CSE 676 Deep Learning

Paper & Code Presentation and Discussions

Part I: Paper Presentation or Code Demo

GraphCast for Medium Range Weather Prediction

Instructor: Alina Vereshchaka

Team Member:
Yavar Khan
Masters in Computer Science
University at Buffalo - NY
yavarkha@buffalo.edu

1. Introduction:

Medium-range weather forecasting plays a vital role in guiding decisions across numerous sectors, from agriculture and transportation to disaster management. Traditionally, Numerical Weather Prediction (NWP) has been the dominant method, relying on solving complex physical equations with supercomputers. While this approach has steadily improved forecast accuracy, it does not directly benefit from the vast historical weather data now available.

In contrast, machine learning-based weather prediction (MLWP) models offer a new paradigm—one where models learn directly from historical reanalysis data, enabling more efficient and scalable forecasting. Among recent advancements, GraphCast stands out as a breakthrough. Developed using graph neural networks, it predicts hundreds of global weather variables at high resolution for up to 10 days—within seconds.

GraphCast not only surpasses traditional systems like ECMWF's HRES on the majority of verification targets but also shows strong performance in forecasting severe events. This marks a significant step toward leveraging machine learning to model complex dynamical systems like Earth's atmosphere more accurately and efficiently.

2. NWP to MLWP: Bridging Physics and AI

For decades, **Numerical Weather Prediction (NWP)** has been the backbone of global weather forecasting. It relies on solving complex mathematical equations rooted in physics, using satellite observations and massive computational power. A prominent example is **ECMWF's High-Resolution Forecast (HRES)**, which produces detailed 10-day forecasts at 0.1° resolution. While highly accurate, these models are resource-heavy and rigid—unable to directly learn from the vast historical datasets we now have.

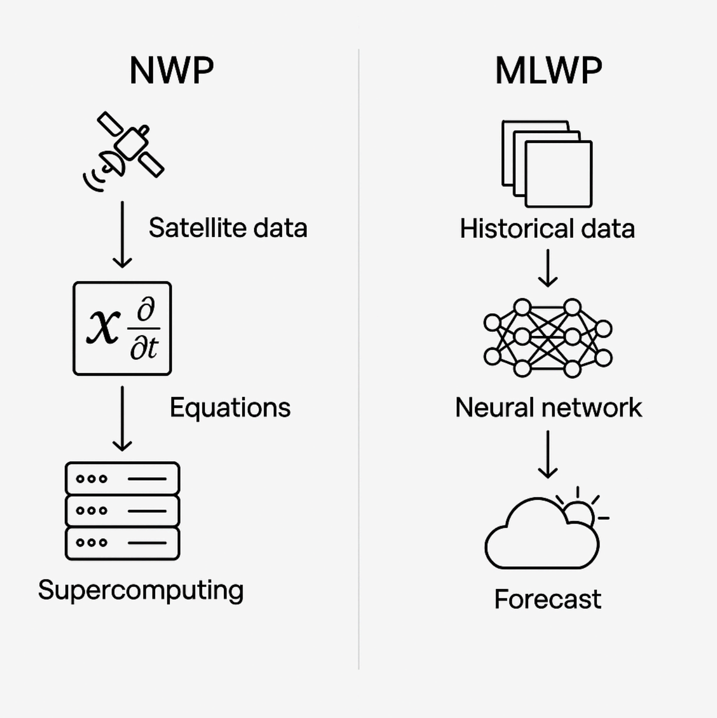
With the rise of big data and deep learning, the field began transitioning toward **Machine Learning-based Weather Prediction (MLWP)**. These models can be trained directly on decades of reanalysis data, uncovering patterns and interactions that are difficult to model with equations alone. They are faster, more adaptable, and often match or exceed traditional methods in specific tasks.

