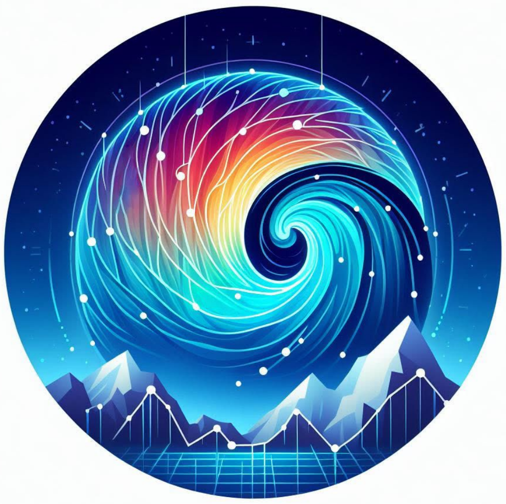


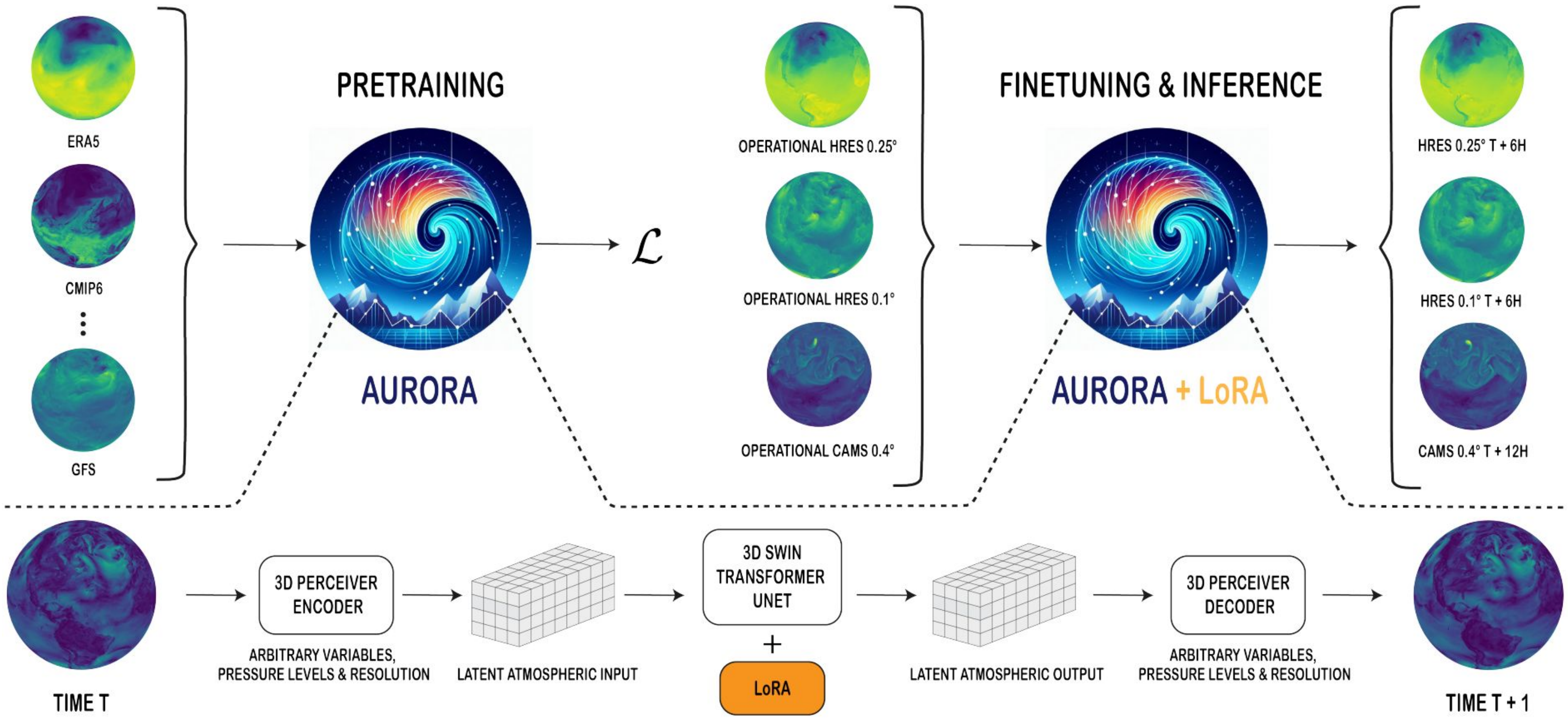
AURORA: A FOUNDATION MODEL OF THE ATMOSPHERE

