Graph Machine Learning for Global Weather Prediction

Graph Residual Transformer + Gated Recurrent Unit (GRU)

Arnold Kazadi

SIParCS & Rice University

Mentors: David John Gagne, John Schreck, Charlie Becker, Gabrielle Gantos, Will Chapman

Machine Integration and Learning for Earth Systems (MILES) - NSF NCAR, Boulder, CO

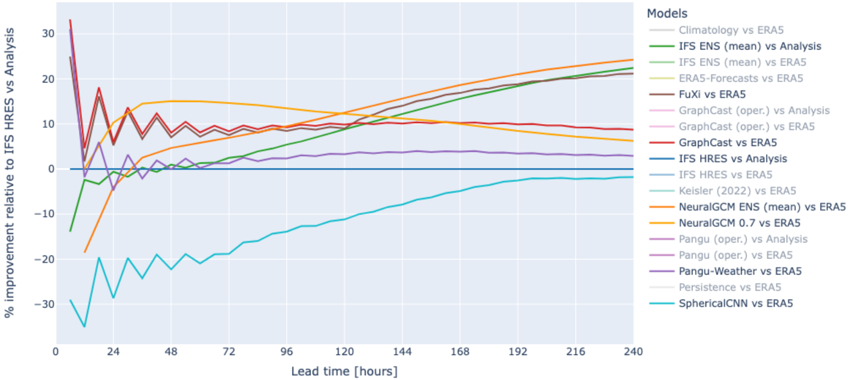
July 31, 2024

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Machine Learning - Potential for Global Weather Forecasting

Machine Learning for global weather prediction

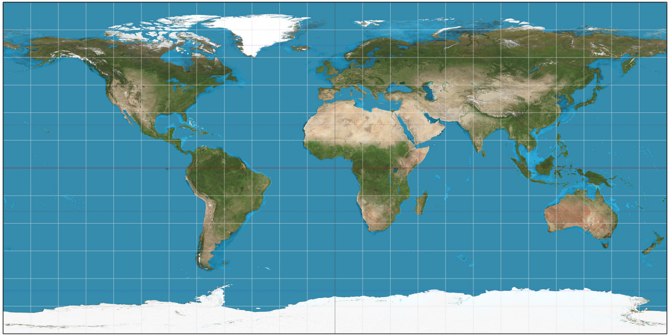
- Competitive with top physics-based models (e.g., IFS-HRES)
- Faster (45,000x)
- Memory efficient



RMSE [m^2/s^2] relative to IFS-HRES for Geopential at 500 hPa. (source: Weatherbench)

Existing Machine Learning Methods for Global Weather Prediction

- Most existing Machine Learning models for weather forecasting were initially developed for images and videos
  - Mostly suitable for rectilinear grid-structured data
- Grids are not suitable for representing spherical objects such as the globe
  - Regions at the poles are overrepresented
  - Requires padding to ensure continuity of the domain
  - 1° in longitude is 111 km at the Equator vs 56 km at 60° North/South.



Source: Wikipedia

Dataset - ERA5

ECMWF Reanalysis version 5 (ERA5)

- Reanalysis of the global climate from 1940 to present
- Hourly estimates of atmospheric variables (e.g., U, V, Q, Z, T)
- 137 pressure levels from surface up to a height of 80 km
- 1° resolution in latitude and longitude

Processed ERA5

Preprocessed data used

- 15 *hybrid sigma-pressure coordinate* (HSPC) levels: upper regions discretized by pressure and lower by sigma vertical coordinate.
- Prognostic variables

Variable	**Long name**	**Level**
U	Eastward wind	HSPC + 500 hPa
V	Northward wind	HSPC + 500 hPa
T	Temperature	HSPC + 500 hPa
Q	Specific humidity	HSPC + 500 hPa
T2m	Temperature	2m from surface
Z500	Geopotential height	500 hPa

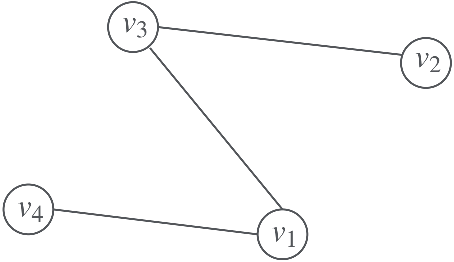
- Static and forcing variables: *Land-sea mask*, *Total Solar Irradiance*

Each lat-lon location is linked to $k(=49)$ nearest lat-lon locations based on the haversine distance.

