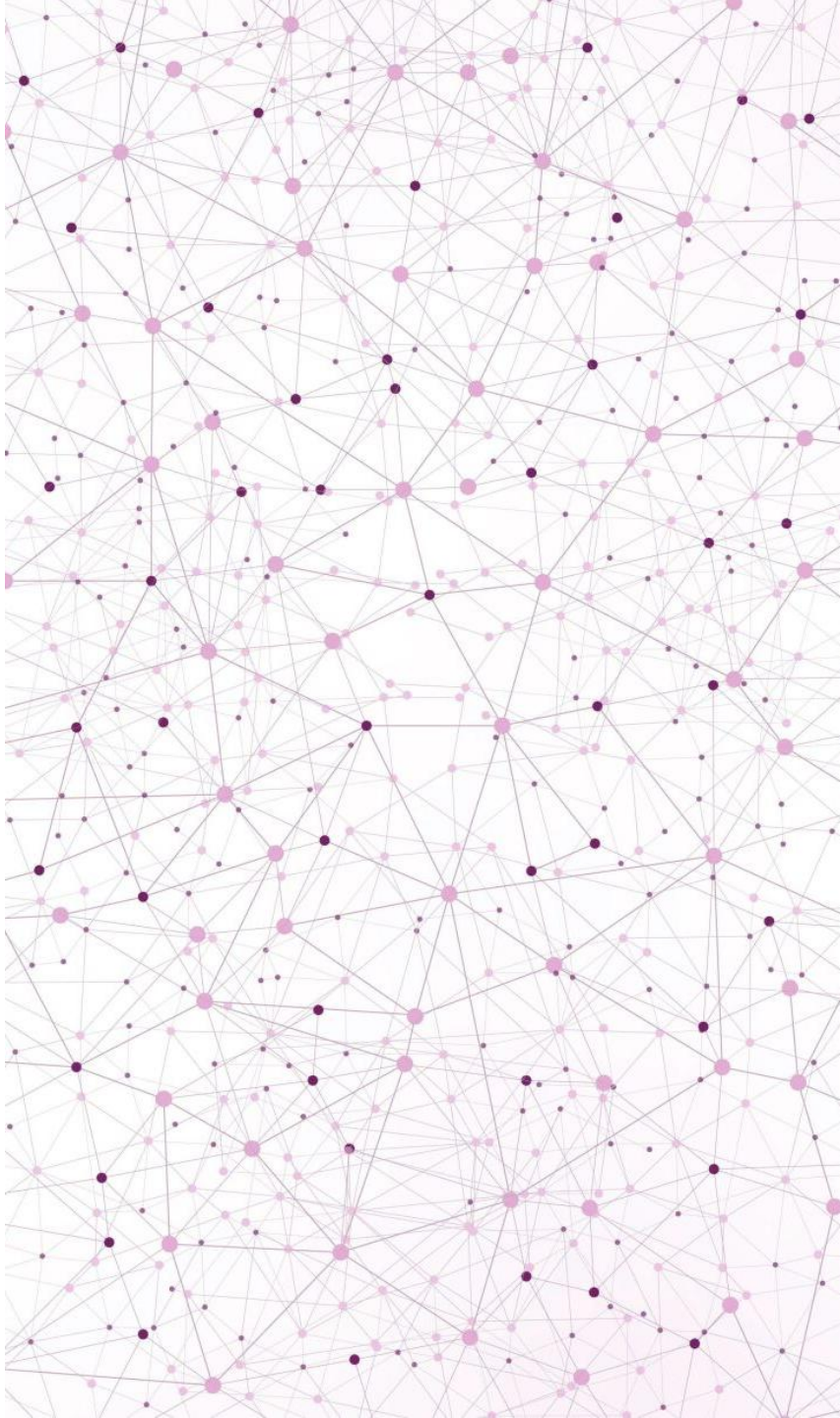
WoFSCast: A Machine Learning Model for Predicting Thunderstorms at Watch-to-Warning Scales

