THE AI REVOLUTION FOR WEATHER AND CLIMATE

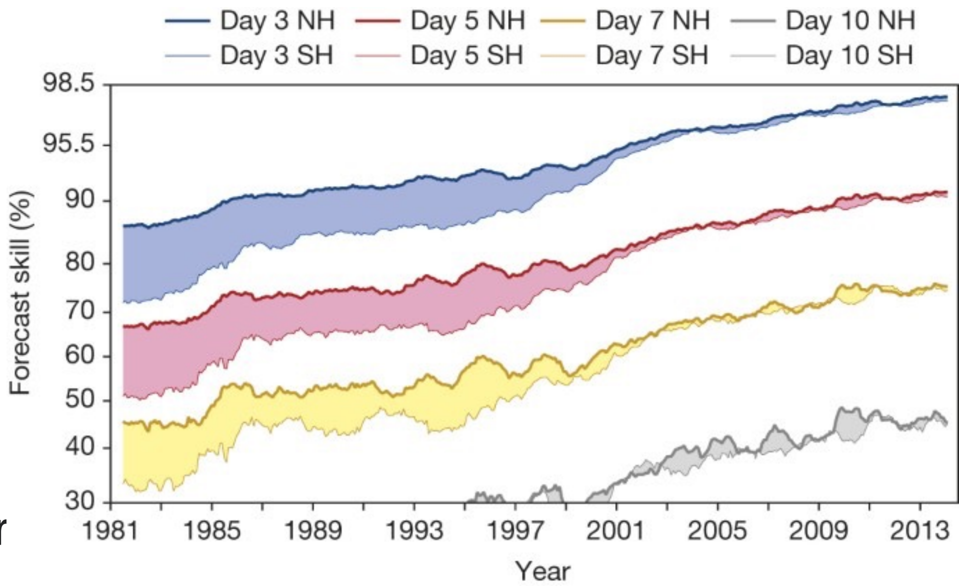
TROY ARCOMANO[1], A. WIKNER[14], T. NGUYEN[6], M. POSKA[13], R. MAULIK[23], S. MADIREDDY[2], I. FOSTER[3], S. FOREMAN[5], A. GROVER[6], R. KOTAMARTHI[1]

[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

Feb 27, 2024

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

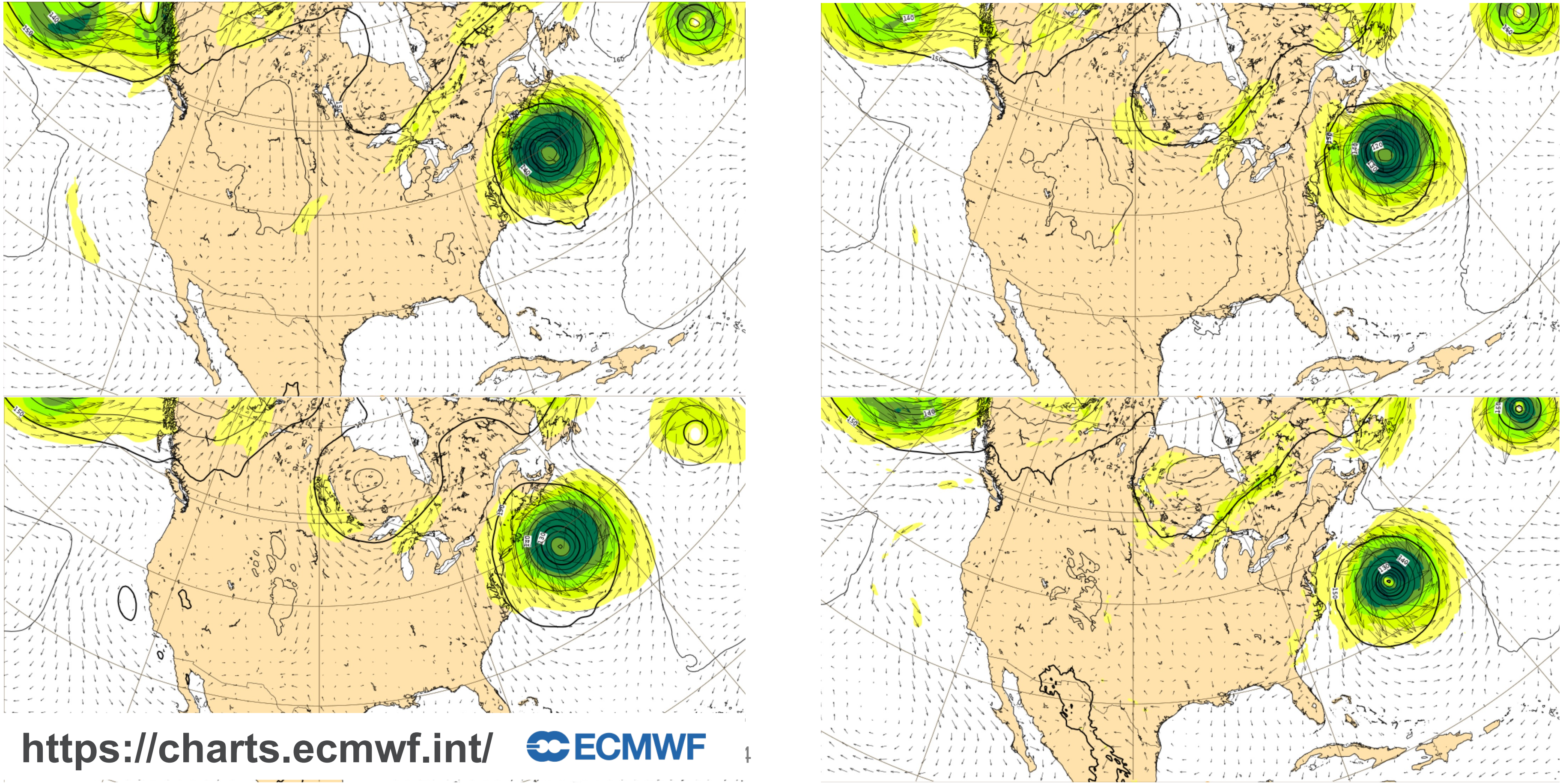
THE RISE OF DATA-DRIVEN WEATHER FORECASTING*

