National Aeronautics and
Space Administration

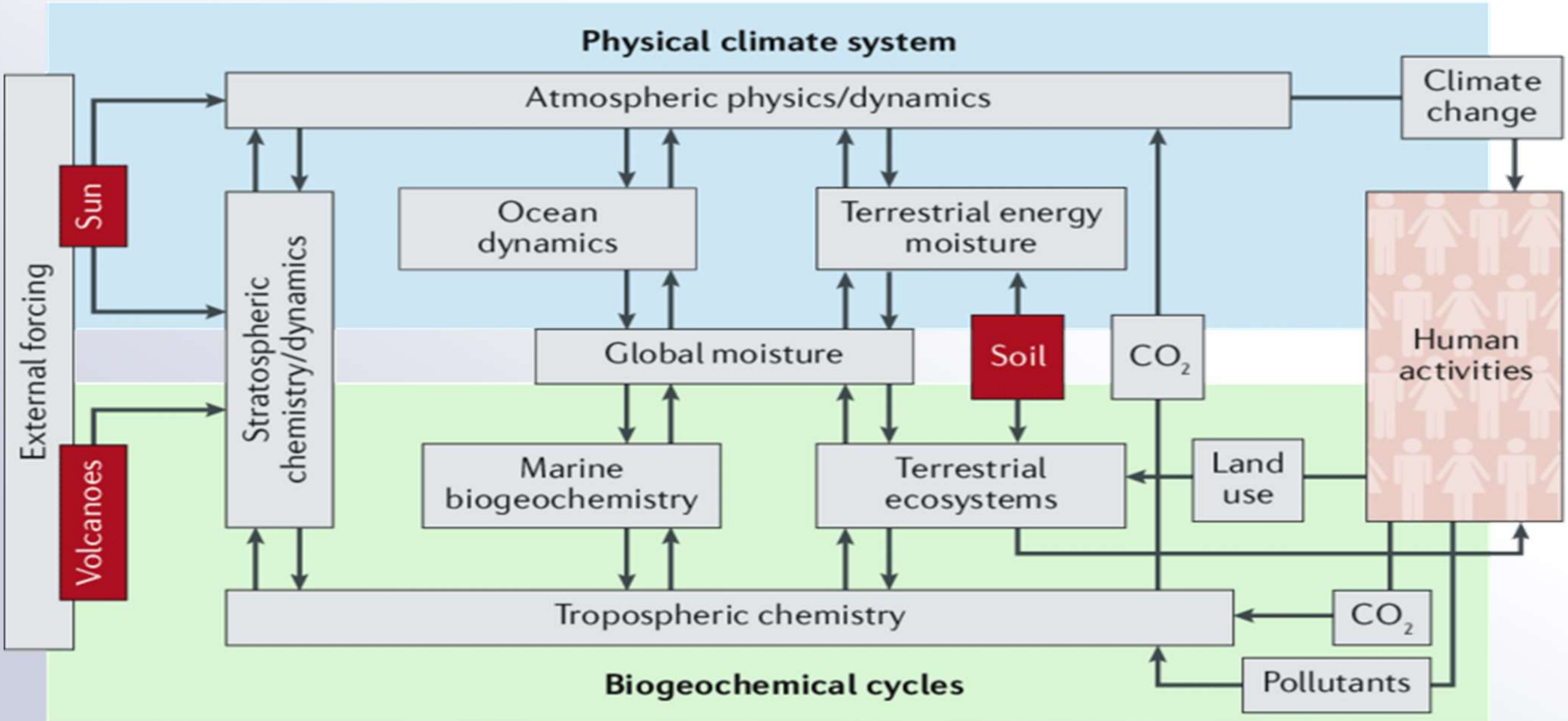
NASA
earth

AI foundation models at NASA
Rethinking the way we do Earth System Science

Tsengdar Lee
Weather Focus Area Lead Scientist
Earth Science Division

