THE AI REVOLUTION FOR WEATHER AND CLIMATE

TROY ARCOMANO[1], A. WIKNER[14], T. NGUYEN[6], H. GUAN[4], M. POSKA[13], R. MAULIK[23], S. MADIREDDY[2], I. FOSTER[3], S. FOREMAN[5], A. GROVER[6], R. KOTAMARTHI[1]

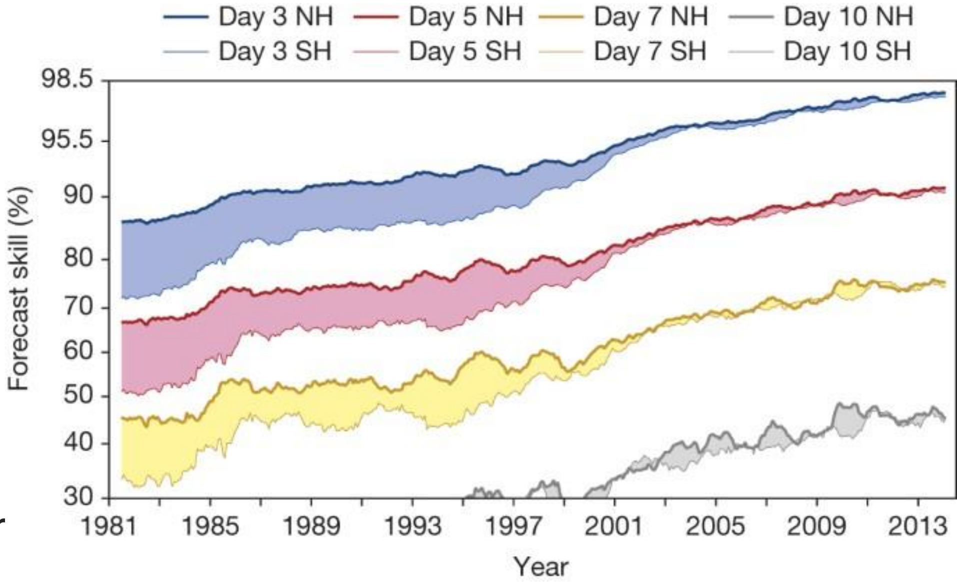
[1]Argonne National Laboratory, Environmental Science Division
[2]Argonne National Laboratory, Mathematics and Computer Science Division
[3]Pennsylvania State University
[4]University of Maryland, College Park
[5]Argonne National Laboratory, Argonne Leadership Computing Facility
[6]University of California, Los Angeles, Department of Computer Science

Oct 1, 2024

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

THE QUIET REVOLUTION OF NUMERICAL WEATHER PREDICTION*

- Weather forecasting is a **multi-billion** enterprise with large socioeconomic impacts
- Currently weather forecasting and climate modeling use **physics-based** numerical models
- **Slow, incremental** but steady progress was been made during the last **40 years** has lead to a quiet revolution for weather forecasting
  - **1 day of forecast skill per decade**
  - Successful predictions of extreme events up to 8 days into the future



Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* 525, 47–55 (2015). https://doi.org/10.1038/nature14956

THE RISE OF DATA-DRIVEN WEATHER FORECASTING*

- Advances in **machine learning architectures**, **hardware**, **big data**, and financial motivation have set the stage for a **paradigm shift** in weather forecasting
  – State-of-the-art machine learning-based models have accuracy on par to operational NWP
    • Success has been demonstrated in operational settings

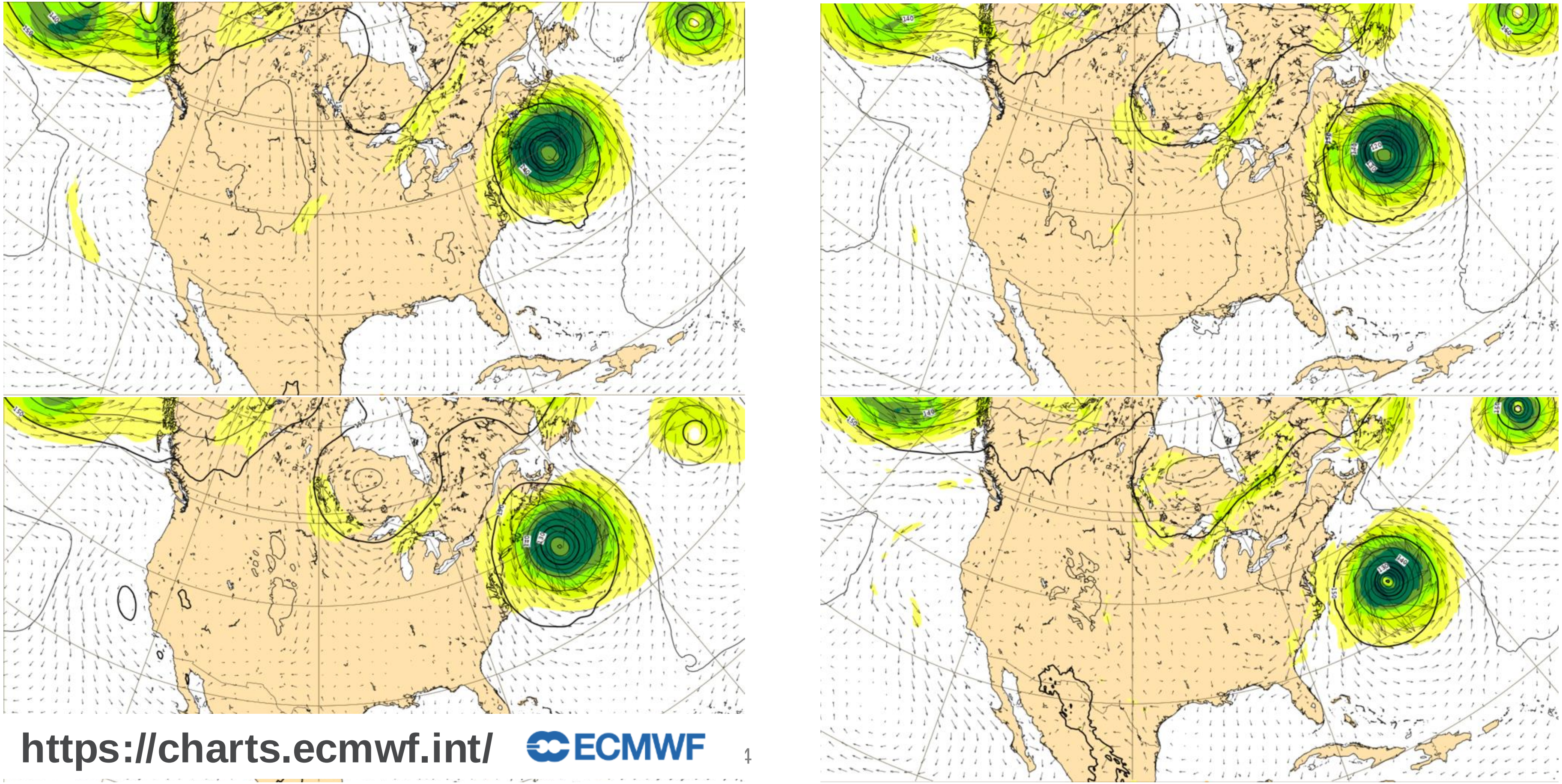
  – The efficiency is orders of magnitude better with 10-day forecasts taking just a few seconds