Paper by Montgomery L Flora, Corey Potvin

Miles Epstein

ML Journal Club 2/15/2025



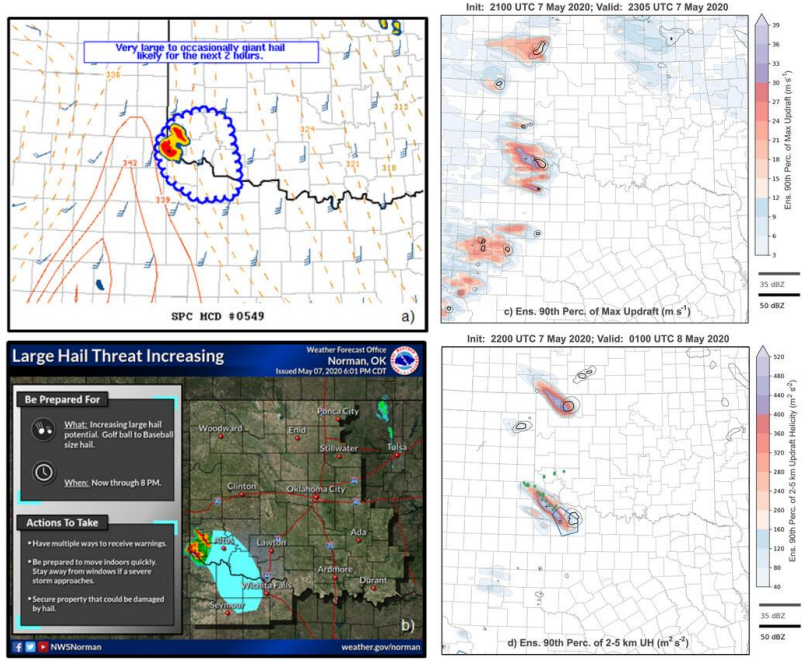

PREDICTING
TIMESCADE
PREDICTING
WARNING
GRAPHIAL
WATCH-TO-WARNING

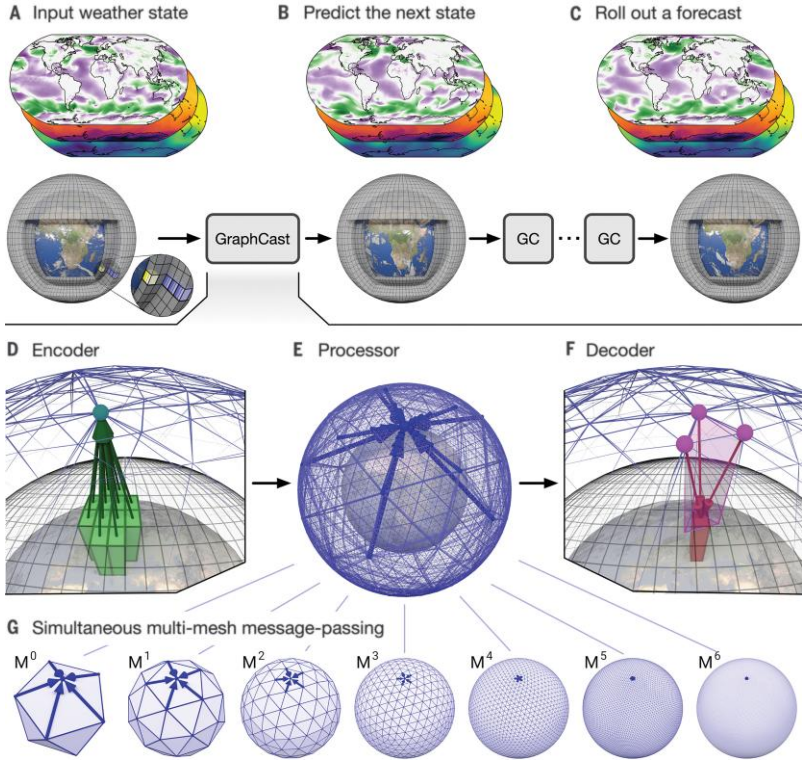
Warn on Forecast System

GraphCast

- Developmental NOAA project with goal of increasing leadtime for tornado, severe thunderstorm, and flash flood warnings (traditionally "warn-on-detection")
- **Convective-allowing ensemble**
- High resolution (3 km/5 min)
- Run out to 6 hours (watch-to-warning timeframe)
- Assimilates radar and satellite data every 15 minutes
- 900x900 km grid, placed on desired region on days of interest

