# AI/ML for NWP Activities at NOAA OAR & NWS (not exhaustive) -- Highlights from NCEP/EMC

NATIONAL WEATHER SERVICE

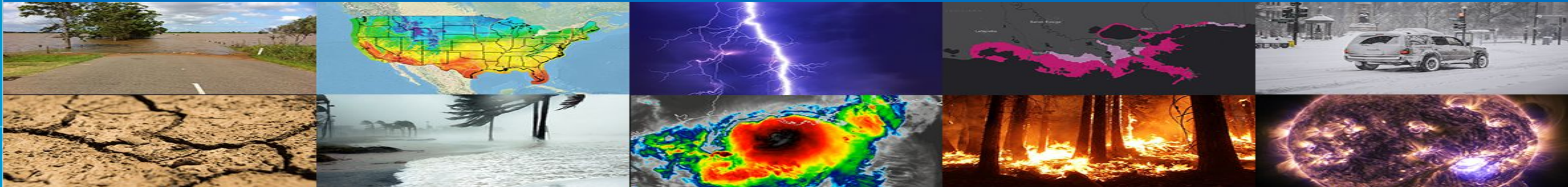
**Fanglin Yang**

Physics and Dynamics Division

NOAA/NWS/NCEP Environmental Modeling Center

Acknowledgment: Most of the slides were copied from a presentation Daryl Kleist gave at the 2nd Workshop on Large-Scale Deep Learning for the Earth System In Bonn, Germany. Additional materials were provided by Jun Wang, Wei Li and other colleagues at NCEP/EMC ......

39th WGNE workshop, November 4-8, *Météo-France, Toulouse, France*



# AI4WP Report -- Workshop Nov 2023

Integration of emerging data-driven models into the NOAA research to operation pipeline for numerical weather prediction

**Authors:** Sergey Frolov[1], Kevin Garrett, Isidora Jankov, Daryl Kleist, Jebb Q. Stewart, John Ten Hoeve.

**Meeting title:** AI4NWP: integrating emerging machine learning tools into NOAA's research-to-operations pipeline for numerical weather prediction.

**What**: Identify a research and development roadmap and priorities for integrating emerging deep learning tools for numerical weather prediction into the NOAA production pipeline.

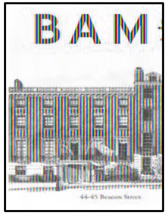
**When**: November 28-30th, 2023.

**Where**: Boulder, Colorado and online.

## Imperative to act

Over the last 18 months, a tidal wave of data-driven models revolutionized the way we think about the future of the weather prediction enterprise (Bauer et.al 2023). Groundbreaking data-

https://doi.org/10.1175/BAMS-D-24-0062.1




**Urgency to act:**

- Technology is developing rapidly
- Potential has been demonstrated

**Opportunities:**

- Ensemble forecasts represent the majority of computational expenses at NOAA R&D and operations
- Replacing ensemble forecasts with AI4NWP tools presents a great opportunity for both skill improvement and cost reduction
- Significantly large ensembles should:
  - Better represent extreme events,
  - Improve initial conditions.
- Reduce cost of operational transitions, operational forecasts, and reanalysis production
- Excitement for NOAA to do things differently, be more agile!

# Training of 'NOAA-native models'

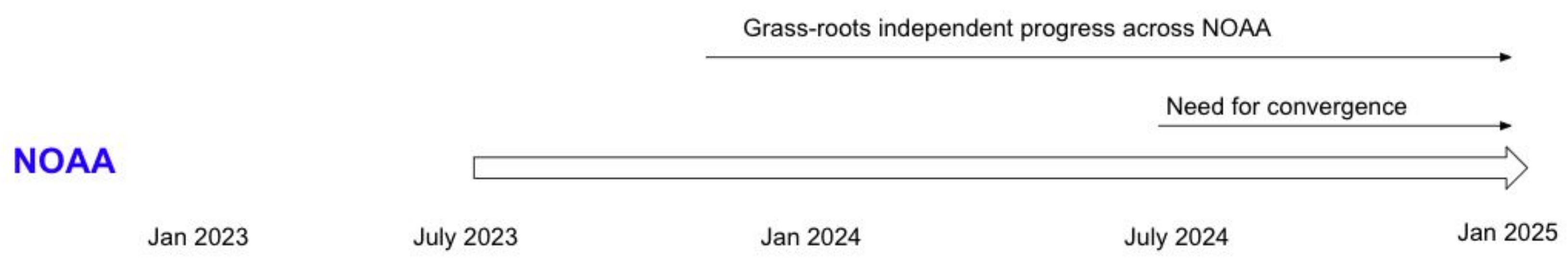
- Most current efforts are based on modification of the **GraphCast** code from **Google DeepMind**.

- To train or fine-tune models :
    - Prepare and stage NOAA data;
    - Develop data loader for NOAA data;
    - Modify GraphCast code to work with NOAA grids and variables;
    - Develop training harness that works on multiple GPUs.