Haversine distance.
Source: https://www.linkedin.com/pulse/haversine-formula-firebird-sql-calculate-distance-between-revelli/

A graph neural network has two main operations: *Message passing* and *Update operation*

Toy example



Messaging Passing

$$
\begin{aligned}
m_{3\to1} &= g(v_3, v_1) \\
m_{4\to1} &= g(v_4, v_1) \\
m_{3\to2} &= g(v_3, v_2)
\end{aligned}
$$

Update Operation

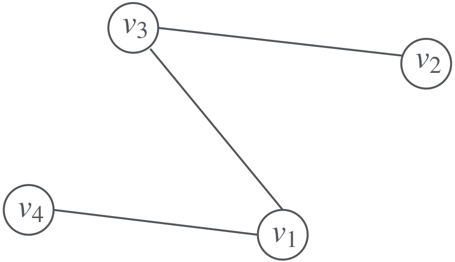
$$
\begin{aligned}
v_1 &= f(m_{3\to1}, m_{4\to1}) \\
v_2 &= f(m_{3\to2})
\end{aligned}
$$

g and f are user-defined functions.

Graph Transformer

A graph transformer defines an attention weight for message passing and summation for update operation.

Toy example



Messaging Passing

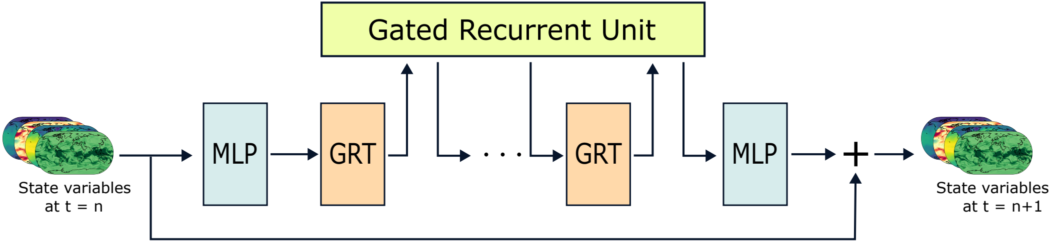
$$m_{3\to1} = \alpha_{3\to1} v_3$$
$$m_{4\to1} = \alpha_{4\to1} v_4$$
$$m_{3\to2} = \alpha_{3\to2} v_3$$

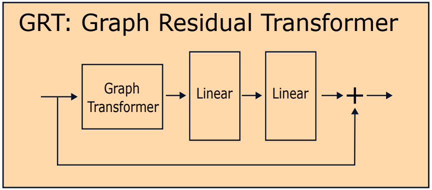
$$\alpha_{i\to j} \geq 0 \qquad \sum_i \alpha_{i\to j} = 1$$

Update Operation

$$v_1 = m_{3\to1} + m_{4\to1}$$
$$v_2 = m_{3\to2}$$

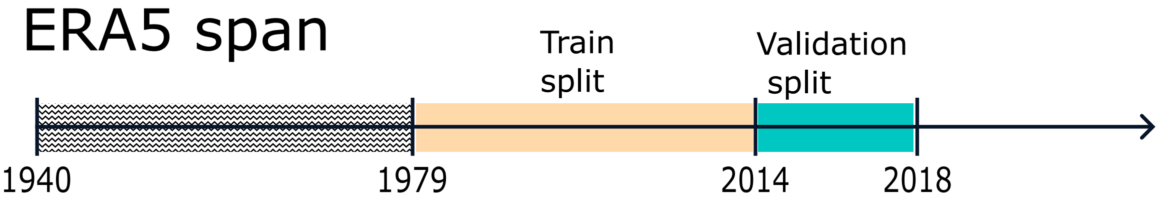
Proposed Method: Graph Residual Transformer + GRU





Top-Overall architecture. **Bottom-**Details of the GRT layer.

Experiments: Setup



Test data

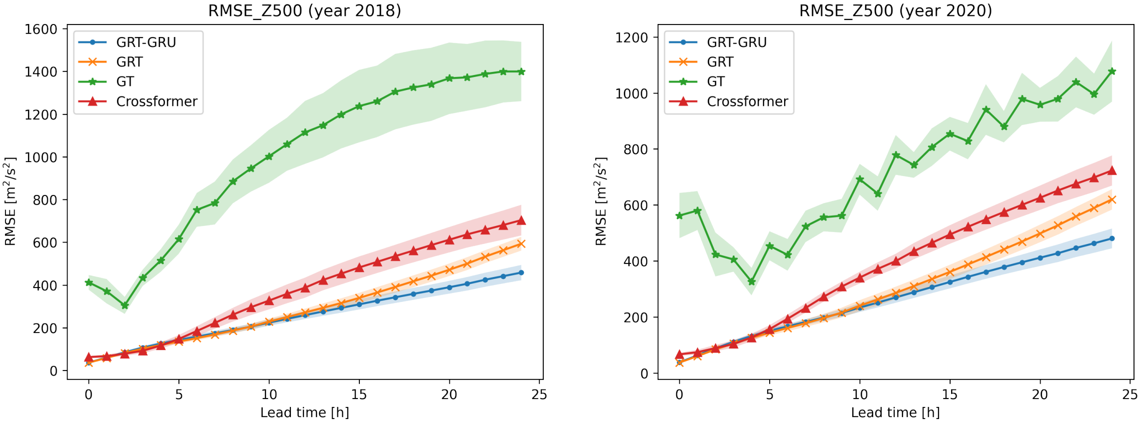
- 2018-**-01: First day of every month of 2018
- 2020-**-01: First day of every month of 2020
- 2022-**-01: First day of every month of 2022