- Google DeepMind's AI medium-range forecast model
- Graph Neural Network (GNN) trained on ERA5 reanalysis
- Coarse resolution (0.25°/6 hour)
- Run out to 10 days (medium-range)
- Global multimesh grid





MOTIVATION/ APPROACH

Existing ML forecast models trained on coarse ERA5 (or HRRR)

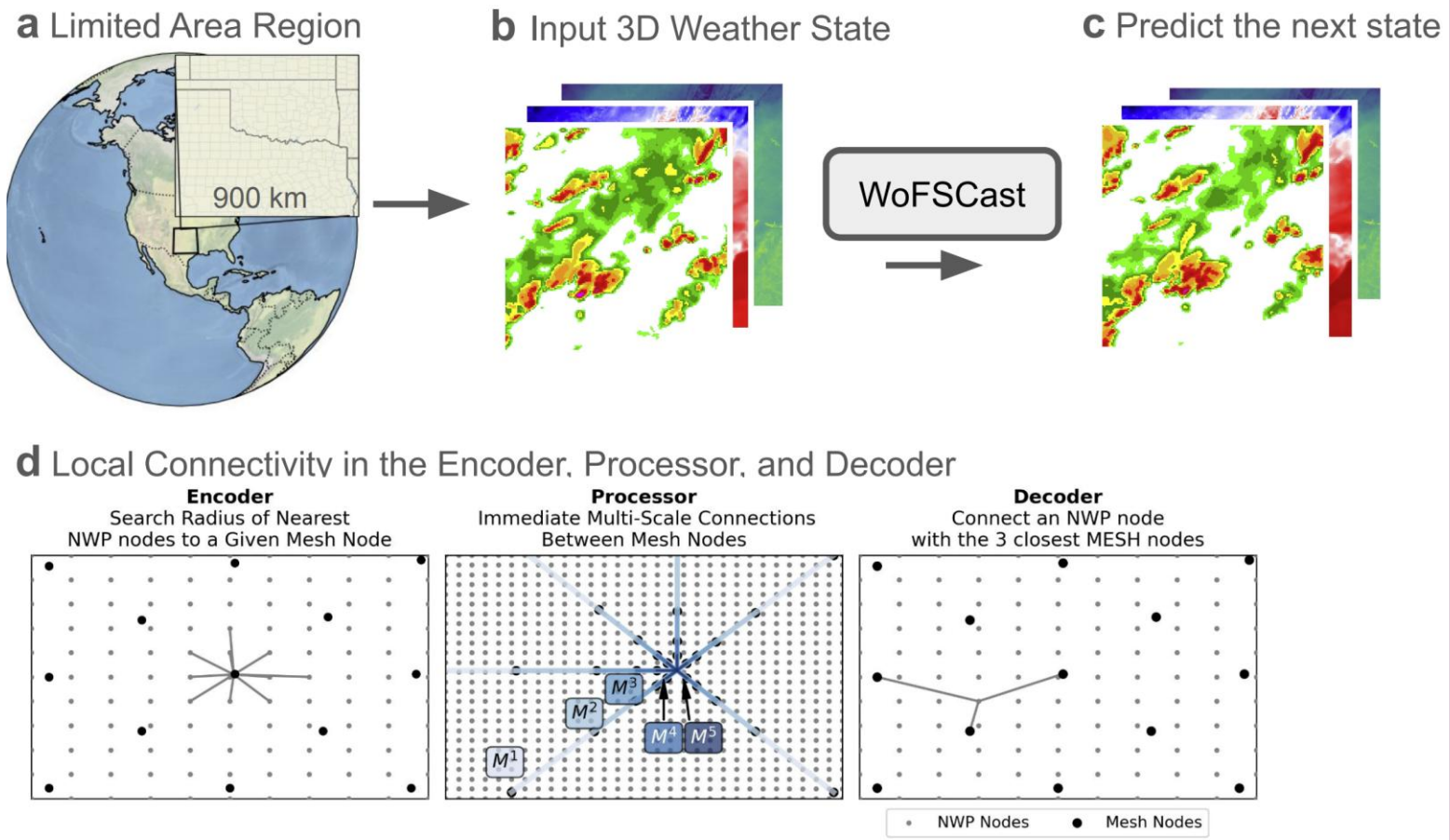
Method: train model with GraphCast framework on archived WoFS data

Result: AI forecast model that works like GraphCast, but used on smaller, higher-resolution domain

Goal: emulate high-resolution WoFS quickly

Discuss:

- Why would we want to emulate WoFS?
- How is training on model output different from training on reanalysis?
- Do short steps improve model ability because changes are more linear?