Ryan Chesler
San Diego Machine Learning

Abstract

- Foundational models have proven to be vital to many domains
  - Text, Video, Drug discovery, Protein Structure prediction
- These models are pretrained on large datasets in a semi-supervised way
  - Ex. next token prediction in text
  - Requires vast amounts of data
  - Especially useful when there is not much labeled data
- Atmospheric forecasting is a great candidate for this kind of approach because of the huge quantity and variety of data
- This paper explores this concept in this domain

Model/Data Overview

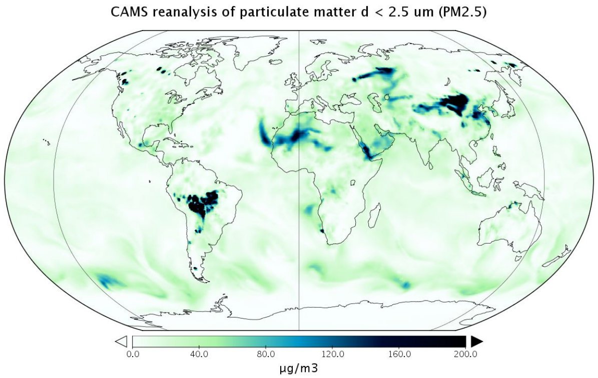


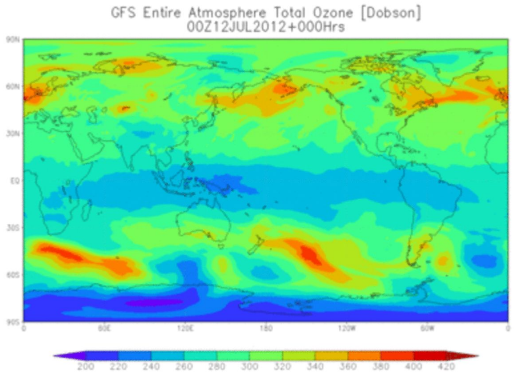
Data

- Mixed variables, spatial and temporal resolution
- ERA5
- CMIP6
- GFS
- HRES 0.1°
- HRES 0.25°
- CAMS 0.4°

ERA5 is available on:

- Single levels
- Pressure levels: 1000/975/950/925/900/875/850/825/800/775/750/700/650/600/550/500/450/400/350/300/250/225/200/175/150/125/100/70/50/30/20/10/7/5/3/2/1
- Potential temperature levels: 265/275/285/300/315/320/330/350/370/395/430/475/530/600/700/850
- Potential vorticity level: 2000
- Model levels: 1/to/137