Experiments: Baselines

Our proposed model, **GRT-GRU** ($\sim$ 3 M parameters), is compared with

- **Crossformer**: Vision transformer-based model ($\sim$ 292 M parameters)
- **GRT**: GRT-GRU without GRU module
- **GT**: GRT without the residual connection

Experiments: Results

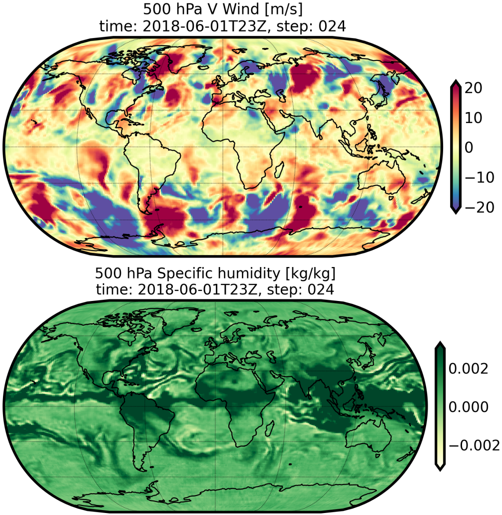


Mean (solid line) and standard deviation (shade) with first day of every month of 2018 (left) and 2020 (right) as initial condition.

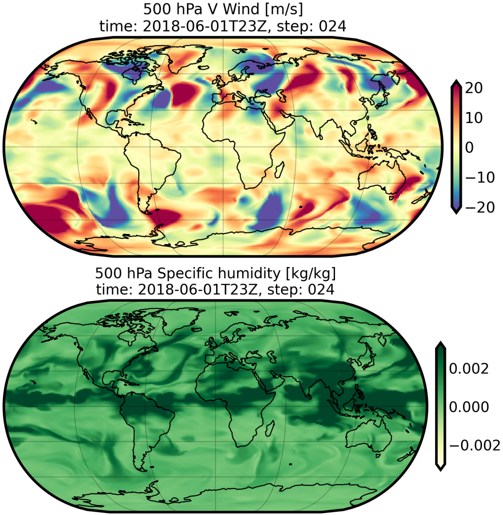
- Residual connection reduces error by factor of 2 (GT vs GRT)
- GRU further reduces error by factor of 1.2 (GRT vs GRT-GRU)

Experiments: Results

Results after rolling out to 24 hours/steps



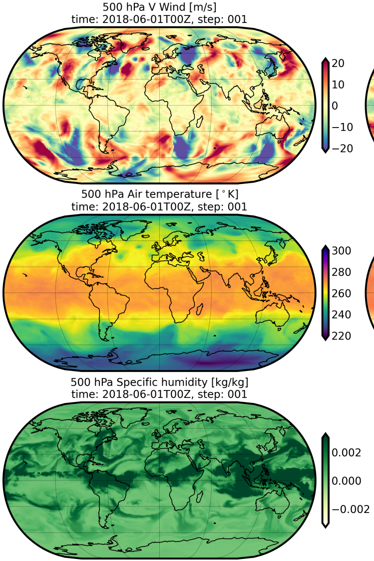
Crossformer



GRT-GRU

Experiments: Rolled out to 40 steps

Link to video



Conclusion

- Proposed Graph Residual Transformer + GRU for weather prediction
  - Relatively small model ($\sim$ 3 M parameters).
  - Trained only on the next 3 steps (auto-regression), can roll out up to 40.
  - Residual connection in state space reduces error by a factor of 2.
  - GRU reduces error even further by factor of 1.2
- Future work
  - Investigate diverse gridding of the globe
  - Explore larger model (trained with fully shared data parallel)