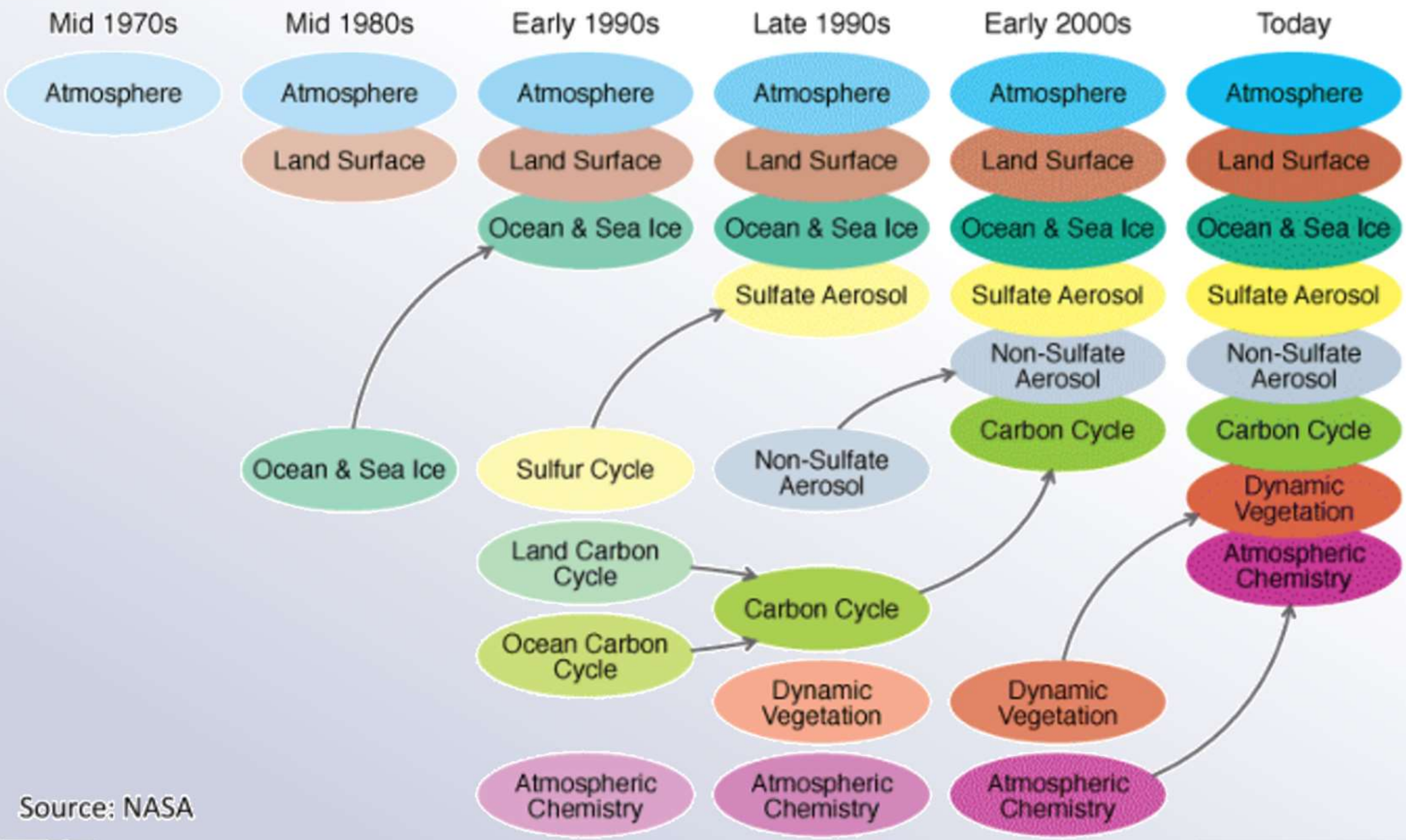
Presented at NAS BDEC meeting, May 27-29, Washington DC

NASA Bretherton of Earth System



NASA (1986) Earth System Science Overview: The NASA Bretherton diagram of the Earth System.

