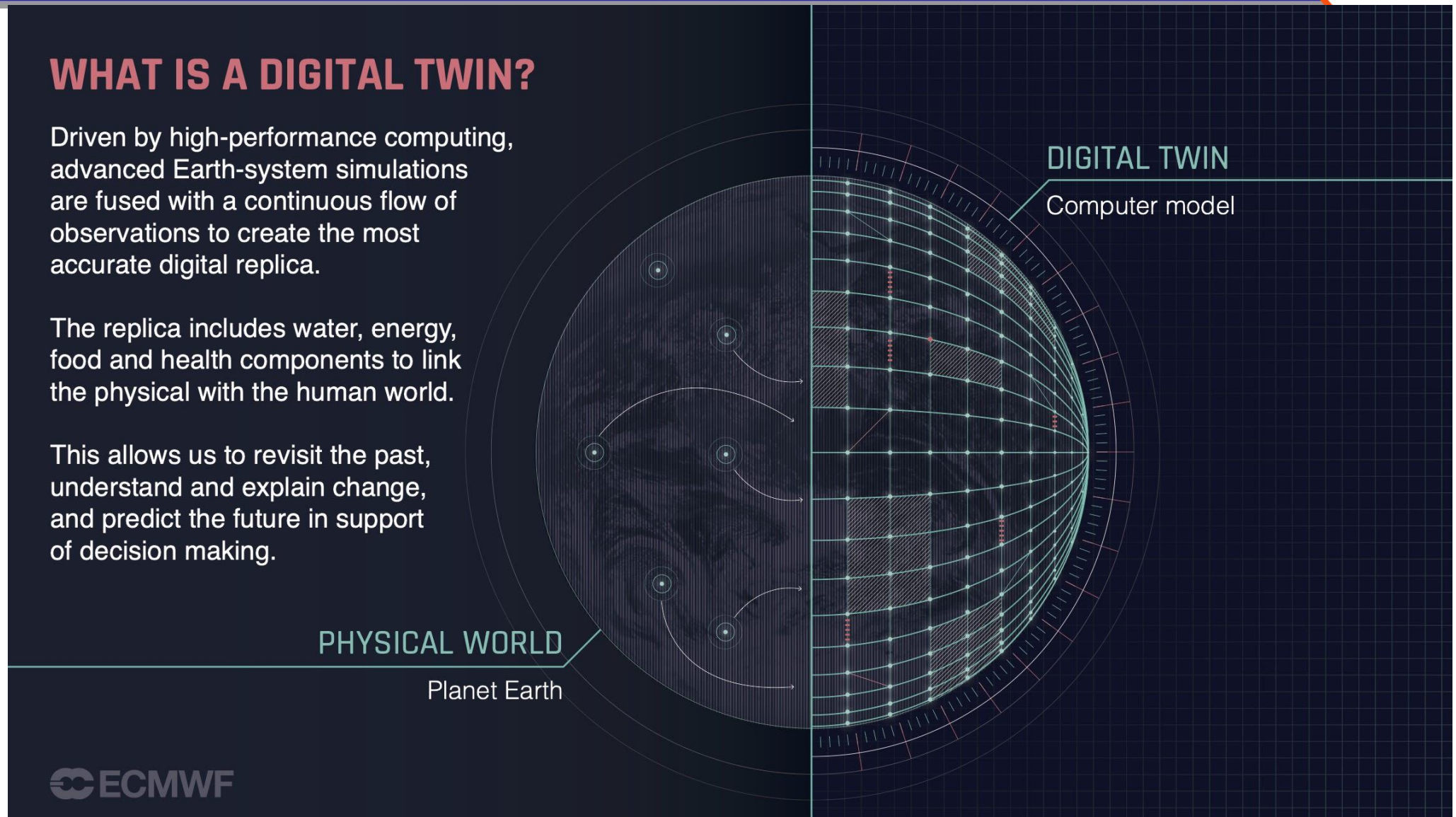
February 1, 2024

Earth Observation Digital Twin

orion
SPACE SOLUTIONS
An Arcfield Company

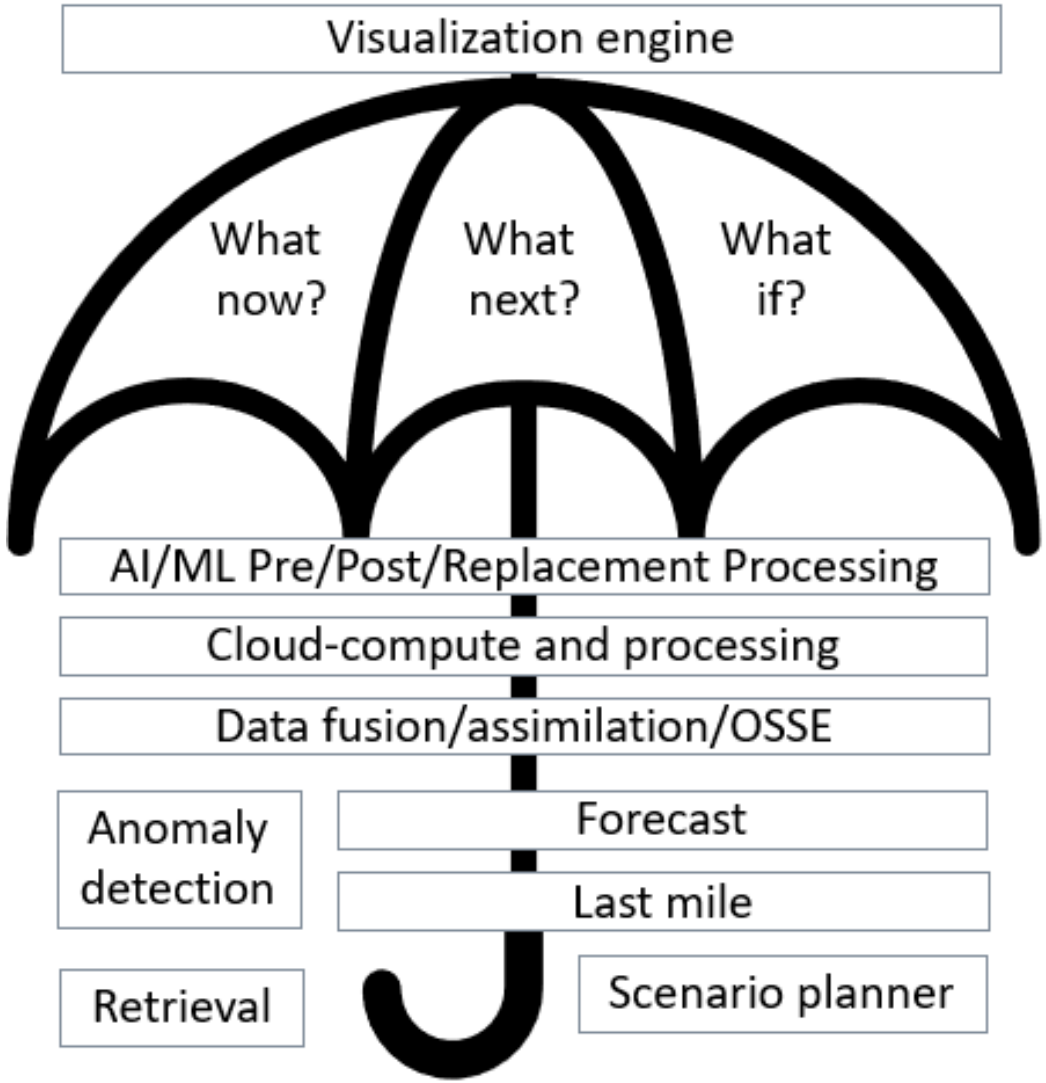


Earth Observation Digital Twin



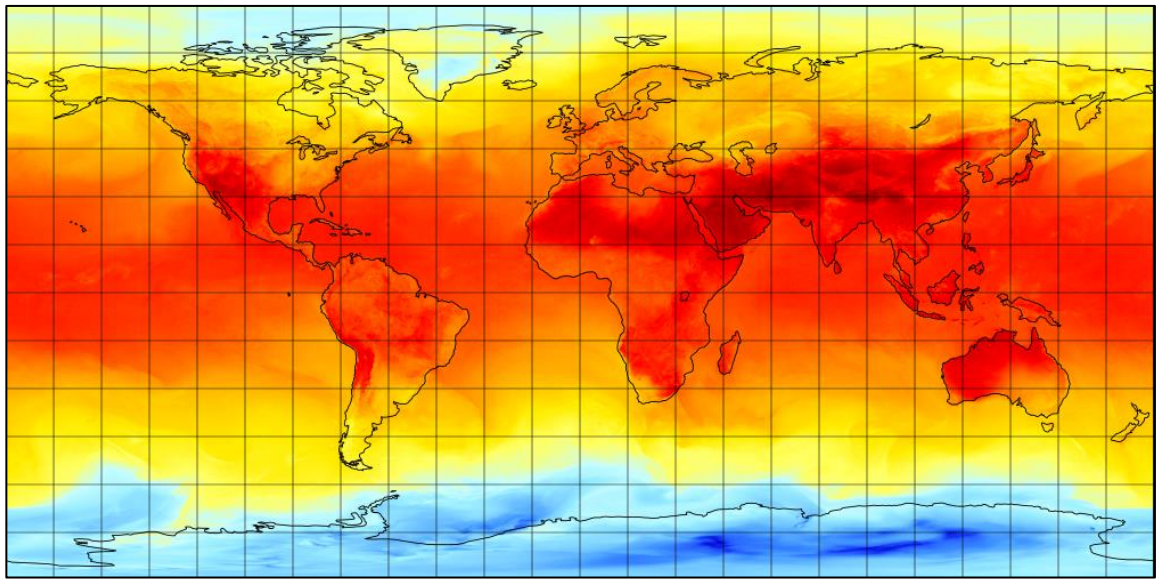
Digital Twin – unifying vision

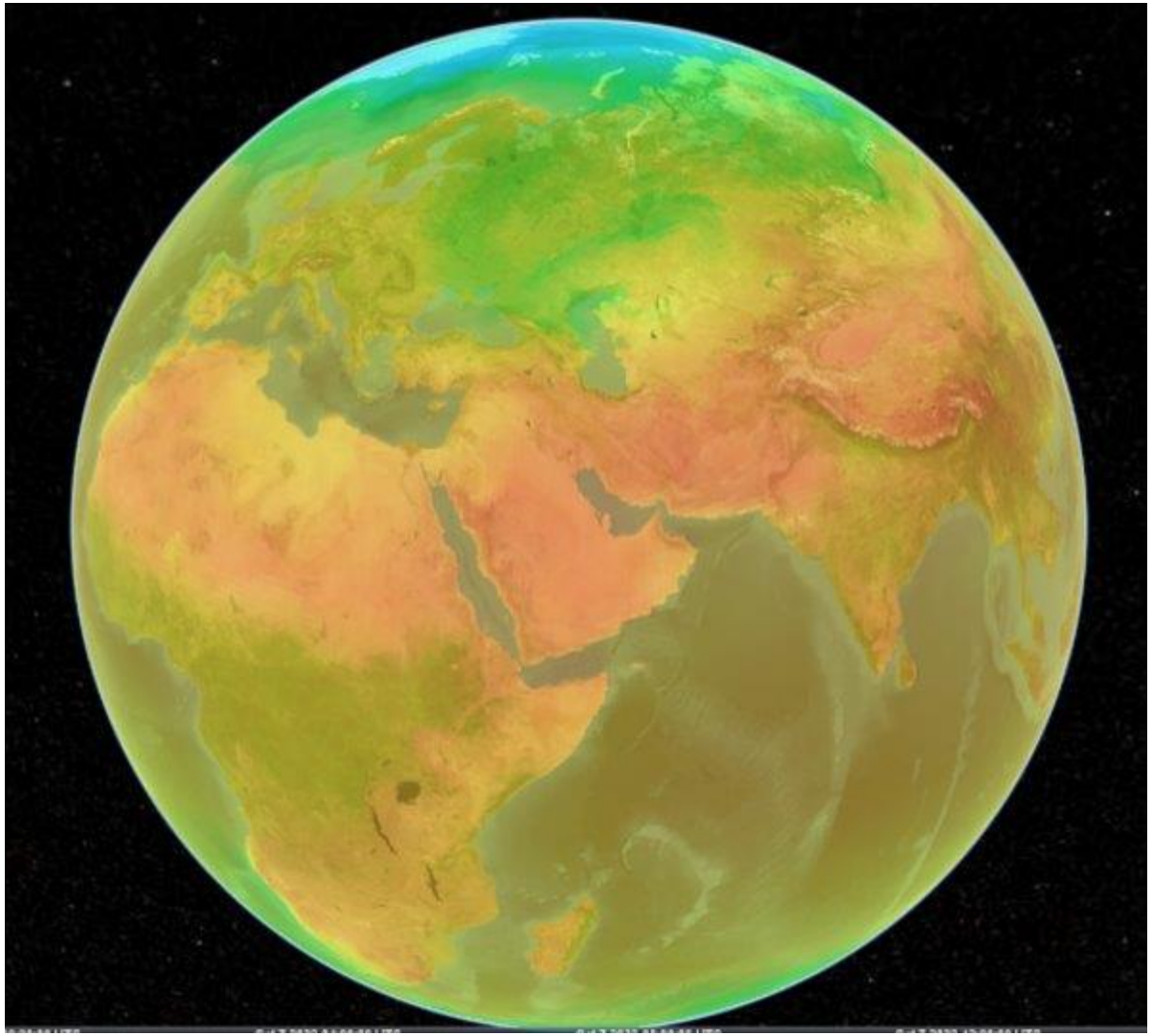
- What is a digital twin?
- Visualization engine/platform
- What now?
  - Current best estimate of the system (Observations and models)
- What next?
  - Forecast of the system
- What if?
  - Interactive exploratory scenario planning
- Computational infrastructure:
  - AI/ML, cloud compute
  - Data fusion and assimilation, OSSE
  - Support decision making, analysis, and other user applications