- 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. 2023**
https://arxiv.org/abs/2307.10128

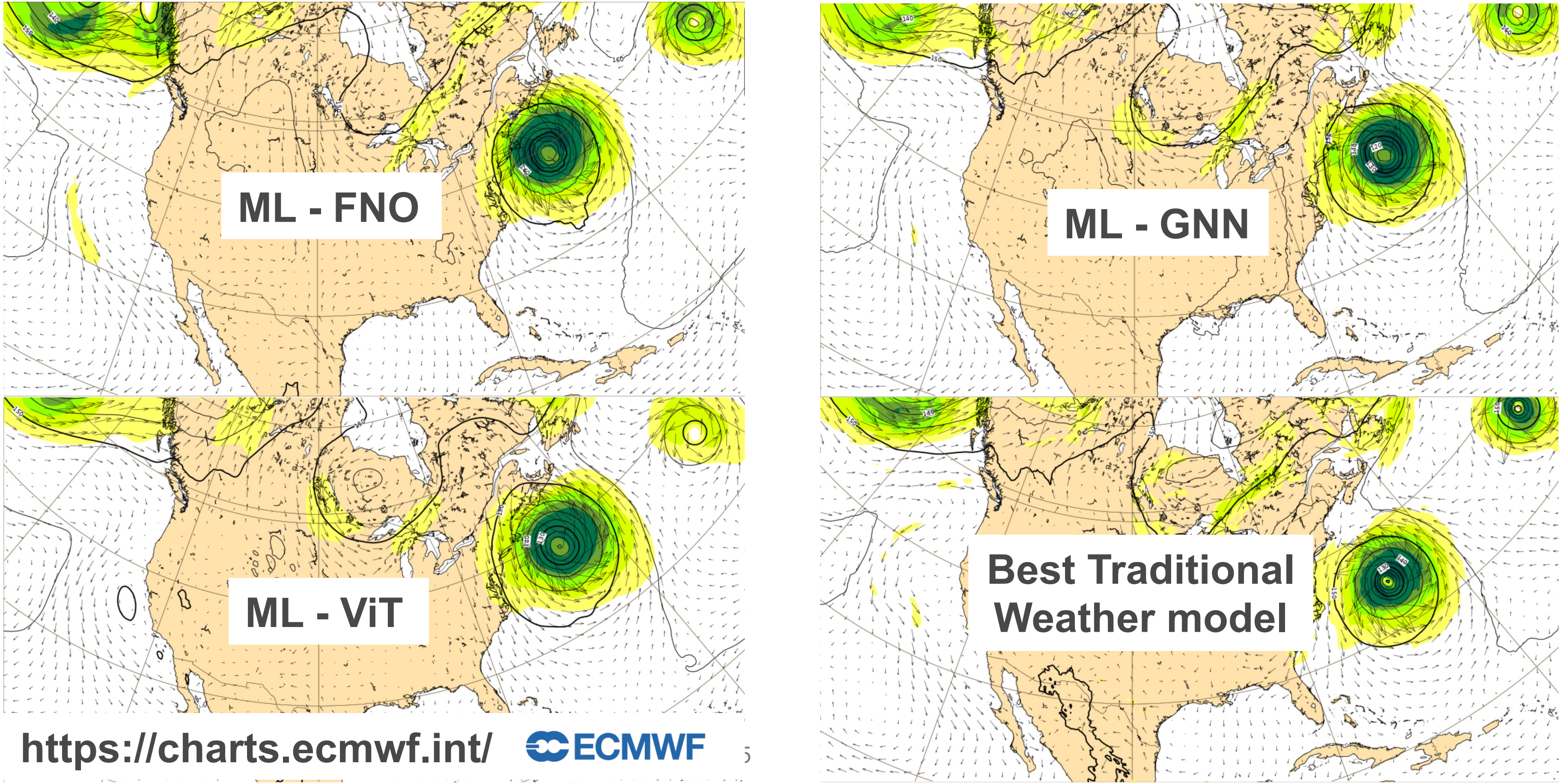
*** ECMWF seminar on data-driven models in operational setting**

