An Evaluation Case Study on the Pre-trained Machine Learning Model FourCastNet-v2

Linlin Cui[1*] (linlin.cui@noaa.gov), Jun Wang[2], Sadegh Tabas[3], Jacob Carley[2]
[1]Lynker at NOAA/NWS/NCEP/EMC; [2]NOAA/NWS/NCEP/EMC; [3]Axiom at NOAA/NWS/NCEP/EMC

Abstract

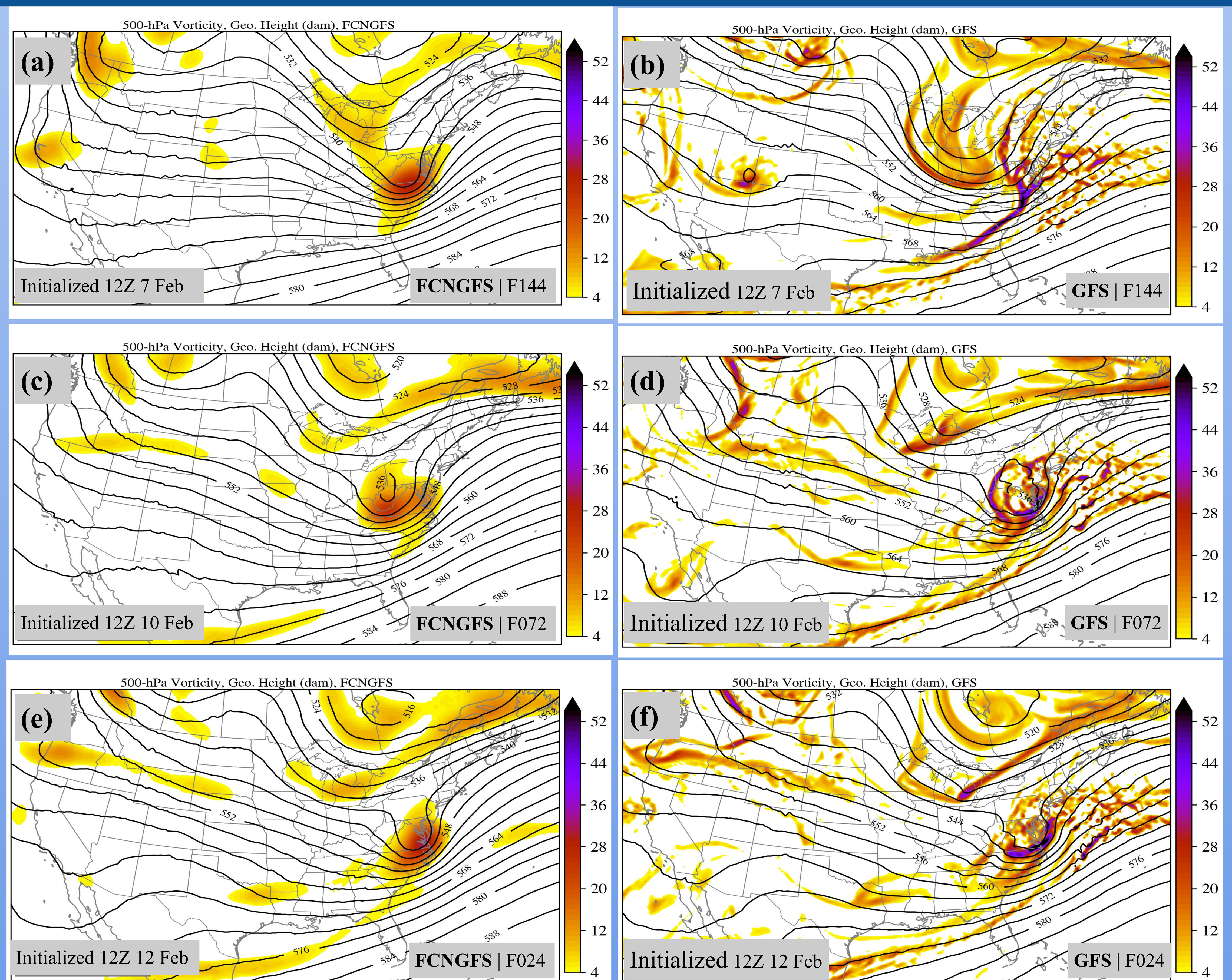
Machine learning techniques are increasingly recognized as viable tools for weather and climate prediction due to their efficiency and competitive performance relative to conventional numerical weather prediction models. This study assesses the effectiveness of FourCastNet-v2, a machine learning model utilizing the vision transformer (ViT) architecture and Spherical Harmonics Neural Operators, designed for modeling non-linear chaotic and dynamical systems on spherical surfaces. The Spherical Fourier Neural Operators (SFNOs) employed in FourCastNet-v2 not only retain the advantage of Fourier Neural Operators (FNOs) in simulating long-range dependencies in spatio-temporal data but also address their limitation in learning operators in spherical coordinates.