Visualization in CesiumJS

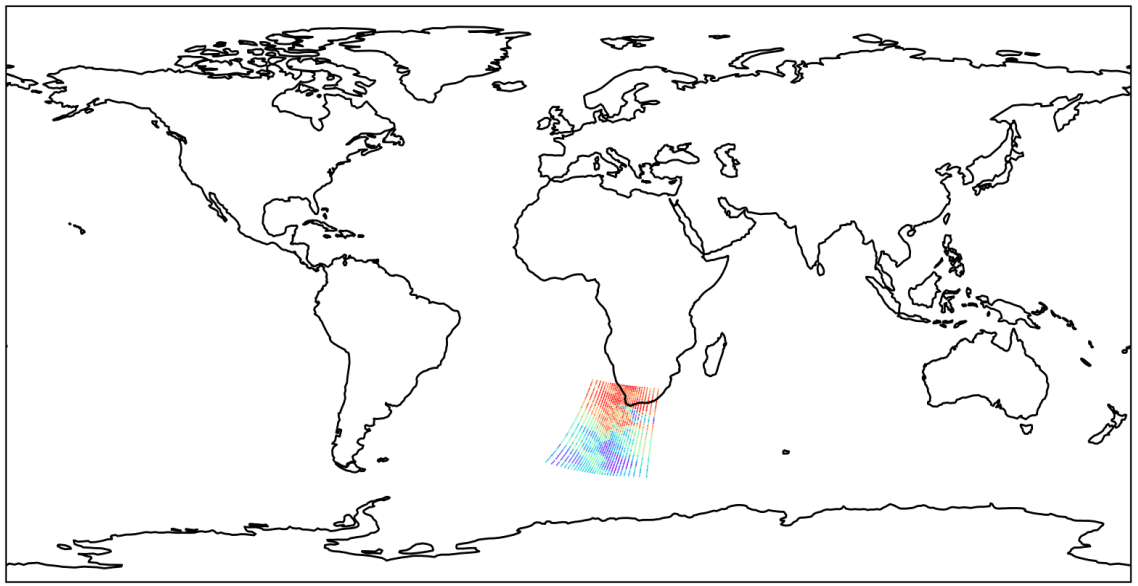
- NOAA Global Forecast System (GFS) weather model
  - Full globe, 0.25° resolution data for dozens of variables
  - Viewing: surface temperature on 9/10/23

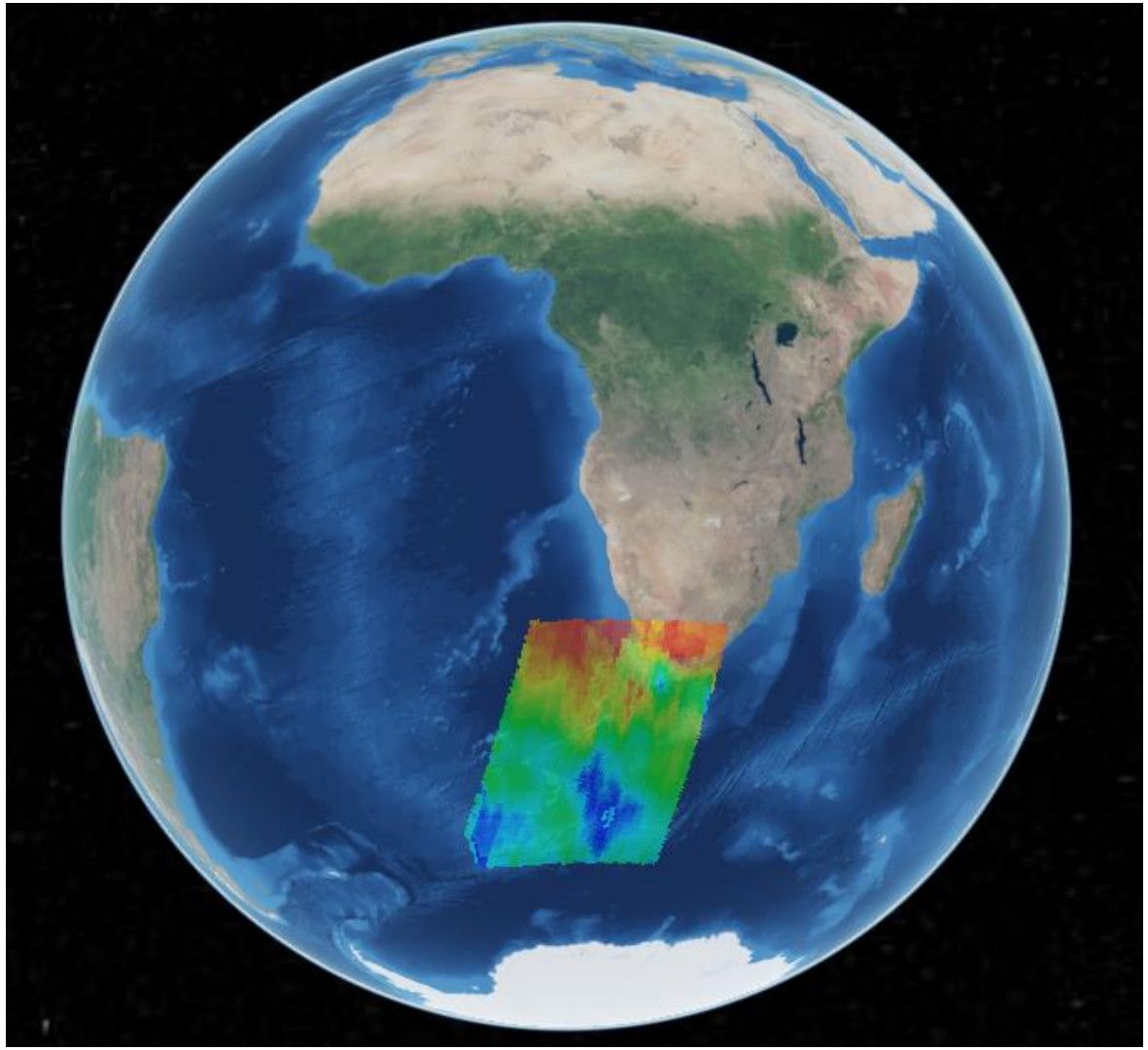




Visualization in CesiumJS

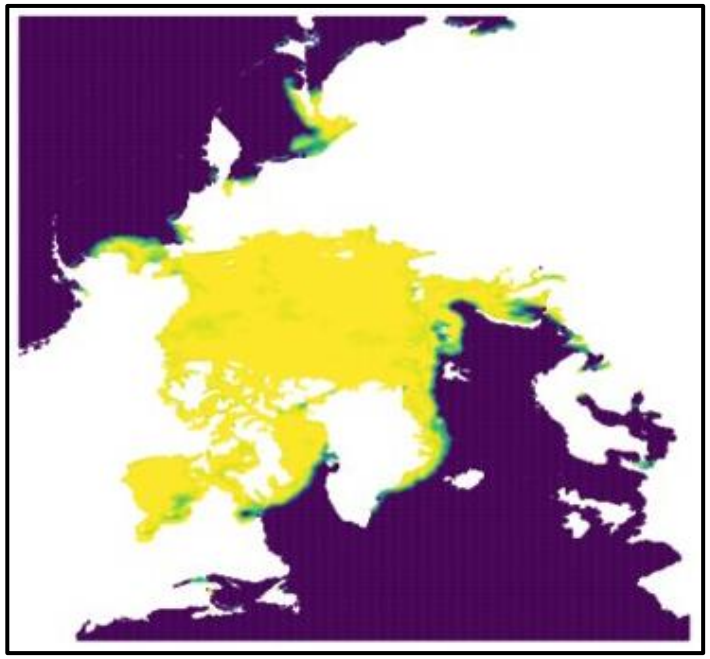
- JPSS NOAA-20 Cross-track Infrared Sounder (CrIS)
  - Swaths of atmospheric temperature and moisture data
  - Viewing: brightness temperature from ch. 551 (992.5cm^{-1}) on 1/9/23