DAY 5 FORECAST



https://charts.ecmwf.int/ ECMWF

DAY 5 FORECAST



https://charts.ecmwf.int/ ECMWF

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

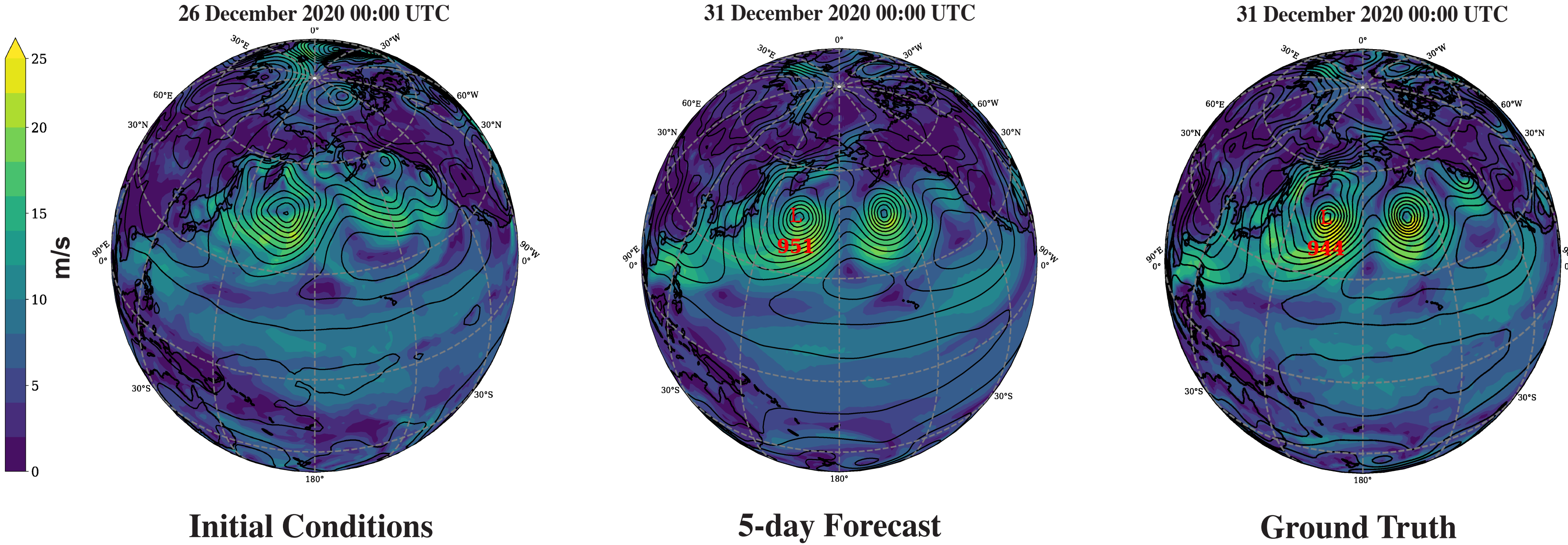
- **Task:** Take a snapshot of the 3-d atmosphere and predict the weather for the next **14 days**

