Graph Machine Learning for Global Weather Prediction: Graph Residual Transformer + Gated Recurrent Unit (GRU)

Arnold Kazadi

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

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

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

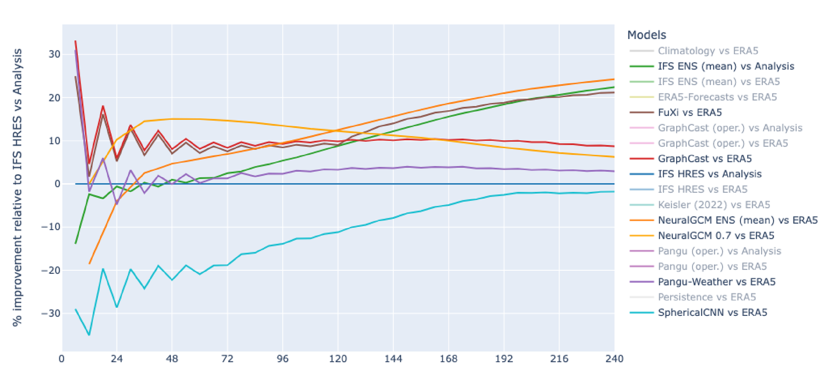


Figure 1:RMSE [m^2/s^2] relative to IFS-HRES for Geopential at 500 hPa. (source: Weatherbench)

Existing Machine Learning Methods

- 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.

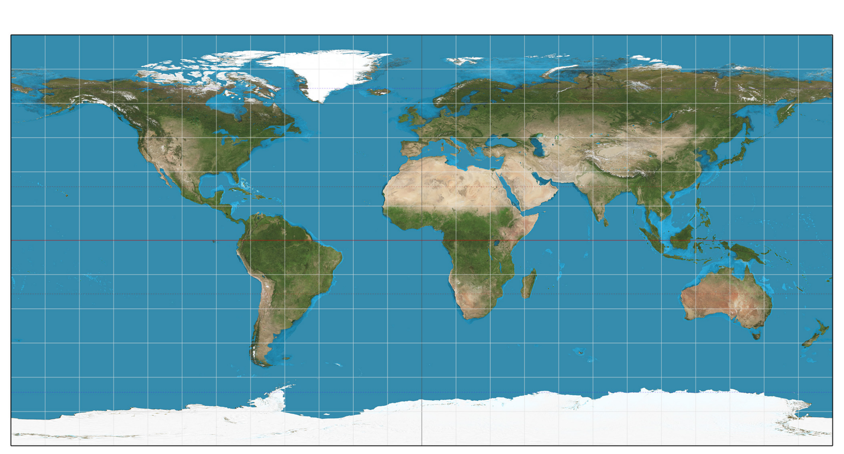


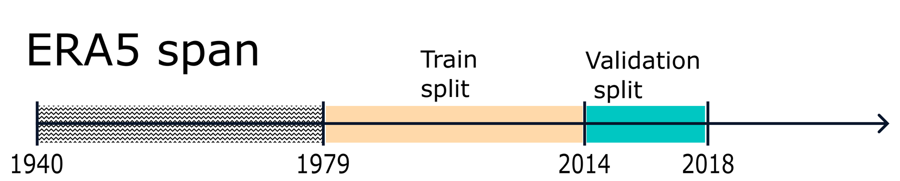
Figure 2:Equirectangular projection of the globe (Source: Wikipedia)

Dataset - ERA5

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

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

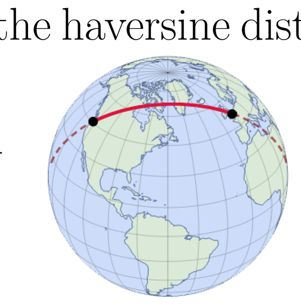
- Static and forcing variables: *Land-sea mask*, *Total Solar Irradiance*



Graph Generation from ERA5

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

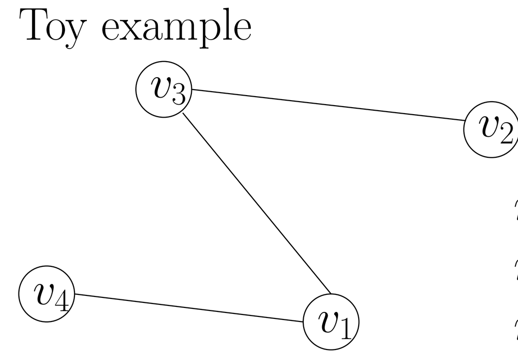
Source: https://www.linkedin.com/pulse/haversine-formula-firebird-sql-calculate-distance-between-revelli/



Graph Transformer

Attention weight for message passing and summation for update operation.