The Development of Earth System Models



Source: NASA

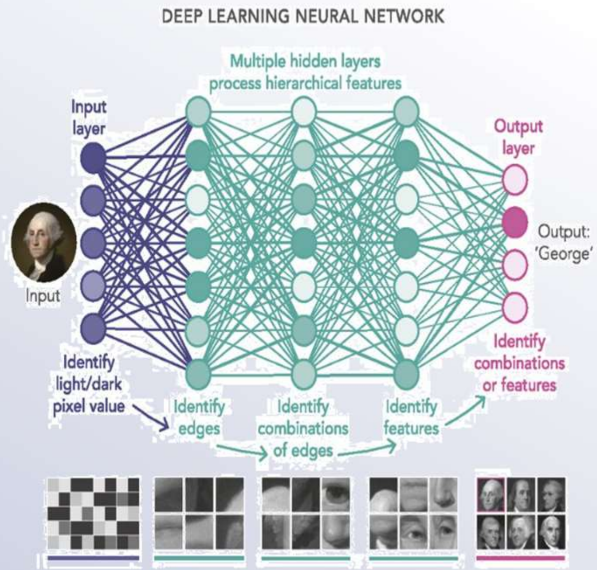
Why Should NASA Be Interested in AI/ML?

1. Earth science programs were organized around the scientific workflow: observe, characterize, understand, modeling, application.
   - Supported by a couple of ROSES elements at 3 years a grant cycle each.
   - Implies a decadal life cycle for the research and development.
2. The data driven paradigm may increase the efficiency tremendously.
   - The coupling of AI/ML algorithms, big data, and big computer platforms has provided the power and opportunity to quickly extract both explicit and latent information from the vast amount of observational data.
   - The deep neural network is able to extract the information from the data, retrieve, summarize, and present the findings quickly and efficiently.
   - AI/ML technologies allow ESD to advance ES2A strategy to create economic outcomes that benefit the society.
3. Supports the 5 focus areas of Administration's American Artificial Intelligence Initiative: (1) increasing AI research investments, (2) leveraging federal computing and data resources, (3) developing AI technical standards, (4) building the AI workforce, and (5) engaging with international allies and the President's recent letter to OSTP Director prioritizing AI/ML to produce economical outcomes.

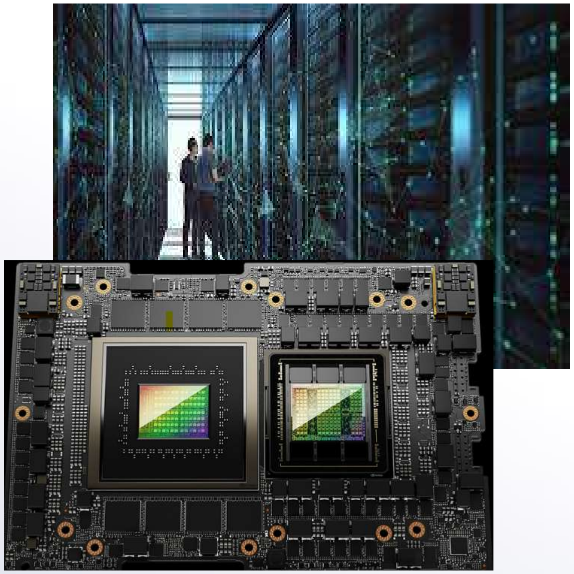
Three Ingredients that Enable the Data Driven Paradigm:

Big Data

Big Model



Big Compute

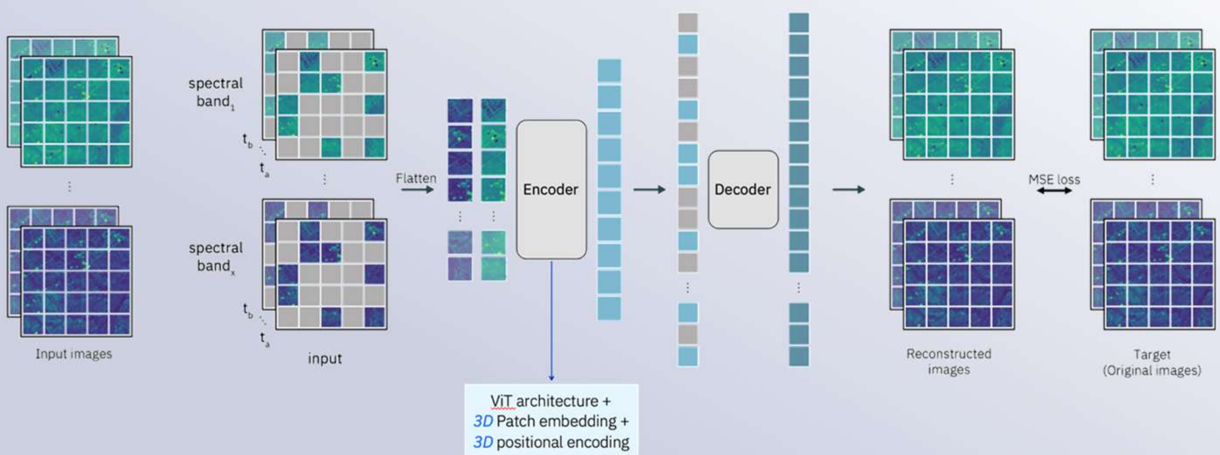


The Three Ingredients Explained

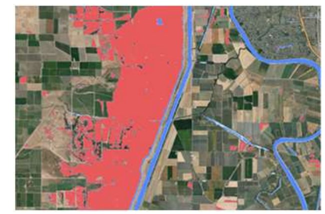
- Big data provides the opportunity for scientists and engineers to quickly extract information from large amount of data.
- Big machine learning model (e.g. neural network or universal function approximator) is an effective way to learn and emulate the behavior of a complex system by fitting to that hyperdimensional data.
- Big compute is the most effective tool to fit a deep NN to the hyperdimensional data.

Prithvi HLS - Geospatial FM for many tasks

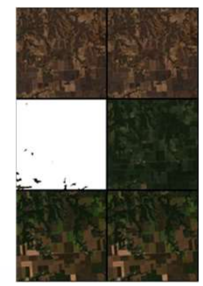
- Jointly developed by IBM Research and NASA
- Prithvi is a first-of-its-kind temporal Vision Transformer trained on NASA's HLS data.
- Uses a self-supervised ViT architecture with a Masked AutoEncoder (MAE) strategy.
- Pre-trained on NASA's HLS V2 L30 data from the contiguous U.S. at 30m resolution.
- Key architecture modifications: 3D patch embedding, 3D positional embedding, and 3D patchify/unpatchify.




Detecting burn scars


Detecting floods


Classifying crop species


Cloud gap filling

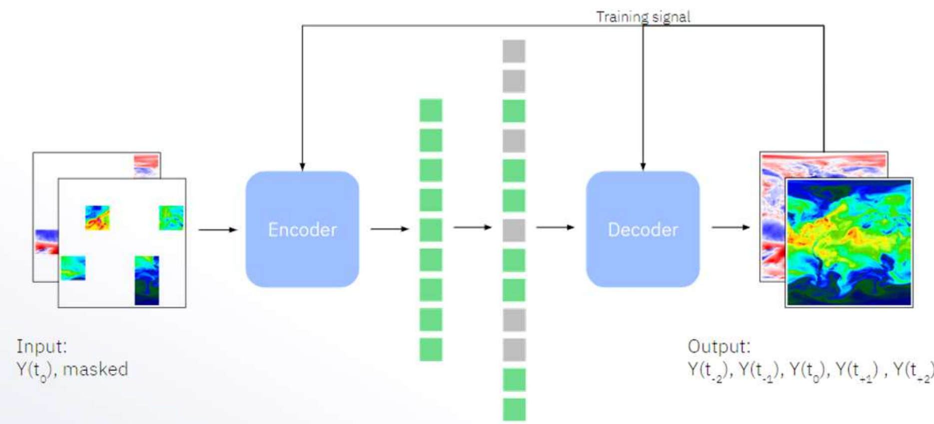
Supports downstream applications like burn scar detection, flood segmentation, and land cover classification.