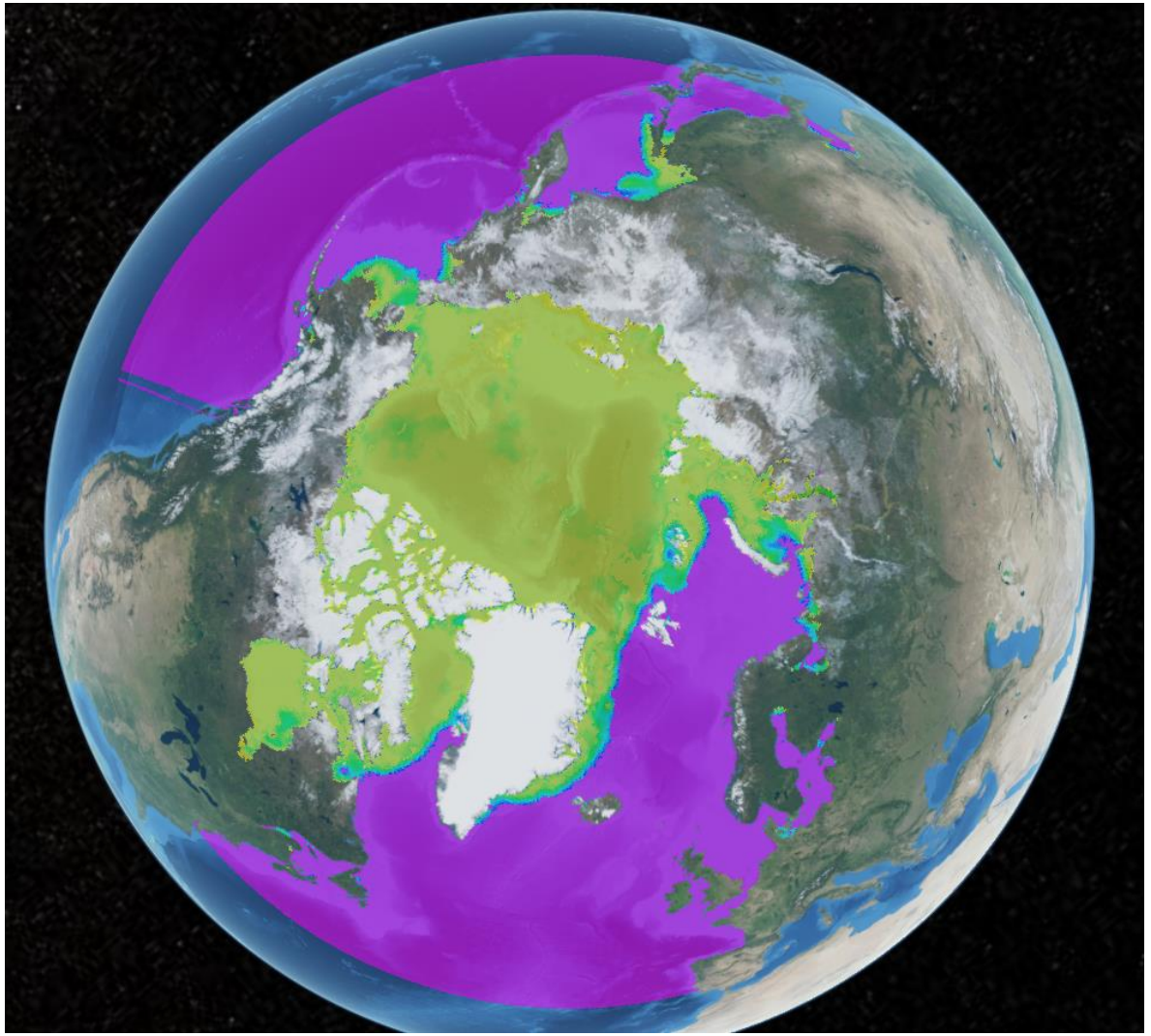




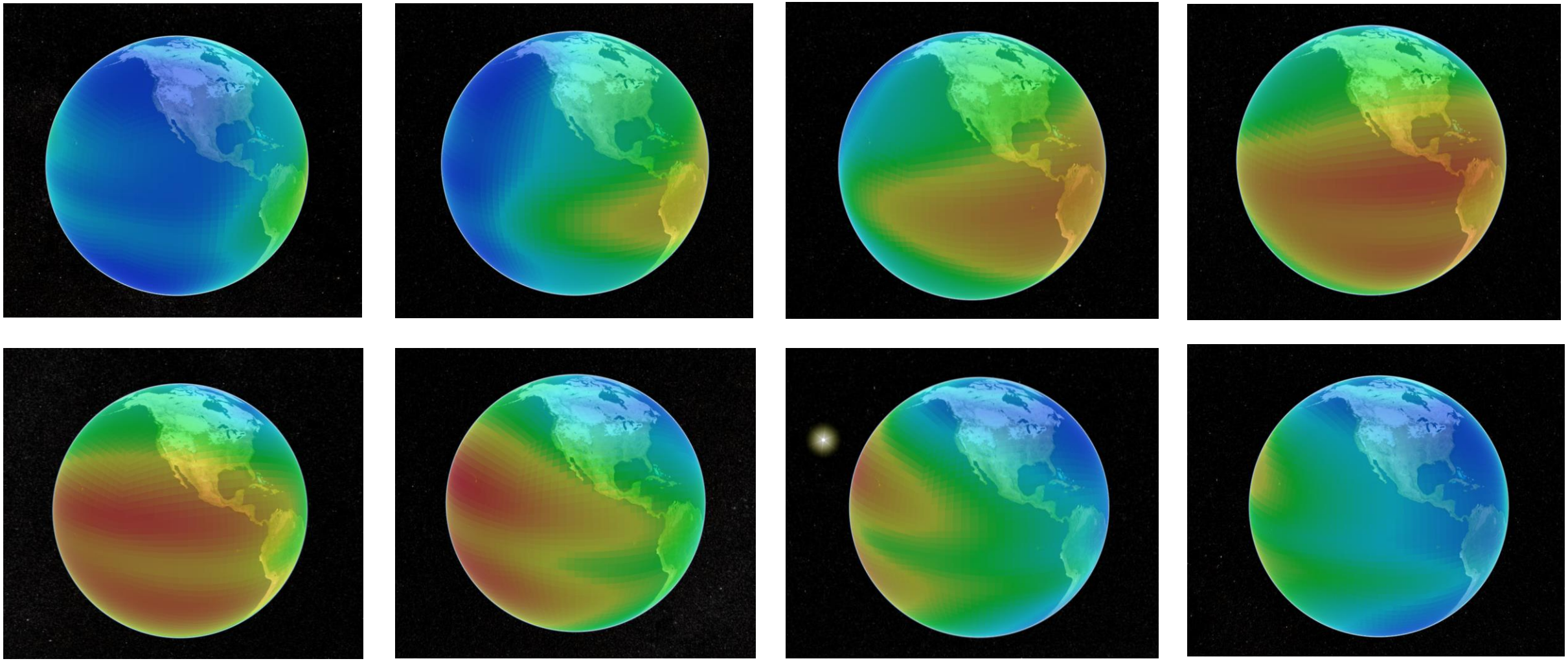
Visualization in CesiumJS

- NOAA/NSDIC Sea Ice Concentration data products
  - Northern hemisphere only with data masked out over land
  - Viewing: sea ice concentration 1/1/23



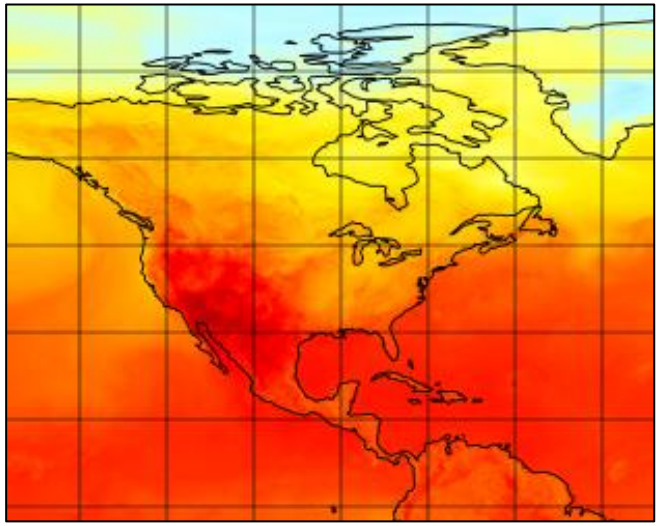


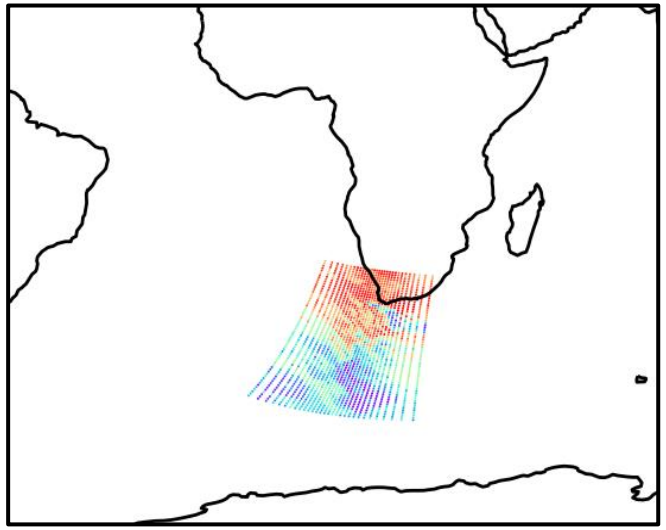
Visualization in CesiumJS

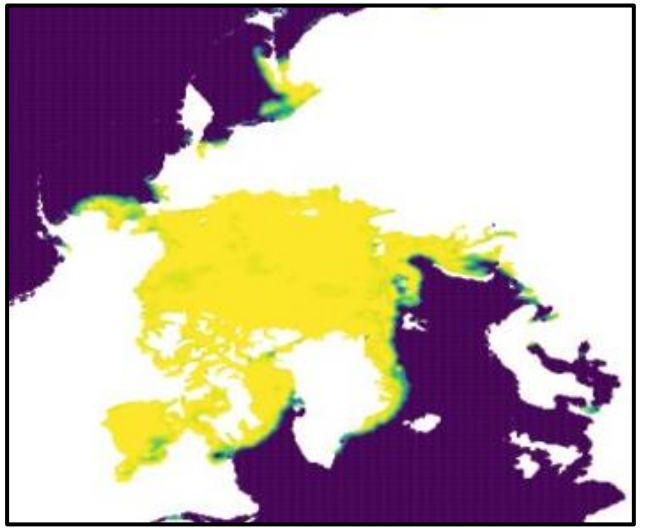


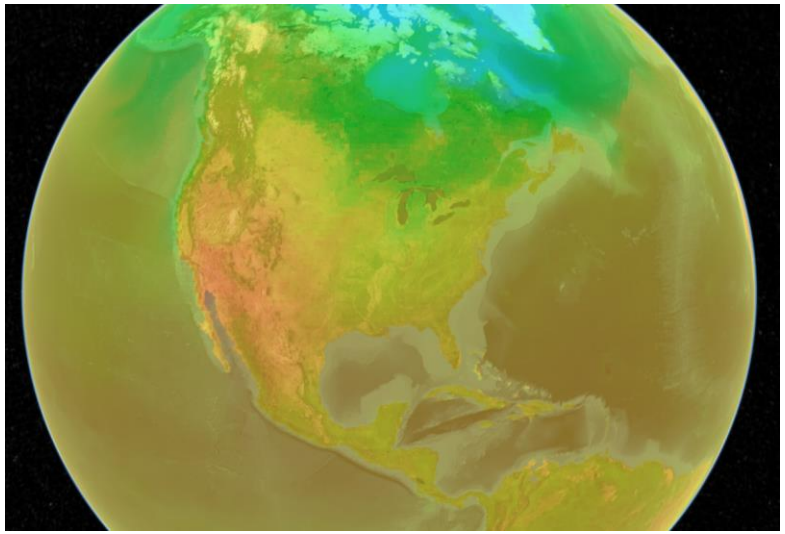
Visualization in CesiumJS

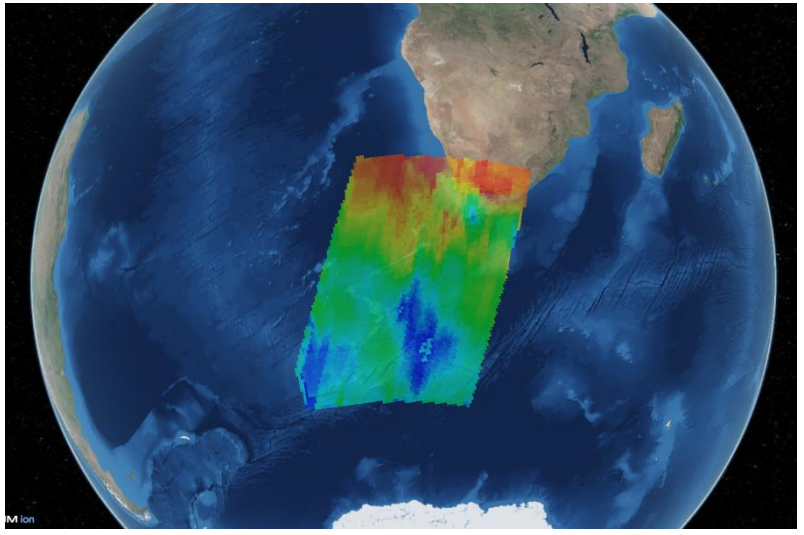
orion SPACE SOLUTIONS
An Arcfield Company