- **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*

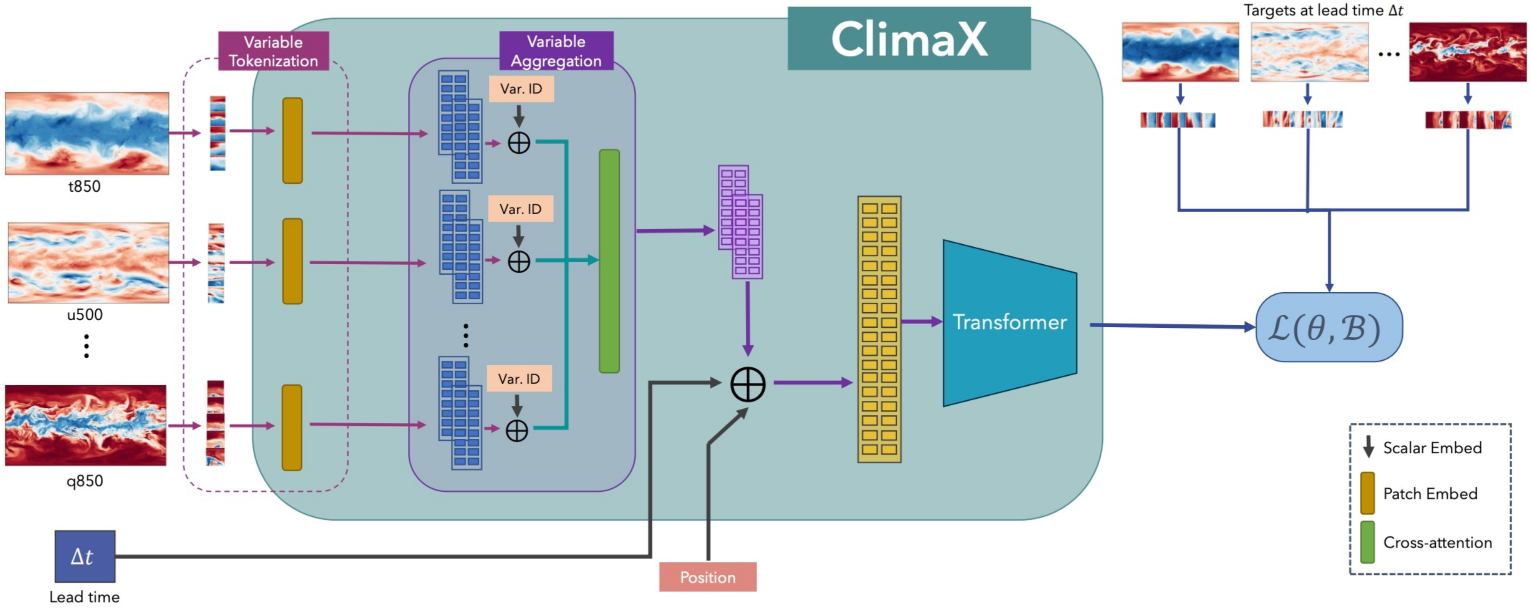


Successful 5-day prediction of an extratropical cyclone in late December 2020 which broke the North Pacific pressure record

