GraphCast: Learning skillful medium-range global weather forecasting

Seminar on Artificial Intelligence 2

Presenter: Matyáš Thér

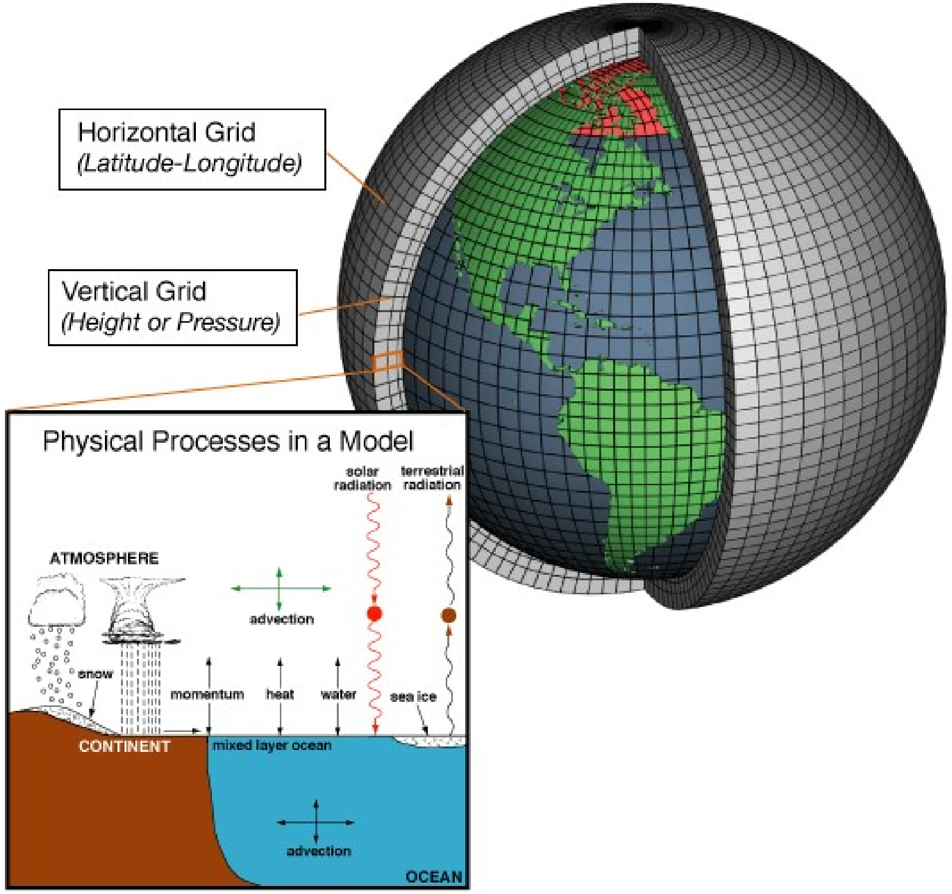
Paper authors: Lam, R. et al., Science (2023)

Presentation Outline

- **Context:**
  - Traditional NWP vs. Machine Learning (MLWP)
- **GraphCast:**
  - Architecture and GNNs
  - Multimesh
  - Training
  - Performance
  - Limitations, Retraining

Traditional Approach (NWP)

- **Numerical Weather Prediction (NWP):**
  - Physical equations solving on supercomputers
  - Partial differential equation systems
- **Accuracy:** more computing power → finer resolution
- **Limitations:**
  - Cannot automatically learn from historical data
  - Manual updates by human experts



NWP Standard Model - HRES

- **ECMWF HRES (High-Res Forecast):**
  - top deterministic operational model globally
  - 0.1° resolution (~9 km), 10-day forecasts, ~1 hour compute time.
  - Key variables:
    - Geopotential (Z)
    - Temperature (T)
    - wind vectors (U, V)
    - Humidity (R, Q) [1)]
- **ECMWF ENS:** multiple stochastic forecasts in ensemble to quantify forecast uncertainty
- **ML baseline for benchmarking**