Toy example



Messaging Passing

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

Update Operation

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

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

Proposed Method: Graph Residual Transformer + GRU

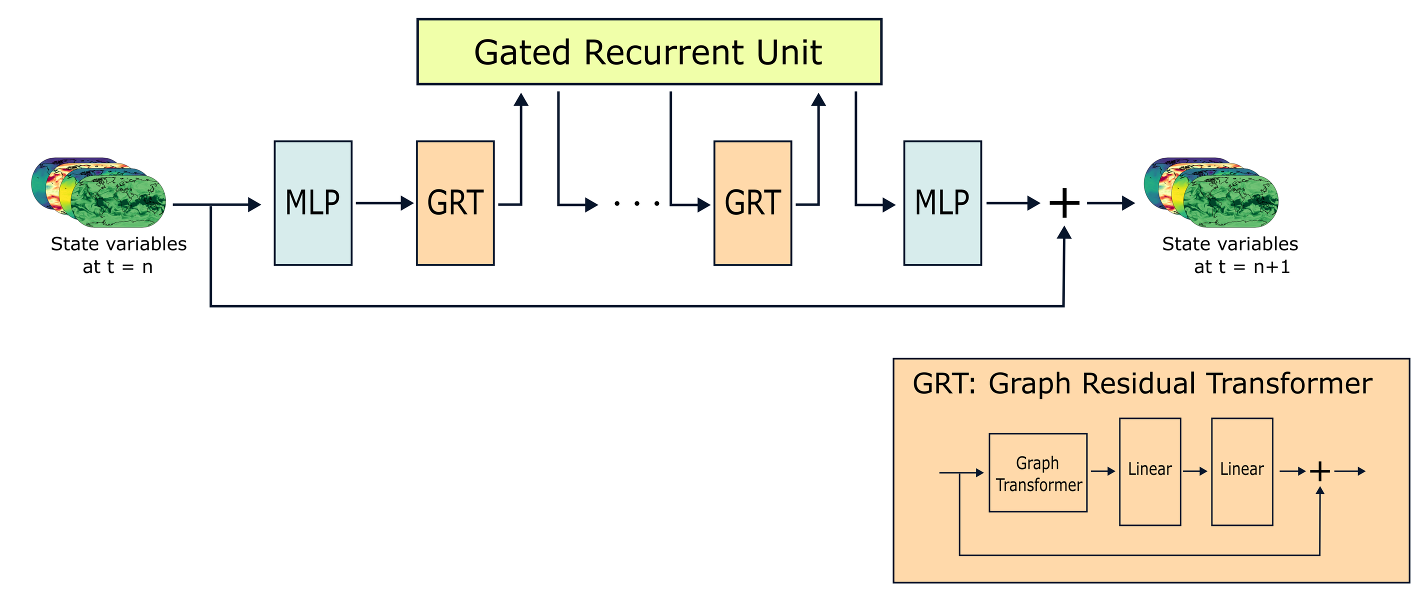
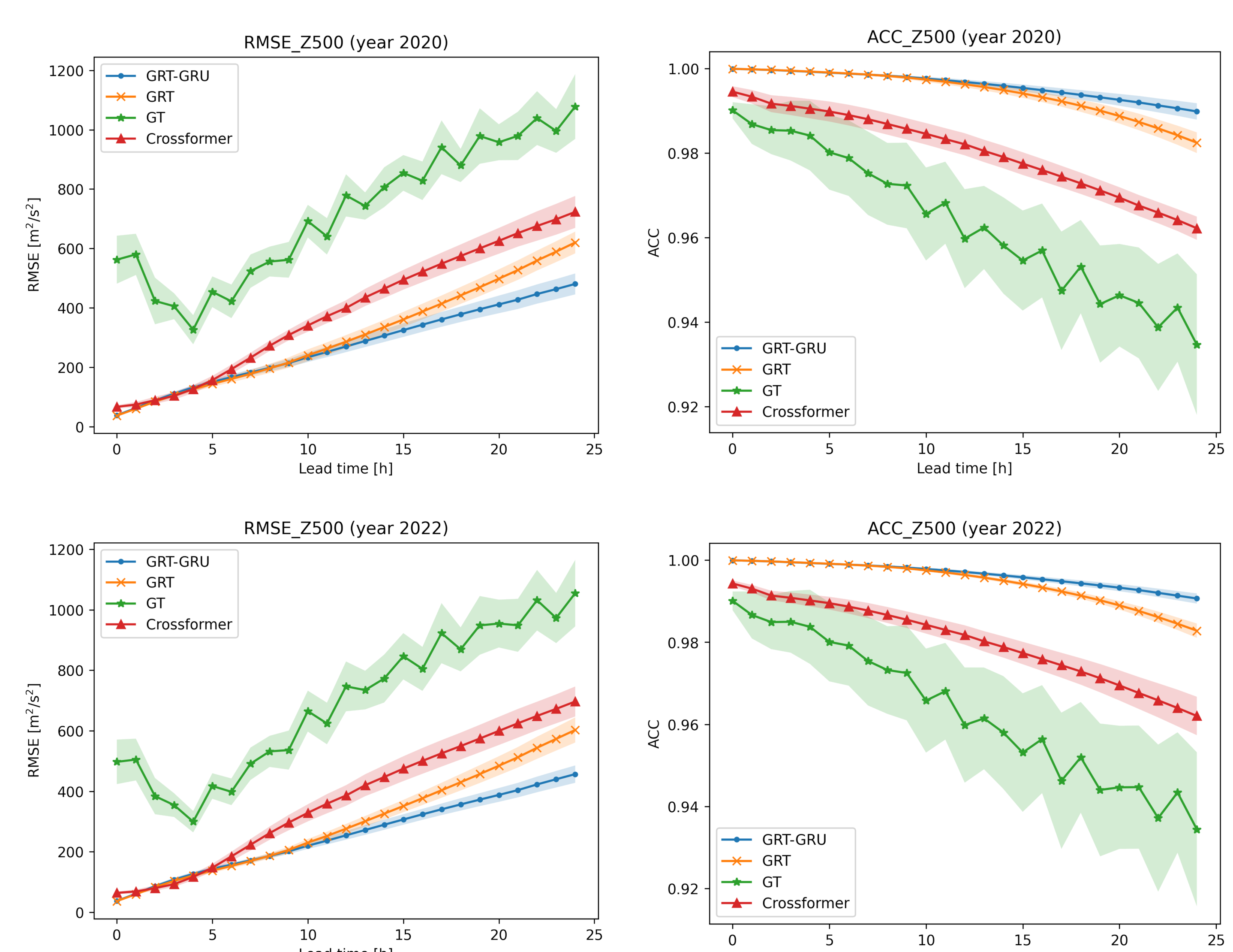


