ICAMS AI/ML Workshop #1 DOE Overview

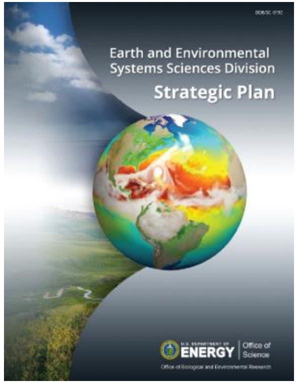
November 4th, 2024

Department of Energy
Office of Science
Earth and Environmental Sciences Division

Gerald Geernaert

U.S. Department of ENERGY | Office of Science

U.S. Department of ENERGY | Office of Science



Biological and Environmental Research (BER)

Earth & Environmental Systems Sciences Division (EESSD)

Gary Geernaert

Atmospheric Sciences

Atmospheric System Research **(ASR)**

Atmospheric Radiation Measurements (ARM) User Facility

Earth and Environmental Systems Modeling (EESM)

Earth System Model Development

Regional and Global Model Analysis

Multisector Dynamics

Environmental System Science (ESS)

Subsurface Biogeochemical Research

Terrestrial Ecosystem Science

Environmental Molecular Sciences Laboratory (EMSL)

Data Management (DM)

Budget: ~445M Roughly Divided Equally Among the Three Groups

Atmospheric Radiation Measurement

Providing the research community with long-term in situ and remote sensing observations of aerosols, clouds and radiation.

Currently using **AI based techniques** to track storms, redirect RADAR and for data quality assurance / quality control.

Three measurement sites: Oklahoma, Alaska, Azores with 24/7 data collection.

ML to streamline data processing

Developed a machine-learning-based pipeline for identifying cloudy periods using the Waggle computer installed at the ARM Southern Great Plains atmospheric observatory.

ARMing the Edge: Designing Edge Computing–Capable Machine Learning Algorithms to Target ARM Doppler Lidar Processing; Jackson, R. C., and Coauthors, 2023

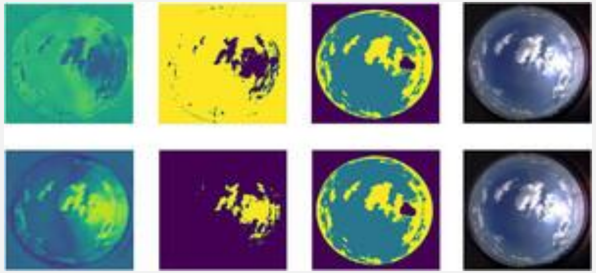
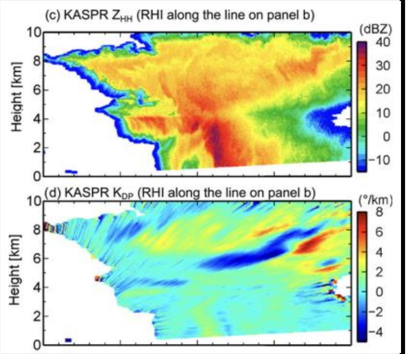


Figure: Two examples of cloud segmentation produced by Distillation No labels (DINO) feature inference model.

ML to control observational instruments

- Radars are automatically guided by external observations from satellite imagery and surface cameras
- **First ever** automated obs. of rapidly evolving shallow cumulus & waterspouts by a phased-array radar and a cloud radar
- Multisensor Agile Adaptive Sampling (MAAS) capitalizes on advancements in communications, edge/fog computing, sensor capabilities, and ML/AI



Agile adaptive radar sampling of fast-evolving atmospheric phenomena guided by satellite imagery and surface cameras; Kollias and Coauthors, 2020

U.S. Department of ENERGY | Office of Science

ML for predicting low cloud fraction (LCF)

ML models (XGBoost) **provide improved prediction of low cloud fraction** across the full spectrum of atmospheric stability and large-scale vertical velocity.

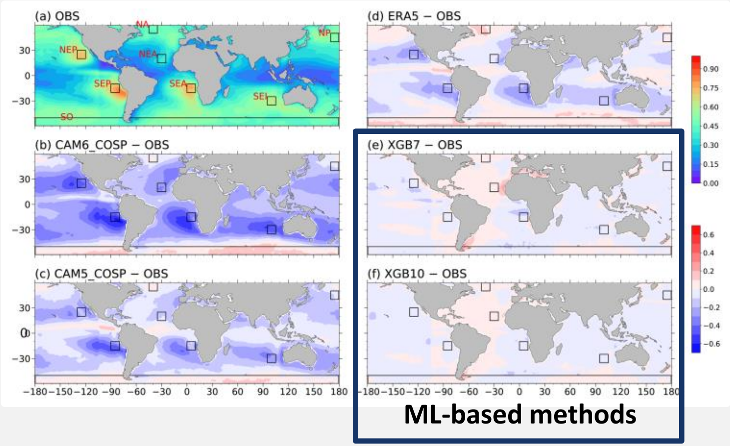


Figure: Global maps of climatological mean LCF for 2004-2005 from CERES and difference in LCF between model and CERES. The ML-based methods match far more closely to CERES than CAM5/6 COSP.

Improving Low Cloud Fraction Prediction Through Machine Learning; [illegible] Zhang, [illegible] Zhang, and Z Li, 2024

Deep-learning driven simulation of boundary layer clouds

Using long-term observations from ARM, a deep-learning model is developed to **simulate the daytime evolution of boundary layer clouds** (BLCs).