However, early ML models came with flaws:

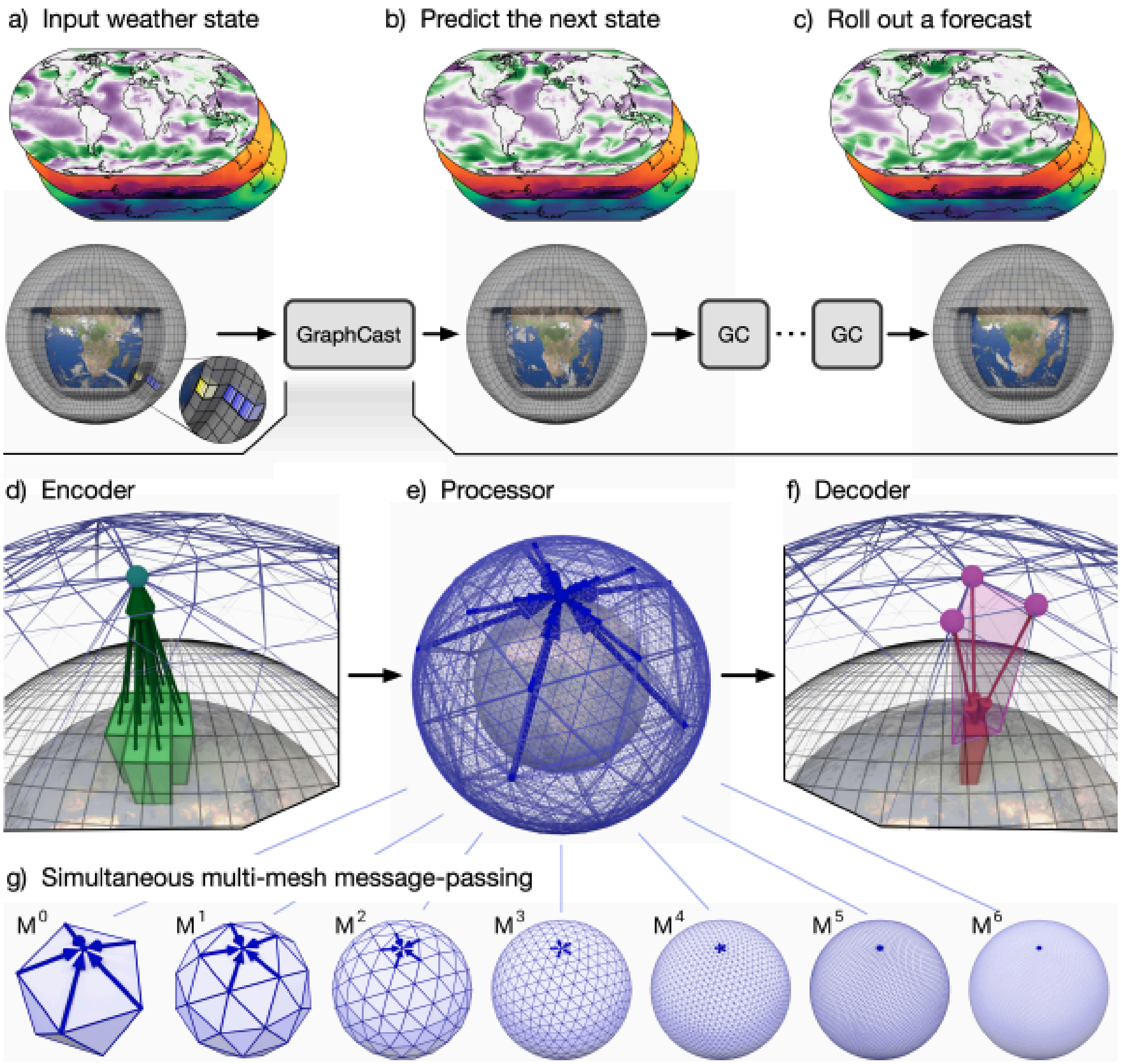
- They sometimes violated fundamental physics laws (e.g., generating unrealistic storms).
- Their predictions often degraded over time as small errors accumulated.
- They lacked the physical grounding of NWP, making them less reliable in certain extreme scenarios.