Figure 3:**Top**-Overall architecture. **Bottom**-Details of the GRT layer.

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

- **Crossformer**: Vision transformer-based model (~ 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 2020 (top) and 2022 (bottom) 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: Visualization

Results after rolling out to 24 hours/steps

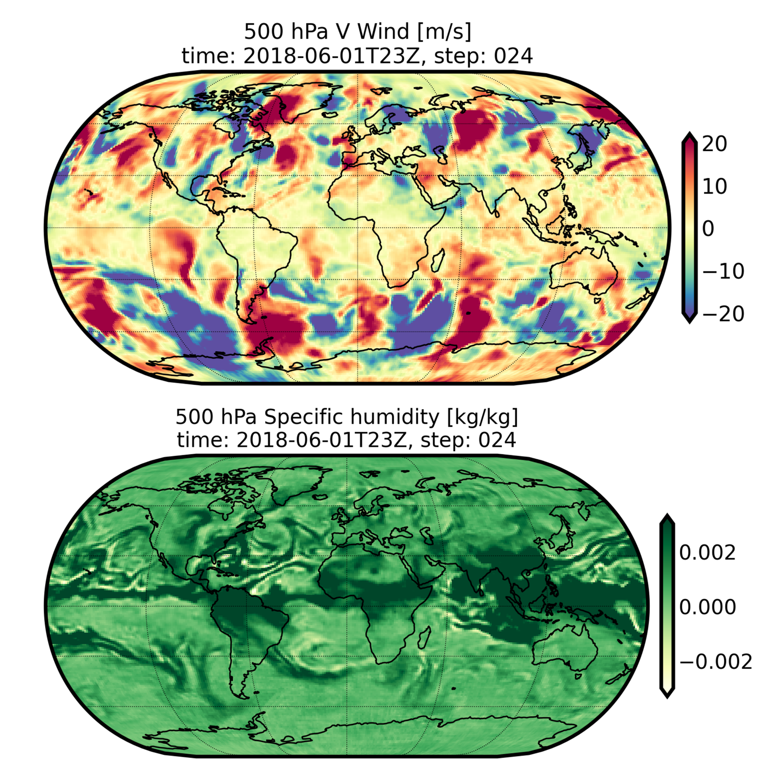


Figure 4:Crossformer