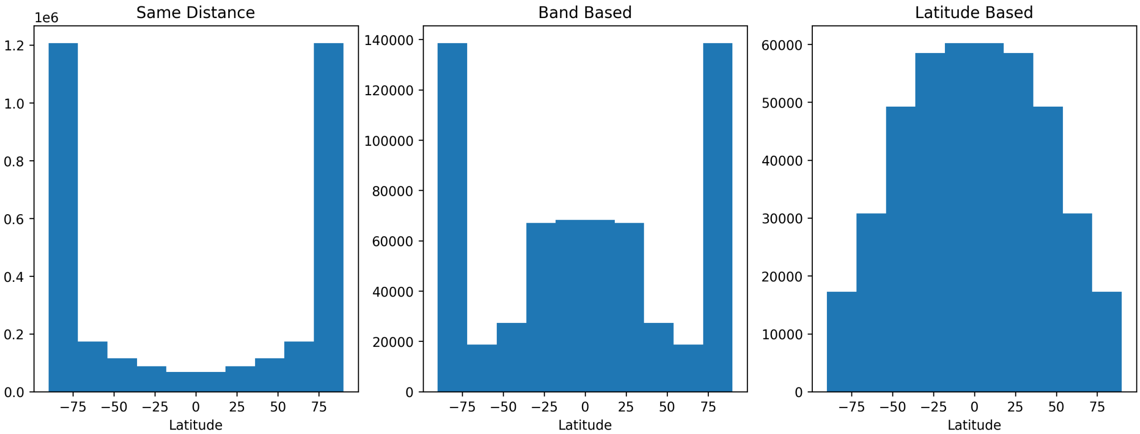
Thank You

Emulation video

akn7@rice.edu/@ucar.edu

Supplementary Material

Edge distribution

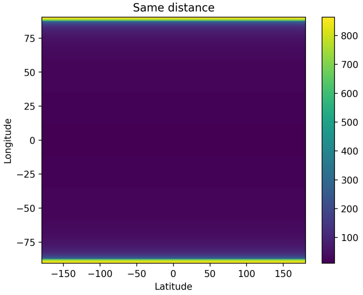


Same radius at all latitudes

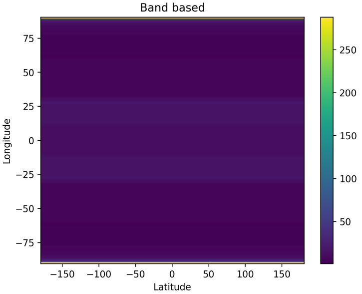
- 0°-30°N/S: keep radius
- 30°-60°N/S: 1/2 radius
- 60°-90° N/S: 1/4 radius

$$\frac{\text{distance at lat}}{\text{distance at equator}} \times \text{radius}$$

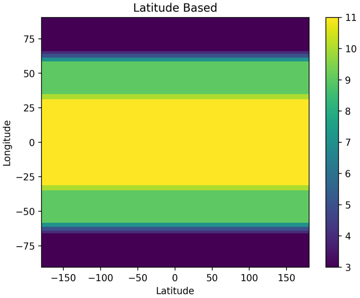
Edge distribution



Too many edges — high memory demand



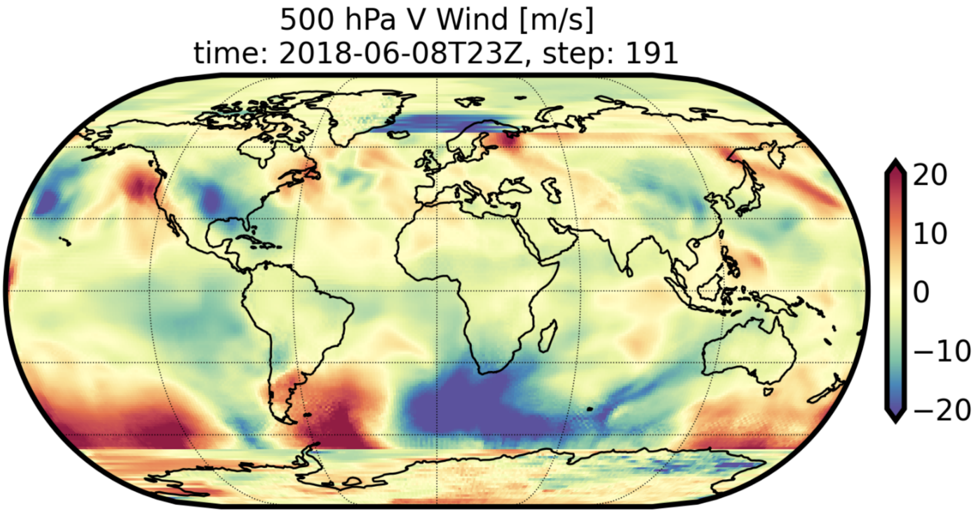
Still memory demanding



Caption

Edge distribution

Bandes at the poles, likely due to insufficient graph connections.