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

STORMER - VISION TRANSFORMER

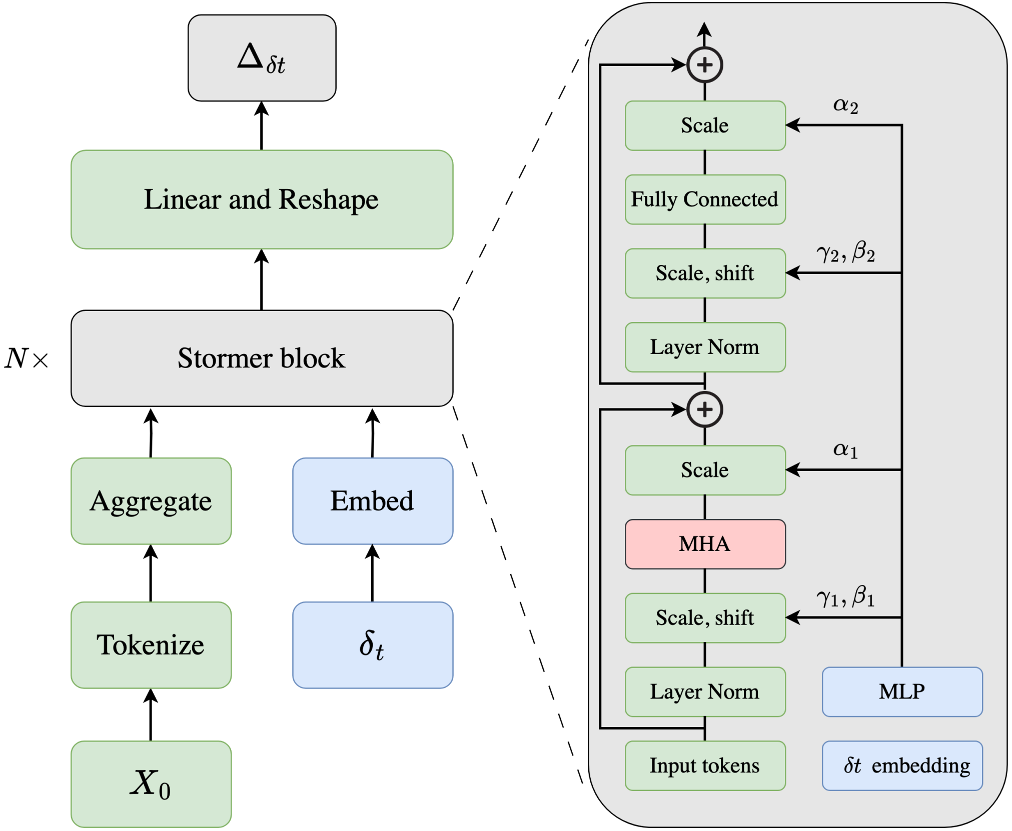
- Using transformer-based machine learning architecture (ViT) based of ClimaX*



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

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

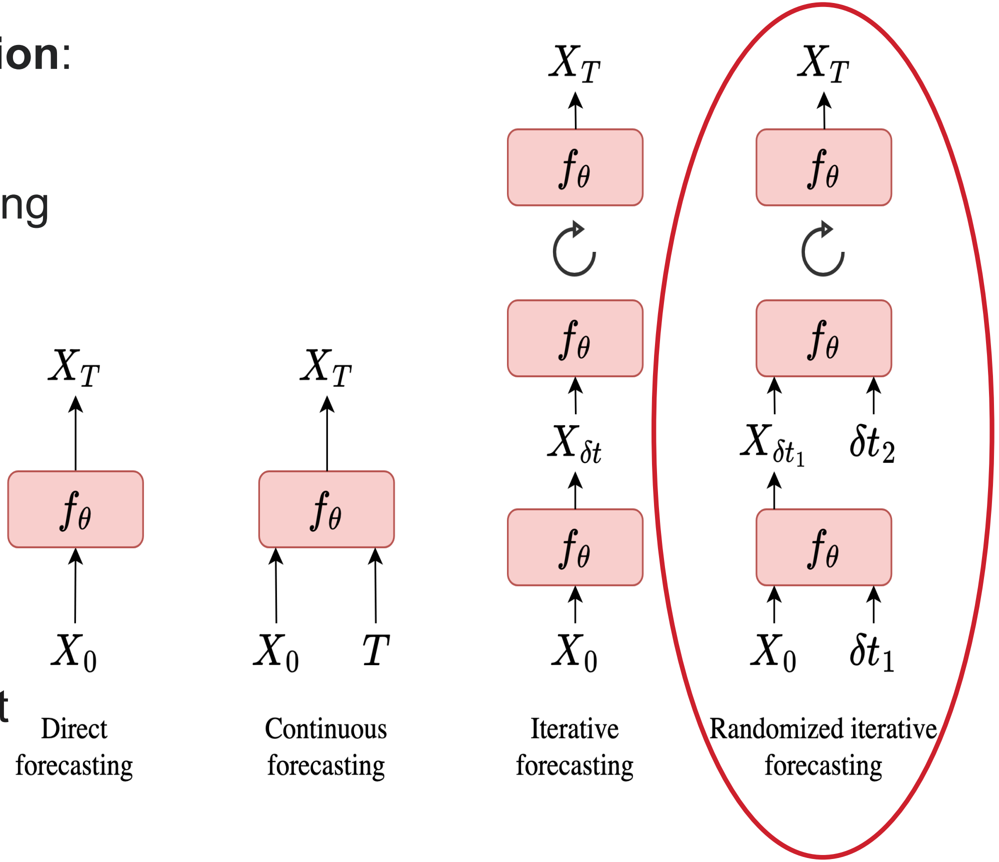


STORMER – TRAINING

- **Time-stepping and Ensemble Generation**:
  - Trained using randomized lead-time embedding [6,12,24] hrs
    - Free model-based ensembles during inference
  - Autoregressive time-stepping