***Ben-Bouallegue et al. 2024**

https://doi.org/10.1175/BAMS-D-23-0162.1

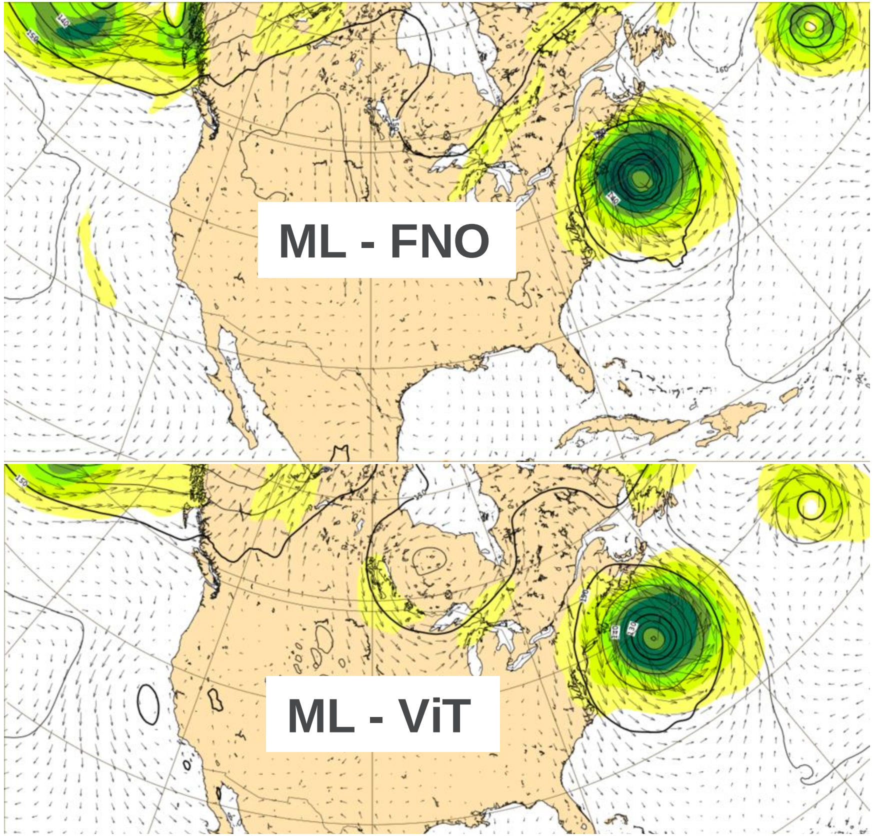
*** ECMWF seminar on data-driven models in operational setting**

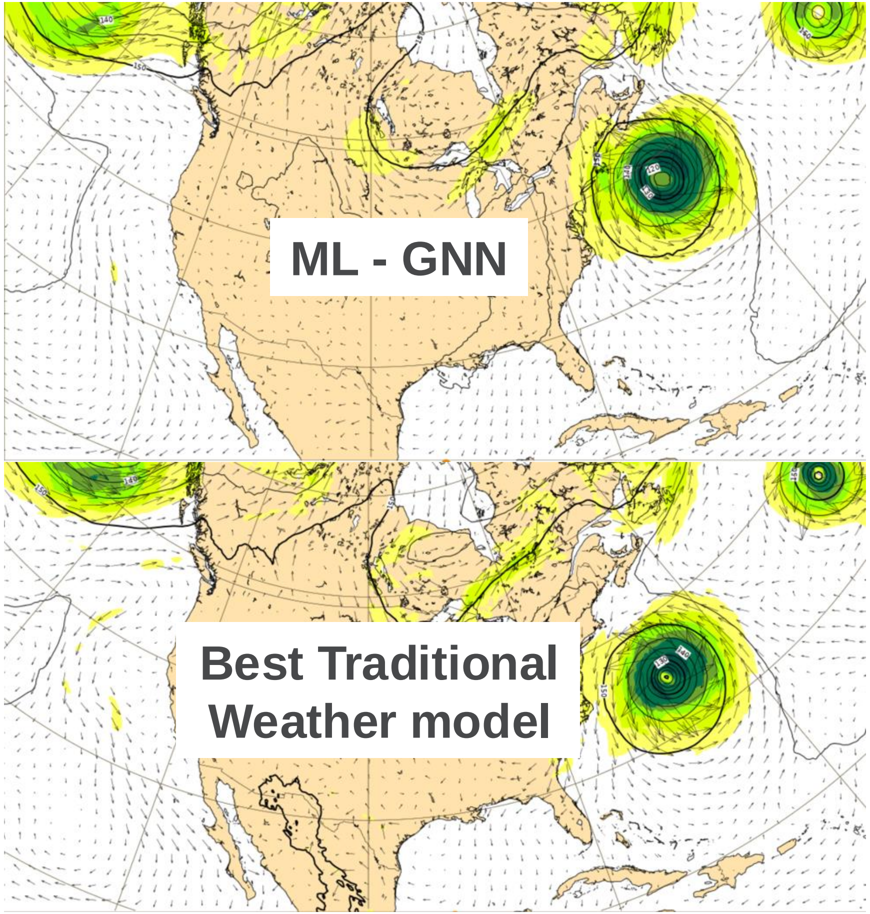
DAY 5 FORECAST



https://charts.ecmwf.int/ ECMWF

DAY 5 FORECAST





https://charts.ecmwf.int/

MACHINE LEARNING APPLICATIONS

- **Data-driven Methods**: Use of data-driven techniques for time-series forecasting
  - Independent of physics-based modeling (typically)
- **Hybrid modeling**: The combination of machine learning with existing traditional, numerical-based models
- **Operational Products**:
  - Severe Weather - Nadocast
  - Ocean Modeling - ENSO Prediction
  - Hurricane intensity forecasting
- **Uncertain Quantification**
- **Basically everything else**