The National Centers for Environmental Prediction (NCEP) has initiated the use of pre-trained machine learning weather prediction models, including GraphCast and FourCastNet-v2, with Global Data Assimilation System (GDAS) data as inputs. The daily forecast results are accessible to the public on AWS. This study presents the model performance evaluation of FourCastNet-v2 through the computation of mean-squared error across several key atmospheric variables. Additionally, the effectiveness of FourCastNet-v2 in predicting a specific winter storm event is detailed.

FourCastNet

- Developed by NVIDIA (Bonev et al., 2023)
- Trained with ERA5 reanalysis data: 1979-2015 as training dataset, 2016-2017 as validation dataset
- Model weights and precomputed stats were used for setting up real-time FourCastNetGFS (FCNGFS) forecast, with GDAS as inputs
- Both input and output are on a 0.25 degree lat-lon grid
- Forecast to 240 hours every 6 hours at 00Z, 06Z, 12Z, and 18Z
- Model output fields include six 2D variables and five 3D variables at 13 pressure levels at 6-h time interval:
  - 2-m temperature, 10-m, 100-m u and v components of wind, mean sea level, surface pressure, precipitable water
  - Temperature, u and v components of wind, geopotential height, relative humidity
  - Vertical pressure levels: 50, 100, 150, 200, 250 300, 400, 500, 600, 700, 850, 925, and 1000 hPa

Model Evaluation - 12 UTC 13 Feb 2024 Winter Storm



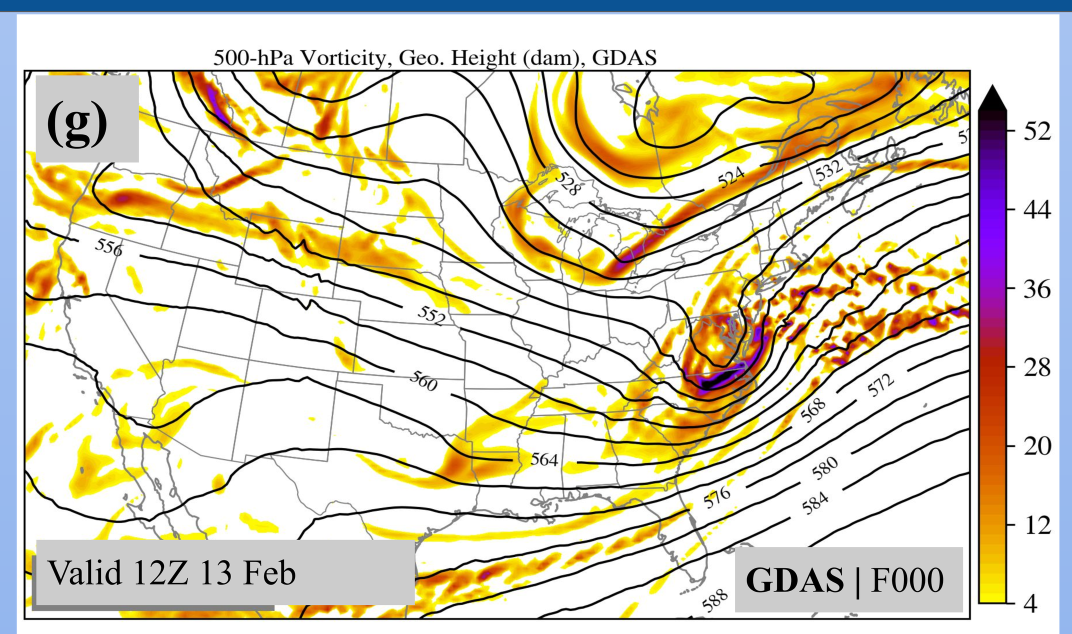


Figure: 500 hPa geopotential height (contours) and vorticity (colors) at 12UTC 13 Feb 2024 for FCNGFS (a, c, e), GFS (b, d, f), and GDAS analysis (g), with (a) and (b) initialized at 12UTC 7 Feb 2024, (c) and (d) initialized at 12UTC 10 Feb 2024, (e) and (f) initialized at 12UTC 12 Feb 2024, respectively. FCNGFS was more consistent run-to-run than GFS at the lead time of 144 hours (Fig. a and b), while GFS started to show this general pattern at the lead time of 120 hours (suppl.). From the lead time of 120 hours to the lead time of 24 hours, both FCNGFS and GFS predicted generally correct trough and cyclone evolution (suppl.).