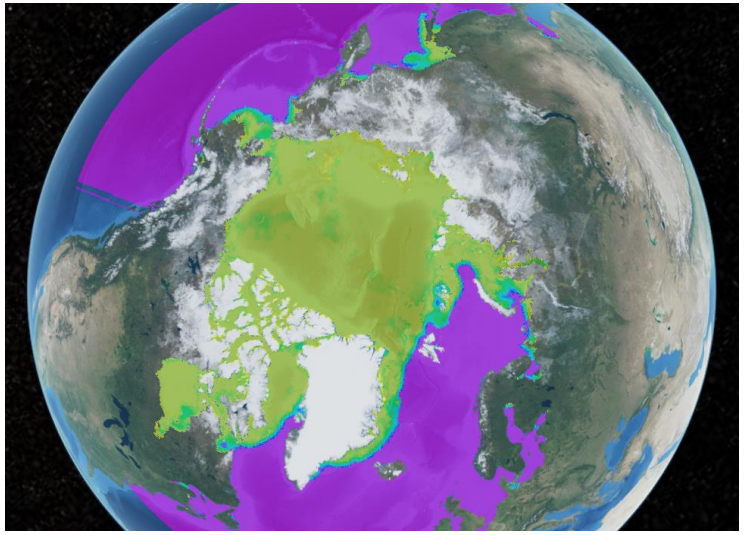






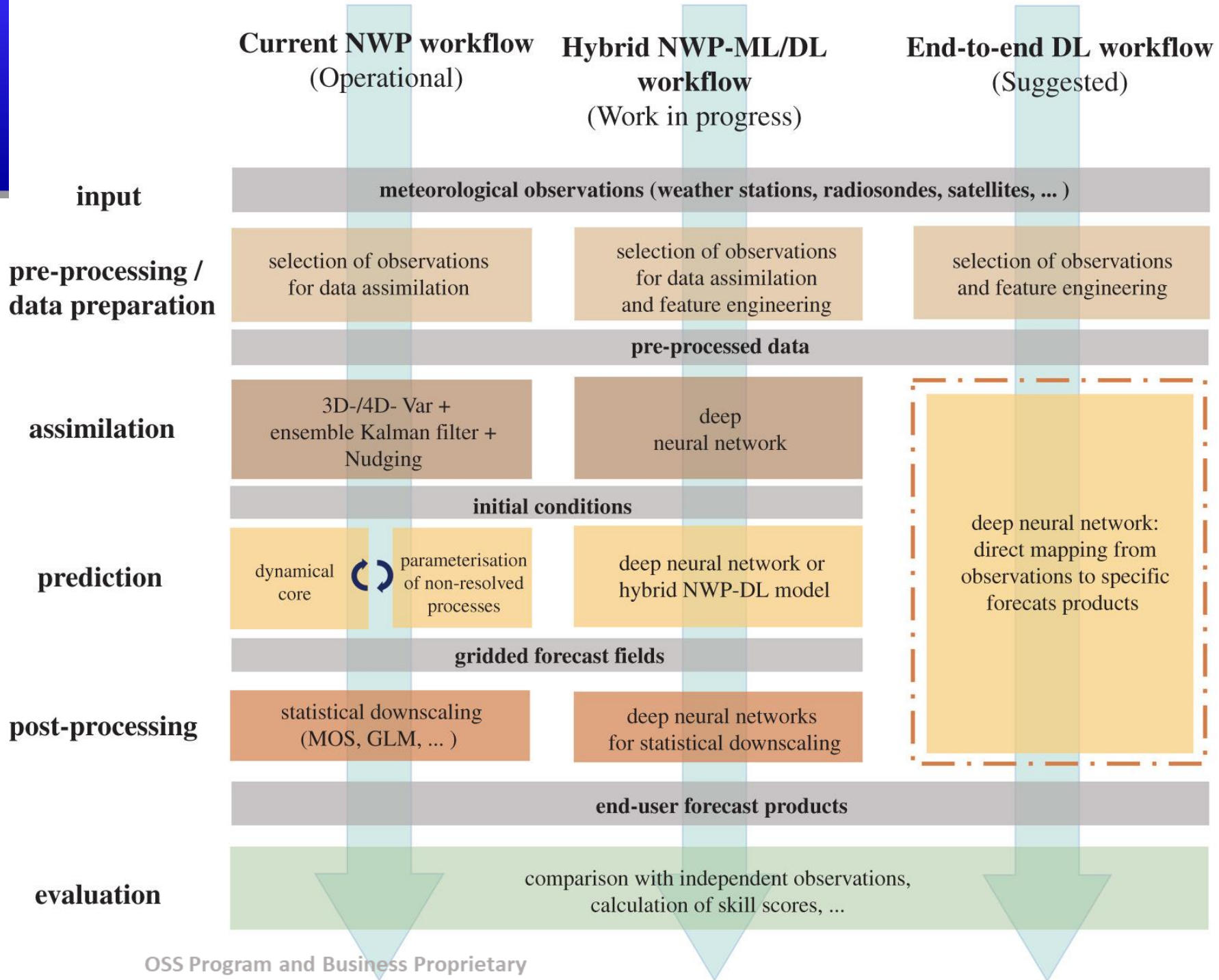




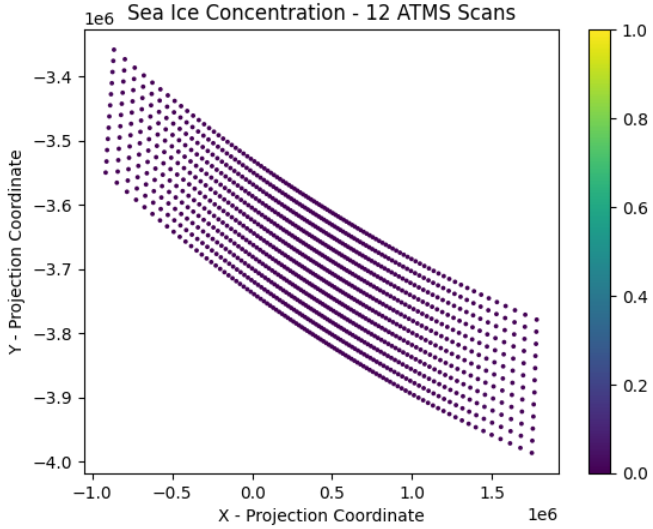


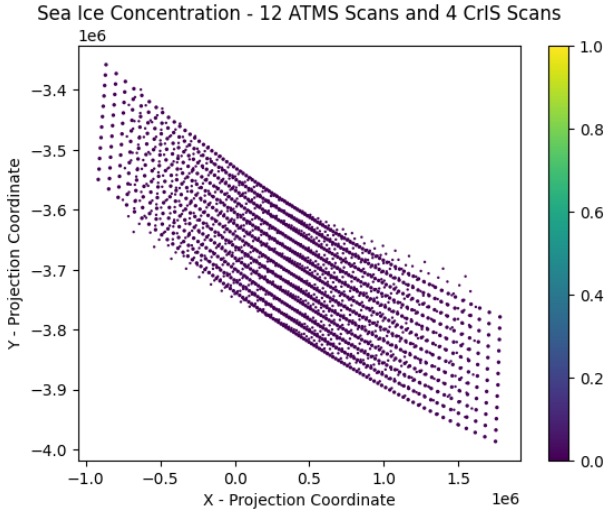
AI/ML

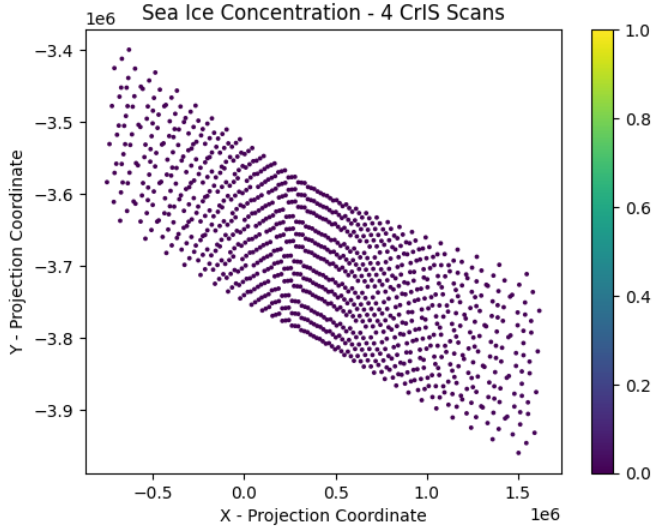
- The role of AI/ML in NWP
- Multiple areas for inclusion
- Need fundamental research, V&V to assure science community
- Graphic from ”Can deep learning beat numerical weather prediction?” Schultz et al 2021 https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0097

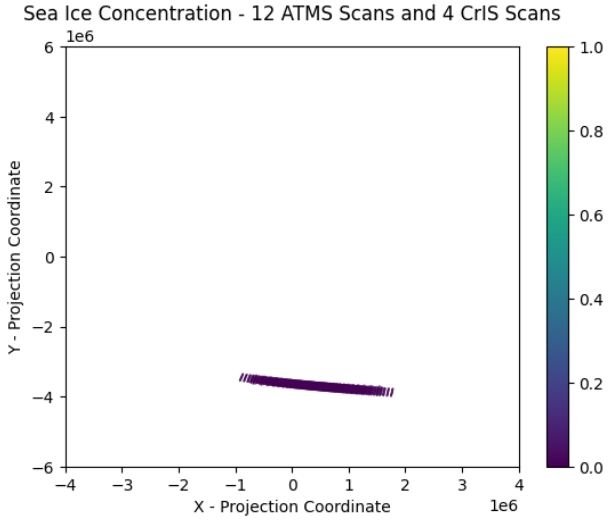


Sea Ice + ATMS + CrIS Paired Data









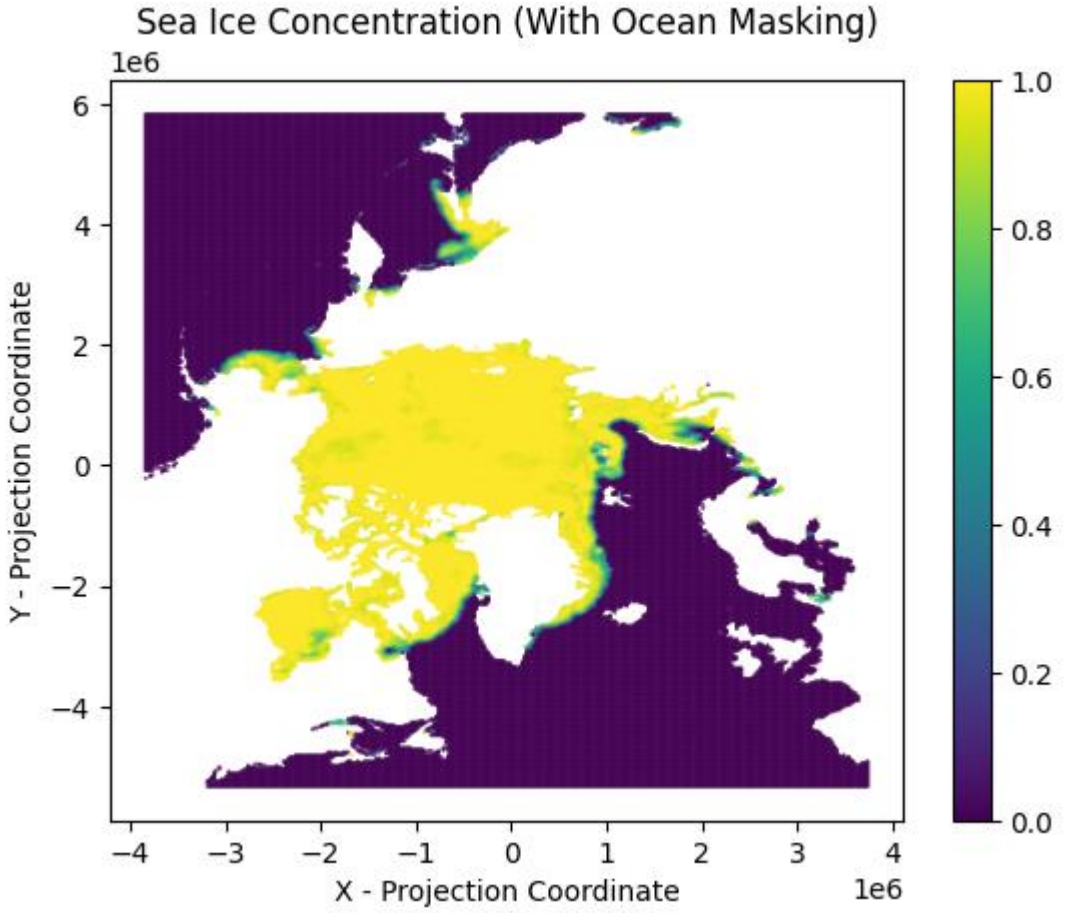
Sea Ice Concentration – Training Labels

Sea Ice Data on January 1st, 2023

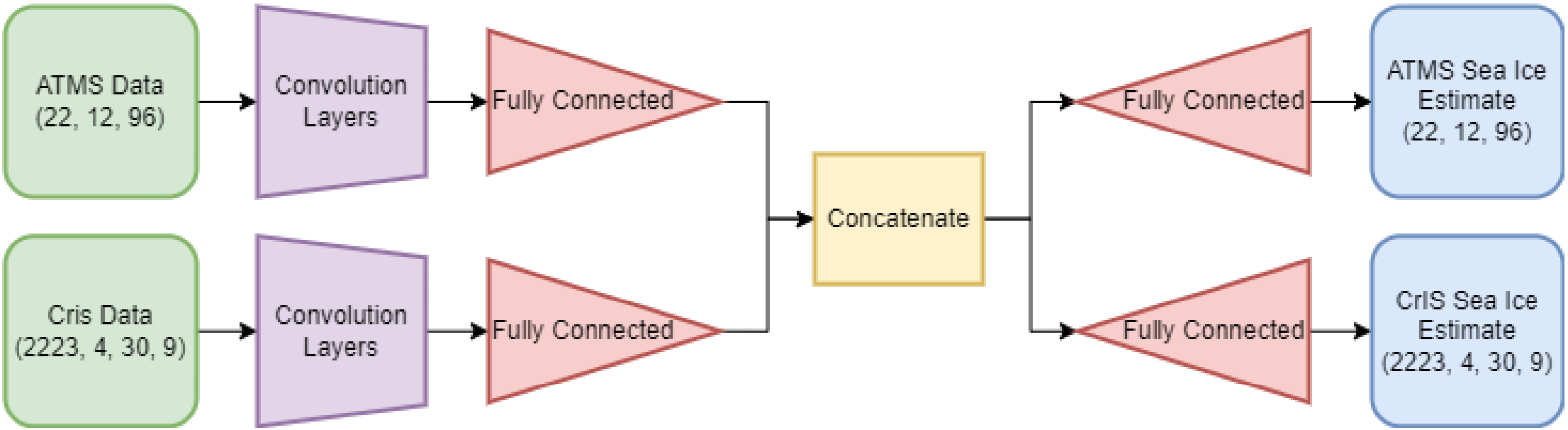
In the Northern Hemisphere

(Stereographic Projection Coordinates)

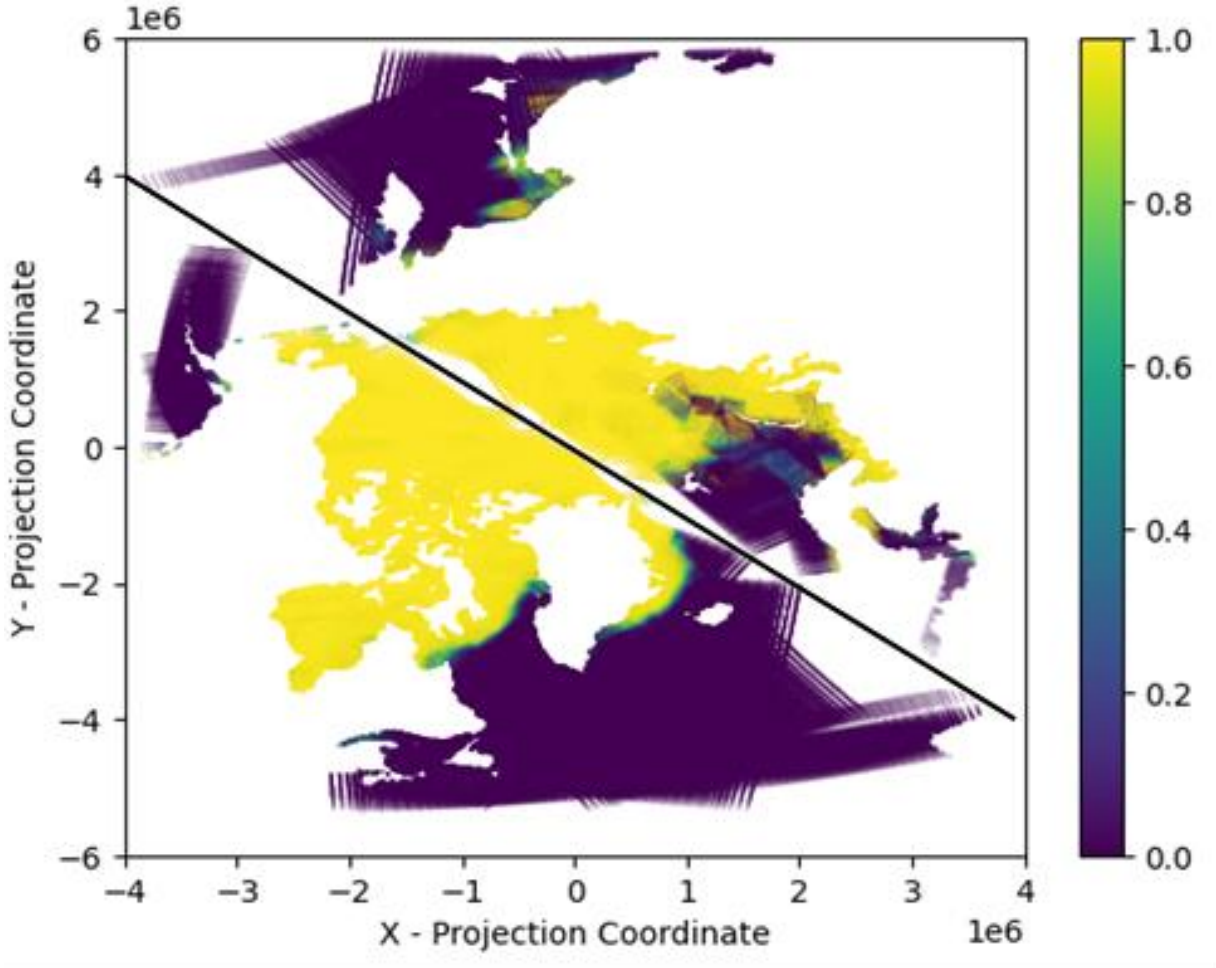
Model trained and tested on this day of data. The train set consists of samples with negative longitude and test set consists of samples with positive longitude.