This created a new challenge: **How can we combine ML's speed and adaptability with the reliability of physics-based systems?**

The answer is emerging in models like **GraphCast**—a hybrid approach that leverages **Graph Neural Networks** to represent Earth's weather as a dynamic system of interlinked nodes. It predicts hundreds of atmospheric variables at a 0.25° resolution for up to 10 days, in under a minute. GraphCast marks a major milestone in blending physics-aware structure with the power of deep learning.



3. GraphCast: A New Era of ML-Based Forecasting



GraphCast is a cutting-edge machine learning model designed to deliver **fast, accurate, and physically realistic weather forecasts** on a global scale. Developed by DeepMind, GraphCast leverages the power of **Graph Neural Networks (GNNs)** to model Earth's atmosphere as a dynamic graph, enabling precise medium-range forecasts up to 10 days into the future.

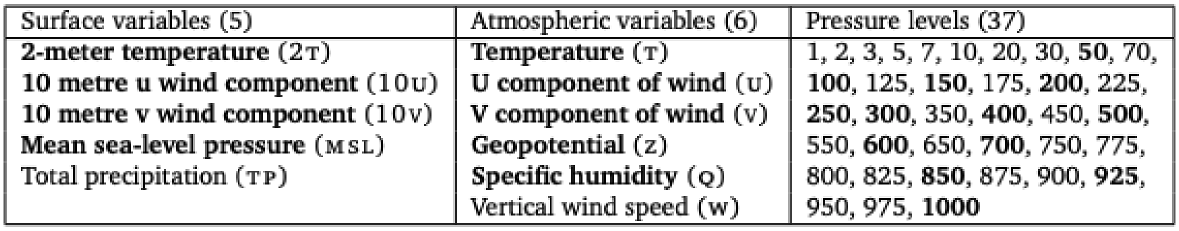
At its core, GraphCast works by:

- Taking an **initial weather state** as input (Fig. a)
- Predicting the **next atmospheric state** (Fig. b)
- And recursively rolling out the forecast over a desired time horizon (Fig. c).

This is achieved through a **graph-based encoder-processor-decoder architecture**:

- The **Encoder** (Fig. d) embeds localized atmospheric information into a graph structure.
- The **Processor** (Fig. e) performs **multi-mesh message passing** across various spatial resolutions (Fig. g), allowing the model to learn both local and global dependencies effectively.
- The **Decoder** (Fig. f) then reconstructs the weather state from the processed graph representation.

GraphCast is trained on **ERA5 reanalysis data**, using two input time steps (6 hours apart) to predict atmospheric conditions 6 hours into the future. By rolling this process forward, it generates full 10-day forecasts at **0.25° resolution**, with each forecast step taking less than one minute—**a major leap in speed and efficiency** compared to traditional methods like ECMWF's HRES.