Model Performance

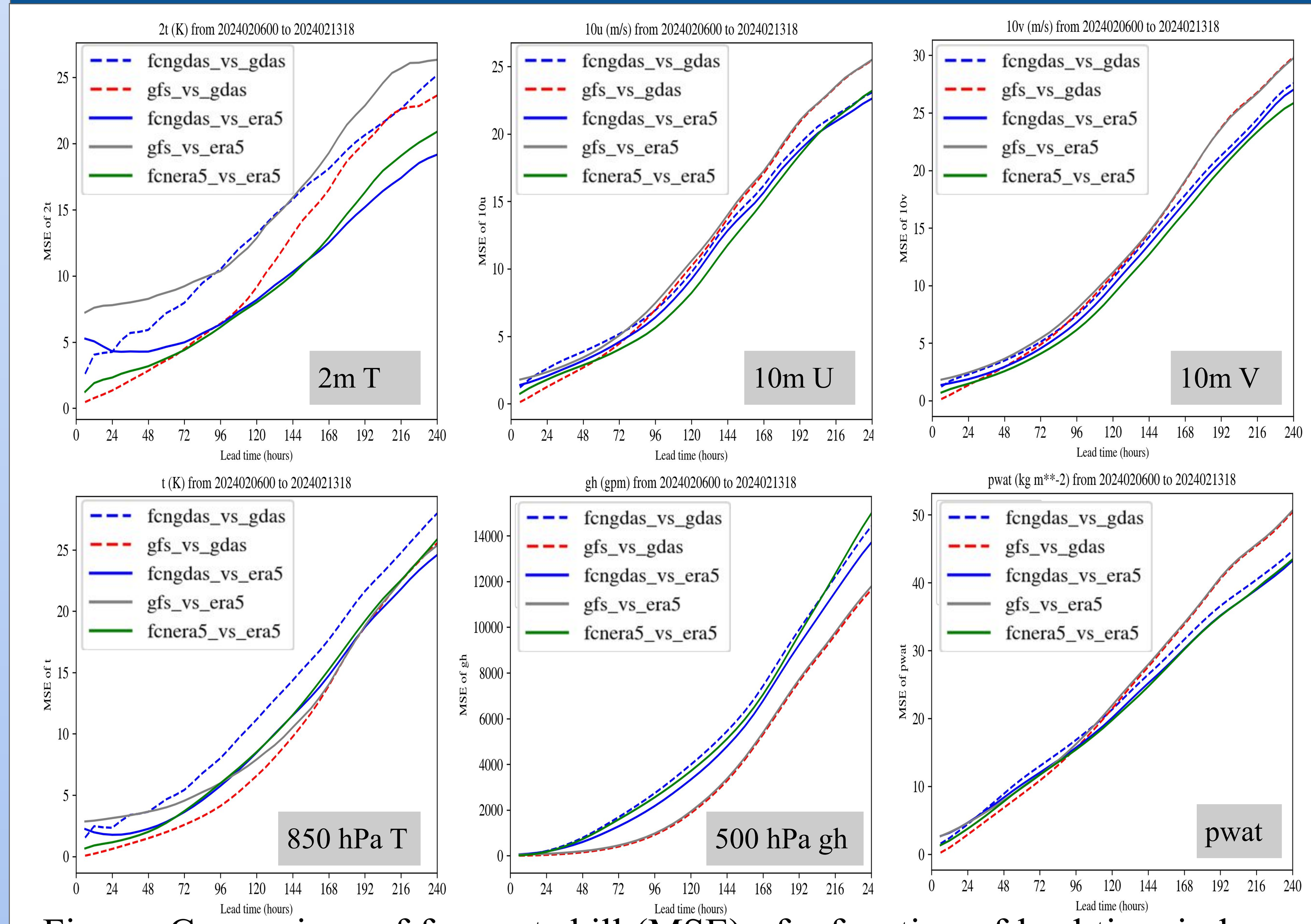


Figure: Comparison of forecast skill (MSE) of a function of lead time in hours for six physical variables: 2-m temperature; 10-m u and v components of wind; 850 hPa temperature; 500 hPa geopotential height; precipitable water. Blue dashed line is the comparison between FCNGFS and GDAS; red dashed line is the comparison between GFS and GDAS; blue solid line is the comparison between FCNGFS and ERA5; gray line is the comparison between GFS and ERA5; green line is the comparison between FourCastNetERA5 (FCNERA5) and ERA5.

Summary

- An experimental ML/AI based global forecast was set up with the pre-trained model and GDAS data as inputs
- The model performance was evaluated by mean squared error and a case study of a winter storm
- The results showed that FCNGFS was able to predict the trough/cyclone position at the lead time of 144 hours, which is one day earlier than GFS
- In terms of MSE, FCNERA5 had better performance than FCNGFS, which is reasonable, because the model is trained with ERA5 data; surface variables have better performance than 3D variables for the FourCastNet-v2 compared with GFS

Supplementary Materials

- FourCastNetGFS documentation:
  - https://fourcastnetgfs.readthedocs.io/en/latest/index.html
- GitHub repository:
  - https://github.com/NOAA-EMC/FourCastNet
- The experimental global forecast product:
  - https://noaa-nws-fourcastnetgfs-pds.s3.amazonaws.com/index.html
- Supplemental plots