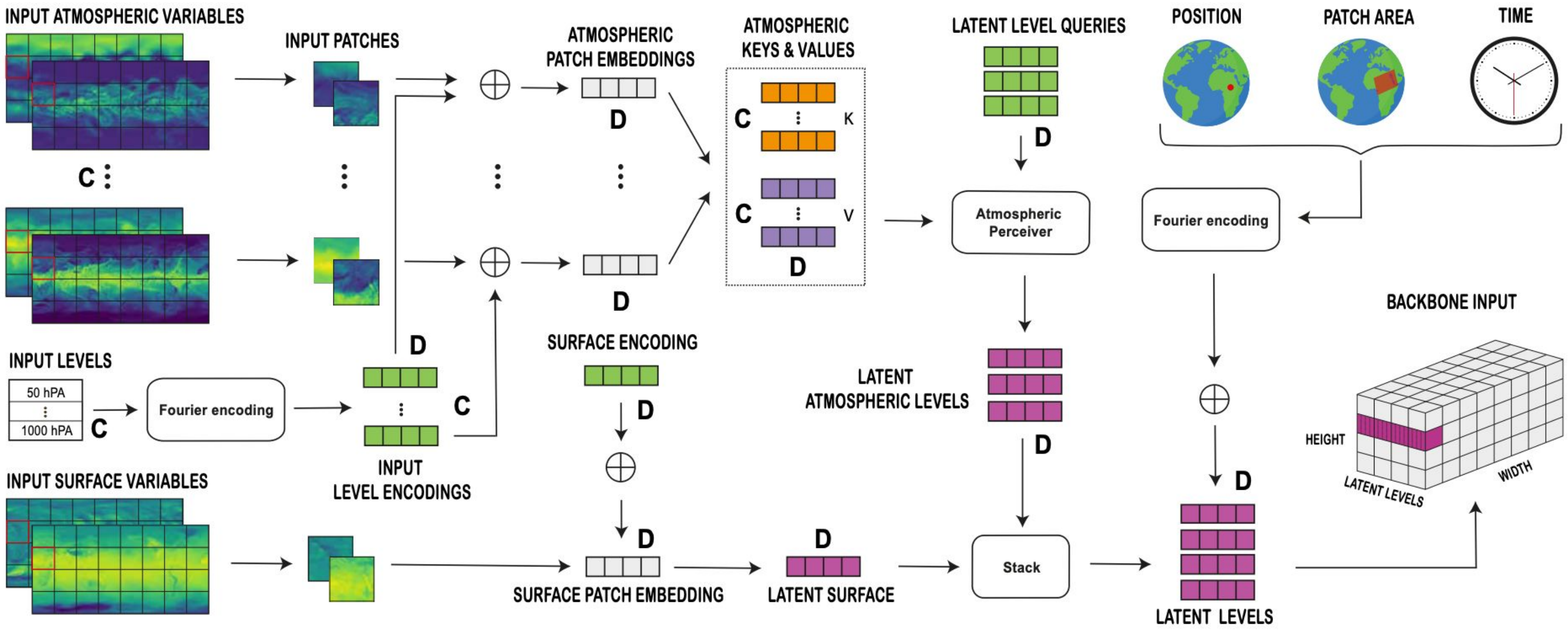




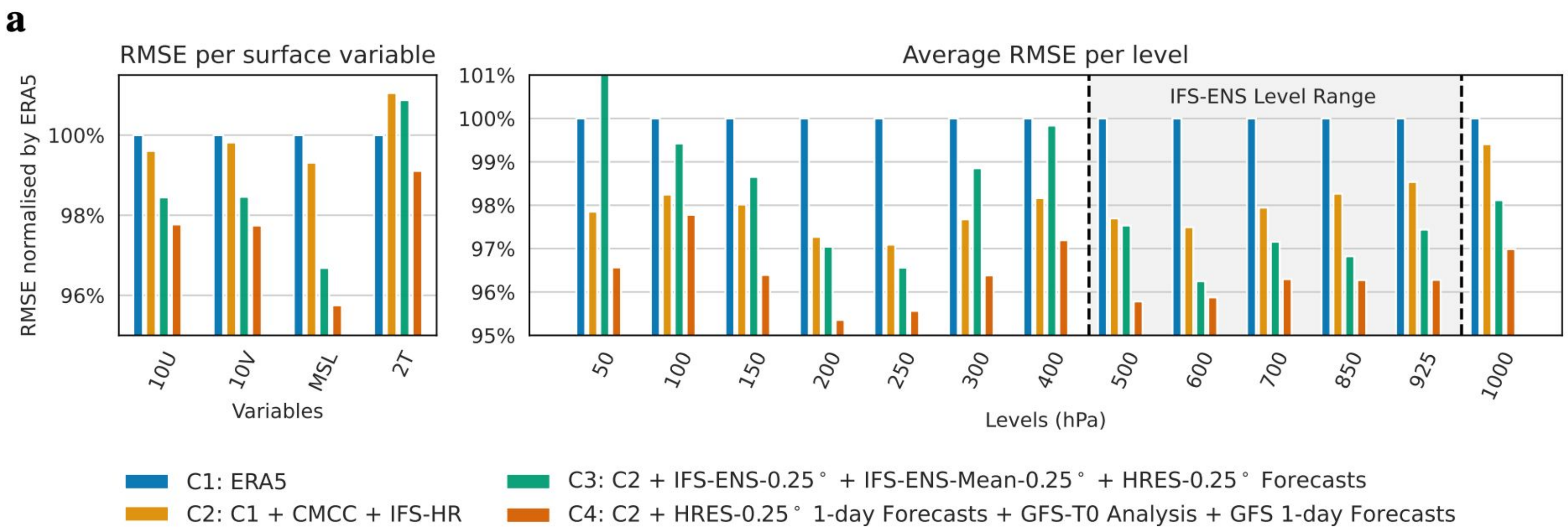
Model

- Architecture
  - 3D Perceiver Encoder - Processes mixed input variables and encodes to a common 3d representation
  - 3D Transformer Unet - Processes 3d interactions across time
  - 3D Perceiver Decoder - Decodes new 3d volume to specific predictions
- Input
  - **VA** x **C** x **T** x **H** x **W** (Variables, Pressures, Time, Height, Width)
  - Static Variables - Local orography, geopotential at surface, land-sea mask, soil-type mask
- Output T+1
- Model can be used autoregressively by passing predictions back to itself to extend forecasting further into the future
- Trained in 3 sizes, 113M, 660M, 1.3B parameters
- 32 A100 GPUs for two and a half weeks
- Compared to some other simulation based systems 5000x more efficient
- With Integrated Forecasting System 10-day forecast takes 65 minutes with 352 x 36 core systems
- With Aurora 1.1s per hour forecast on a single A100