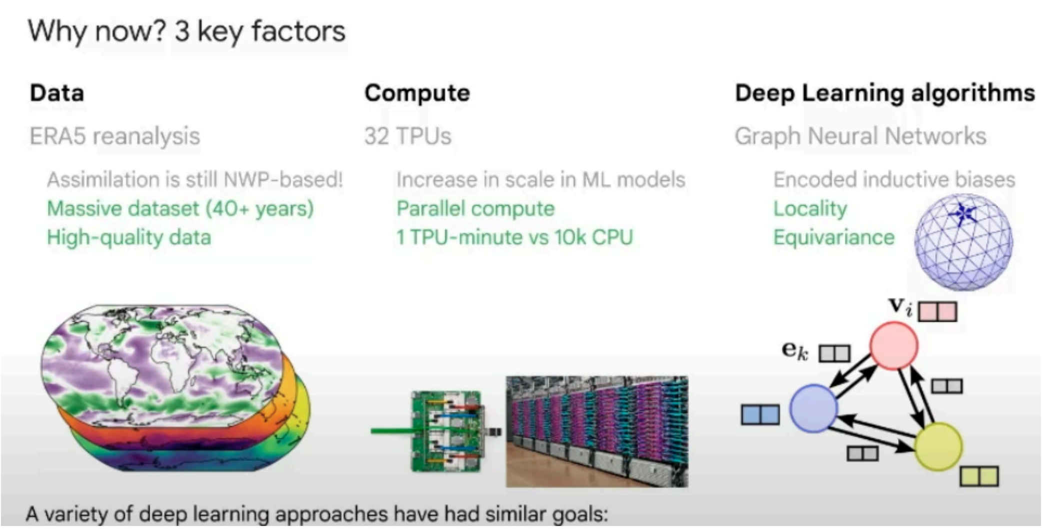
Surface variables (5)	Atmospheric variables (6)	Pressure levels (37)
2-meter temperature (2T)	**Temperature** (T)	1, 2, 3, 5, 7, 10, 20, 30, **50**, 70,
10 metre u wind component (10U)	**U component of wind** (U)	**100**, 125, **150**, 175, **200**, 225,
10 metre v wind component (10V)	**V component of wind** (V)	**250**, **300**, 350, **400**, 450, **500**,
Mean sea-level pressure (MSL)	**Geopotential** (Z)	550, **600**, 650, **700**, 750, 775,
Total precipitation (TP)	**Specific humidity** (Q)	800, 825, **850**, 875, 900, **925**,
	Vertical wind speed (W)	950, 975, **1000**

Table 1 | **Weather variables and levels modeled by GraphCast.** The numbers in parentheses in the column headings are the number of entries in the column. Boldfaced variables and levels indicates those which were included in the scorecard evaluation.

In terms of variables, GraphCast models a rich set of atmospheric and surface parameters. These include:

- **Surface variables** such as temperature, wind, sea-level pressure, and precipitation.
- **Atmospheric variables** like geopotential, humidity, and vertical wind.
- Coverage across **37 pressure levels**, enabling accurate vertical profiling of weather systems (see Table 1).

Three factors enable GraphCast's breakthrough now:



Why now? 3 key factors

Data

ERA5 reanalysis

Assimilation is still NWP-based!
Massive dataset (40+ years)
High-quality data

Compute

32 TPUs

Increase in scale in ML models
Parallel compute
1 TPU-minute vs 10k CPU

Deep Learning algorithms

Graph Neural Networks

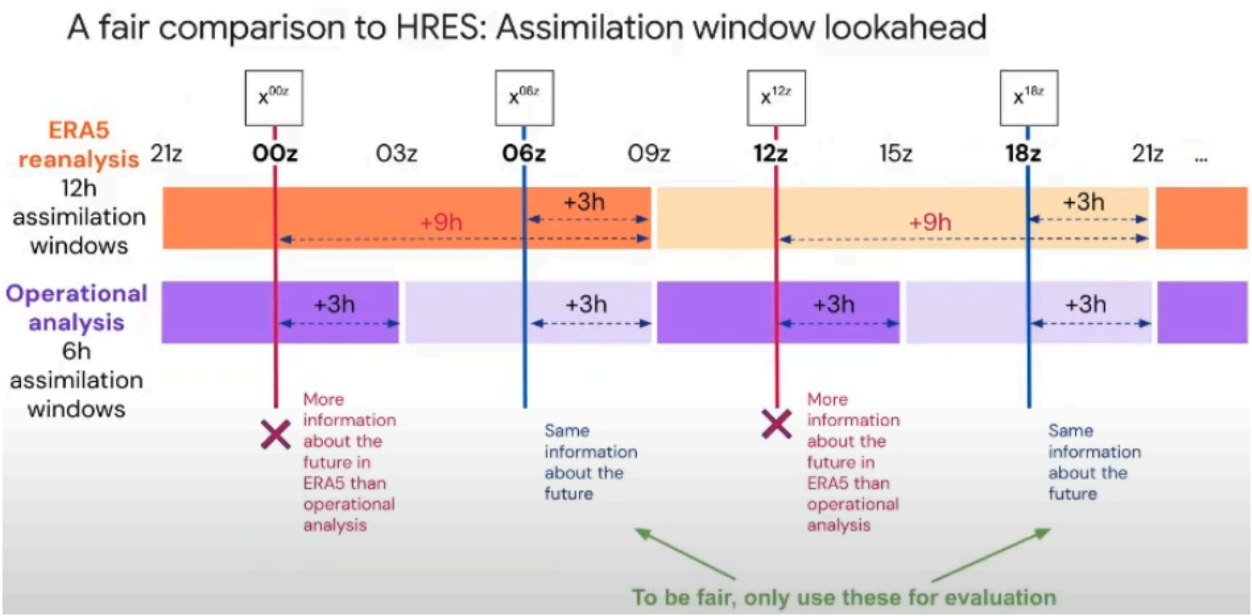
Encoded inductive biases
Locality
Equivariance