MODEL ARCHITECTURE

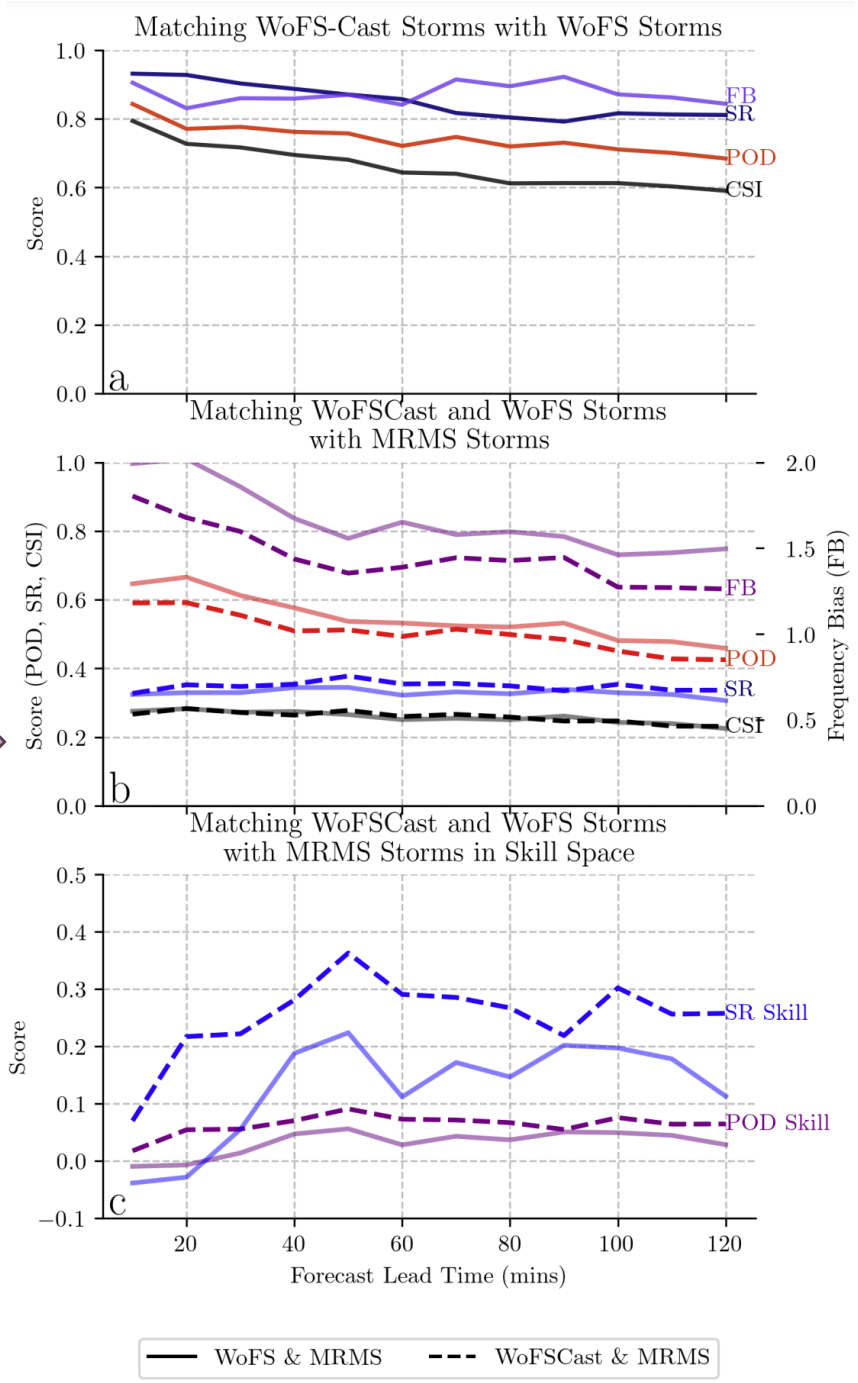
- Steps 105 state variables forward at 10-minute increments (thinned in time and space)
  - Implications?
- 6/3 Fully connected layers to encode/decode
- 16 GNN layers in processor
- Grid is just triangles in small domain
- Training set: 131 Spring WoFS cases 2019-2020
- Test set: 100 random WoFS cases in 2021
- Thoughts/critiques?