DATA-DRIVEN APPROACH

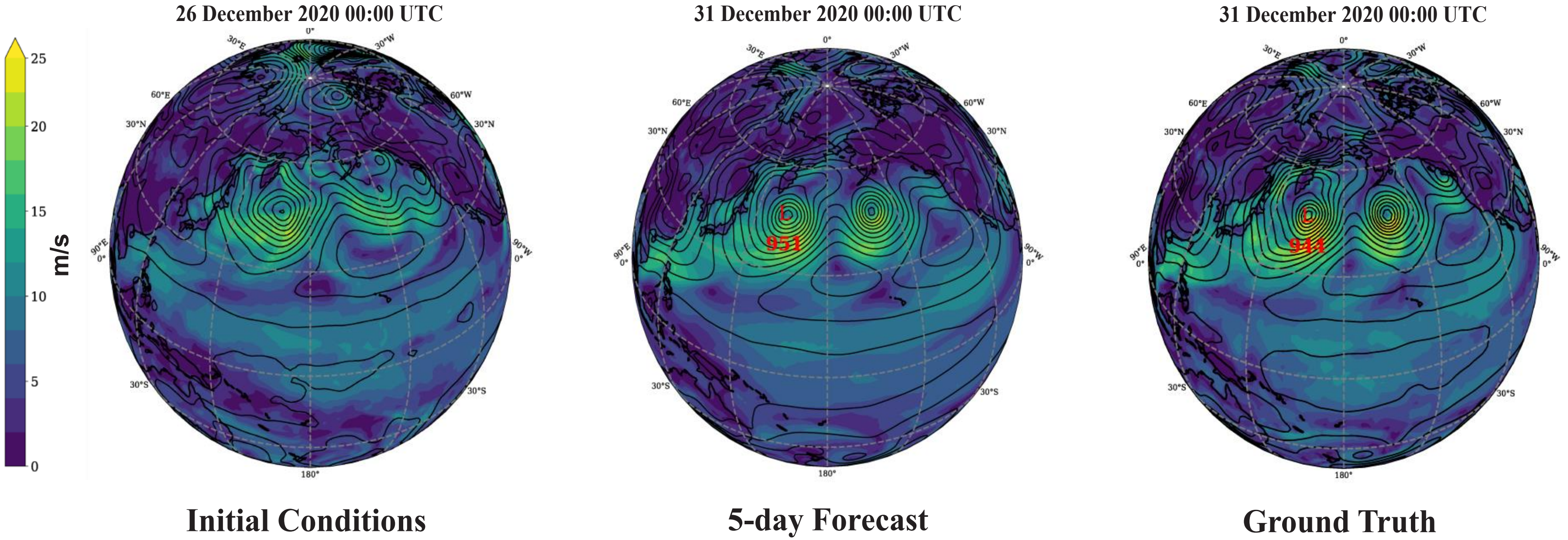
- **Task:** Take a snapshot of the 3-d atmosphere and predict the weather for the next **14 days**

- **Dataset:** Use observation-based reanalysis (best guess of the atmosphere)
  - ERA5

- **Challenges:**
  - **Image size** – 721 x 1440
  - **Channels** – 100s to 1000s of channels (each channel represents a 2d field)
  - Adaption software and hardware to these datasets
    - E.g. Complicated loss functions, using ViT for image translation, etc

- **Currently** using a weather specific ViT to predict the weather

MACHINE LEARNING-BASED WEATHER FORECASTING MODEL – STORMER*



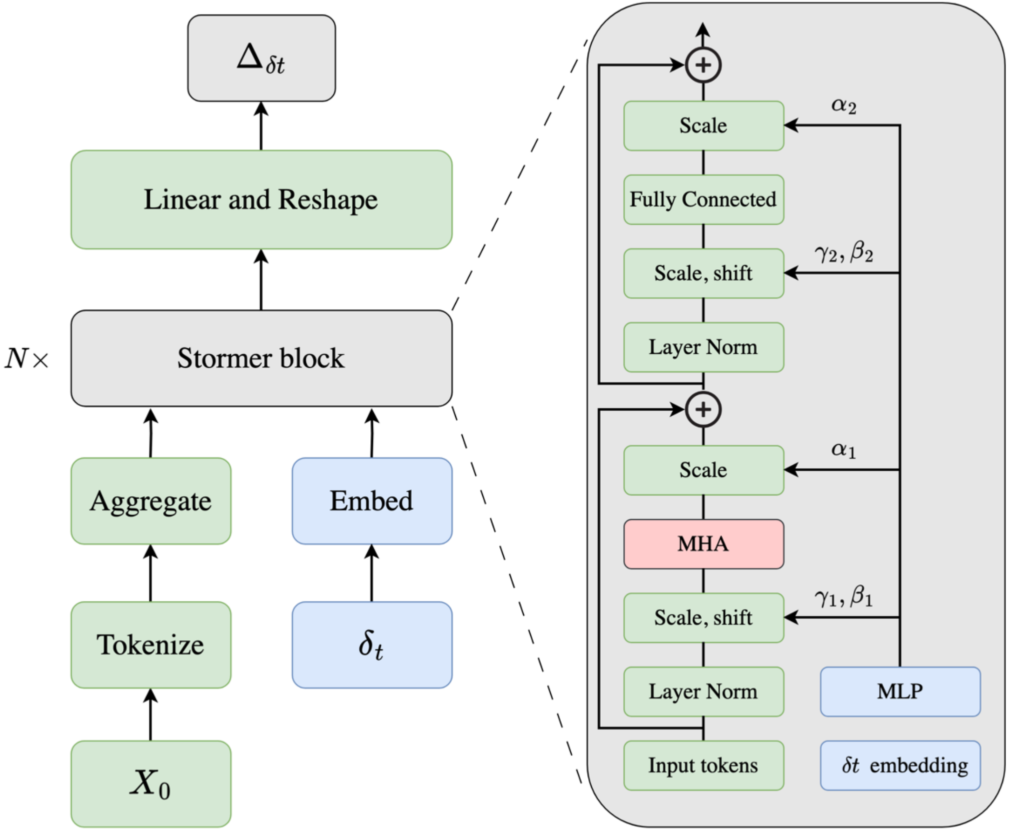
Successful 5-day prediction of an extratropical cyclone in late December 2020 which broke the North Pacific pressure record

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

*Nguyen, T., et. al. , 2023: Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. 2312.03876