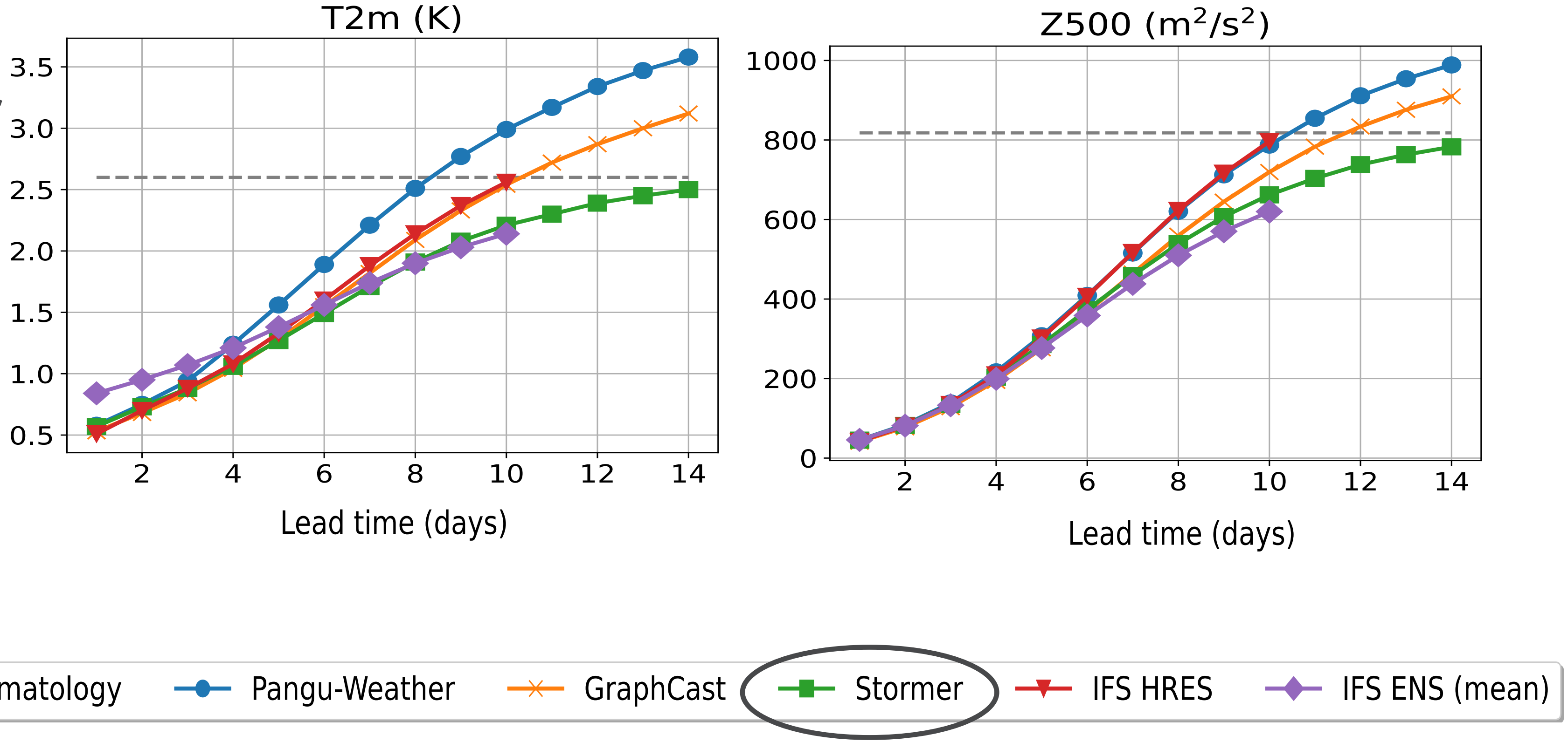
- **Loss**:
  - Pressure weighted loss
  - Scaling in the output layer
  - Training on the ”deltas”
  - 2 stage training:
    - Optimizing 1-step predictions
    - **Fine-tuning** using a multi-step lost function (up to 7 days)

- **Training**: Model size ~400M and trained using 126 A100 GPUs taking ~24 hours



STORMER - PERFORMANCE

RMSE (Lower is better)



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