Model Details

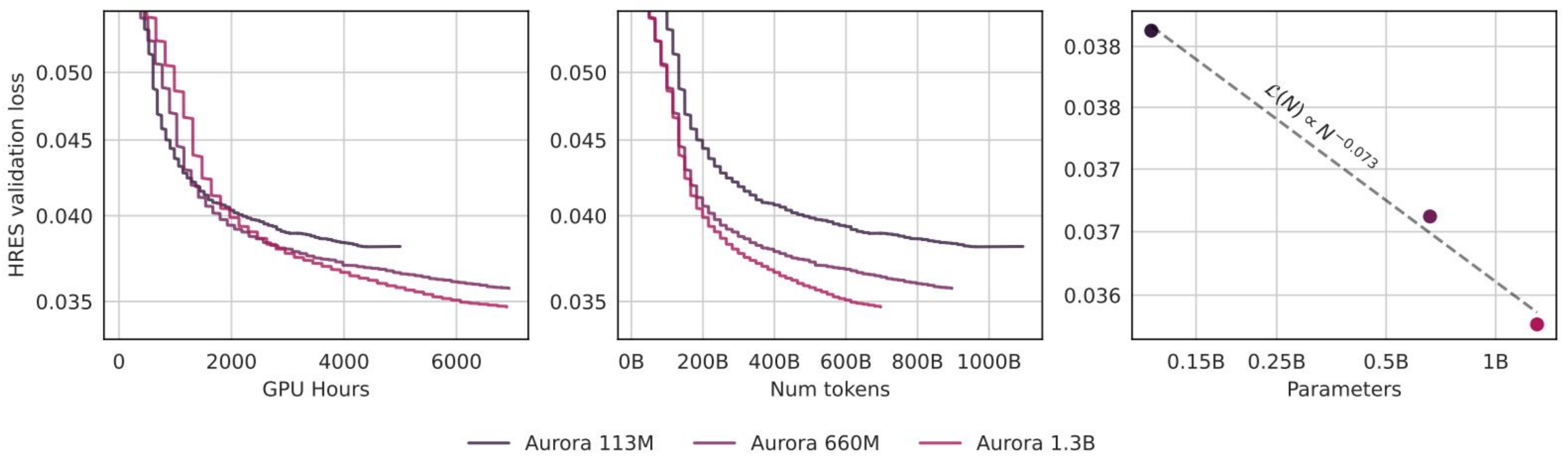


Model(Data Variety)



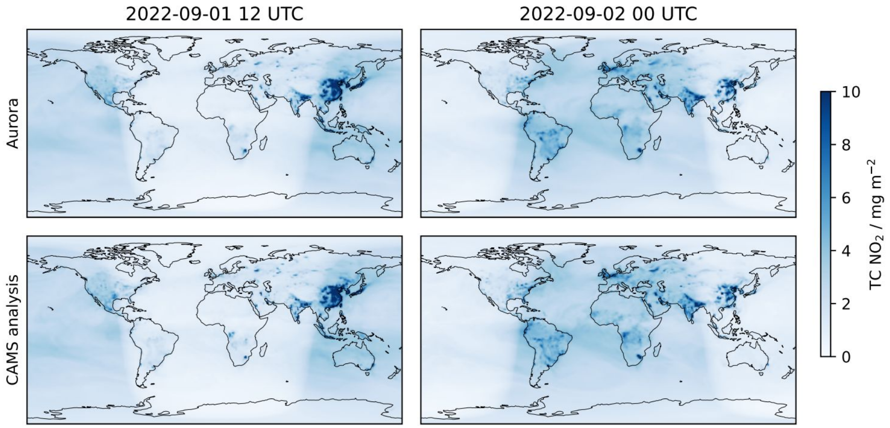
Model(Scaling)