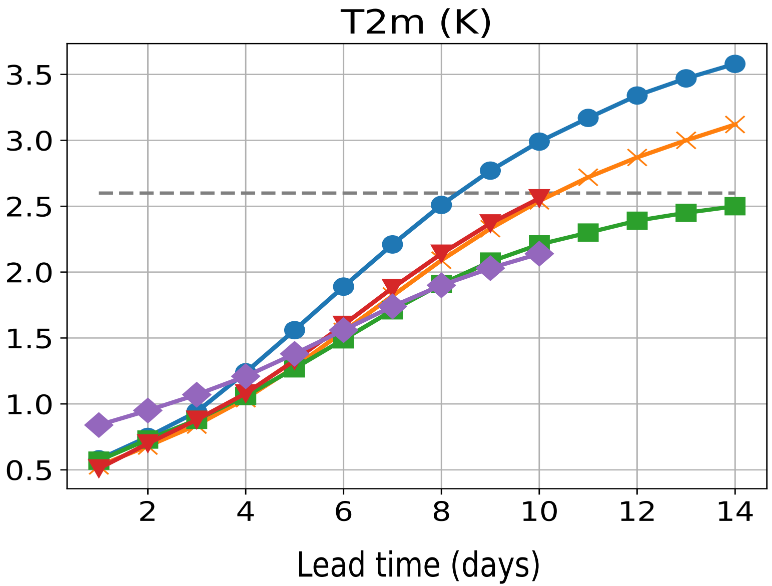
STORMER - VISION TRANSFORMER

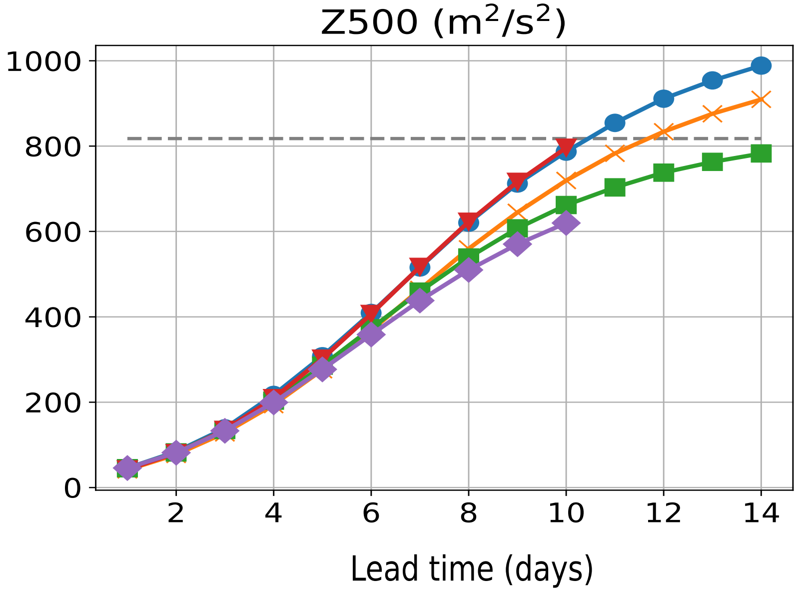
- **Model** :
  - Vision transformer backbone
    - adaptive layer normalization (**adaLN**)
  - Variable aggregation and tokenization
    - single-layer cross-attention mechanism
    - Model does not scale by number of channels



STORMER - PERFORMANCE

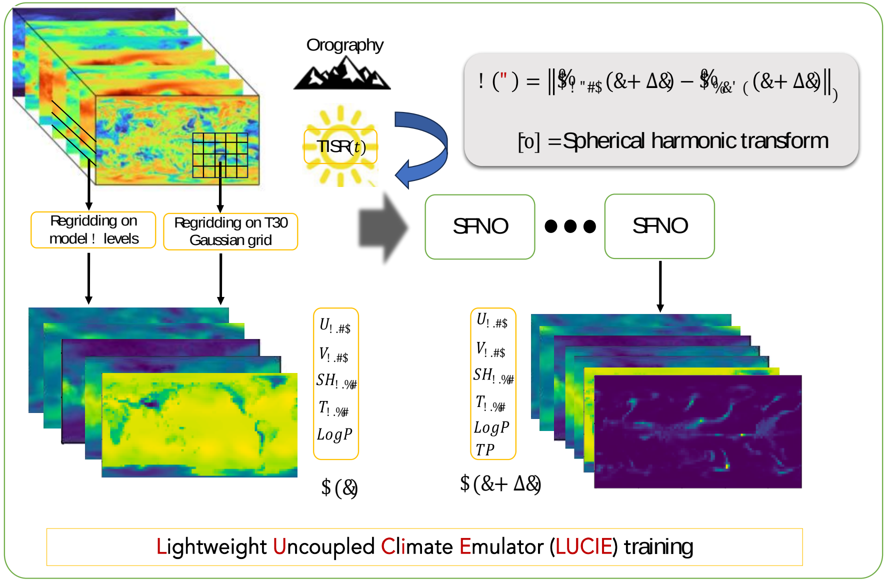
RMSE (Lower is better)

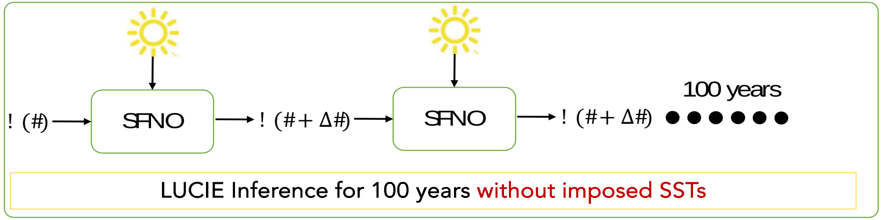


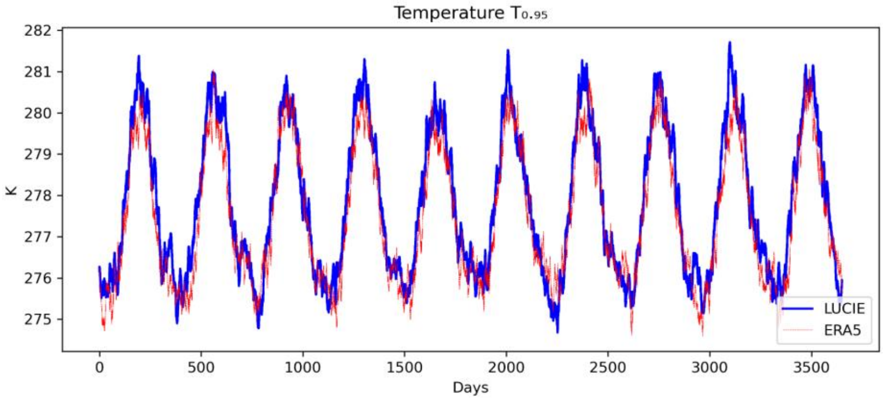


Climatology | Pangu-Weather | GraphCast | Stormer | IFS HRES | IFS ENS (mean)