The model predicts hourly estimates or cloud occurrence, cloud boundaries and vertical profile of cloud fraction.

Model outputs agree better with observations than cloud fields from ERA5 and MERRA2.

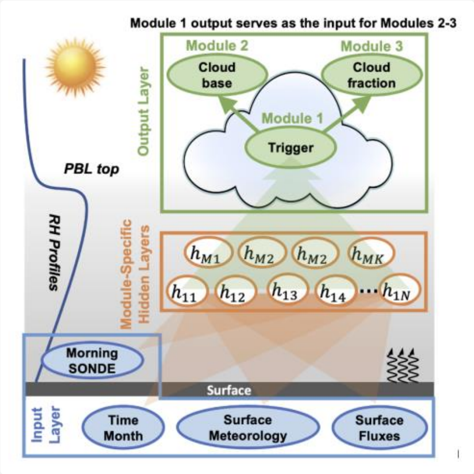


Figure: Conceptual diagram of the deep-learning framework for simulating boundary layer cloud (BLC) characteristics over the US Southern Great Plains.

Deep-learning-driven simulations of boundary layer clouds over the Southern Great Plains; T Su and Y Zhang, 2024

Vision for AI in E3SM

E3SM Energy Exascale Earth System Model

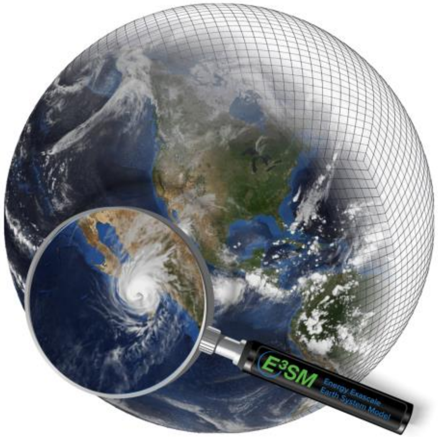
AI is used in E3SM development. Some examples are highlighted below

1. **Emulation** will accelerate simulations, enabling quantification of natural variability

2. **Downscaling, data compression**, and **online feature tracking** will solve local-scale data availability problems

3. Fast **emulation** and **prediction** will solve model spin-up problems

4. **Chat interfaces** will make model configuration and analysis easier



5. **Autotuning** will improve model skill and accelerate model development

6. **ML-based parameterizations** may improve model accuracy and efficiency *if trained on all possible futures*

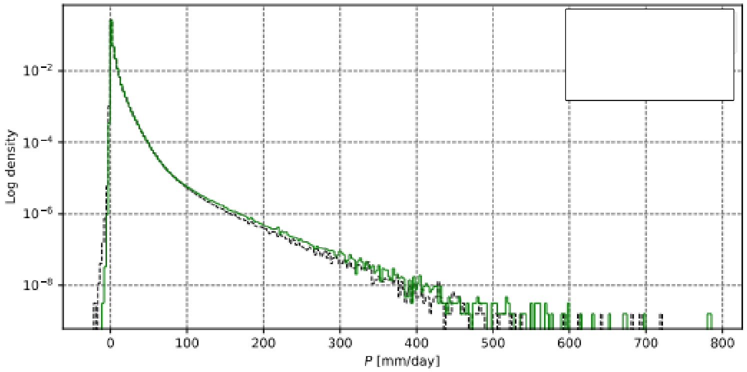
7. **Uncertainty quantification** is critical for informed adaptation planning

8. Data science will permeate **model analysis**

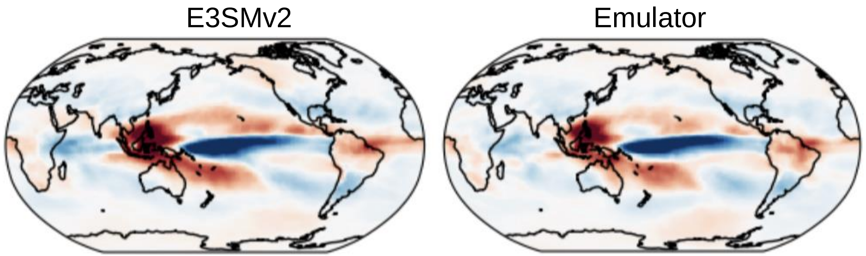
Developing a ML emulator of E3SM

- Emulators make climate models fast and cheap, allowing us to:
    - exhaustively explore natural variability
    - obtain excellent statistics for extreme events
    - choose dx and dz for accuracy rather than affordability
- In collaboration with the **Allen Institute for AI (AI2)**, we have created an emulator for the E3SM atmosphere model which:
    - is 100x faster and 100x more efficient than its physical counterpart
    - runs stably for hundreds of years
    - reproduces E3SM's rainfall intensity almost perfectly
    - captures E3SM's response to SST changes
- We are now working on ocean and sea ice emulators

Histogram of precipitation intensity sampling 10 yrs of daily mean data for each grid column globally. (From Duncan et al. 2024 JGR-ML)



Regression slope between Nino3.4 index and local outgoing-longwave radiation (a proxy for convection)



Automated Calibration of SCREAM

- Calibrating uncertain physics parameters accounts for a huge fraction of labor in climate modeling centers
- AI-based calibration promises improved model accuracy, faster model development, and improved objectivity
- We tuned the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM) with dx=3 km globally based on 150 2-day forecasts