OBJECT-BASED VERIFICATION

- WoFS and WoFSCast reflectivity compared to each other and to MRMS (observation)
- Presence of composite reflectivity above threshold (40 dBZ for MRMS, 47 dBZ for WoFS[Cast])

	Observed Yes	Observed No
Forecast yes	a	b
Forecast no	c	d

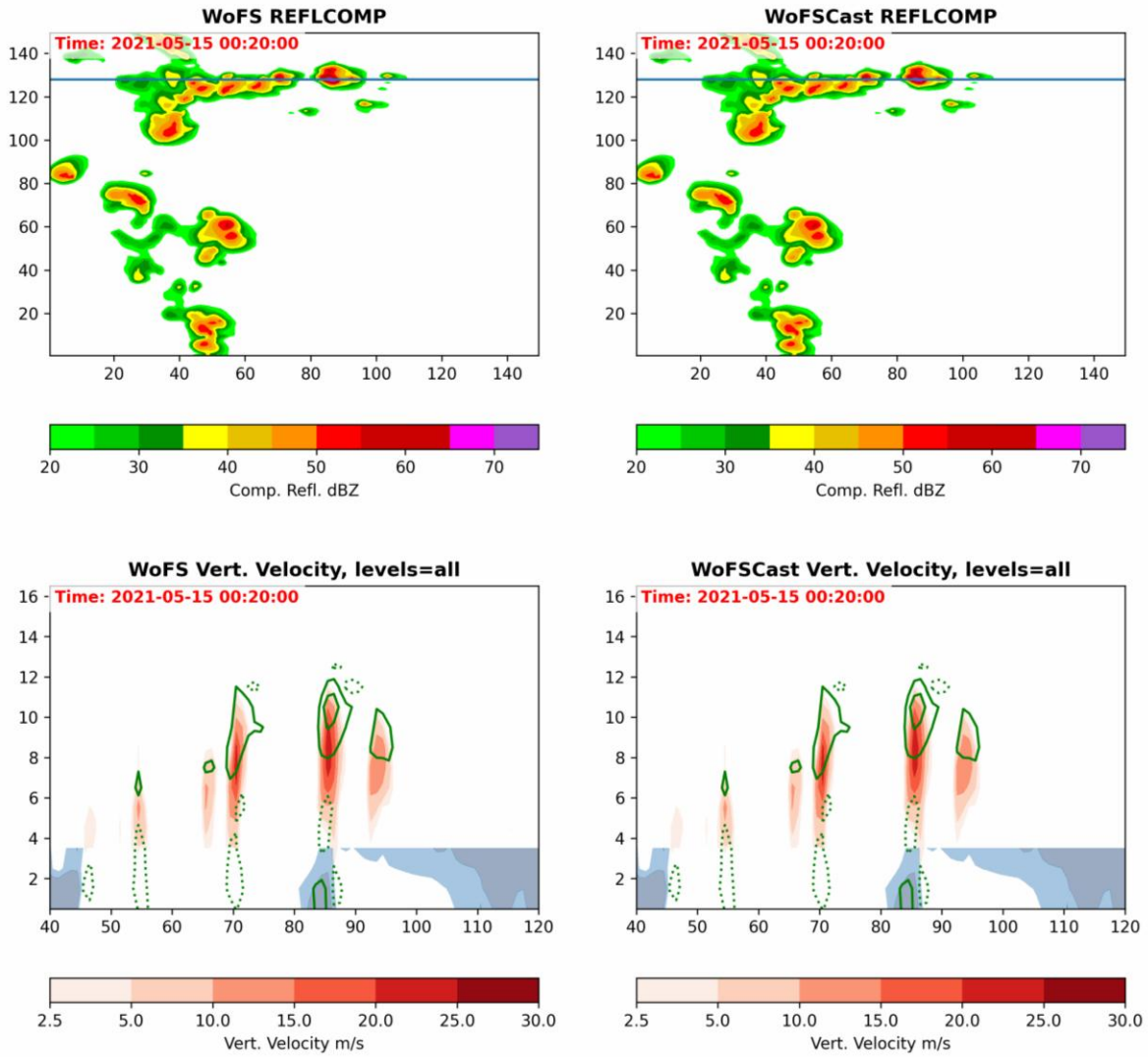
Figure 2