LUCIE - CLIMATE

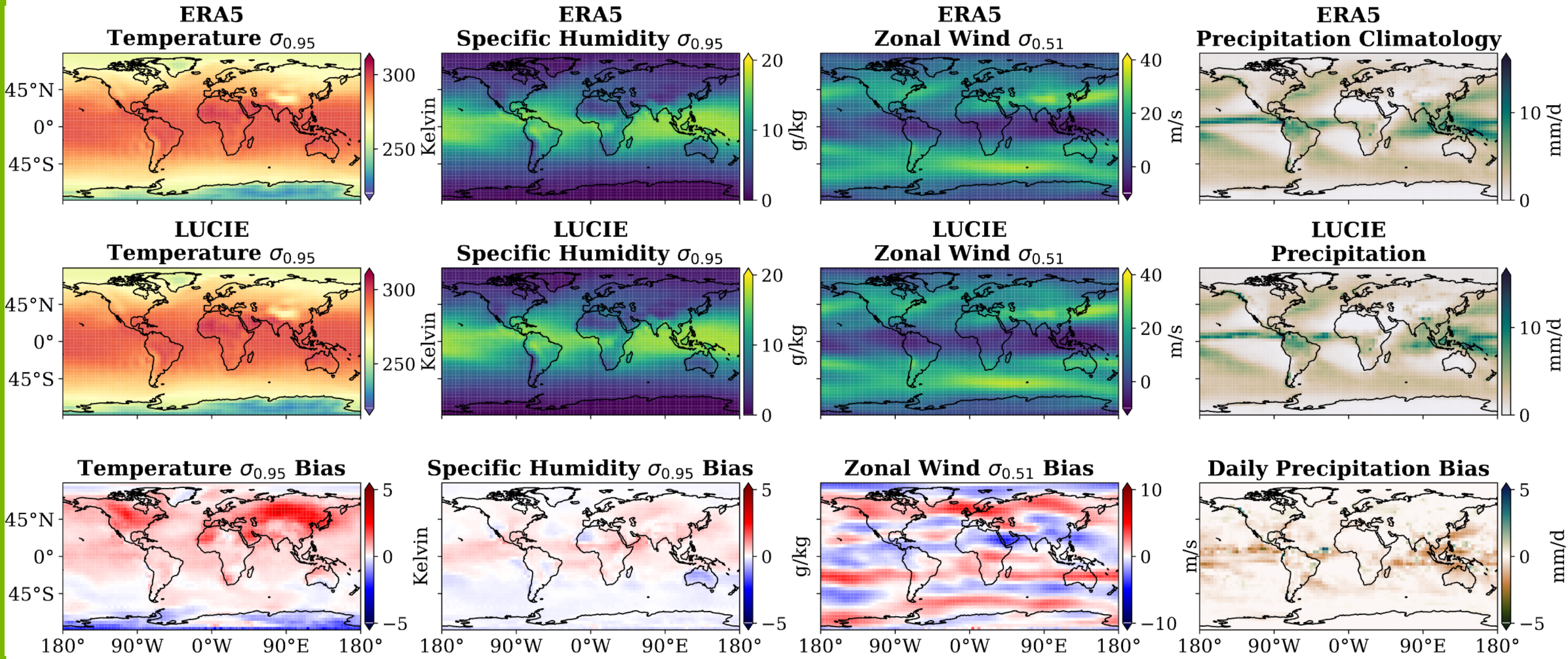






- **Dataset** – ERA5
- **Inference** - **6000 years** of simulation **per day** (**stable** for at least **1000 years**)
- **Architecture** – Spherical Fourier Neural Operator (SFNO)

LUCIE – CLIMATOLOGY



LUCIE can reproduce the general circulation with minimal biases

CONCLUSIONS

- The advent of scalable machine learning architectures, vast amounts of quality data, and access a large number of GPUs/TPUs is leading to a paradigm shift for weather forecasting

- Climate modeling may soon undergo a similar paradigm shift

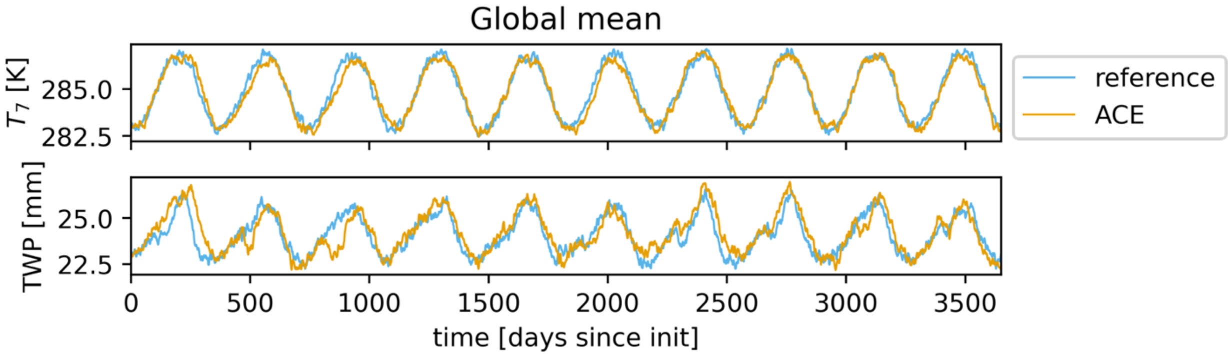
- Weather is a great test bed for newly developed ML architectures
  – Large data (PetaBytes)
  – Pushing limit of current hardware and software