- Short simulations and multi-resolution techniques are needed to tune SCREAM
- Overall, we were able to improve many aspects of SCREAM simulations (blue vs red areas in graphic)

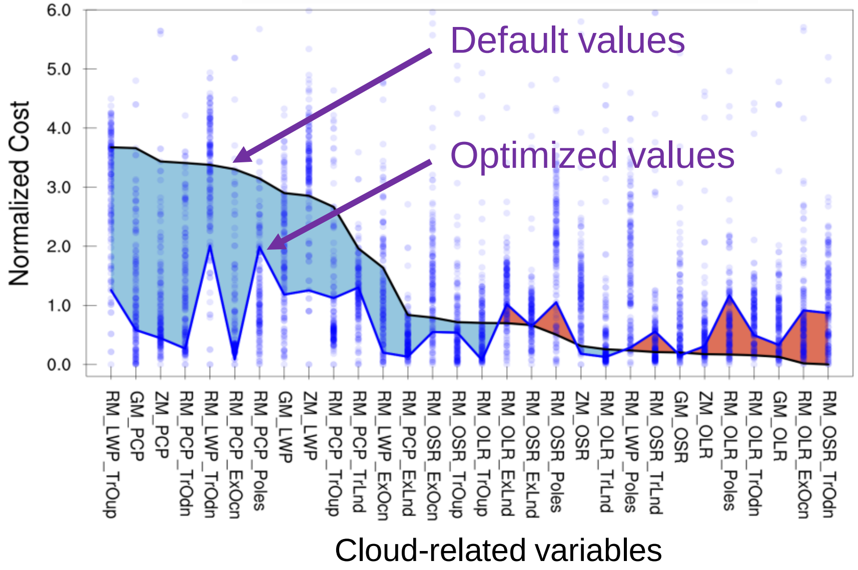
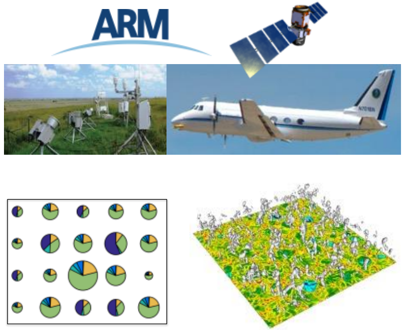


Figure: SCREAM error (normalized cost) in variables listed on x axis for a large variety of parameter choices (blue dots). The black line is the default tuning and the blue line is our optimized value. The blue area indicates improvement and the red area shows variables that got worse after tuning

Improving the understanding and predictability of aerosol and Aerosol Cloud Interactions in the Earth system

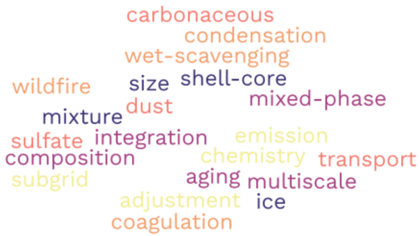
EAGLES
Enabling Aerosol-cloud Interactions at GLobal convection-permitting scalES

Big Data



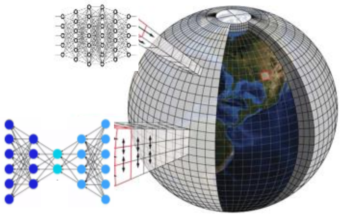
- Big data consisting of both observational and high-resolution process modeling data
- Data curation to facilitate AI applications

Better Physics



- Understand intrinsic connections in the nonlinear Earth system
- Represent complex adjustments and feedbacks
- Emulate multiscale physics and subgrid features

Process Surrogate



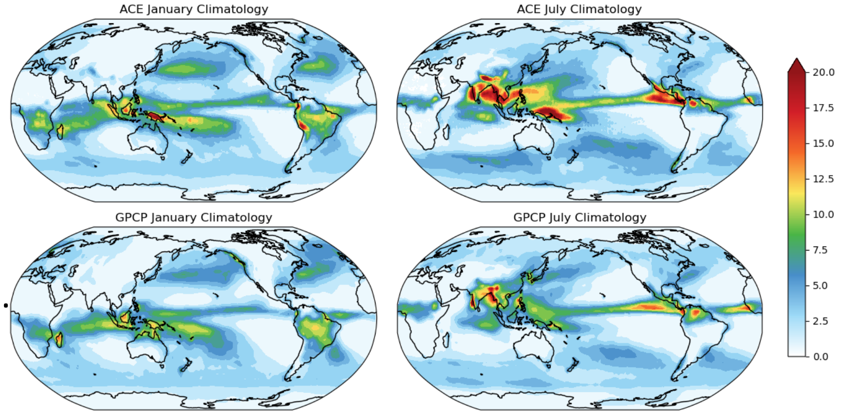
- Software infrastructure for deploying AI/ML within Earth system models
- Accelerate numerical calculations
- Improve understanding and trustworthiness with explainable AI techniques