1) WeatherBench notation

Machine Learning Weather Prediction (MLWP)

- **Data-Driven:** Trained directly from historical observation and analysis data.
- **Implicit Learning:** Captures complex patterns
- **Usage of GPUs**
  - Better speed-accuracy trade-off
  - Global medium-range forecasts in minutes

Influential MLWP Models

- **Architectural Shift**:
  - CNNs at coarse grids (>1.0°) → Transformer based architectures
- **FourCastNet (2022):** [1]
  - Adaptive Fourier Neural Operators (AFNO)
  - Modelling of fluid dynamics at high resolutions (0.25°)
- **Pangu-Weather (2022):** [2]
  - 3D Earth-Specific Transformers (0.25°)
  - ML baseline for GraphCast comparison
- **ClimaX (2023):** [3]
  - A weather “foundation model“ using ViT
  - Pre-trained on heterogeneous data to learn general Earth system dynamics
  - Fine-tuned for diverse downstream tasks

[1] Pathak, J. et al. (2022). "FourCastNet: A Global Data-driven High-resolution Forecasting Model."
[2] Bi, K., et al. (2023). "Accurate medium-range global weather forecasting with 3D neural networks."
[3] Nguyen, T., et al. (2023). "ClimaX: A Foundation Model for Weather and Climate."

GraphCast

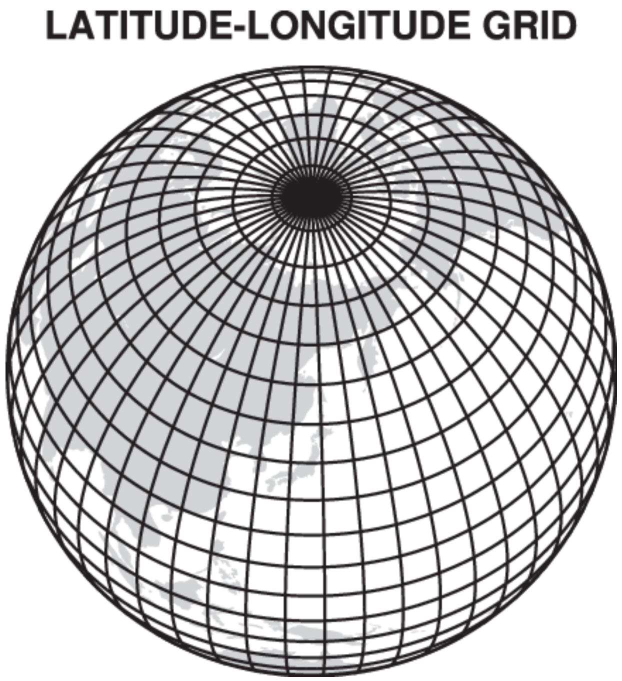
GNNs for Physical Systems

- **Graph Neural Networks:**
  - **Nodes:** Represent local weather states
  - **Edges:** Represent spatial interactions
  - **Message Passing:** Iterative update of node and edge states through separate node and edge FC neural networks
- **Why GNNs for Weather?**
  - Effective modelling of partial differential equation systems from the data
  - Capture complex interactions in fluid systems [1]
    - **→ good fit for modelling the Earth's atmosphere**

[1] Sanchez-Gonzalez, A., et al. (2020). "Learning to Simulate Complex Physics with Graph Networks."