https://huggingface.co/ibm-nasa-geospatial/Prithvi-100M

Prithvi-WxC :Foundation model for weather and climate

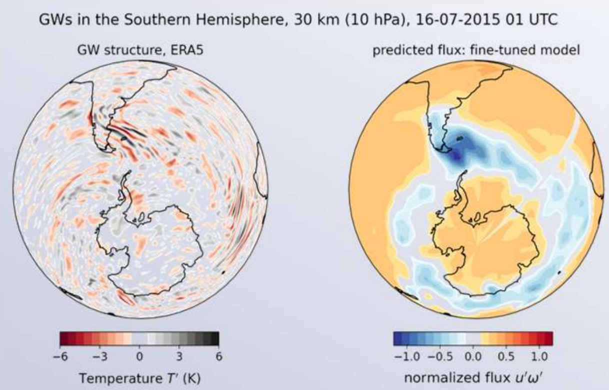
- Prithvi WxC is a 2.3 billion parameter model trained on 160 variables from MERRA-2 data.
- Jointly developed by NASA, IBM Research, ORNL and several Universities
- Pretrained for both forecasting and masked reconstruction tasks.
- Capable of reconstructing atmospheric states from partial data and forecasting future states.

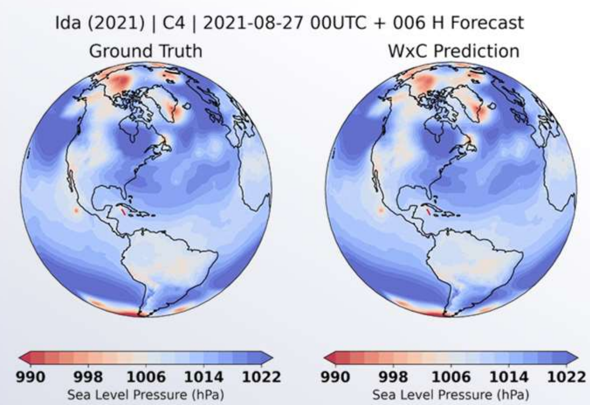
Developed to address applications that aren't focused solely on forecasting (parameterization, downscaling, etc)



Mixture of forecasting & data assimilation / state estimation.
We are using [-24, -12, -6, -3, 0, +3, +6, +12, +24] hours ahead.

- Weather & climate Forecasting
- Downscaling
- Model output statistics/bias correction
- Multi-model ensembling
- Parameterization
- Data Assimilation
- Detection
- Tracking
- Impact Modeling





Considerations:

NASA's observations are inherently heterogeneous in wave-lengths, spatial and temporal resolutions, physical data types (temperature, humidity, precipitation, aerosol, and composition etc.), and dimensionalities (point, line, profile, and image).

- Balance between generalization and specialization
  - The more general and global the more difficult and costly to train
  - The more specific and local the less costly to train but the applications would be more limited
- Well defined training objectives
- Training data set must include the [explicit or latent] information we are seeking
- Model architecture which may take multimodal data sets
- Efficiency of computing algorithms
- Flexibility in adding or removing satellite datasets to accommodate NASA satellite missions.

This would require a "whole of Earth Science" approach to accomplish.

Adopted and expanded from ClimaX: A foundation Model for Weather and Climate

Thank You!