Using domain-aware metrics for deploying AI/ML in weather/climate applications

Peetak Mitra, Ph.D.
Founding Scientist, Excarta
(peetak@excarta.io)

Excarta

NSF Meta-Workshop on AI to accelerate Science and Engineering Discovery (AI2ASED)
October 2, 2023

Challenges in translating AI metrics to scientific disciplines

- **Data**:
  - Multiscale, multimodal/heterogenous, high-dimensional, sparse, noisy, class imbalance (for example in case of long-tail /extreme weather events), and shifting distributions
  - Physical laws <u>that need to be respected</u> *and* unknown physics that <u>needs to be discovered</u>!
    - Models should **respect or incorporate physical laws**, constraints, and other domain knowledge!
    - In discovery-oriented tasks, **ground truth is unknown** and benchmark data sets are unavailable
  - Data privacy and silos (private vs open-sourced)!

- **Algorithmic/Interpretability**:
  - Reproducibility, robustness and ability to quantify uncertainty are required for scientific rigor
  - Extracting new scientific insights from data requires **human-interpretable** models or outputs.

Synopsis of AI/climate workshop:

Tackling Climate Change with Machine Learning at NeurIPS 2022

High level overview:

- **Goal**: To catalyze important work at the intersection of AI and climate change
- Theme on **climate change-informed metrics for AI**
- Multiple panels and keynotes from academic and industry leaders organized
- Over 250 original research articles and proposals submitted
  - About 100-odd papers and proposals accepted, including 10 spotlights!
  - Most innovative articles encouraged to be published in the Environmental Data Science journal

Highlights of key takeaways

On translating AI-centric metrics to science and engineering

- **Deployment-in-the-wild** remains a major challenge (across domains)
  - ML-centric metrics works great in a sandbox, but often **translate poorly to real world**
  - **Generalization** to *long-tail events* remains an issue
  - Very **high cost of making errors** (ex. in power grid management, could lead to blackouts or high penalties)
- **Domain specific metrics** (for guardrails)
  - Each domain has their own optimization requirements **beyond standard ML metrics**
    - Ex. in model control, metrics such as energy efficiency, cost optimization, constraint violations are key
    - Ex. in forest management, the ecosystem services the asset provides to humanity (ex. fresh water, carbon sequestration capacity) cannot be directly optimized but are critical
  - Transferability of ML metrics to domain understanding (== physically meaningful) isn't often clear
    - Ex. in weather forecasting, a low global RMSE *does not guarantee* physical laws are being respected
- For safety-critical engineering systems, **fail safe implementations are critical**!

AI is revolutionizing weather and climate science!