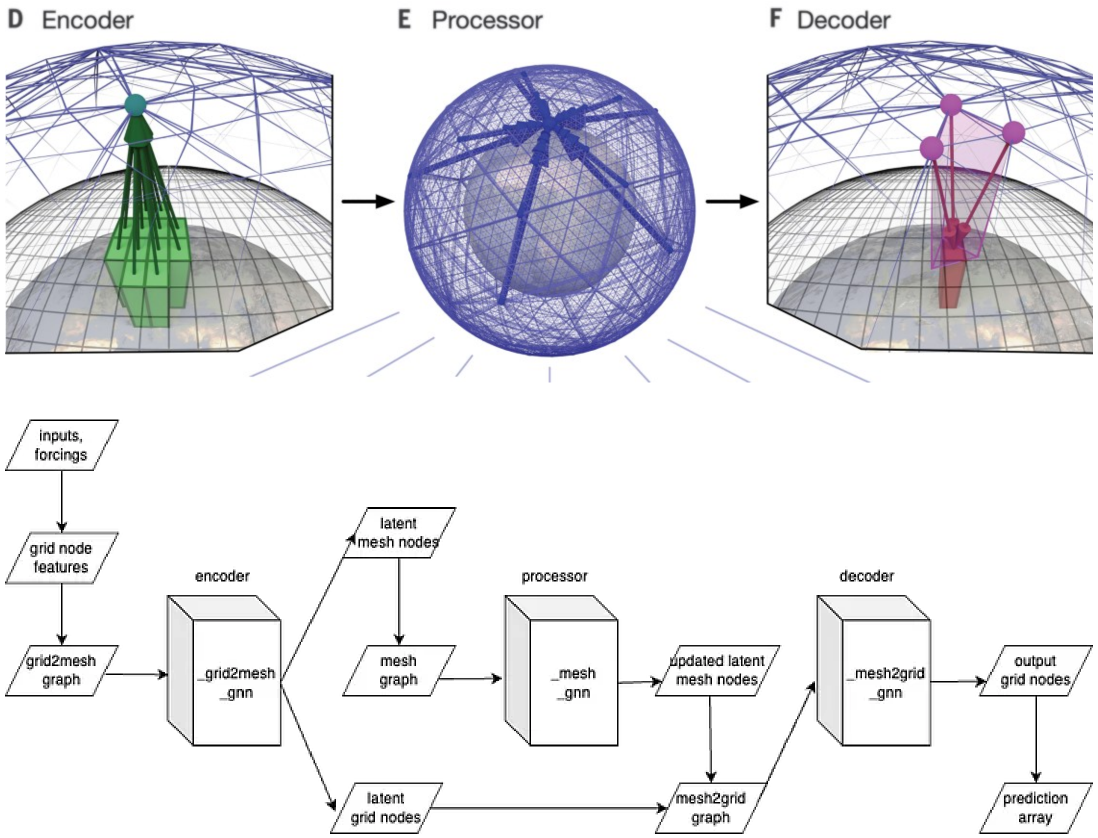
Earth Representation

- **Standard 2D Grids (CNN style)**:
  - Earth is spherical
  - Lat-lon grids are distorted
    - Real distance between points in the grid is inconsistent
- **Graph Solution**:
  - Flexible representation of spatial data
  - Removes uniform grid constraints
  - Enables spatially homogeneous resolution globally



Architecture Overview

- **Encoder-Processor-Decoder:** Mapping grid data to a latent graph and back.
  - **Encoder:** grid variables → internal multimesh representation
  - **Processor:** 16 unshared GNN layers for message-passing the multimesh graph
  - **Decoder:** processed multimesh → lat-long grid.
- **Residual Learning:**
  - The decoder predicts the change (residual) to the current state

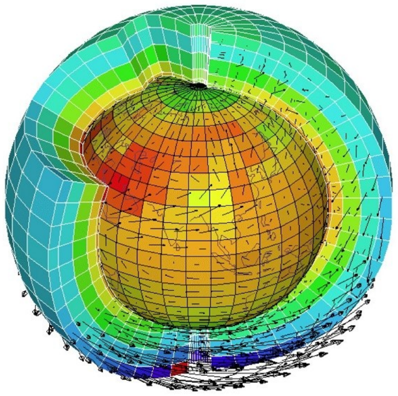


Model Input

- **Temporal Input**
  - 2 consecutive weather states: *t* and *t-6 hours*
- **State Representation**
  - 0.25° lat-lon grid (~28 km)
  - 721 × 1440 points
  - 227 features per pixel
    ~ 235.6 million features per state