A variety of deep learning approaches have had similar goals:

1. Data: High-quality, 40+ years of ERA5 reanalysis data (NWP-processed).
2. Compute: Powerful TPUs train models 10,000x faster than CPUs.
3. Algorithms: Graph Neural Networks (GNNs) model weather's spatial interactions naturally.

Code for GraphCast class implementation:
https://github.com/google-deepmind/graphcast/blob/main/graphcast/graphcast.py

4. Verification Method



A fair comparison to HRES: Assimilation window lookahead

More information about the future in ERA5 than operational analysis

Same information about the future

More information about the future in ERA5 than operational analysis

Same information about the future

To be fair, only use these for evaluation

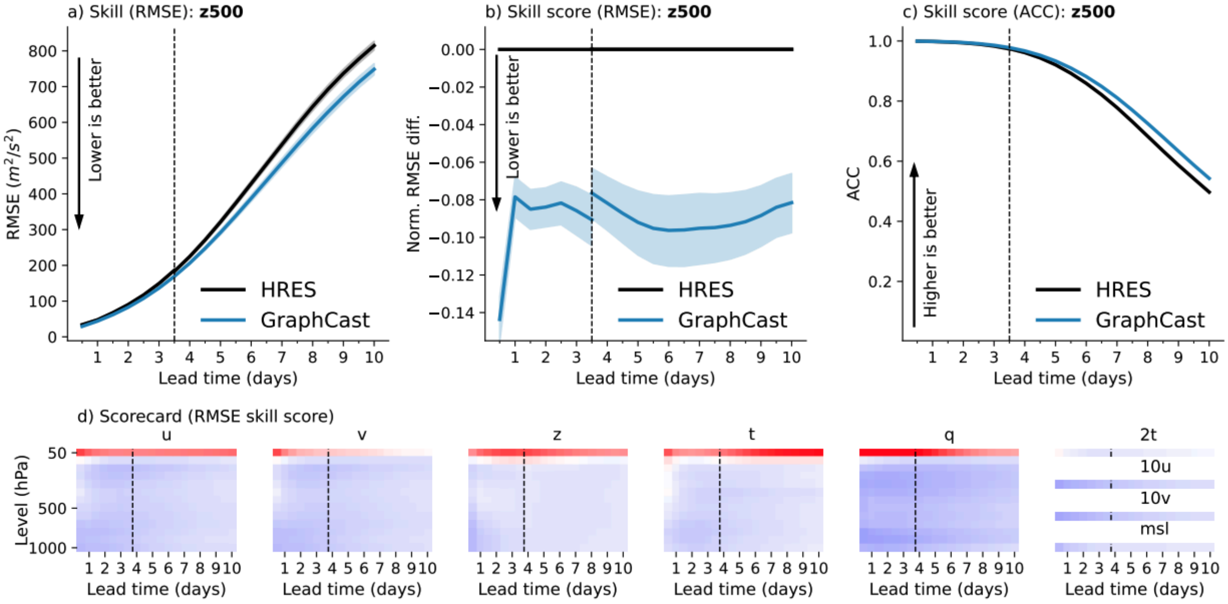
To evaluate GraphCast's forecasting skill, the authors compared it with ECMWF's HRES model using two metrics: **RMSE** and **Anomaly Correlation Coefficient (ACC)**. They tested 69 key variable-level combinations, excluding total precipitation due to known issues in the data.