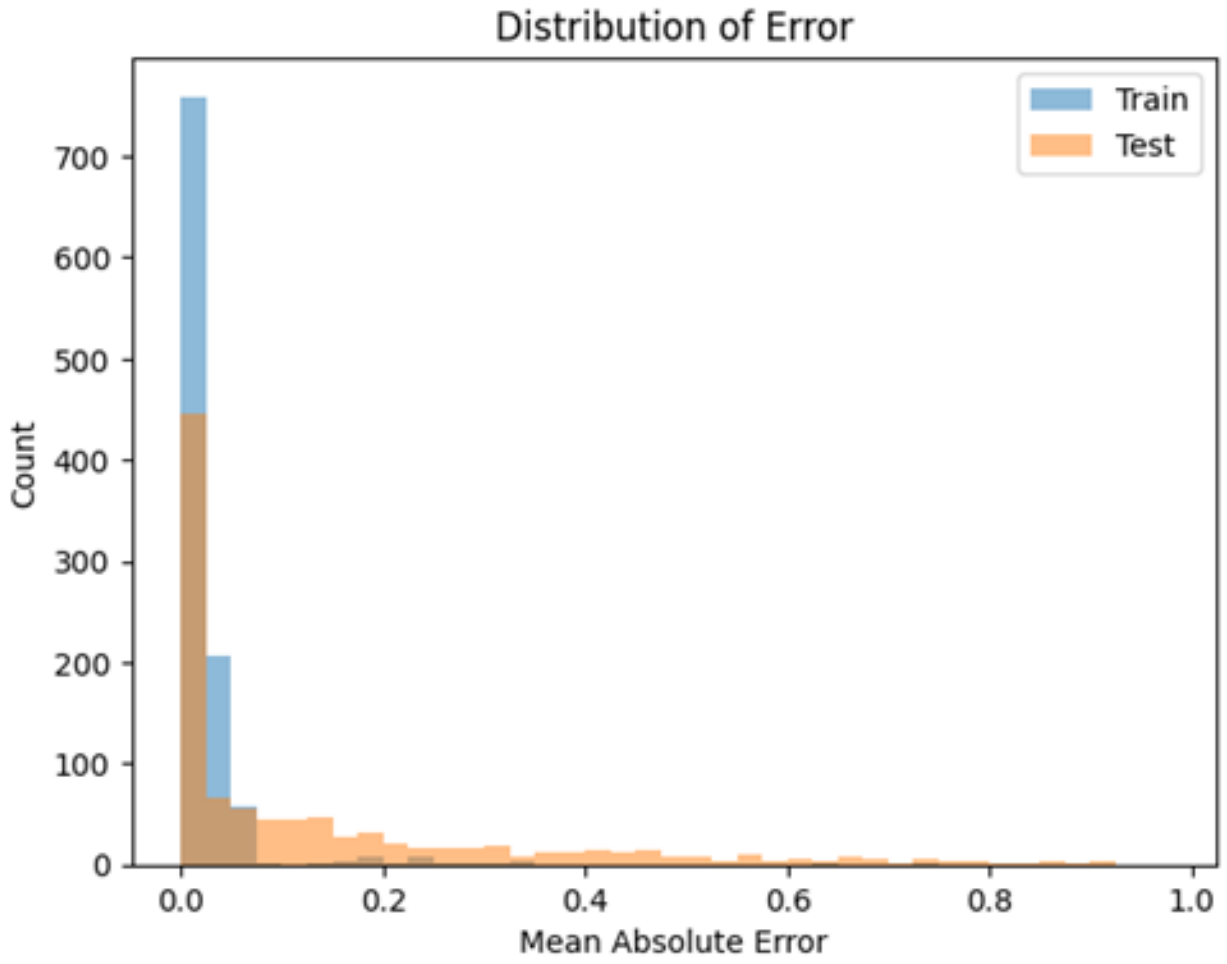


Sea Ice Model Structure



Sea Ice Model Output

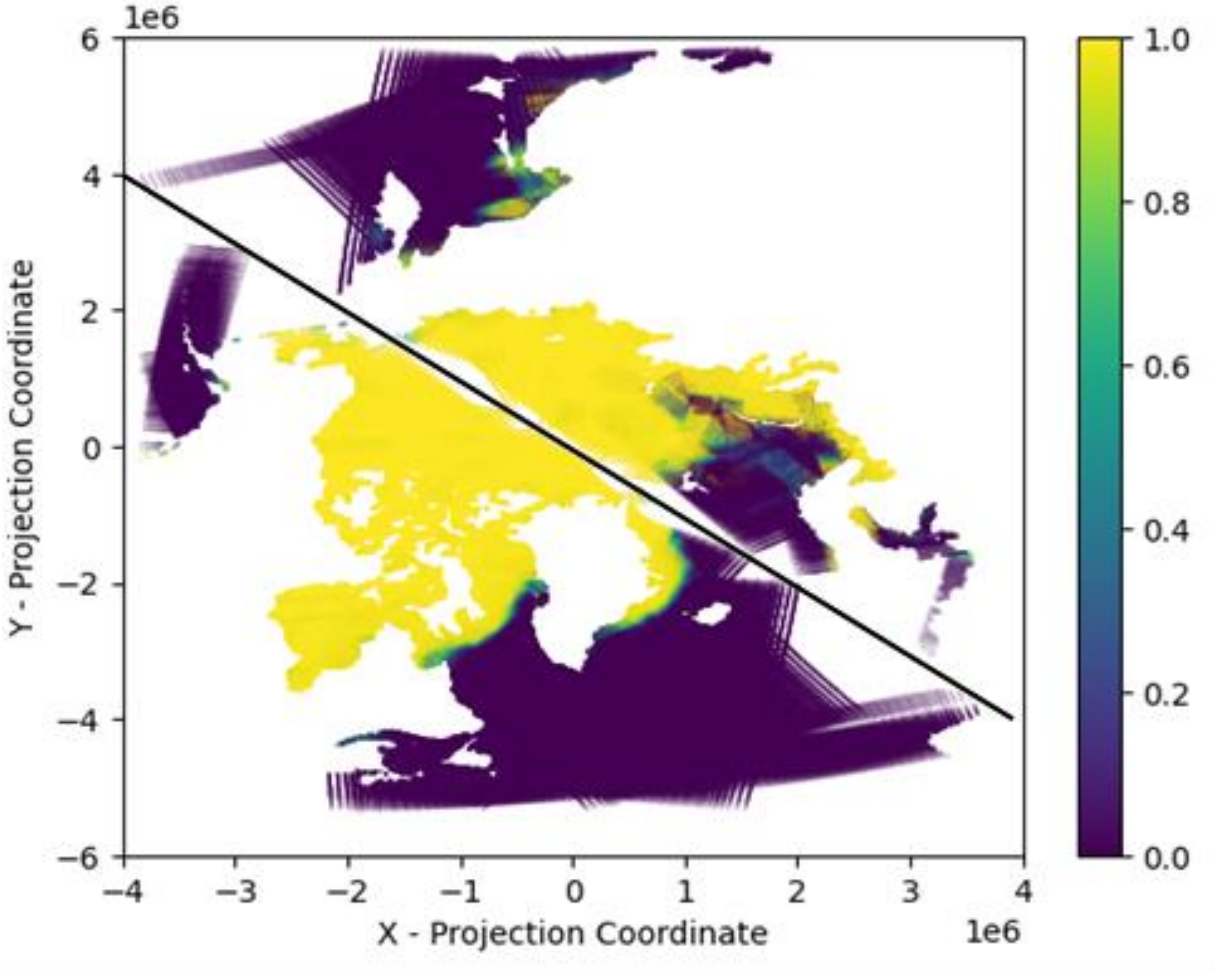




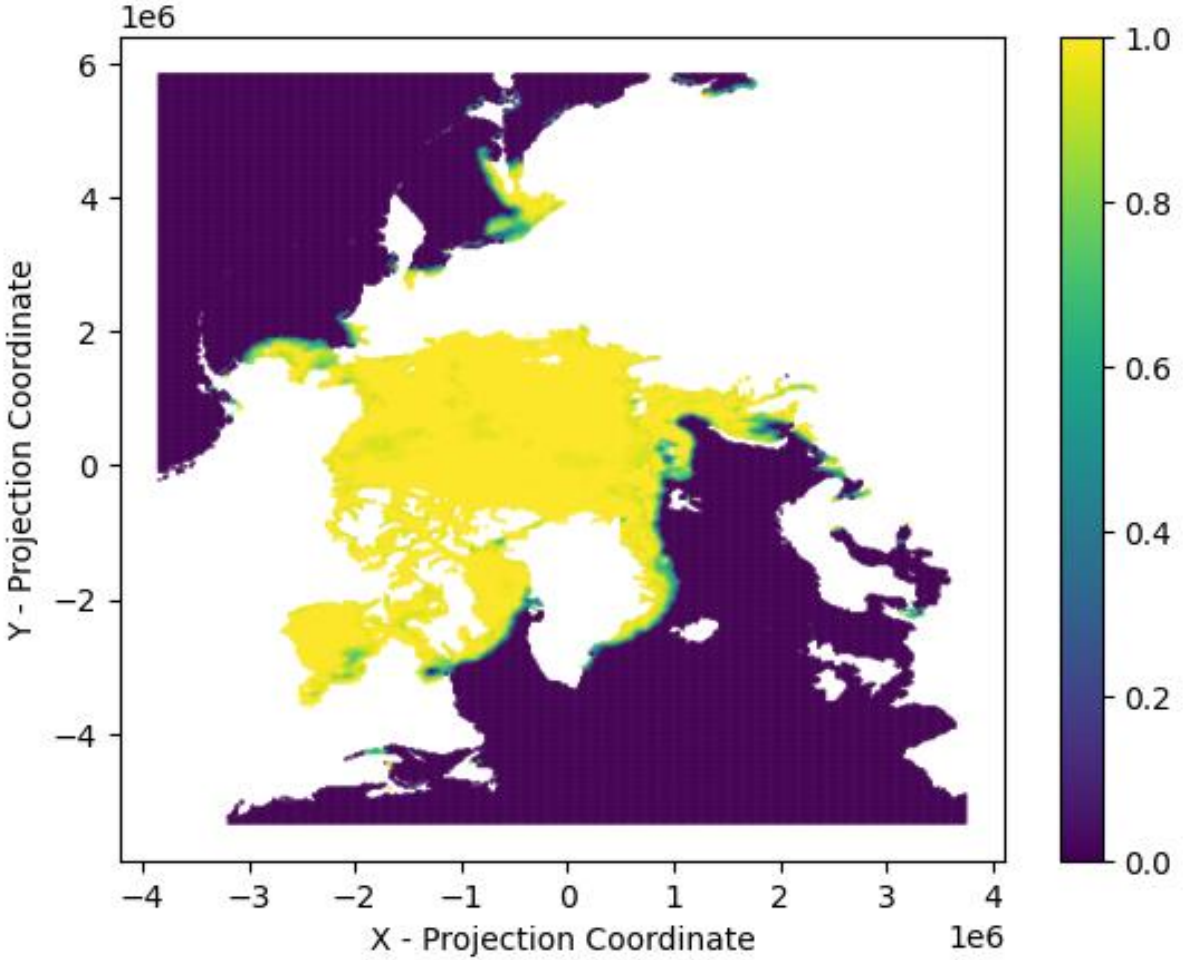
Sea Ice Model Output

orion SPACE SOLUTIONS
An Arcfield Company

Model Estimated Sea Ice Concentration

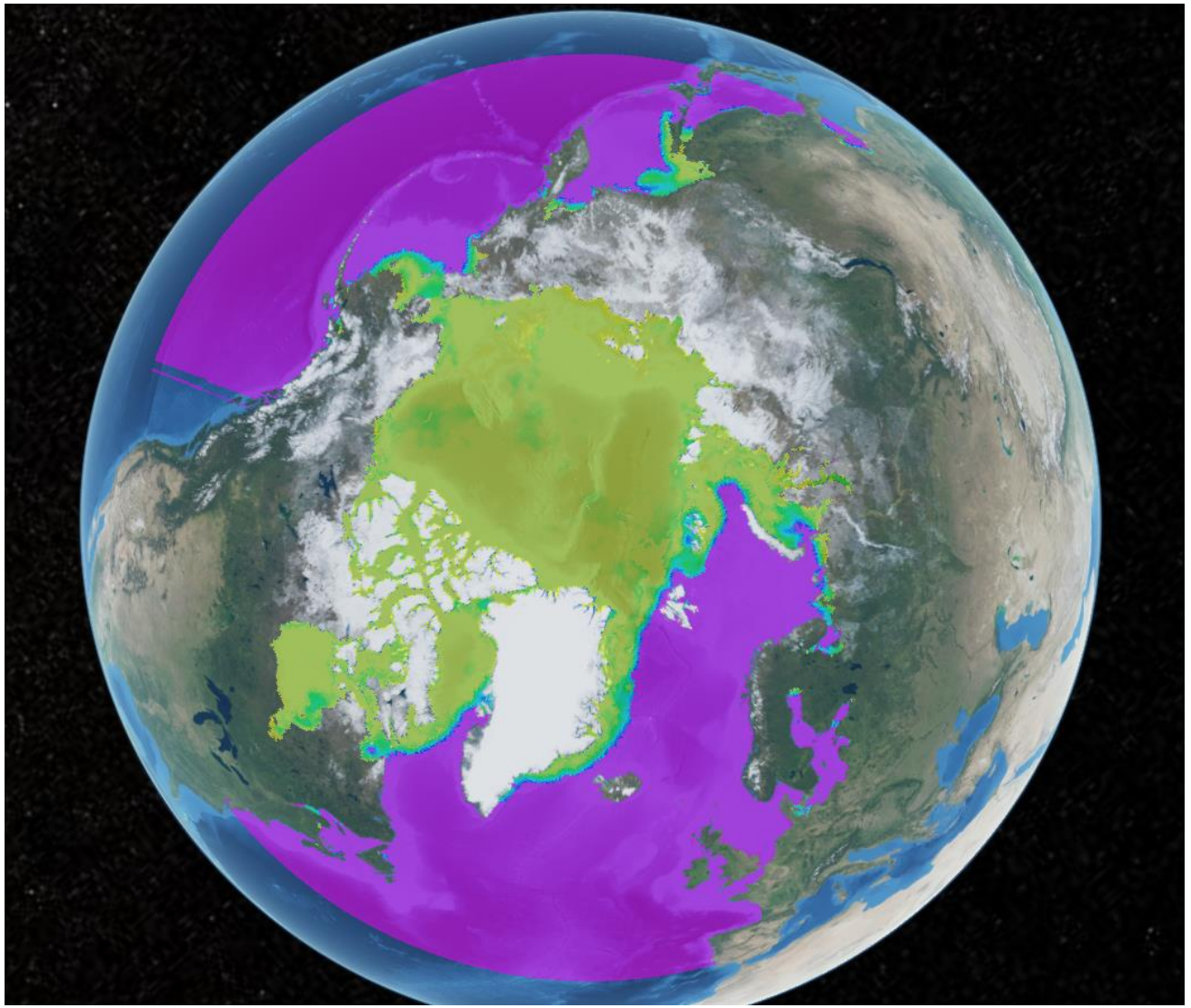


Independent Validating Sea Ice Concentration



Sea Ice Model Ground Truth Data – Visualization

- NOAA/NSDIC Sea Ice Concentration data
  - Source: https://nsidc.org/data/g02202/versions/4
  - Concentration ranges from 0 (purple) to 1 (green/yellow)



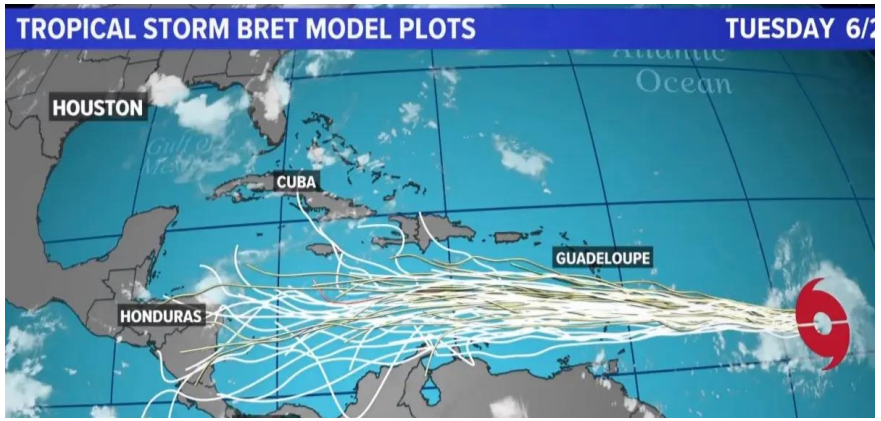
Sea Ice Model Ground Truth Data – Visualization

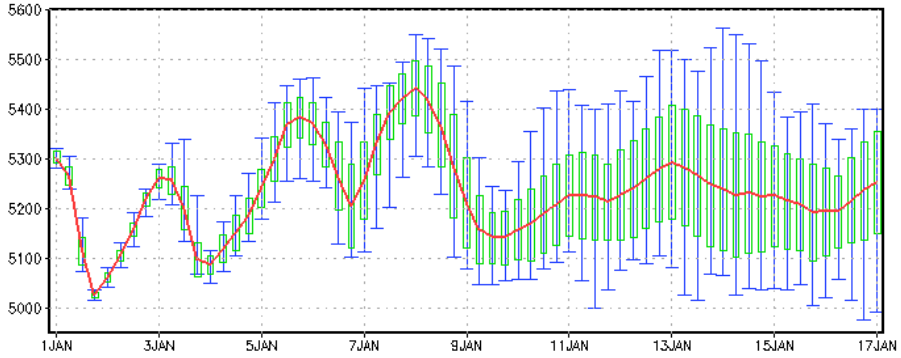
orion SPACE SOLUTIONS
An Arcfield Company

What if?

- Create a hypothetical situation, run it, visualize it. **User interactive.**
  - **Example**: run model potentially many times with different inputs, use variety in outputs to quantify uncertainty or compare experiments

- **Compute**: many nodes w/ MPI
  - **Ideal solution**: start a job from within the digital twin, monitor progress, visualize/analyze/download output





Running Models on Clusters

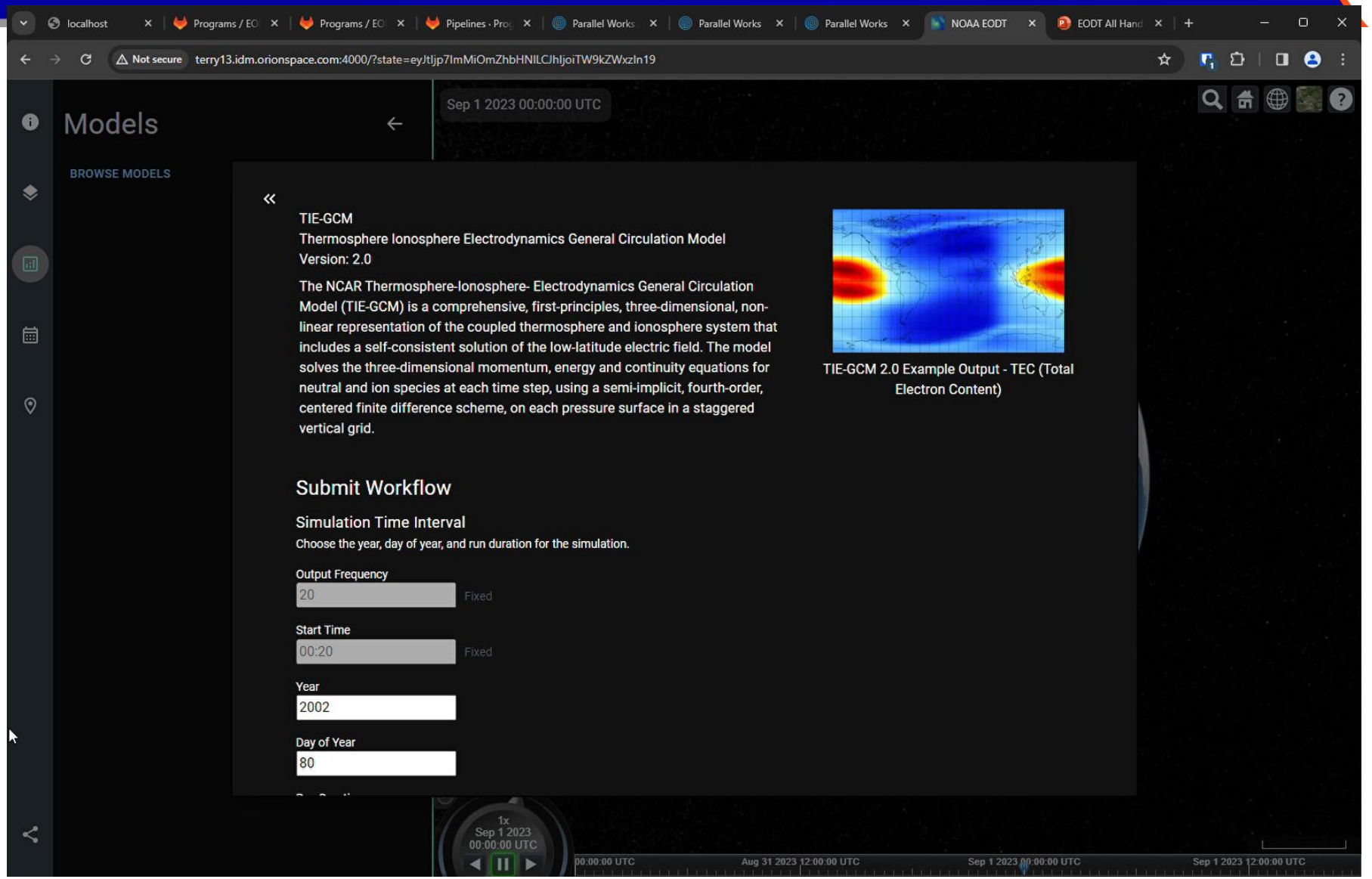
- Integrate Parallel Works REST API into our existing node.js server