QUESTIONS

BACKUP SLIDES

FUTURE OF CLIMATE MODELING

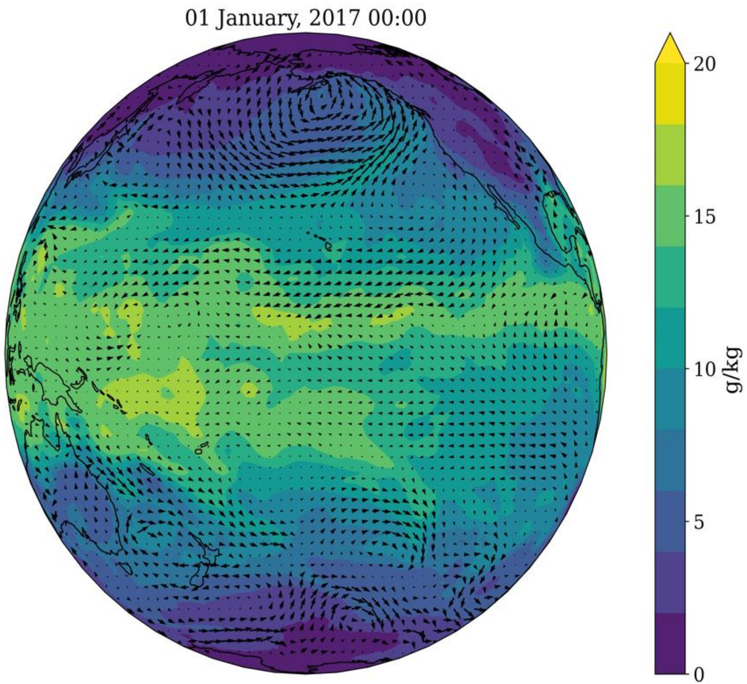
- New research demonstrates machine learning only emulators can produce stable simulation for a decade
  - **ACE Model**
- Biases, especially for precipitation, are better than SOTA numerical-based climate model



Watt-Meyer, O., Dresdner, G., McGibbon, J., Clark, S. K., Henn, B., Duncan, J., ... & Bretherton, C. S. (2023). ACE: A fast, skillful learned global atmospheric model for climate prediction. arXiv preprint arXiv:2310.02074.

MACHINE LEARNING-BASED WEATHER FORECASTING MODEL - CLIMAX

- Using transformer-based machine learning architecture based off **ClimaX***
- Model size:
  - **Training Data**: right now we are using a data set of 28 variables from ERA5 at 1.4 degree resolution (**WeatherBench** Rasp et. al 2020)
    - Temperature, UV-wind, geopotential and specific humidity at 250, 500, 700, 850, 925mb
    - Plus U10, V10, 2m Temperature
  - **Parameters**: ~ 100M and fine-tuned using 40 A100 GPUs
    - Takes ~8 hours



7-day ML-based forecast showing a series of atmospheric rivers impacting the west coast

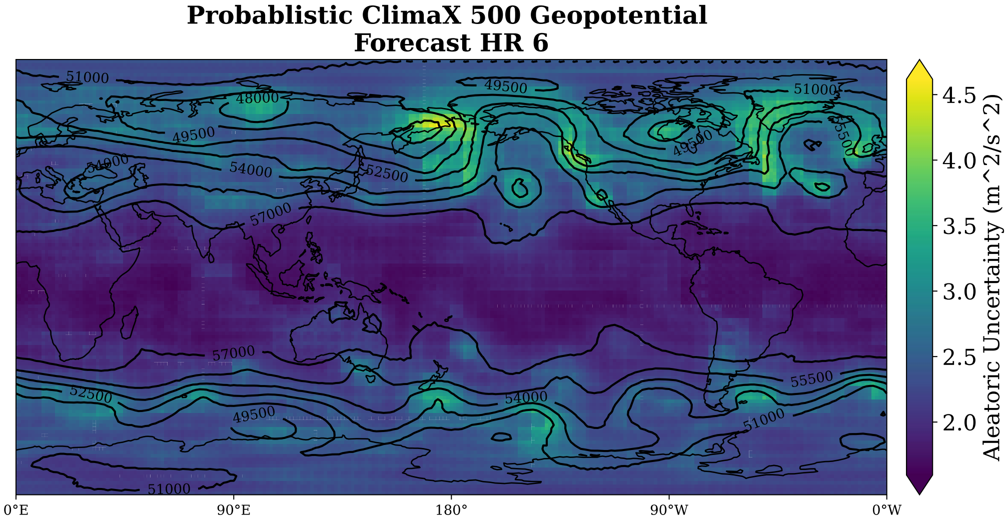
*Nguyen, T., J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, 2023: Climax: A foundation model for weather and climate. 2301.10343.

UNCERTAINTY QUANTIFICATION USING PROBABILISTIC OUTPUT

- Taking advantage of the flexibility of a ClimaX foundation model, we replace the normal output layer with a probabilistic one
  - Training can still be achieved in a reasonable amount of time
- Creating an output layer that is parameterized by a Gaussian provides assumption of aleatoric uncertainty.

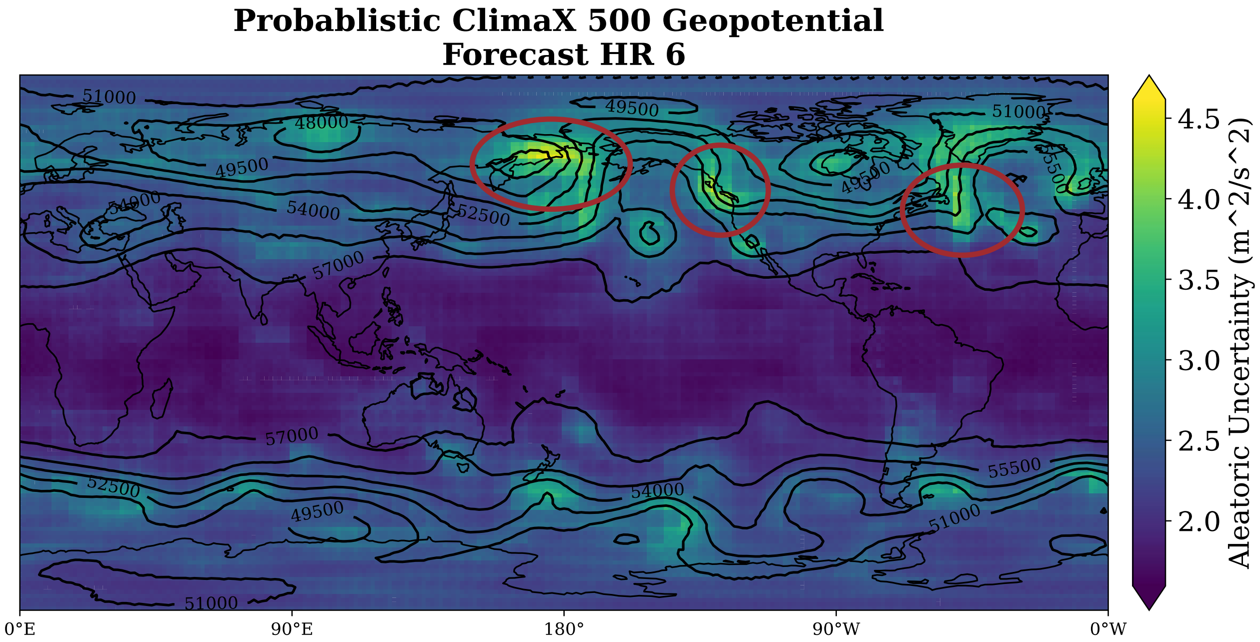
$$\mu_{\mathcal{E}} = \frac{1}{K} \sum_{\theta \in \mathcal{E}} \mu_\theta$$

$$\sigma^2_{\mathcal{E}} = \underbrace{\frac{1}{K} \sum_{\theta \in \mathcal{E}} \sigma^2_\theta}_{\text{Aleatoric Uncertainty}} + \underbrace{\frac{1}{K-1} \sum_{\theta \in \mathcal{E}} (\mu_\theta - \mu_{\mathcal{E}})^2}_{\text{Epistemic Uncertainty}},$$