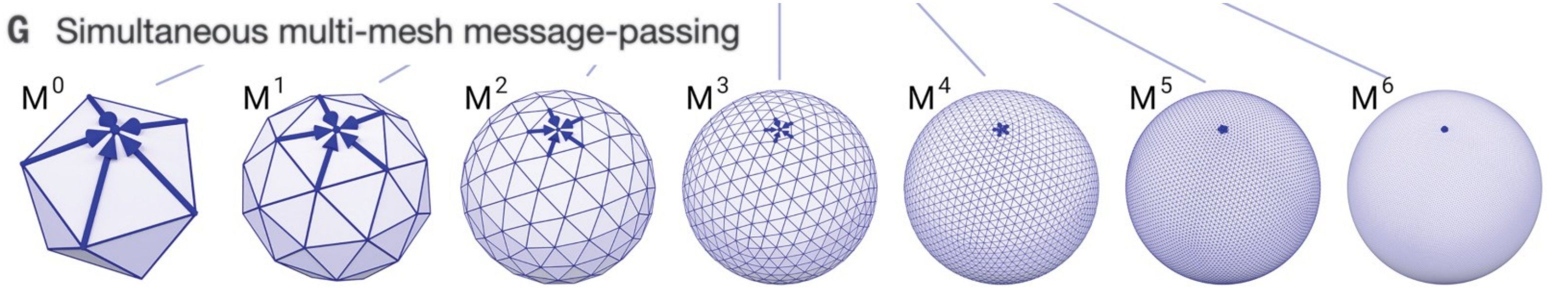
Table 1. Weather variables and levels modeled by GraphCast. The numbers in parentheses in the column headings are the number of entries in the column. Boldfaced variables and levels indicate those that were included in the scorecard evaluation. All atmospheric variables are represented at each of the pressure levels.

Surface variables (5)	Atmospheric variables (6)	Pressure levels (37)
2-m temperature (*2T*)	**Temperature** (*T*)	1, 2, 3, 5, 7, 10, 20, 30, **50**, 70,
10-m u wind component (*10U*)	**U component of wind** (*U*)	**100**, 125, **150**, 175, **200**, 225,
10-m 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, and **1000** hPa