U.S. Department of ENERGY | Office of Science

Systematic and collective evaluation of ML-based ESMs

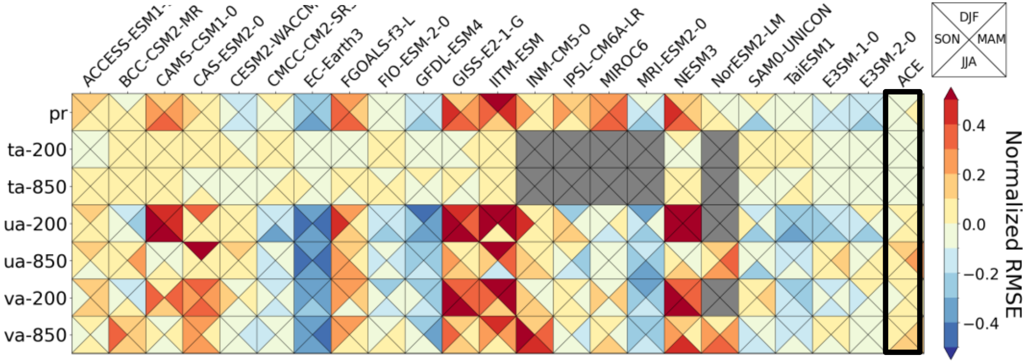
PCMDI
Earth System Model Evaluation Project

- Address the need within the ML-based ESM community on a common evaluation framework for these models.
- PCMDI is leveraging its past experience in evaluating GCMs to develop these standards.
- Multi-pronged evaluation strategy:
    - Existing evaluation packages
    - Conservation laws
    - Idealized test cases
    - Emergent constraints
    - Leveraging the model ecosystem



AI-ESM (ACE

OBS

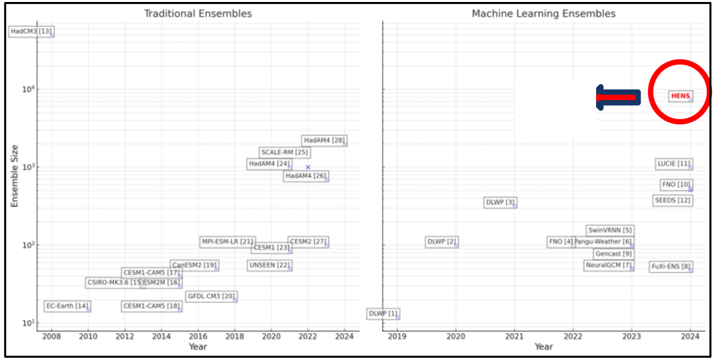


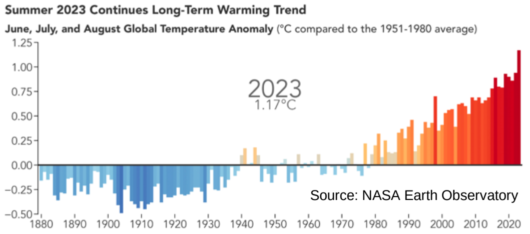
U.S. Department of ENERGY | Office of Science

ML-Enabled Huge Ensembles (HENS) of Extremes

CASCADE
CALIBRATED & SYSTEMATIC CHARACTERIZATION, ATTRIBUTION, & DETECTION OF EXTREMES

- **Why HENS? (using NVIDIA's FourCastNet)**
  - Characterizing low-likelihood high-impact extremes (LLHI) requires adequate statistics.
  - Adequate statistics need huge ensembles of simulations.
  - Huge ensembles can only be generated with ML emulators.
  - We studied near-miss LLHIs in ultra-large counterfactuals from June-August 2023, the 2nd hottest summer in 2000+ years.
- **HENS drastically reduces hindcast uncertainties:**
  - >10X smaller than NWP's
- **HENS doubles sampling of distribution "tails":**
  - Extends 4σ into tails, compared to 2σ for NWP.
  - This extends rarity of LLHIs from $1/10^2$ to $1/10^4$.
- **HENS hindcasts LLHIs missed by NWP:**
  - HENS captures 96% of LLHIs missed by NWP.





Source: NASA Earth Observatory

Mahesh et al. (2024a,b)

U.S. Department of ENERGY | Office of Science

ML-Based ID of Extremes in Climate Model Output

CASCADE
CALIBRATED & SYSTEMATIC CHARACTERIZATION, ATTRIBUTION, & DETECTION OF EXTREMES

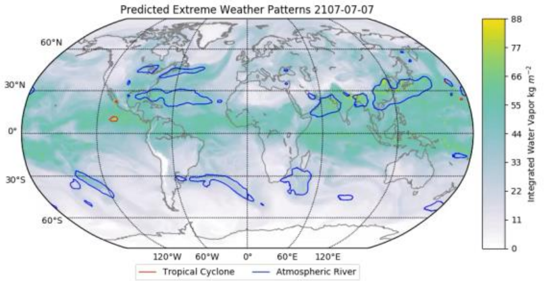
World's 1st Climate Exascale Application – 2018 Gordon Bell Prize

- **Training Input: Cropped, Centered, Multi-variate patches with Labels***
  - Tropical Cyclone (TC)
  - Atmospheric River (AR)
  - Weather Front (WF)

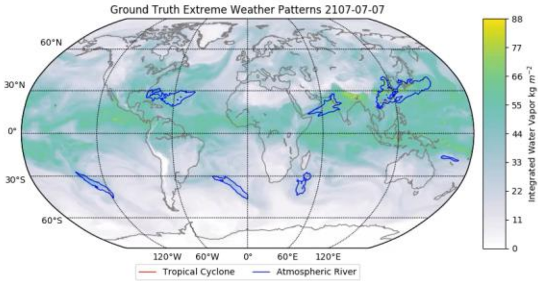
**Labels are provided by human-specified criteria*

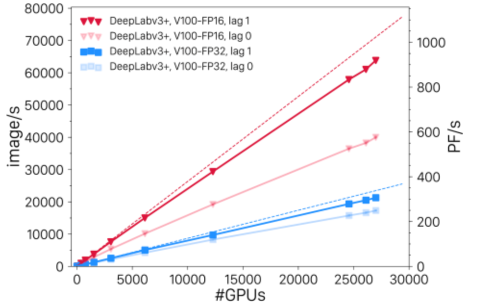
- **Machine Learning Output: Binary (Yes/No) on Test patches**
  - Is there a TC in the patch?
  - Is there an AR in the patch?
  - Is there a WF in the patch?