c

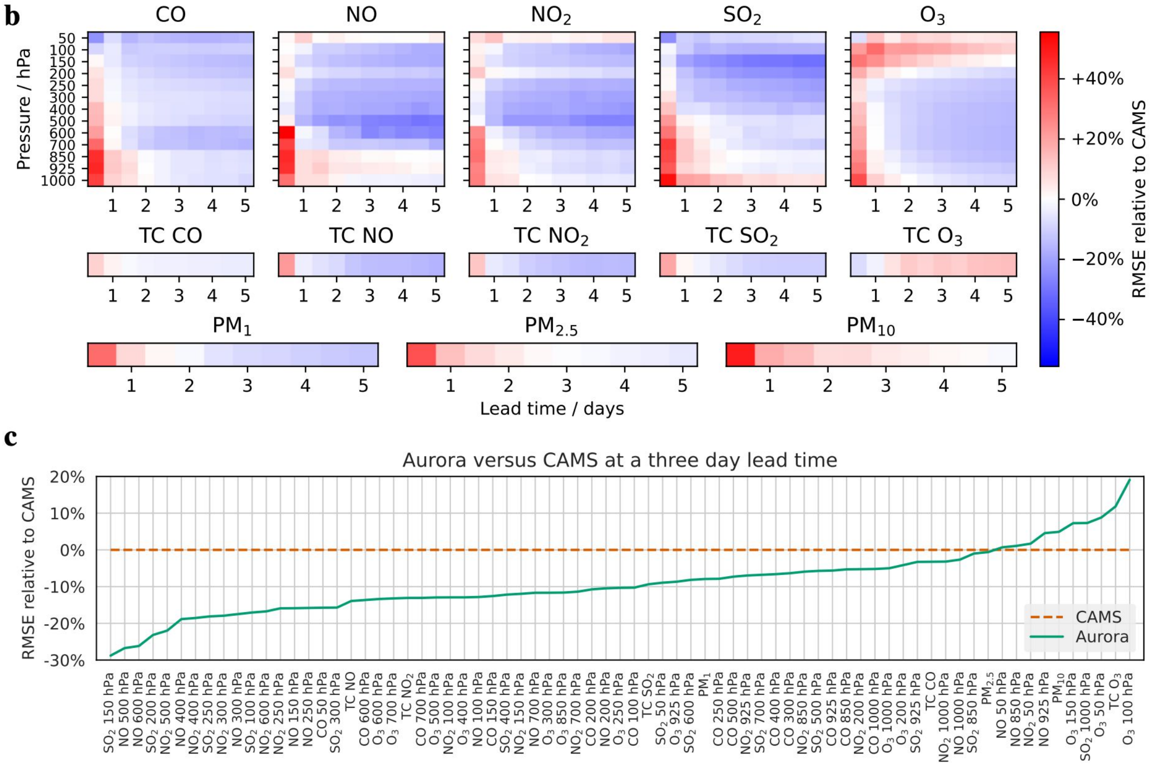


Applications(Atmospheric chemistry and air pollution)

- Complex and important problem
- Interaction between chemicals and movement of weather patterns
- Larger impacted by anthropogenic emissions
- Copernicus Atmosphere Monitoring Service (CAMS)
  - Forecasts
  - Analysis
  - Reanalysis
  - Extensions of IFS, even slower and 10x more computationally costly
- Aurora can be fine tuned on CAMS analysis data to create forecasts that are better on 74% of targets without additional emissions data at orders of magnitude less compute

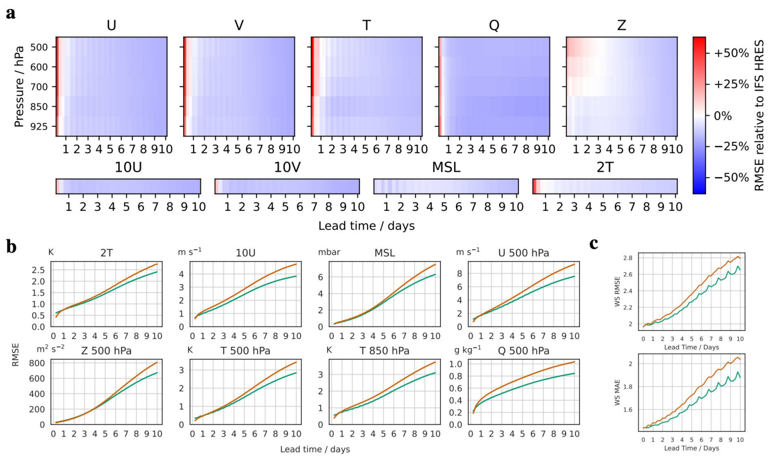


Applications(Atmospheric chemistry and air pollution)