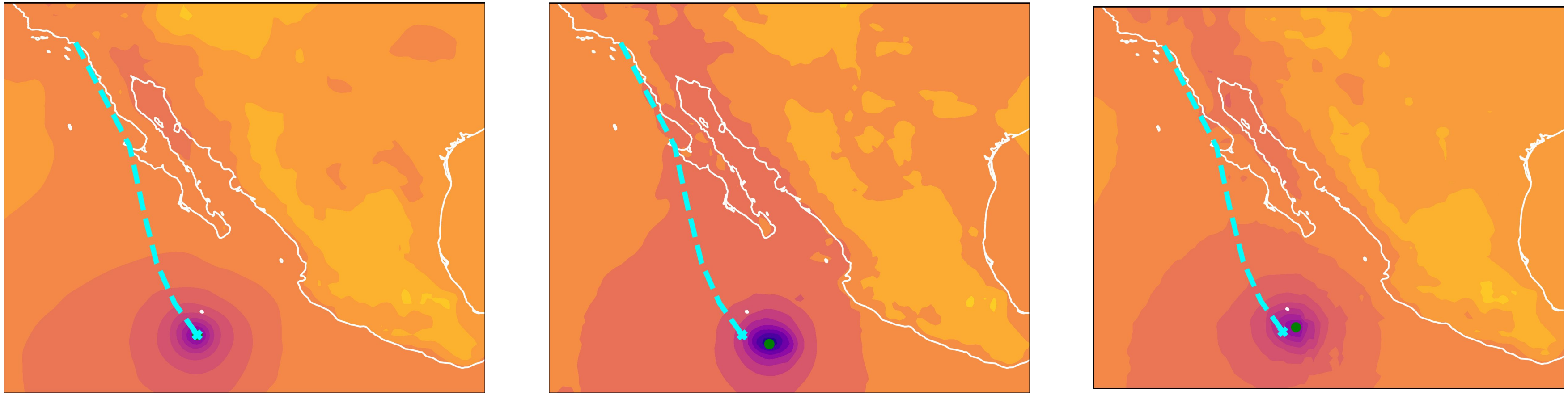
Surface pressure

Forecasts issued on August 15, 2023 at 00 UTC, landfall occurred August 20, 2023

Actual conditions: 2023-08-18 12:00 UTC

GFS: Issued 2023-08-15 00:00 UTC,
Valid 2023-08-18 12:00 UTC

Excarta: Issued 2023-08-15 00:00 UTC,
Valid 2023-08-18 12:00 UTC



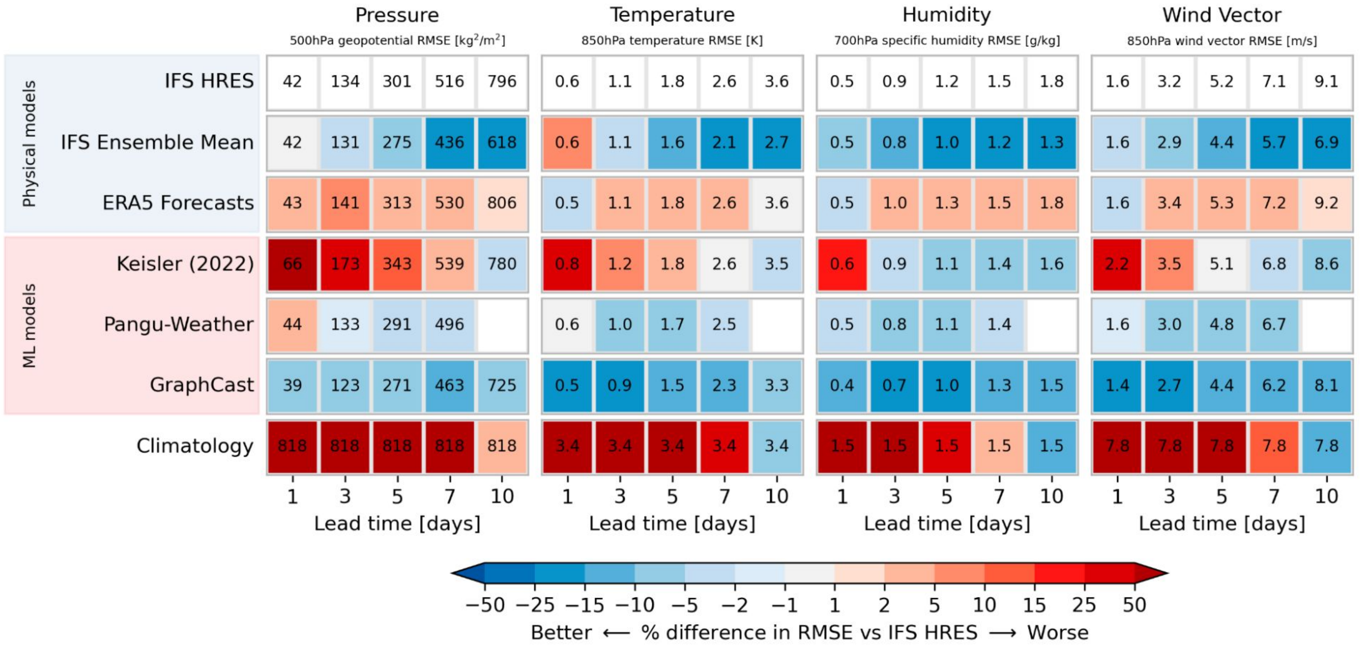
Tropical Storms and Hurricanes making landfall on the western seaboard of the CONUS are considered extremely rare