  - Provision, start, and stop clusters in the cloud on AWS, Azure, GCP, or on-prem
  - Create and execute workflows to run on the clusters
  - Real time monitoring on the frontend achieved by polling the PW endpoint for job status
  - Workflow outputs available for download from the AWS S3 Bucket
    - Users can download from frontend – available once workflow polling succeeds
    - Or users can download using AWS CLI using their group's credentials
  - Initial workflow executes initialization run of TIE-GCM 2.0

Running Models on Clusters

End-to-End Workflow Submission

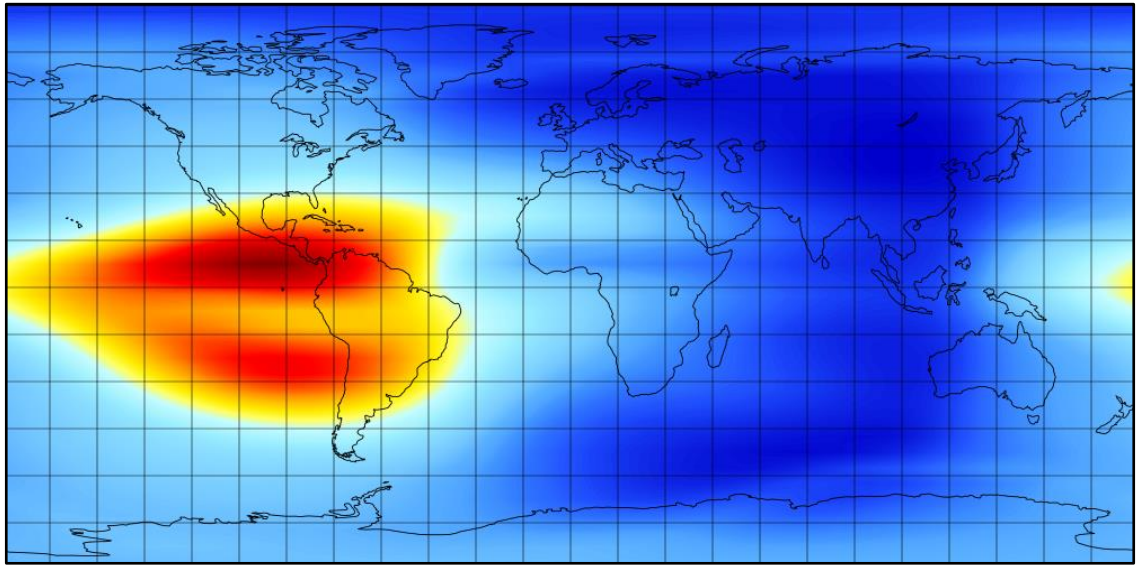
- Run multi-node SLURM ensemble
  - Implement the model using singularity containers, executed using mpi
  - Each compute node runs an instance of TIE-GCM
    - Input values are perturbed for each simulation using a python script (truncated normal distribution)
    - Compute nodes: c5n.2xlarge (8 vCPUs, 21 GB RAM amd64)
  - Submit TIE-GCM Workflow from UI
    - Node.js server will start the cluster if not already on
    - Submit the workflow to run on cluster once active
    - Poll the workflow job from frontend to monitor status
  - Once complete, the workflow script will export outputs to AWS S3 Bucket
  - On polling success, show output on frontend as list of result filenames
    - Click to download from S3 Bucket