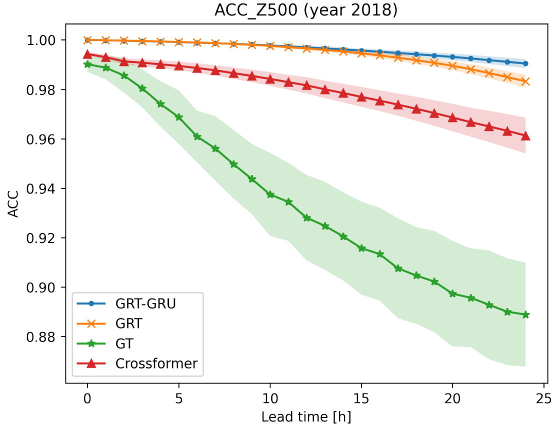


Correlation between the forecast anomaly and the verifying analysis anomaly with respect to the climatology.

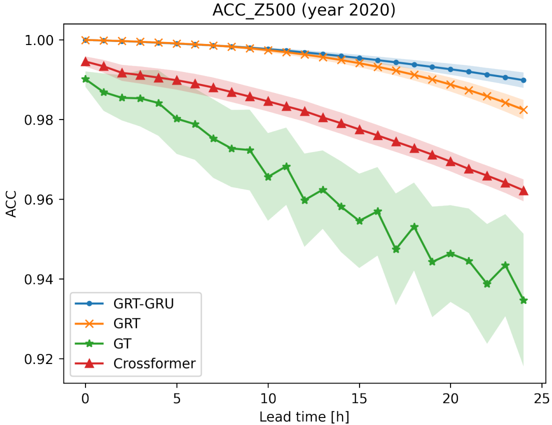
$$ACC = \frac{1}{N} \frac{\sum_{i,j} w_j (x_{i,j} - c_{i,j})(\hat{x}_{i,j} - c_{i,j})}{\sqrt{\sum_{i,j} w_j (x_{i,j} - c_{i,j})^2} \sqrt{\sum_{i,j} w_j (\hat{x}_{i,j} - c_{i,j})^2}}$$

- w_j: latitude-based weight
- $x_{i,j}$: true variables at lat j and lon i
- $\hat{x}_{i,j}$: predicted variables at lat j and lon i
- $c_{i,j}$: climatology at lat j and lon i

More evaluation metrics: ACC and RMSE



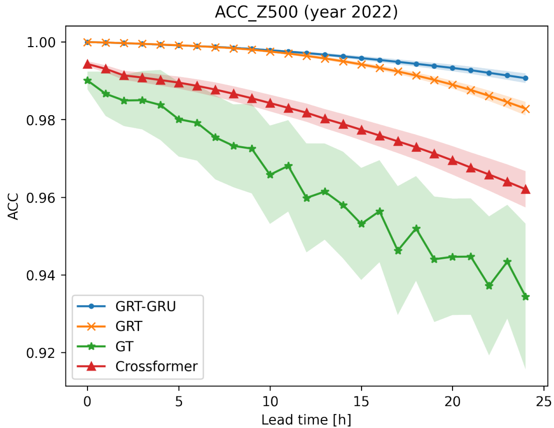
Higher is better



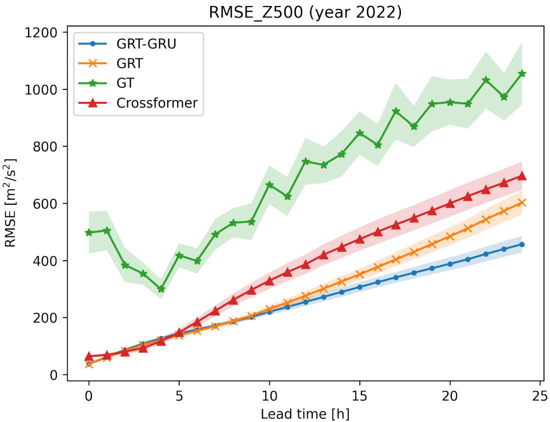
Higher is better

Mean (solid line) and standard deviation (shade) with first day of every month of 2018 (left) and 2020 (right) as initial condition.

Results for year 2022



Higher is better



Lower is better

Mean (solid line) and standard deviation (shade) with first day of every month of 2022 as initial condition.

Hybrid Sigma-pressure Coordinate

Source: https://www2.cesm.ucar.edu/models/atm-cam/docs/usersguide/node25.html