Machine Learning



Ground Truth





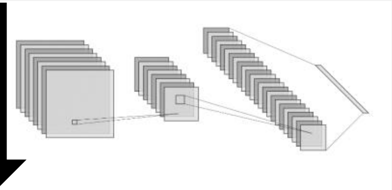
- 4560 Summit nodes, 27,360 Volta GPUs
- 1.13 EF peak performance (16-bit)

Kurth et al. (2019)

Unsupervised Learning for Signal Extraction

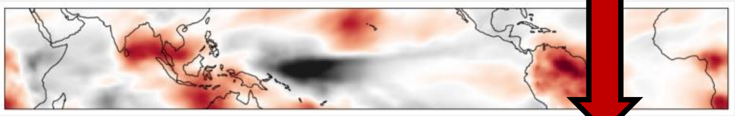
Explainable AI techniques are being developed to identify regions/processes that contribute to prediction skill.

Step 1: **Train ML**
Physically relevant upstream fields (SSTs, OLR)

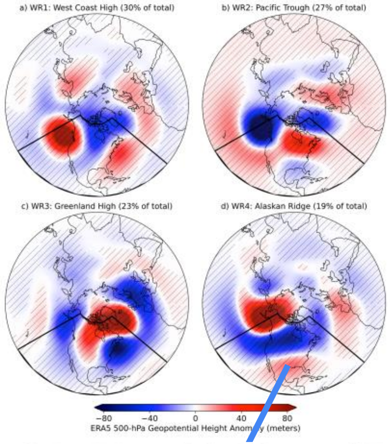


S2S/S2D Prediction
Modes of variability (MJO, ENSO)
Impacts (temperature, precip.)

Step 2: **Explainable AI**
Understand the relevant features behind impacts

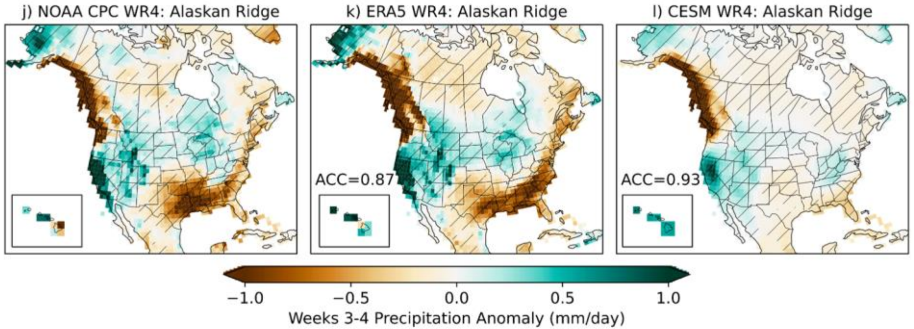


Generate heatmaps (saliency maps, LRP) using input fields (e.g., Barnes et al. 2020) to identify regions/processes that contribute to prediction skill.



If predicted accurately, dominant patterns of atmospheric circulation (top) can produce derived patterns of predicted precipitation (bottom).

Molina et al., 2023, AI for Earth Systems



Using LLMs to Quantify Floods in Crowd-Sourced Data

HYPERFACETS — U.S. Department of Energy | Office of Science — A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales

- Accurate observations of flash floods are essential for quantifying and validating flood risk.
- **However, flood observations are rarely available in urban areas due to a lack of gauges.**
- Can crowd-sourced data, including citizen and government reports, social media, and news outlets be used?

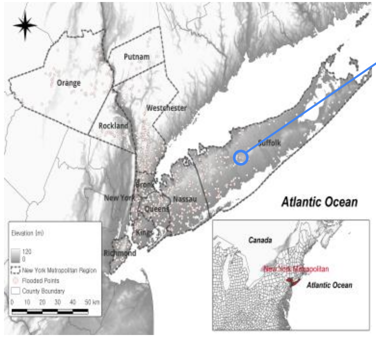


Figure: Distribution of flash flood crowd-sourced data over the New York Metropolitan Area

Report: "The bus was flooded up to the floorboards."

Prompt (User): "Could you estimate the inundation depth indirectly described in this report based on the average dimensions of buses?"

ChatGPT: "The ride height of buses is 27.6 inches. Therefore, the inundation depth in the report is at least 27.6 inches."

Figure. Example of LLM application to flood data retrieval

- Most crowd-sourced data, such as video, pictures, and textual descriptions, are unstructured and not immediately suitable for engineering or scientific applications.
- Large Language Models have recently gained attention in processing and extracting information from unstructured data.
- LLMs have potential to retrieve flood observations from textual descriptions.