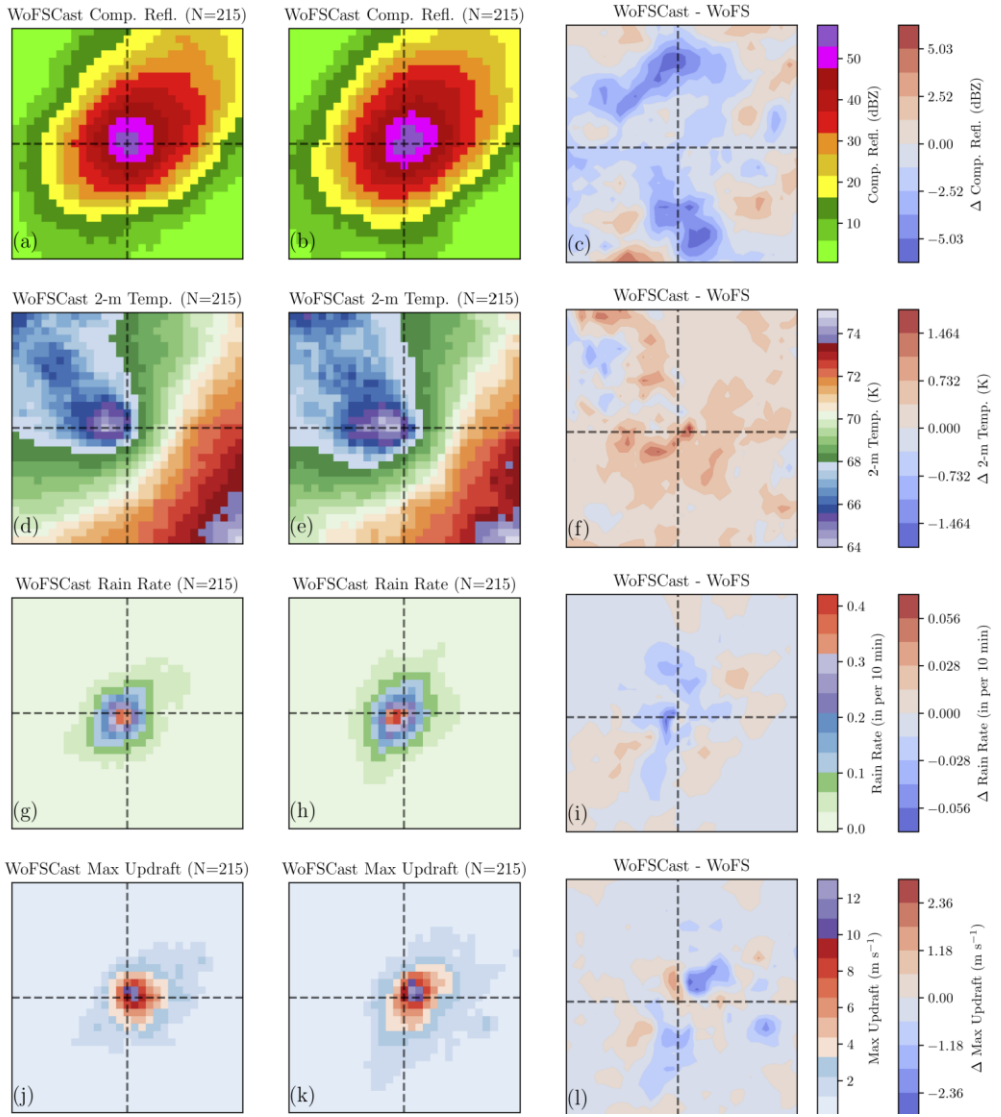
Statistic	Value	Range	Meaning
POD	a / (a + c)	[0, 1]	Fraction of storms forecast
SR	a / (a + b)	[0, 1]	Fraction of forecasts that verify
FB	(a + b) / (a + c)	[0, ∞]	<1: underforecast 1: same number of storms forecast as observed >1: overforecast
CSI	a / (a + b + c)	[0, 1]	Skill score