Example output from a probabilistic model. Contours are mean prediction and color-fill is the aleatoric uncertainty

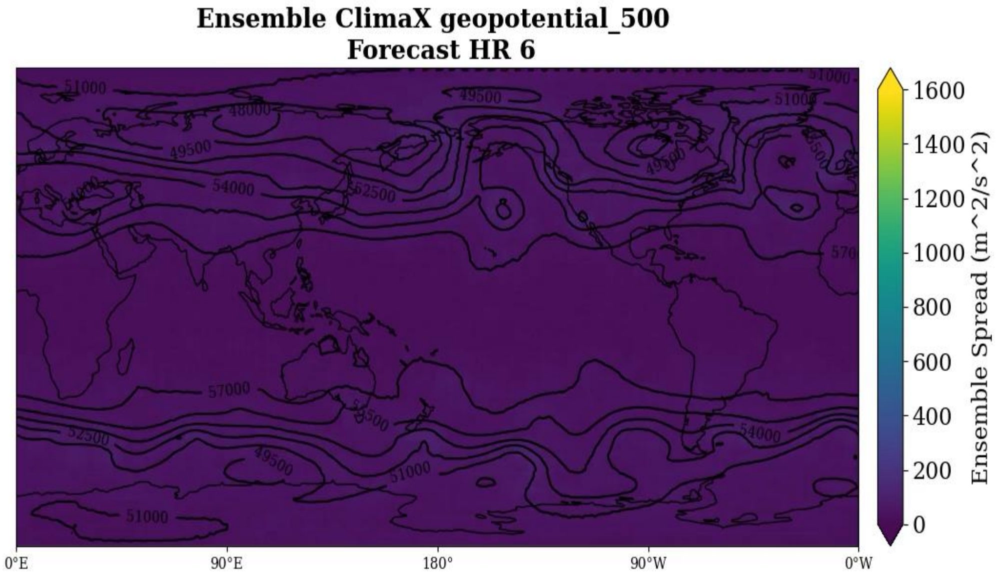
UNCERTAINTY QUANTIFICATION USING PROBABILISTIC OUTPUT



Largest uncertainty is contained in areas with synoptic features (e.g. short wave and deep cyclone)

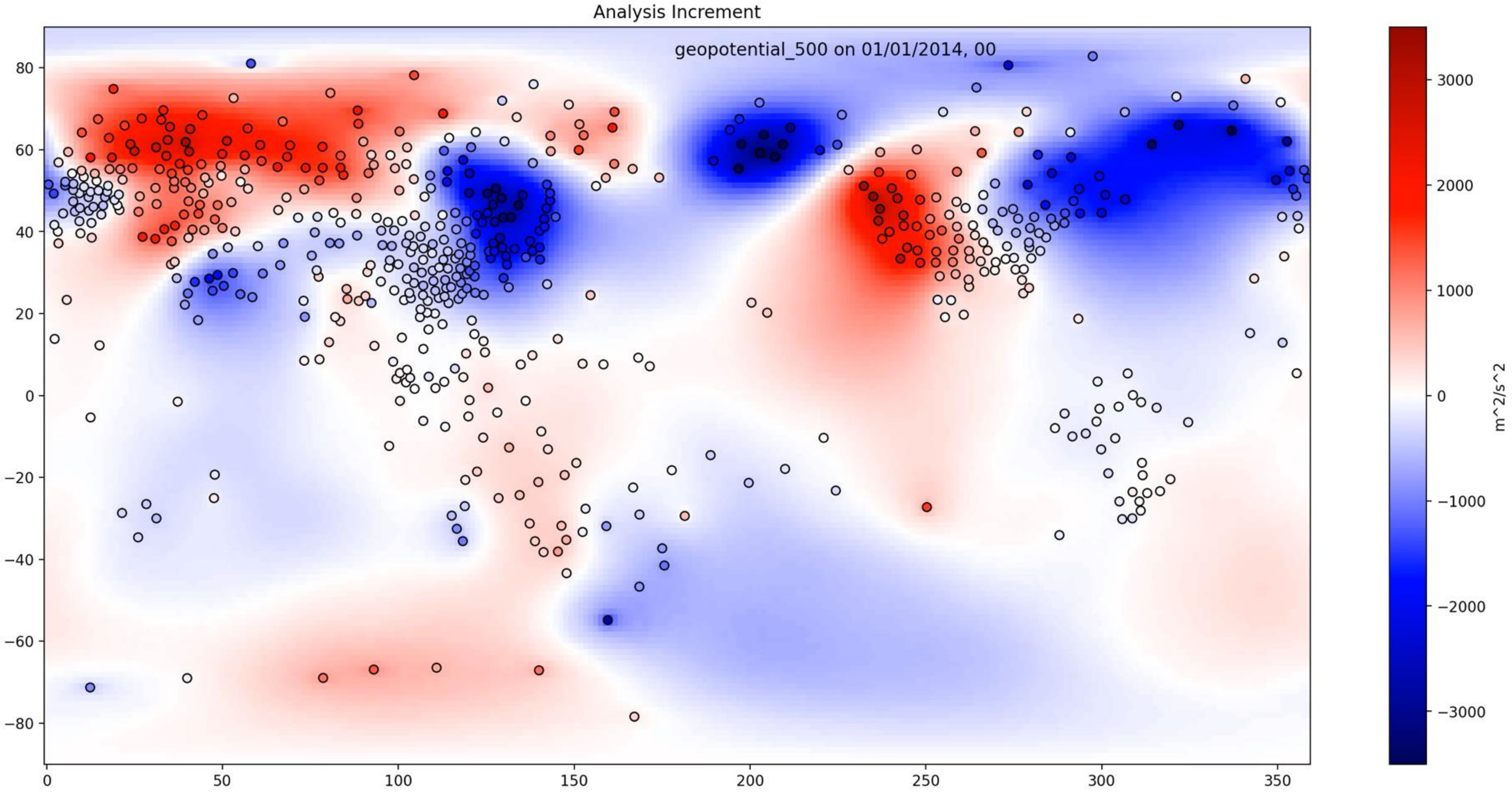
UNCERTAINTY QUANTIFICATION USING PROBABILISTIC OUTPUT

- **Ensemble System**
  - 40-member ensemble system 7 day takes 30s on a single A100
  - Use aleatoric uncertainty to perturb IC's for ensemble system
  - Can be used to estimate **B** matrix in a 4d-var DA system