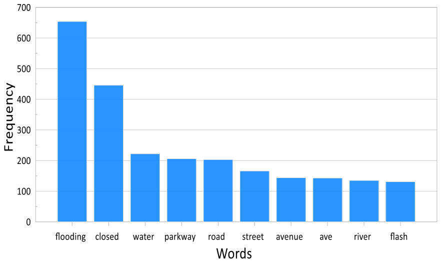


Figure: Most frequently mentioned words in flood damage descriptions

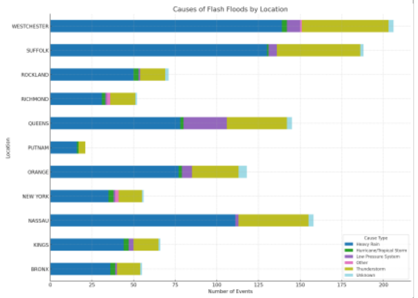


Figure: Most frequently mentioned words as flash flood causes

Kim, S.H., Yavari, Y., Devineni, N., and Pei, T. (2025). Can Large Language Models Quantify Floods from Crowd-Sourced Data?, World Environmental and Water Resources Congress 2025. American Society of Civil Engineers.

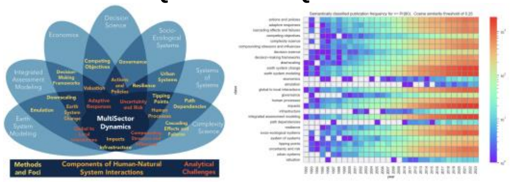
U.S. Department of ENERGY | Office of Science

MSD-FUTURES: Foresights for Understanding Thematic Unity in Reviews of Emergent Science

Activities/Approach

- Gain an understanding of the MultiSector Dynamics (MSD) literature landscape through an AI-integrated framework
- Keep the **expert-in-the-loop** by integrating 24 MSD Community of Practice (CoP) members of diverse expertise throughout the process to train models and provide guidance
- Corpus of literature spanning **105,336** publications each assigned probabilities of relevance to MSD and research area characteristics
- Explore **communities of researchers**, institutions, etc. that conduct highly-aligned research
- Predict emerging research topics and potential growth patterns within MSD at large

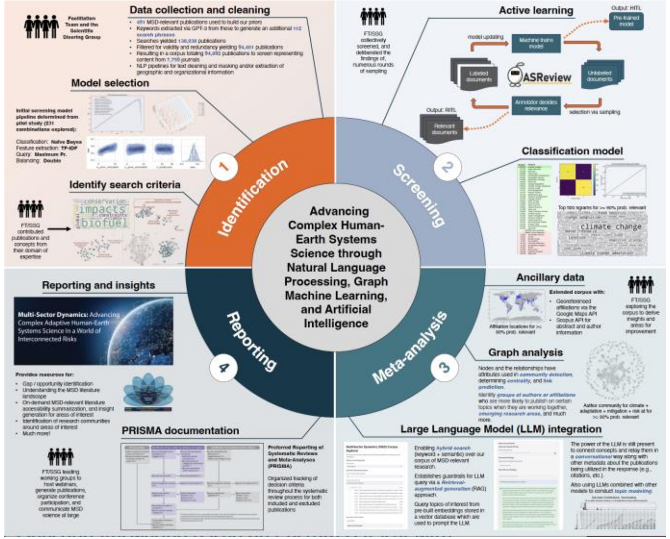
From Qualitative to Quantitative



Vernon, C.R., Reed, P.M., Hadjimichael, A., Monier, E., et al. (in prep)

Key Capabilities

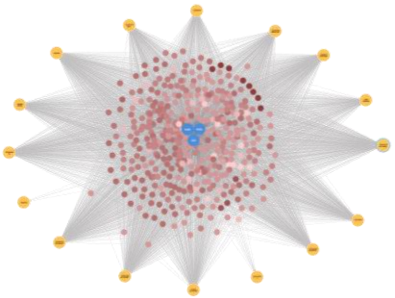
A living, transferable AI-integrated framework using Large Language Model, Graph, Natural Language Processing, and other novel ML advancements (e.g., semantic classifier for MSD) to accelerate collaborative research and strategic planning.



Future Work

- Individual working groups exploring focused expansion, growth patterns, and terminology and methodology evolution from our corpus
- Automated update protocol for real-time literature harvesting, classification, and graph community expansion to enable a highly-relevant resource to the community
- Extension to explore MSD CoP impact through time

Leiden Communities across Research Areas



Total ELectricity Loads (TELL) Model: Projecting Multiscale Future Hourly Electricity Loads in the U.S.

The **TELL machine learning model** is able to effectively predict hourly electricity demand using a multi-layered perceptron (MLP) neural network.