GraphCast was evaluated using **ERA5** as ground truth (since it was trained on it), while HRES was evaluated using a special dataset (**HRES-fc0**) that ensures zero error at the start time for fairness.

Because ERA5 and HRES use different **data assimilation windows** (time ranges for collecting observations), only forecasts from **06z and 18z** were used. These times ensure both models have access to the same amount of recent data (+3h future observations).

Evaluations were done every **12 hours**, and since HRES at 06z and 18z runs only for 3.75 days, comparisons beyond that used 00z and 12z forecasts instead. A dashed line in the graphs shows where this transition happens.

5. Forecast Verification Results:

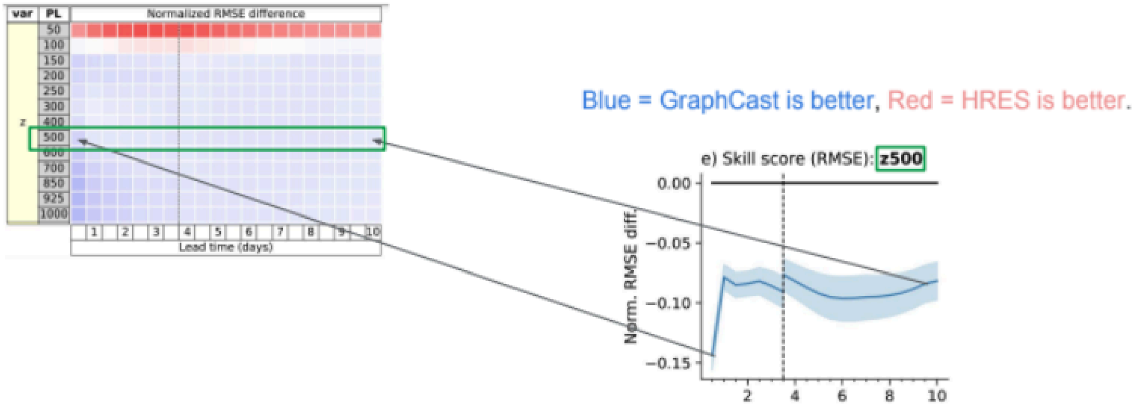


GraphCast was evaluated against ECMWF's HRES model on 10-day global forecasts at 0.25° resolution and 13 pressure levels. As shown in the above figure, GraphCast consistently outperformed HRES in terms of RMSE, RMSE skill score, and Anomaly Correlation Coefficient (ACC) on the key variable z500. The RMSE skill score shows an improvement of around 7–14% across lead times.