Full blog: https://excarta.io/2023/09/08/case-study-tropical-storm-hilary/

Excarta

Weather Bench 2 evaluation

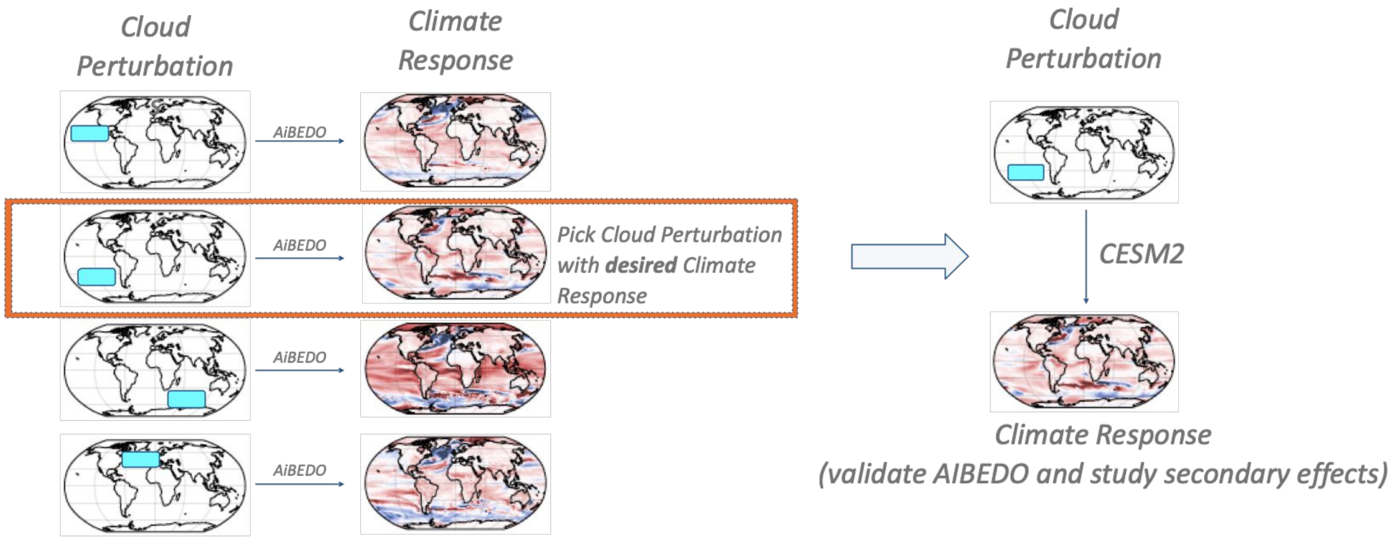
Skill of different ML forecasts compared to ECMWF's Integrated Forecast System (IFS)



Source: Google AI blog post

Climate emulators can help in better understanding of our shared future!

Surrogate Based Optimization for problems such as exploring climate intervention scenarios or running large ensembles with perturbations



AIBEDO inference on single NVIDIA V100 is is 1-2 seconds per model year
CESM2 run on 200 cores are 28 hours per model year

DARPA parc Palo Alto Research Center University of Victoria W

Paper accepted in GRL

In conclusion: word of caution for the use of AI!

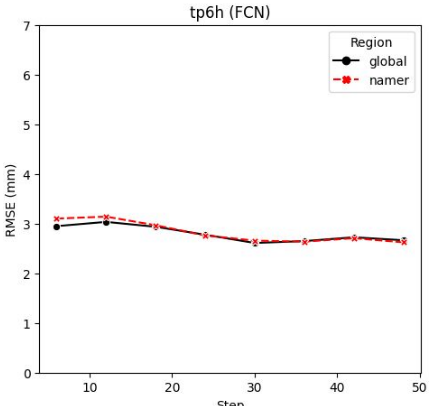
- Does AI really add value to the problem?
  - Is it just mimicking what's already present or is it providing a new source of information

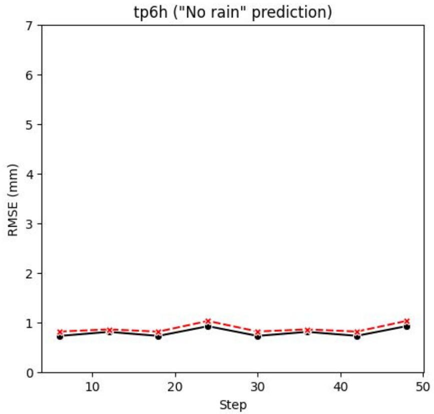
- Choosing the right metrics
  - Are standard ML metrics (ex. RMSE) appropriate?
  - Domain specific metrics might yield to better generalization!

- Care about data quality
  - **Training**: Models are very sensitive to data
  - **Validation**: You can't value what you can't measure!

- Aligning inductive biases of data to algorithm is important
  - Helps choose right models

- Deployment in the wild is a major challenge
  - Continuous monitoring and improvement remains critical





Ramavajjala, V, and Mitra, P. "Verification against in-situ observations for Data-Driven Weather Prediction." *arXiv preprint arXiv:2305.00048* (2023)