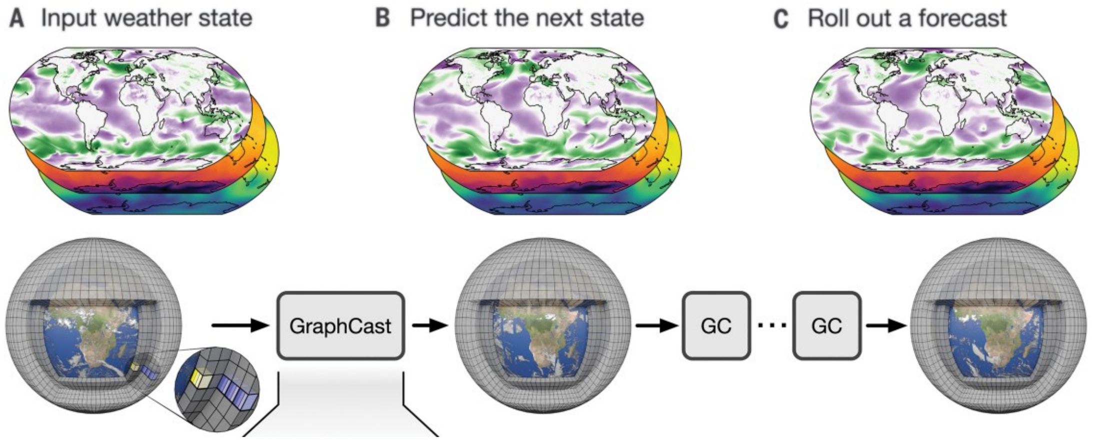
Multimesh & Message Passing

- **Multimesh Construction**:
  - Iterative refinement of a regular icosahedron (M^0 – M^6)
  - **M^0:** 12 nodes, 20 faces, 30 edges
  - **M^6** : 40,962 nodes (M^0 – M^5 nodes included as subsets)
- **Message Passing**:
  - **Processor:** 16 GNN layers for node updates
    - **Local**: M^6 edges
    - **Long-Range:** M^0 – M^5 edges - shortcuts for global message spreading



Time Stepping – Autoregressive Rollout

- **Single Step Output**: weather state at *t+6 hours*
- **Autoregressive Rollout**:
  - Outputs used as new inputs for subsequent steps
  - 40 iterations → final 10-day global forecast
- **Error Propagation**: challenging to mitigate compounding errors