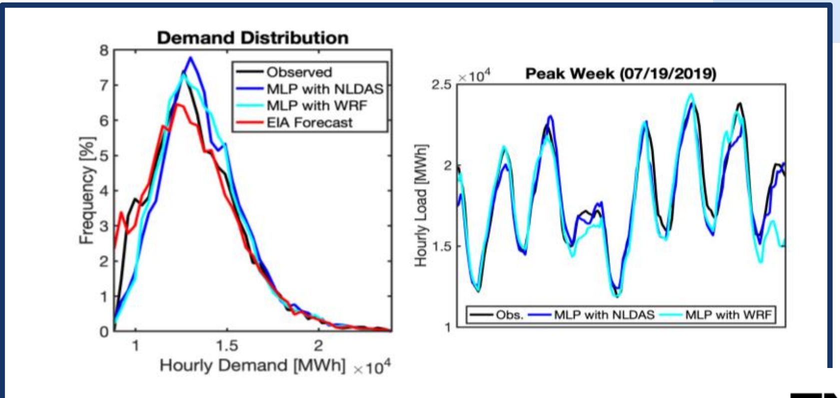
Predictive Variables	Units
Temperature	K
Specific Humidity	kg kg^{-1}
Shortwave Radiation	W m^{-2}
Longwave Radiation	W m^{-2}
Wind Speed	m s^{-1}
Day of Week	Weekday/ Weekend
Federal Holiday	Yes/No
Hour of Day	00-23 UTC

Able to capture the distribution of demand and the dependence of load on temperature, seasonal variations and weekday-weekend variations, diurnal variations during peak demand weeks.

Each balancing authority (electricity grid region) is trained and evaluated independently. TELL performs better in larger, summer-peaking balancing authorities.

Projects like the National Transmission Planning Study (NTPS) are using TELL to generate hourly load profiles used to test the resilience of their projected system to extreme weather events.

https://immm-sfa.github.io/tell/



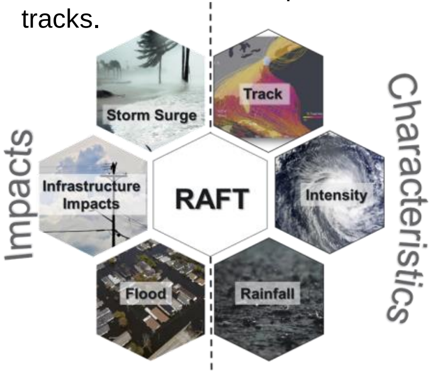
McGrath, C., C. D. Burleyson, Z. Khan, A. Rahman, T. Thurber, C. R. Vernon, N. Voisin, and J. S. Rice, 2022: tell: a Python package to model future electricity loads. *Journal of Open Source Software*, 7(79), 4472, doi:10.21105/joss.04472.

Risk Analysis Framework for Tropical Cyclones (RAFT)

WACCEM WATER CYCLE AND CLIMATE EXTREMES MODELING
ICOM

Activities/Approach

- **RAFT** is a hybrid model that combines physics, statistics and ML to produce large ensembles of Tropical Cyclones (TCs) for robust characterization of risk from TCs.
- While a physics-based approach is used to produce TC tracks and rainfall, **deep neural networks** are used to predict intensity along tracks.

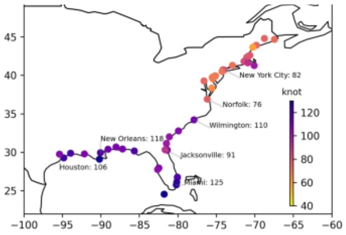


Xu, W., Balaguru, K., Judi, D. R., Rice, J., Leung, L. R., & Lipari, S. (2024). A North Atlantic synthetic tropical cyclone track, intensity, and rainfall dataset. *Scientific Data*, *11*(1), 130. https://doi.org/10.1038/s41597-024-02952-7

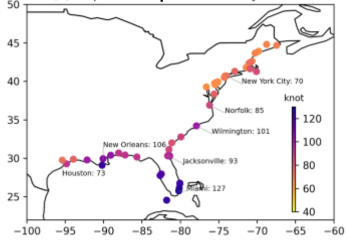
Key Capabilities

- Produces large numbers of synthetic TCs consistent with observed historical storms.
- Can be integrated with climate models to project TC risk into the future.
- Has been combined with an ML-based model of the power grid to evaluate resilience to TCs
  https://interactive.epri.com/readi-storymap/p/1

Observed 99th percentile landfall intensity



RAFT simulated 99th percentile landfall intensity

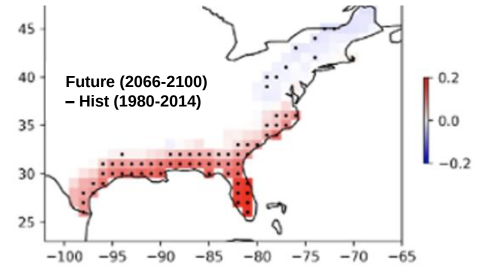


Balaguru, K., Xu, W., Chang, C.C., Leung, L.R., Judi, D.R., Hagos, S.M., Wehner, M.F., Kossin, J.P. and Ting, M., 2023. Increased US coastal hurricane risk under climate change. *Science advances*, *9*(14), p.eadf0259. https://doi.org/10.1126/sciadv.adf0259.

Future Work

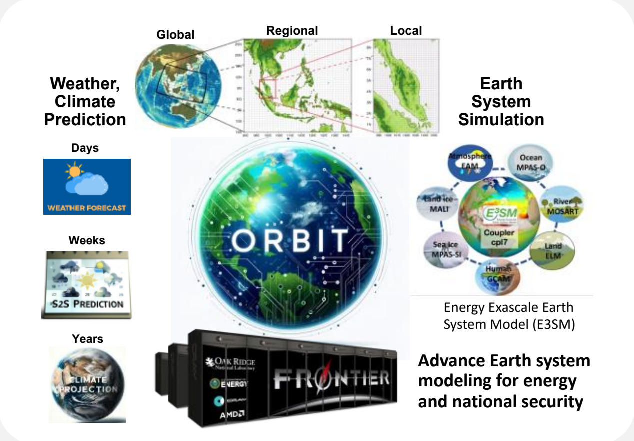
- Expand the domain of RAFT globally, include multiple climate scenarios and time slices to better understand uncertainty in TC projections.
- Develop a new rainfall model within RAFT that uses **GAN** to improve accuracy.
- Application to various types of energy infrastructure (eg. offshore wind)