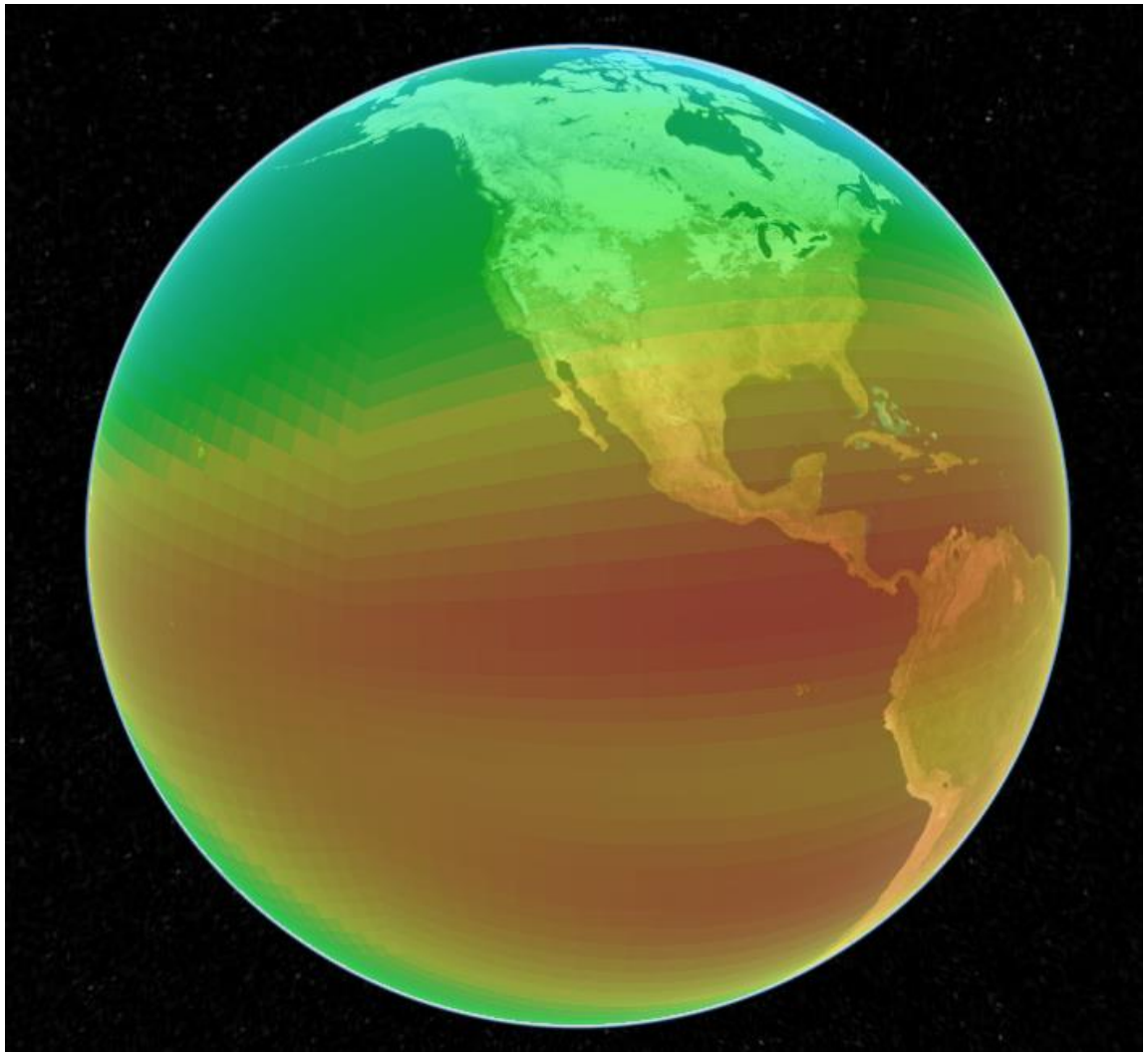
Running Models on Clusters



TIE-GCM

- Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM) output
  - Physics-based model at 5° resolution
  - Viewing: total electron content (TEC) on 10/10/23



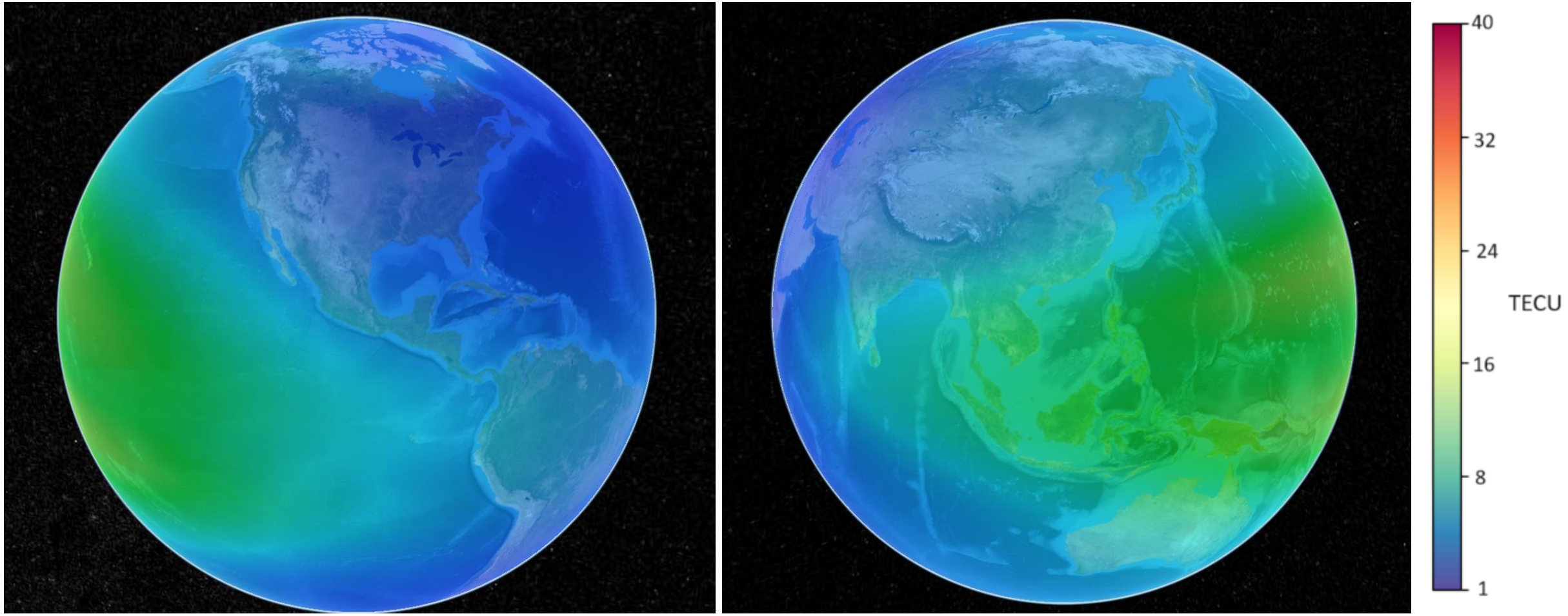


Parallel Works Integration – Visualization

- When a model run finishes, get it from the S3 bucket using the job ID and process the file for visualization
  - Using boto3 python library to access S3 buckets provisioned via PW
  - Secondary history files (sech) include the data we are interested in
  - Can do this for each ensemble member if we do an ensemble run

```
TIE-GCM ensemble run outputs found for job ID 00410:
mem001: tiegcm2.0_res5.0_mareqx_smin_prim_001.nc
mem001: tiegcm2.0_res5.0_mareqx_smin_sech_001.nc
mem002: tiegcm2.0_res5.0_mareqx_smin_prim_001.nc
mem002: tiegcm2.0_res5.0_mareqx_smin_sech_001.nc
mem003: tiegcm2.0_res5.0_mareqx_smin_prim_001.nc
mem003: tiegcm2.0_res5.0_mareqx_smin_sech_001.nc
Downloaded model-outputs/tiegcm/tiegcm2.0/ens/00410/mem001/tiegcm2.0_res5.0_mareqx_smin_sech_001.nc
```

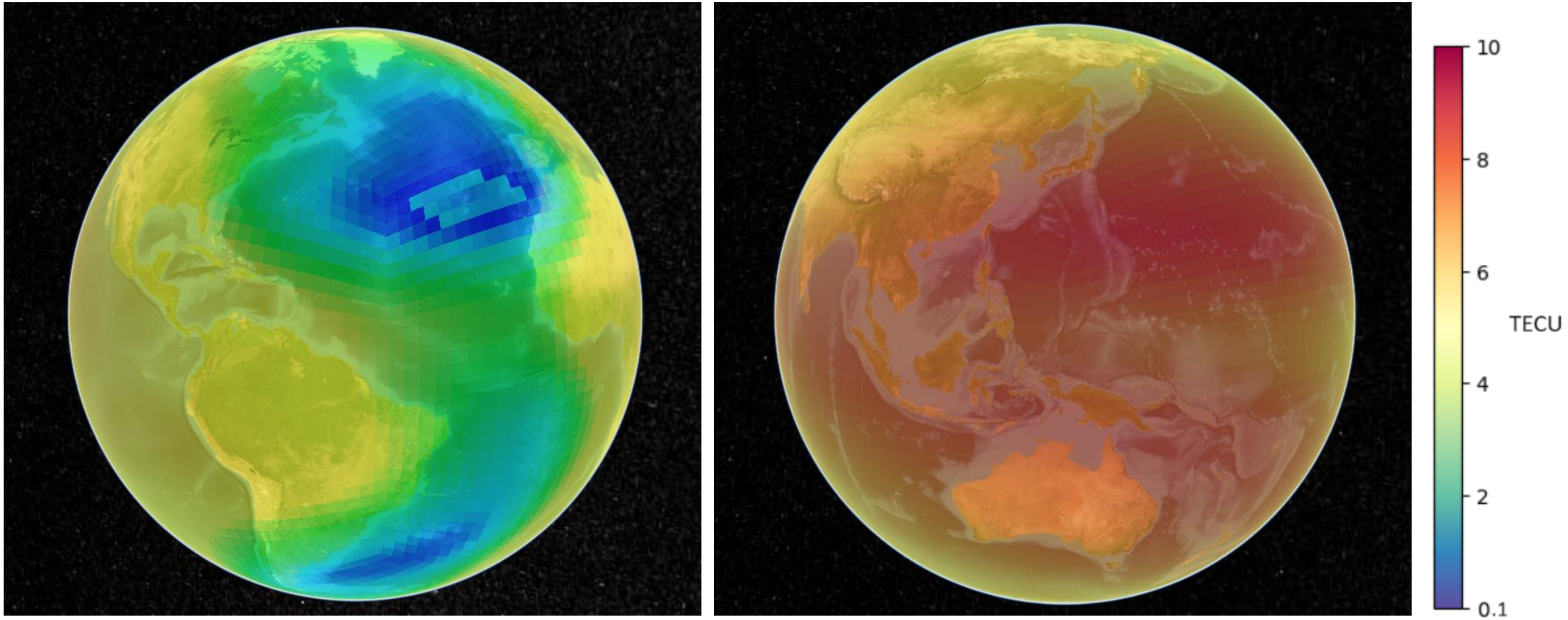
What if Exploration: Visualization



TIE-GCM TEC (Total Electron Content) from one member of an ensemble run

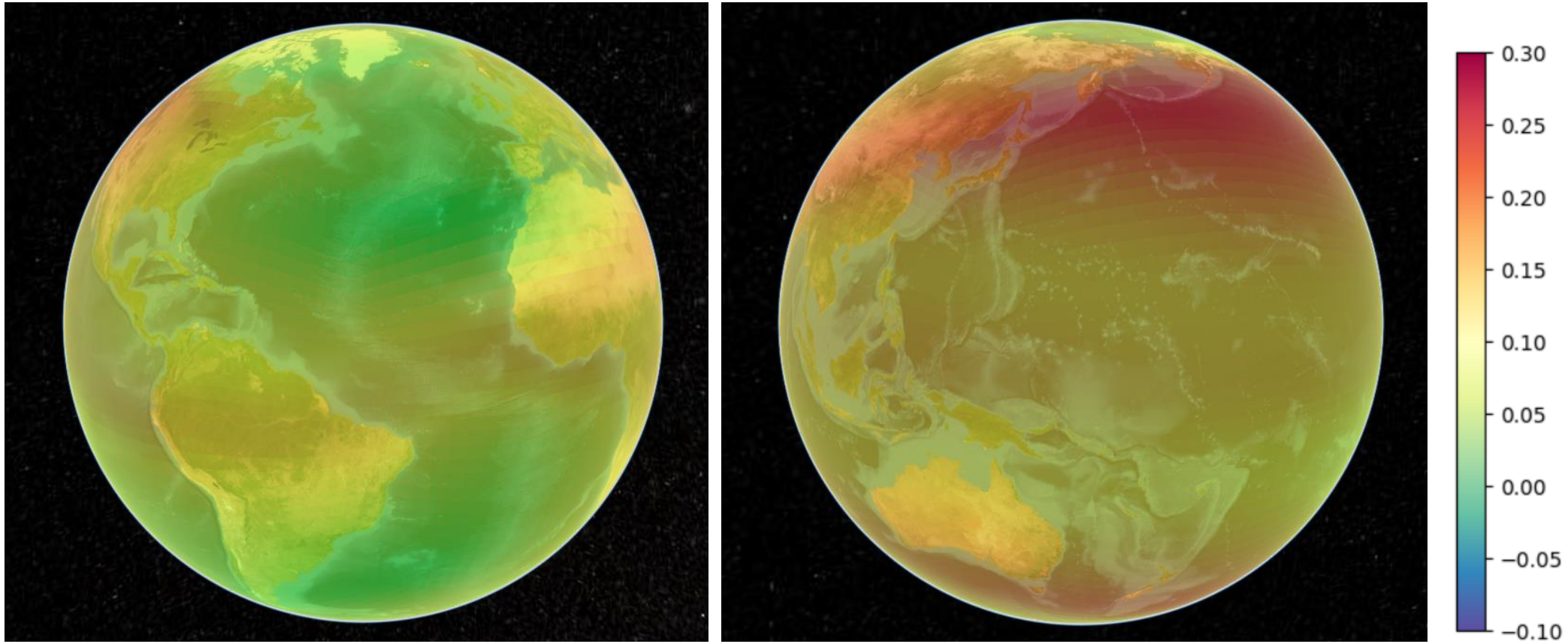
What if Exploration: Visualization

- Difference between two runs of TIE-GCM with perturbed solar drivers
  - SWPC-posted F10.7 and Kp for the time of the run were each multiplied by 1.1 and 0.9