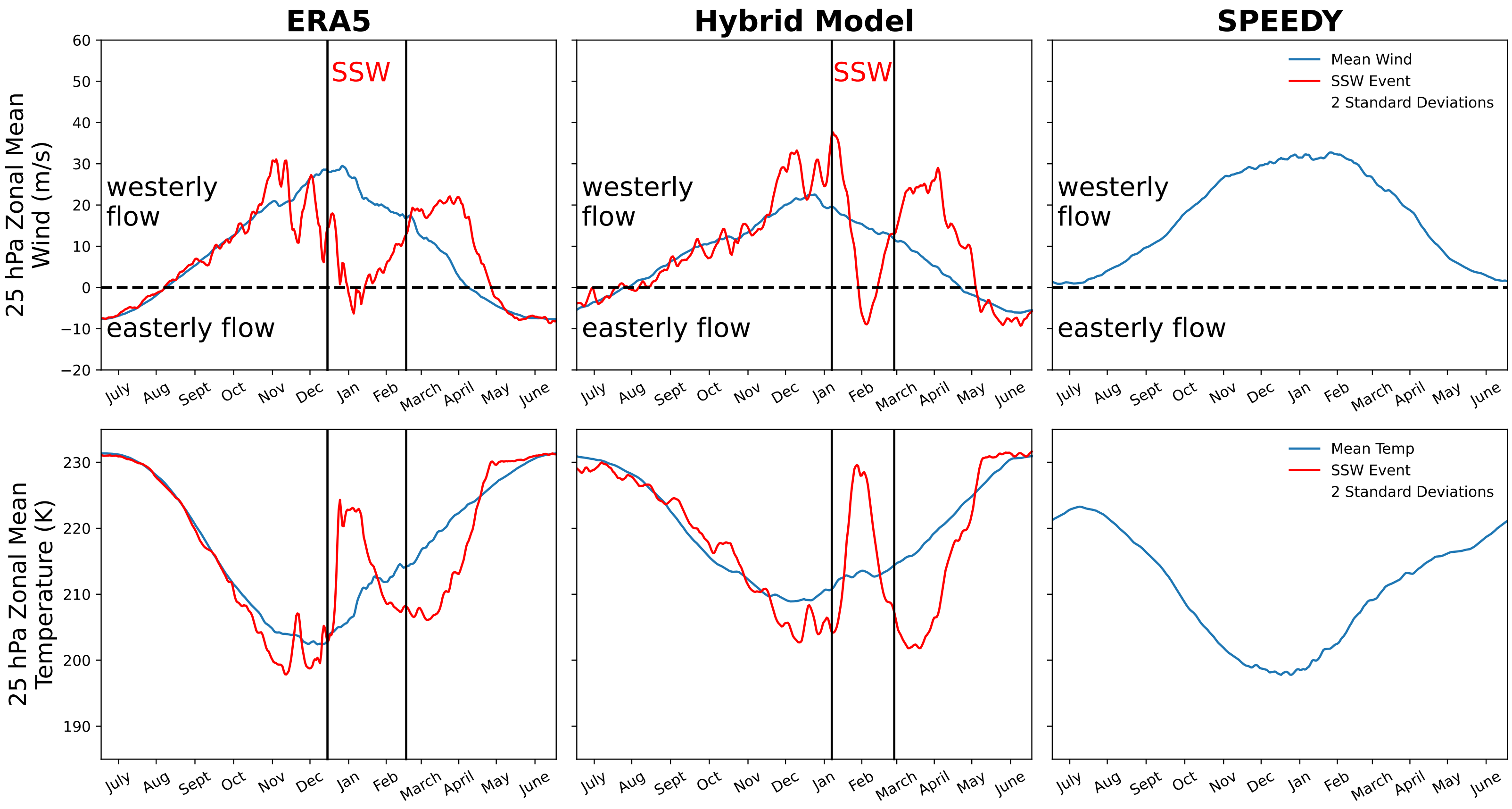


Example forecast using a ClimaX-based ensemble system with the black contours being ensemble mean and color-fill is ensemble spread

ANALYSIS INCREMENT

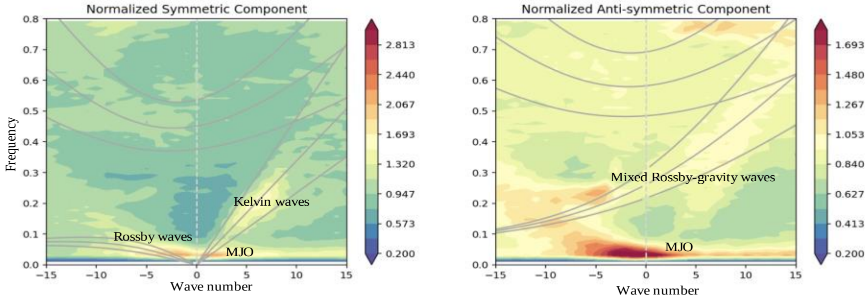


CLIMATE VARIABILITY – STRATOSPHERE

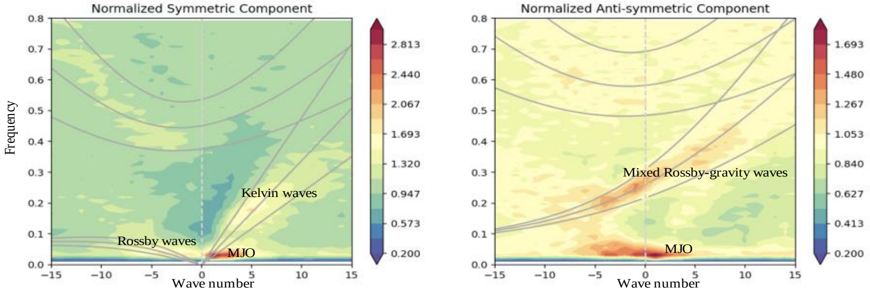


CLIMATE VARIABILITY – PRECIP

Hybrid Model



ERA5



SPEEDY Model