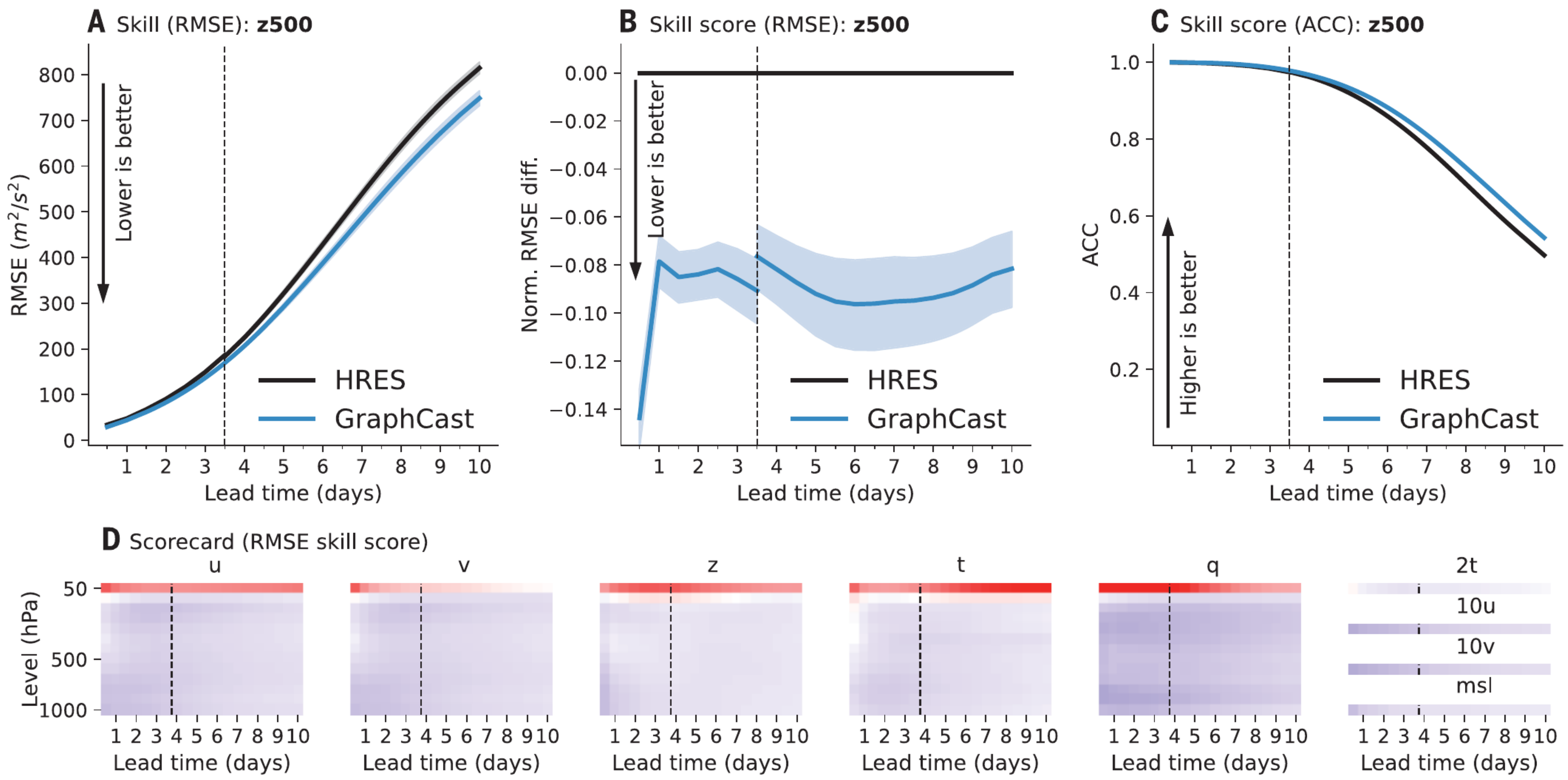
Training GraphCast

- **Training Dataset:** 39 years of ERA5 reanalysis data (1979–2017).
- **Training Objective:**
  - **Minimize MSE between** predicted and actual states
  - **Mitigating Compounding Errors:**
    - **Backpropagation Through Time (BPTT).**
    - **Curriculum Learning:** Incremental increase of the number of training steps (*N*) from 1 to 12 (6 hours to 3 days).