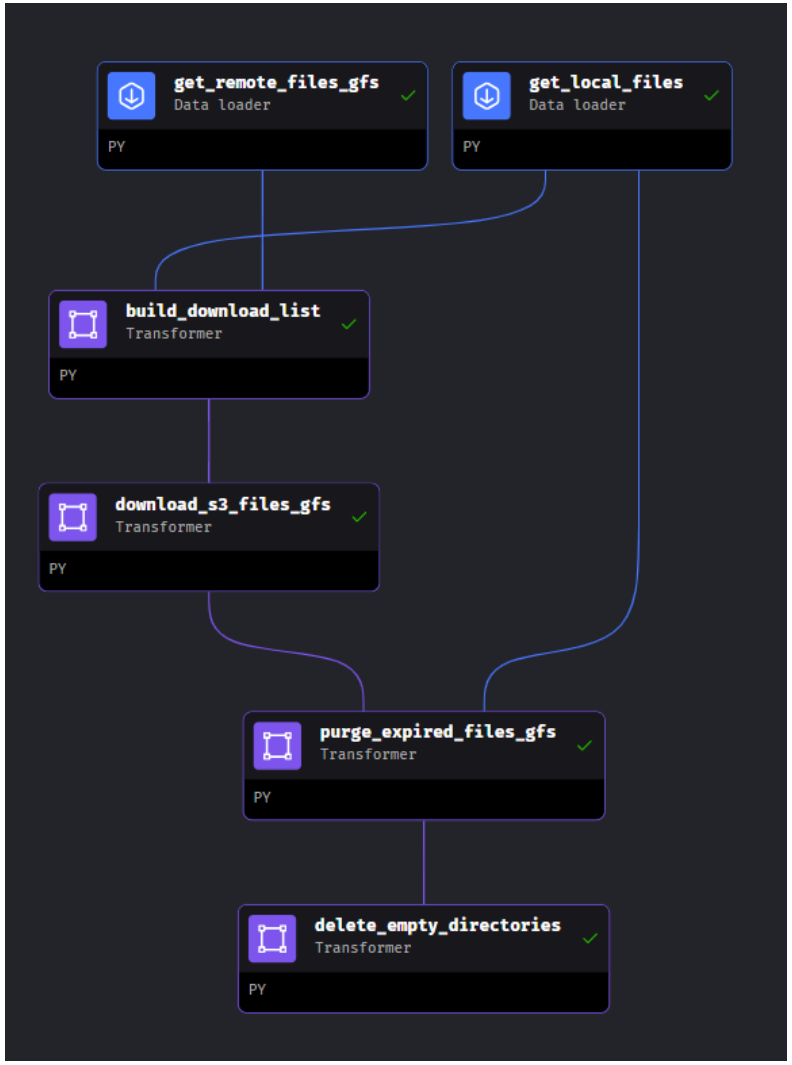


What if Exploration: Visualization

- Difference between two runs of TIE-GCM with perturbed solar drivers
  - Percent difference from average value at each location reached 25%

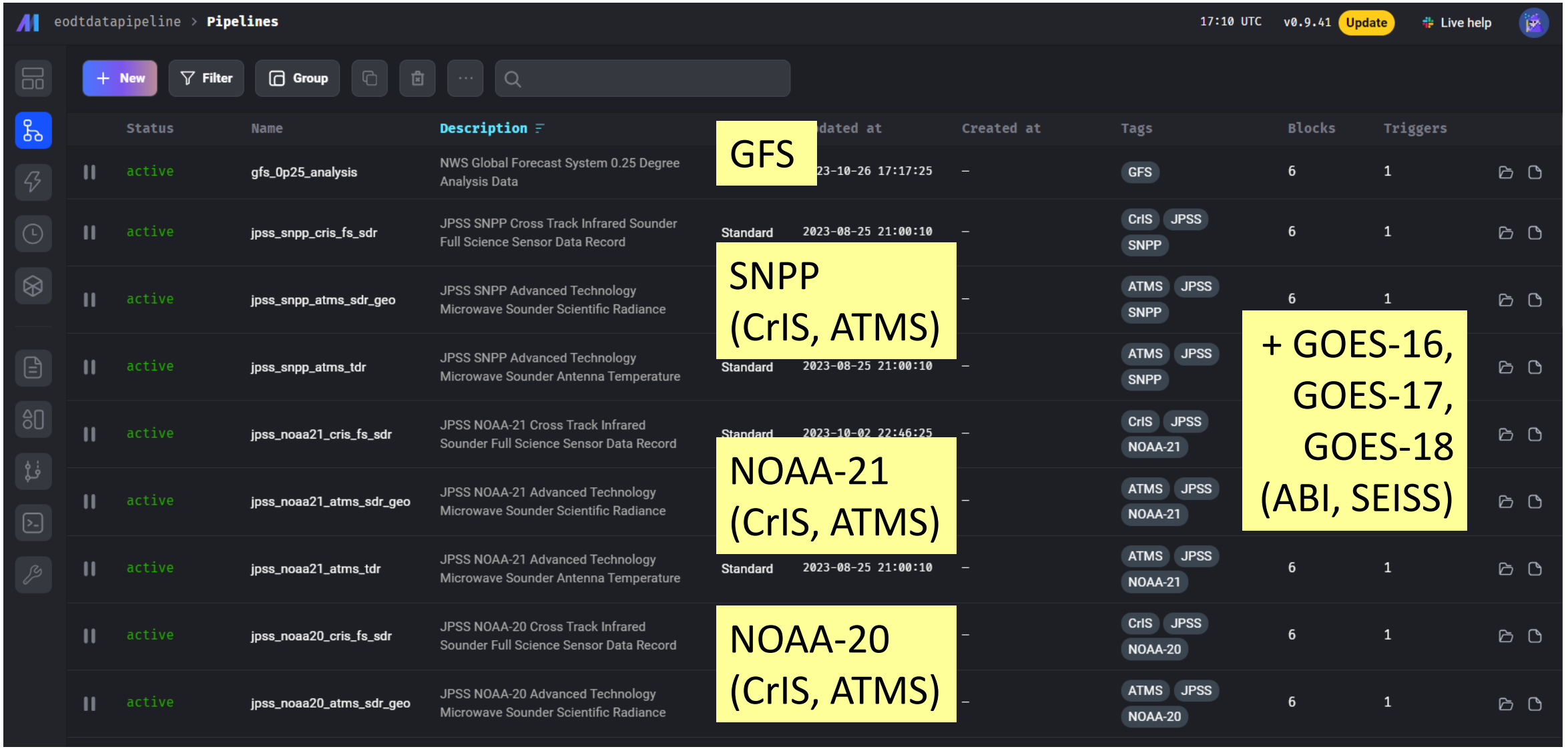


Data Pipeline using Mage.ai

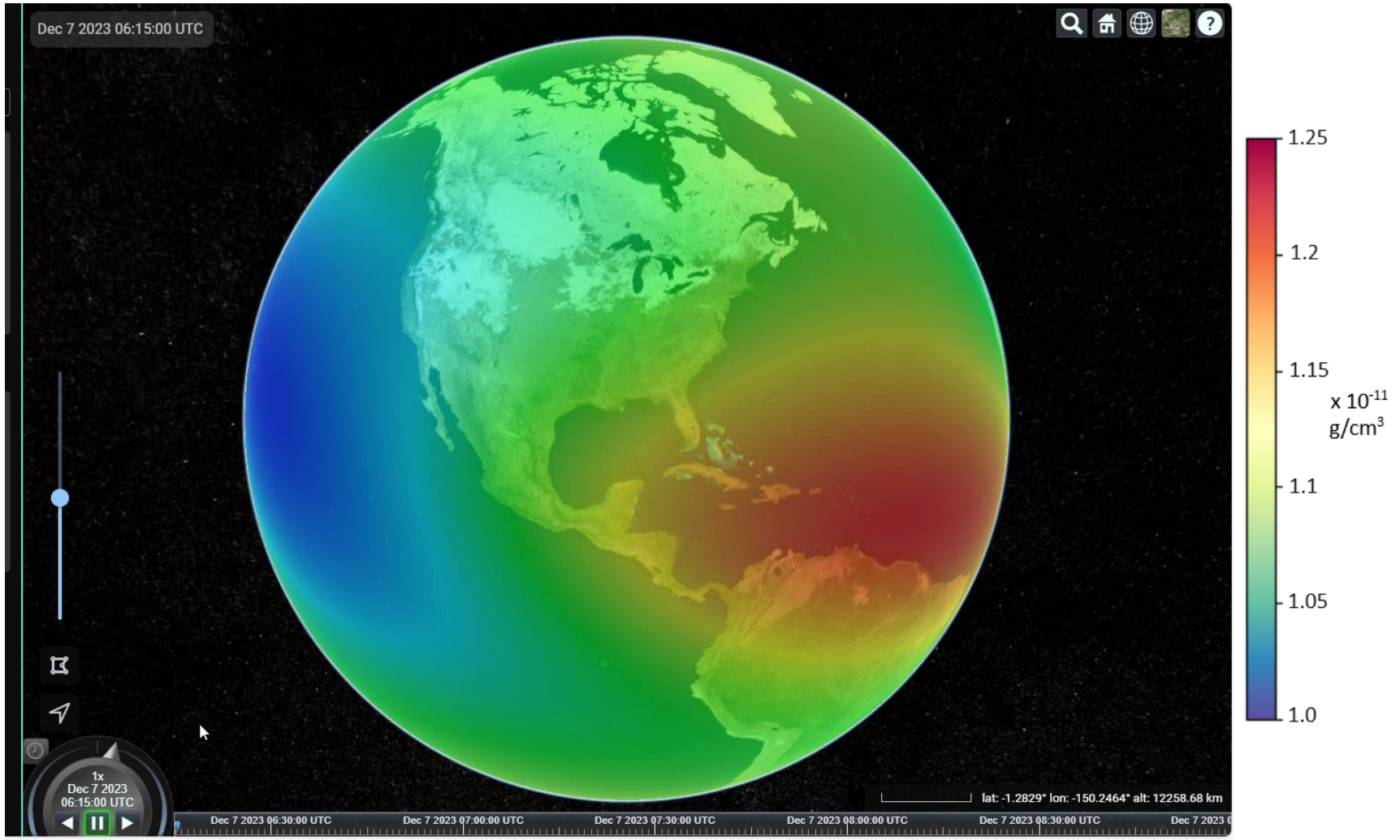


- Mage has built-in capability to connect to data storage locations such as AWS S3 buckets
- We are maintaining a database with a rolling window of data on Terry13
  - Window varies based on size of data from each instrument, from 24 hours (CrIS/ATMS) to 30 days (SEISS)

Data Pipeline using Mage.ai



Time Slider



The user can visualize multiple tilesets for a given dataset generated over a range of time

TIE-GCM total density output at 15-minute intervals over a 3-hour period on 12/7/23

EO-DT: Final thoughts

- Work done to support NOAA's prototype EO-DT effort
- **What now**, **what next**, and **what if** capabilities to visualize, analyze, explain, and predict the earth system model
- Place AI/ML on even-footing with traditional models, support validation and verification efforts
- Incorporate cloud-based computing with on-prem using same workflows on AWS, Azure, GCP, etc. Avoid vendor lock-in.
- Accelerate the speed of science and R2O2R, what if exploration
- Enable decision makers utilize best available info in real-time

Q&A