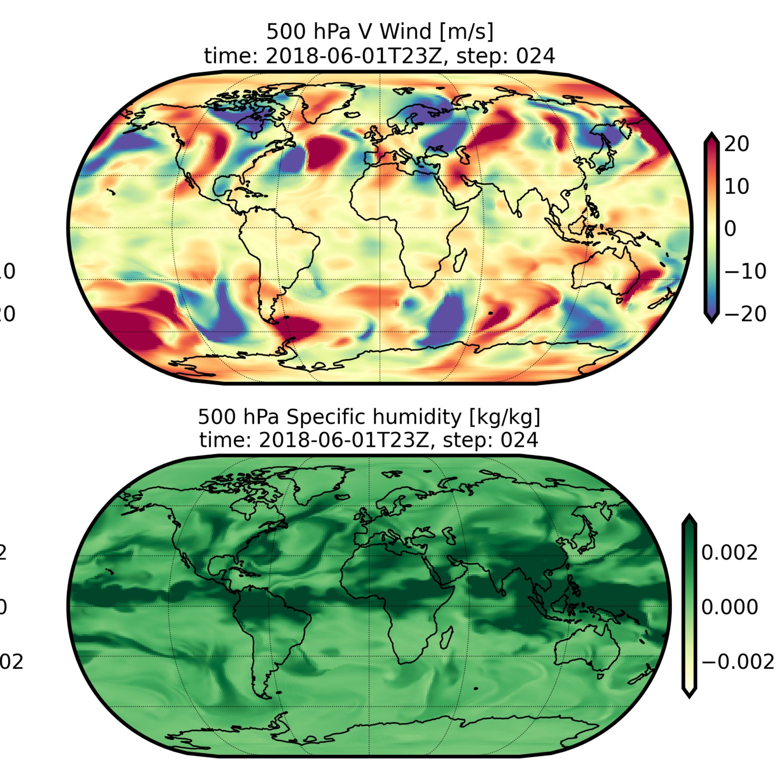
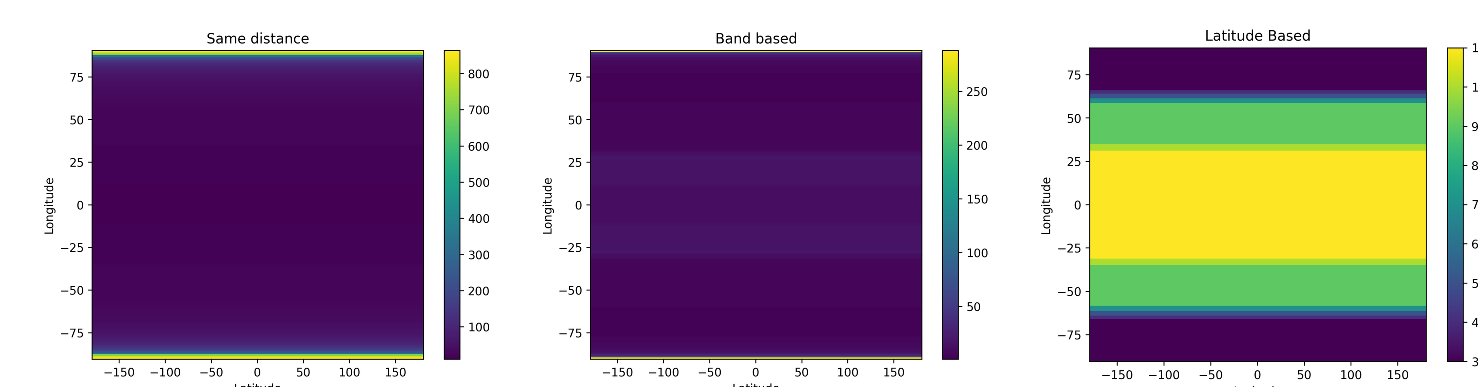
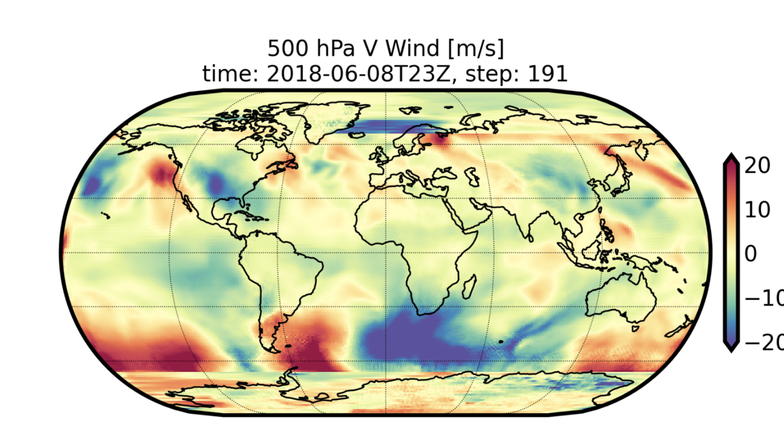


Figure 5:GRT-GRU

Appendix: Other Graph Approaches



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



Future Work

- Investigate diverse gridding of the globe
- Explore larger model (trained with fully shared data parallel)

Contact Information

- Email: akn7@rice.edu/@ucar.edu

Emulation video