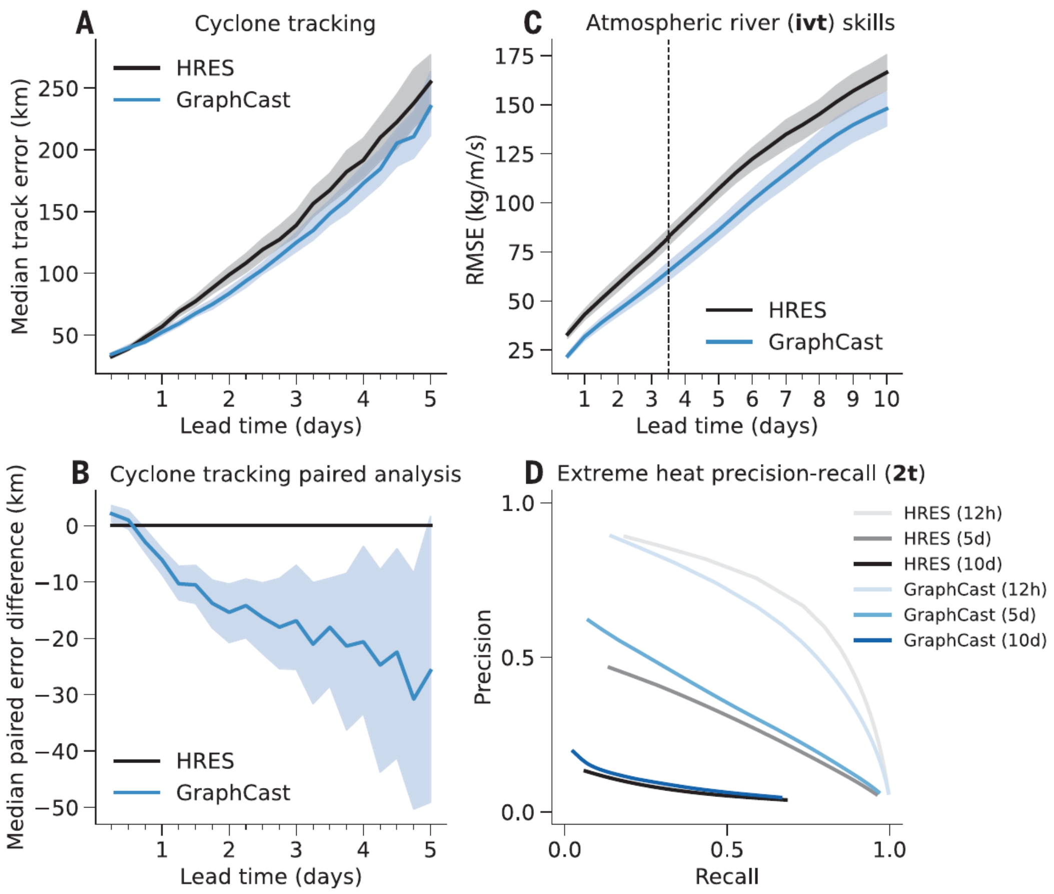
Evaluation Methodology

- **Evaluation Period:**
  - Held-out test data from 2018 onward
- **Forecast Skill Metrics:**
  - RMSE
  - Anomaly Correlation Coefficient (ACC)
- **Ground Truth:**
  - **ERA5** (GraphCast)
  - **HRES-fc0** (HRES)
- **Strict Assimilation Constraint:**
  - **GraphCast** initialized only at 06z/18z cycles to prevent unfair future-data leakage