STRUCTURE-BASED VERIFICATION

Single-storm animation (Movie S1)



Composites (Fig 3)



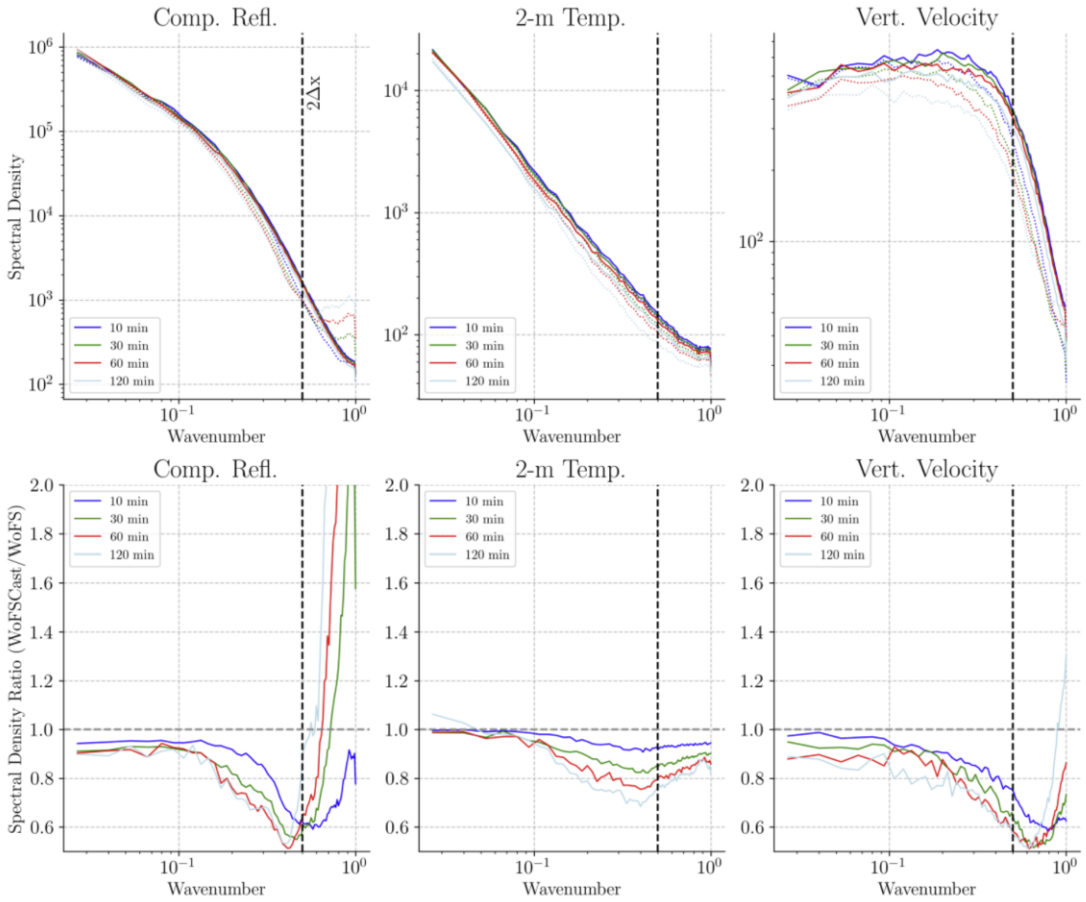


Figure S3. Top row: Energy spectra of select variables at four forecast lead times (blue=10 min, green=30 min, red=60 min, and light blue=120 min). WoFS is shown in solid lines, and WoFSCast in dashed lines. Bottom Row: Energy spectra ratio of WoFSCast to WoFS.

GRID-BASED VERIFICATION

- RMSE increases (Figure S1) and FSS decreases (Figure S2) over time between WoFSCast and WoFS
- Energy spectra (left) mostly retained over time in both models

CONCLUSIONS

- WoFSCast is good proof-of-concept for high resolution AI NWP emulators
- Main benefit: speed, and thus application in ensemble
- Next steps: train with more data, higher resolution, or analyses; run out to 6h
- Does WoFSCast emulate WoFS well? How useful is this?
- Is it reasonable to use as extra ensemble members?
- Is this a promising direction of research?