RAFT projected change in TC exposure probability



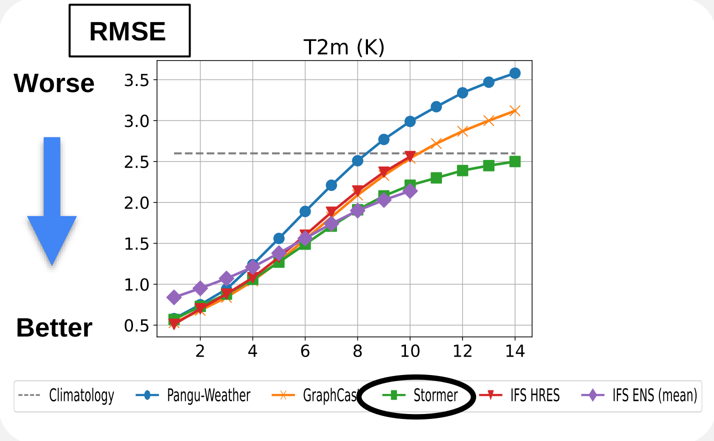
Foundation Models developed by Labs (outside of EESSD)

ORBIT: Oak Ridge Base Foundation Model for Earth System Predictability



- ORBIT is a 113-billion-parameter AI model, advancing Earth system predictability with accuracy, efficiency, and versatility.
- ORBIT achieves 1.6 exaflop performance on the Frontier exascale supercomputer, winning Top Supercomputing Achievement Award and selected as Gordon Bell Prize Finalist.
- ORBIT demonstrates superior accuracy in weather forecasting up to 30 days ahead.

STORMER: Foundation Model for Medium Range Weather Forecasting



- A simple transformer model that achieves state-of-the-art performance on weather forecasting with minimal changes to the standard transformer backbone.
- Able to make 14-day forecasts in 2 seconds at 30km resolution.

Summary of Top Challenges

1. AI-ready Data: High-fidelity, globally covered AI-ready data is often lacking, making data collection and curation resource-intensive. Training on incomplete data restricts model performance

- Finding, compiling, and curating high-quality AI-ready data is resource- and time-consuming
- Need for high-resolution data to capture important fine-scale features and for out-of-sample testing
- Need for all earth system domain data including atmosphere, land, ocean, sea-ice, ice-sheet, and biogeochemistry
- Address uncertainty in climate model data that can propagate to ML models

2. Credibility evaluation of AI models: AI models must align with physical laws, yet current validation standards are underdeveloped, and explainability is a challenge.

- Demonstrating credibility of AI/ML generated data products
- Establishing requirements for checking consistency of AI models with physical laws to ensure trustworthy predictions
- Enforcing desired structural properties like equivariance and mass conservation in models

3. Out-of-Sample use of AI/ML and lack of rigorous theory: The climate system has components that evolve on timescales of centuries to millennia, much longer than the quantitative observational record.

- Understanding the limitations of AI/ML to be used out-of-sample (e.g., unprecedented extremes)
- Lack of rigorous theory for what the necessary and sufficient conditions are for an ML emulator (or ESM) to be skillful at prediction

4. Expanding AI Applications: Need to build on limited success in using CNNs for climate systems, incorporating AI for better understanding of predictability and data assimilation.

- ML-based ESMs are still nascent, with many issues that make them difficult to use alongside physics-base

Top Areas of Interest for Cross-Agency Collaboration

1. Model Development and Calibration

- **Model Parameterization and Emulation**: Use AI/ML to develop model surrogates, parameterizations, and emulators that improve model calibration, climate sensitivity analysis, and Earth system predictability.
- **Physics-ML Hybrid Models**: Develop AI/ML hybrid models across various Earth system components to improve model stability and generalizability.

2. Data Sharing and Standardization

- **AI-Ready Datasets and Data Assimilation**: Develop AI/ML compatible datasets and integrate observational data across processes, resolutions, and instrument platforms.
- **Multi-Source Observational Data**: Develop diverse observational data, such as remote sensing and ground-based sources, to enrich ML model training and Earth system insights.

3. Predictive Capabilities across scales and Real-World Applications

- **Subseasonal–Multidecadal Predictability:** Leverage AI/ML for refining predictions at various scales, from subseasonal to long-term climate change impacts.
- **Climate Resilience and Infrastructure Planning**: Develop interdisciplinary datasets, including socio-economic factors, to tailor ML climate forecasts and projections for operational applications in disaster preparedness and infrastructure planning.

4. Model Interpretability and Trustworthiness

- **Interpretability and Consistency with Physical Laws**: Develop interpretable models that align with physical principles to ensure reliable results, especially when addressing overfitting and generalizability challenges.
- **Evaluating Model Credibility**: Develop/apply diagnostics and metrics to validate AI-based GCMs and emulators, supporting their credibility for climate research.

Thank You

November 4th, 2024

Department of Energy
Office of Science
Earth and Environmental Sciences Division

Gerald Geernaert

U.S. Department of ENERGY | Office of Science