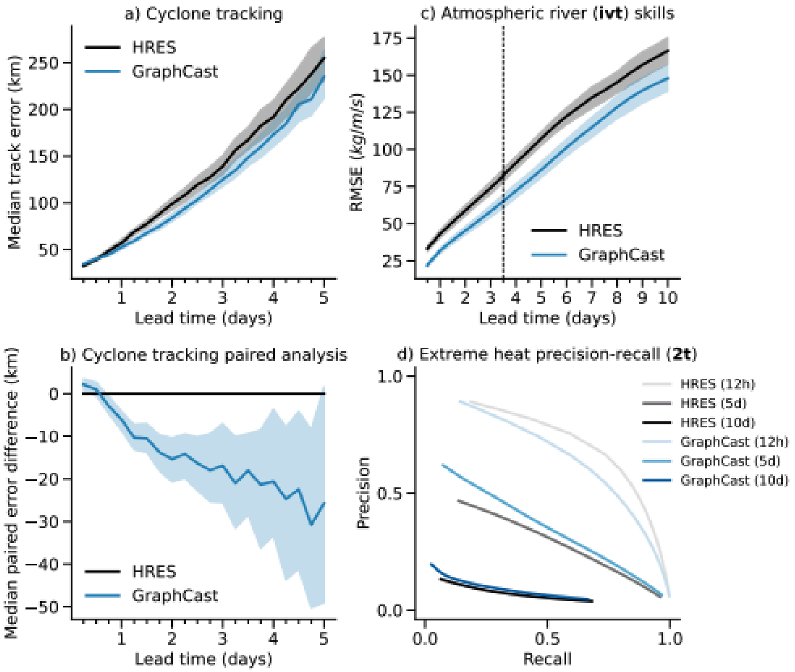
The evaluation covers 1380 variable-level combinations. GraphCast outperformed HRES on 90.3% of these targets and significantly ($p \le 0.05$) on 89.9%. Most of HRES's advantages were limited to the stratosphere (e.g., 50 hPa level), which had lower training weight in GraphCast. When excluding the upper levels, GraphCast's performance advantage rose to 96.9% and 99.7% respectively.

Even when both models were blurred to simulate forecast uncertainty, GraphCast maintained its lead on 88.0% of targets. It also outperformed the previous best ML model, Pangu-Weather, on 99.2% of shared targets.

ECMWF-style "scorecard" for comparing GraphCast to HRES



6. Severe Event Forecasting Results



In addition to general forecast accuracy, GraphCast was evaluated for its ability to predict high-impact weather events such as tropical cyclones, atmospheric rivers, and extreme temperature events.

- Tropical Cyclones: GraphCast showed lower median track error than HRES from 2018 to 2021 and significantly outperformed HRES for lead times between 18 hours and 4.75 days.
- Atmospheric Rivers: Over coastal North America during cold months, GraphCast achieved higher accuracy in predicting integrated water vapor transport (ivt) than HRES, with 25% improvement at short lead times and 10% at longer ones.
- Extreme Heat/Cold: When forecasting extreme temperature events (top 2% anomalies), GraphCast had better precision-recall curves than HRES for 5-day and 10-day lead times, while HRES performed slightly better at 12-hour forecasts.

These results highlight GraphCast's strong performance even in specialized severe event scenarios, despite not being explicitly trained for them.

CODE DEMO:
IPYNB file shared.

7. Conclusion:

Graphcast represents a breakthrough in weather forecasting by using graph neural networks to model global weather patterns as interconnected nodes. Unlike traditional methods, it efficiently captures complex atmospheric relationships, offering faster and scalable predictions.

In the code demo, we saw Graphcast implementation via the remote-graphcast library, generating forecasts through cloud computing (Runpod) and comparing them to ERA5 and Open-Meteo datasets. The code automated data retrieval, merging, and visualization, showing Graphcast's ability to match reanalysis trends, though minor errors highlighted room for improvement.

Looking ahead, Graphcast could transform real-time forecasting for agriculture, disaster response, and climate research. Future steps include refining accuracy, supporting more weather variables, and reducing reliance on external tools. By simplifying access and enhancing its design, Graphcast may soon empower communities worldwide with reliable, accessible weather insights.

References:

https://arxiv.org/pdf/2212.12794
https://cds.climate.copernicus.eu/datasets
https://github.com/google-deepmind/graphcast/blob/main/graphcast/graphcast.py
https://pypi.org/project/remote-graphcast/
https://open-meteo.com/
https://www.runpod.io/console/login?redirect=%2Fconsole%2Fuser%2Fbilling
https://www.youtube.com/watch?v=PD1v5PCJs_o&t=1212s