Outperforming HRES: Global Skill

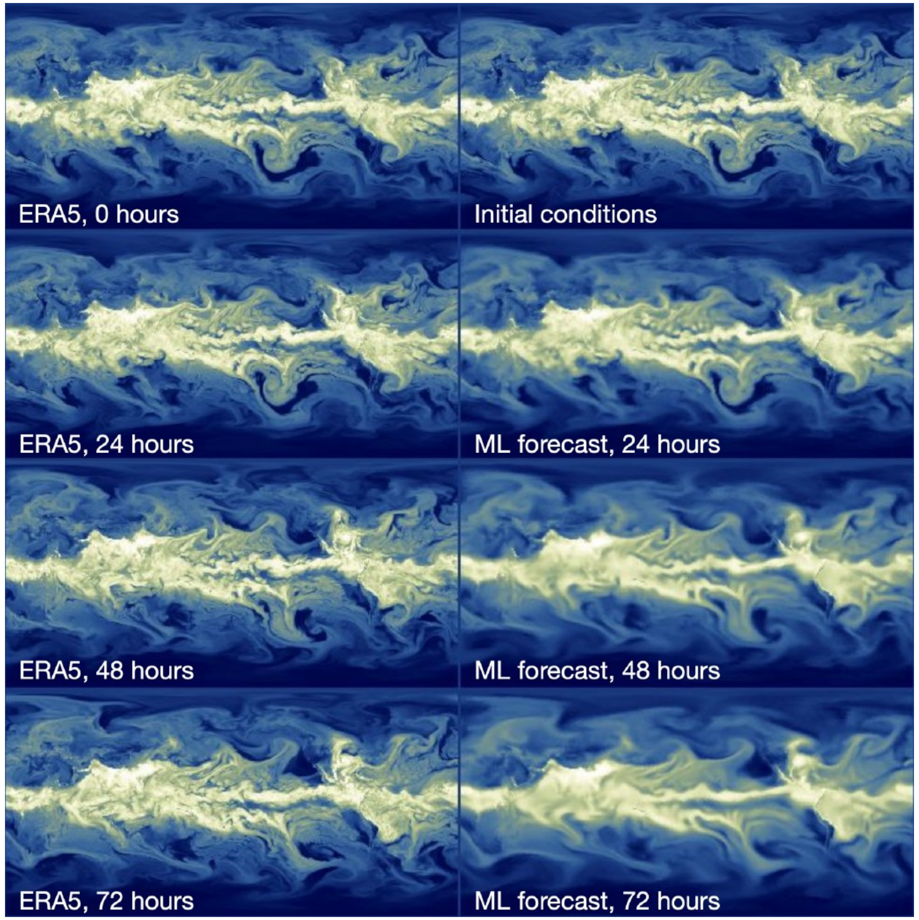


Outperforming HRES: Severe Events



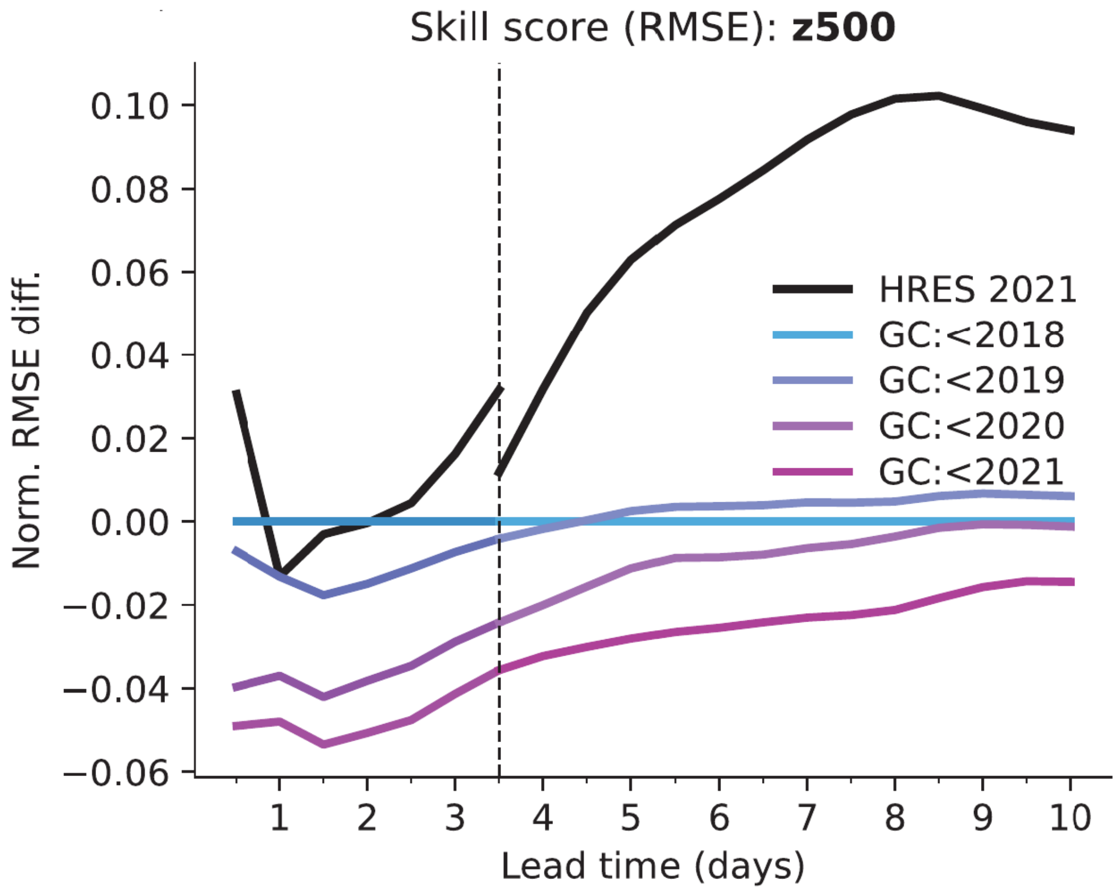
Limitations: Blurring & Uncertainty

- **Dealing with Uncertainity:**
  - **GraphCast:** a single trajectory
  - **ENS:** models predictive distributions using multiple stochastic models
- **The "Blurring" Effect:**
  - The **MSE objective** forces the model to average out uncertain future states
  - Spatially blurred features at longer lead times



GraphCast: Learning skillful medium-range global weather forecasting

Retraining



GraphCast: Learning skillful medium-range global weather forecasting

Conclusion

- **MLWP** can equal and surpass **NWP** in both **skill** and **compute efficiency**
- GraphCast is a powerful **complement** to NWP, **not a total replacement**
- Demonstrates the usage of GNNs for weather predictions

Q&A