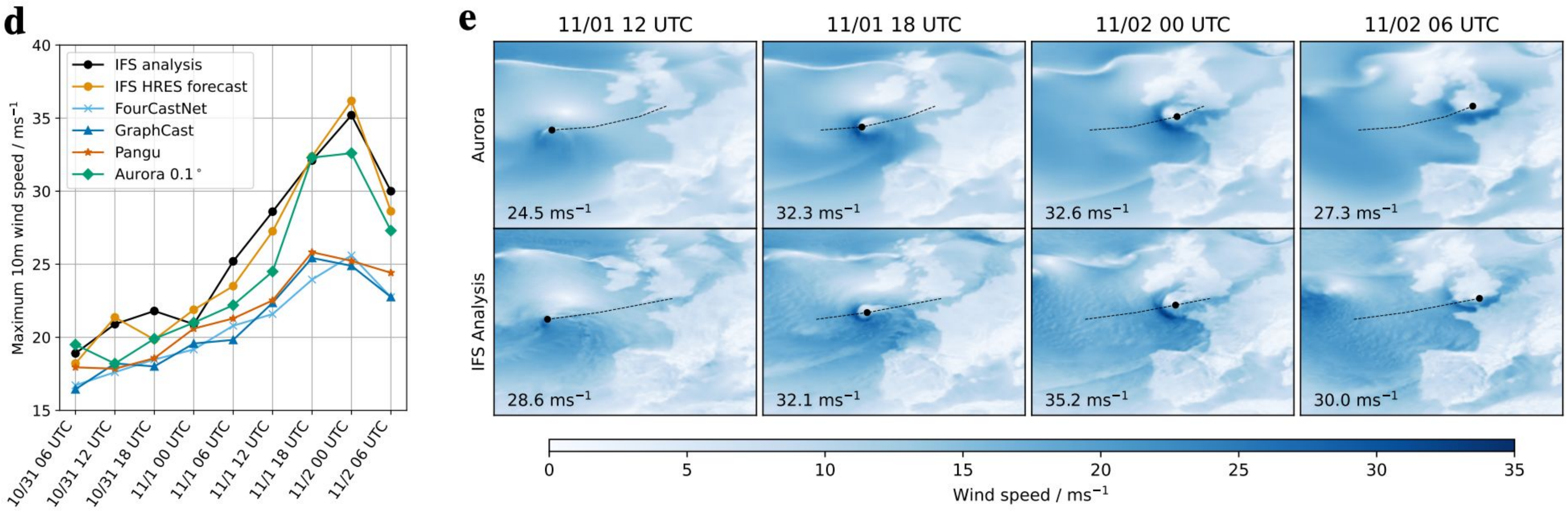


Applications(Weather forecasting)

- There is a huge wealth of historic weather recordings at 0.25° spatial resolution going back to 1950, but only 0.1° data from 2016 onward
- Aurora can shows great transfer with fine-tuning to this new higher resolution
- When pretrained on this larger pool of historic data the model can better handle extreme cases



Applications(Weather forecasting)



Rollout Fine-tuning

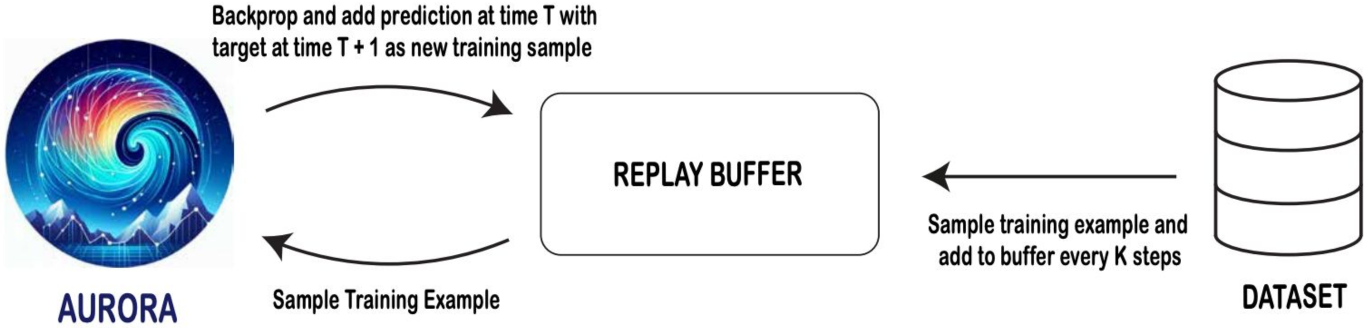


Figure 9: Diagram of the Rollout finetuning procedure. The replay buffer is initially populated with samples from the dataset. At each fine-tuning step, the model fetches a training sample from the replay buffer, performs a training step, and then it adds this new prediction (together with its next step target from the dataset) to the replay buffer. Every K steps, the replay buffer is refreshed with a new training sample from the dataset. Since the replay buffer is generally much smaller than the dataset size, this ensures that enough samples from the dataset are seen. Overall, this procedure allows the model to train fast on an evolving distribution of auto-regressive rollout steps coming from a mixture of model versions. This avoids the need to run expensive rollout procedures at each step.