- This has been undertaken in parallel by at least 3 groups at NOAA: **EMC, NSSL, PSL/GSL.**

# AI4WP Cross-Line Office Team

## Balancing fast progress and collaboration

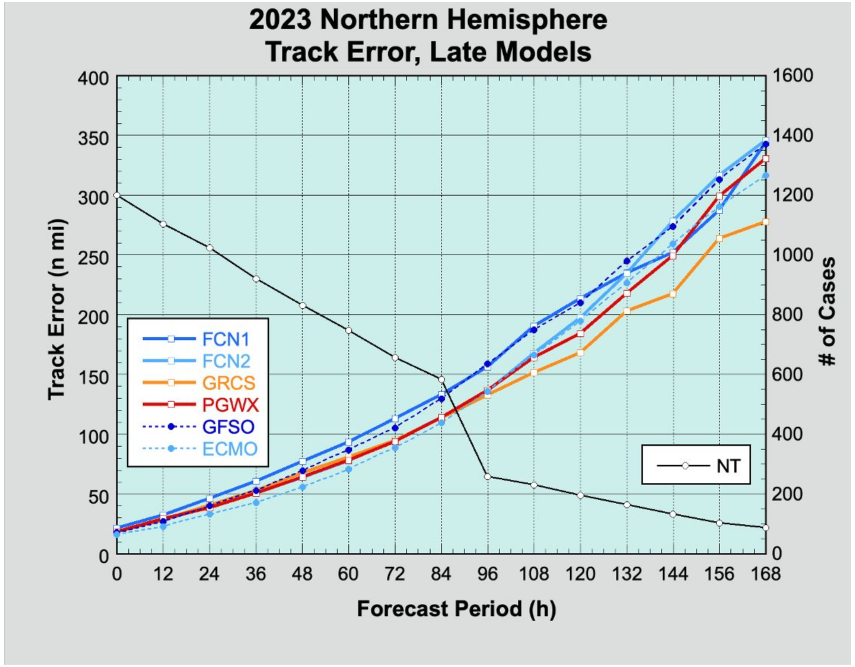



- EPIC is developing website to highlight and connect activities across NOAA.
- EPIC/PSL/GSL: evaluating suitability of Anemoi/AIFS as platform for convergence of AI4NWP activities.
- GSL/PSL/NSSL/EPIC: developing verification pipeline for models over CONUS.
- EMC/EPIC/GSL/PSL/?: could we converge on the official R2O pipeline for training and evaluation?
- Investigating path forward to generate high-resolution, high-quality training data.

# CIRA/GSL TC Evaluation

Research Team: CIRA/CSU Tropical Cyclone and AI Groups, and GSL
Lead: Mark DeMaria (CIRA)

**Verification of May-Nov 2023 Tropical Cyclone Track Forecasts using the National Hurricane Center Verification Methodology**




FCN1 = FourCastNet original
FCN2 = Updated FourCastNet
GCRS = GraphCast
PGWX = PanguWeather
GFSO = NCEP GFS global Model
ECMO = ECMWF-IFS global model

## A) Track results (on left):

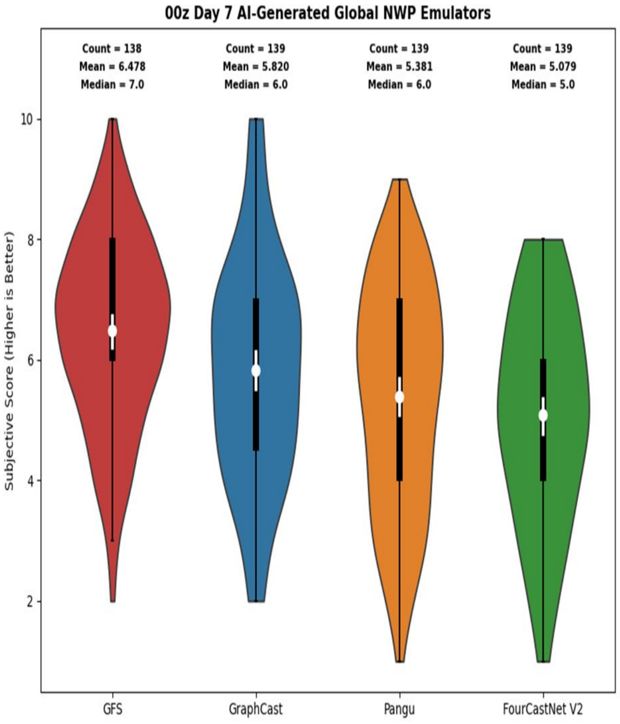
Model track errors very similar between NWP and AI4WP models, except:

- FCN1 a bit worse through 96 h,
- GraphCast was better than all models after 96 h.

## B) Intensity results (not shown here):

- All AIWP models have extreme low bias.

# NSSL: Spring forecasting experiment results

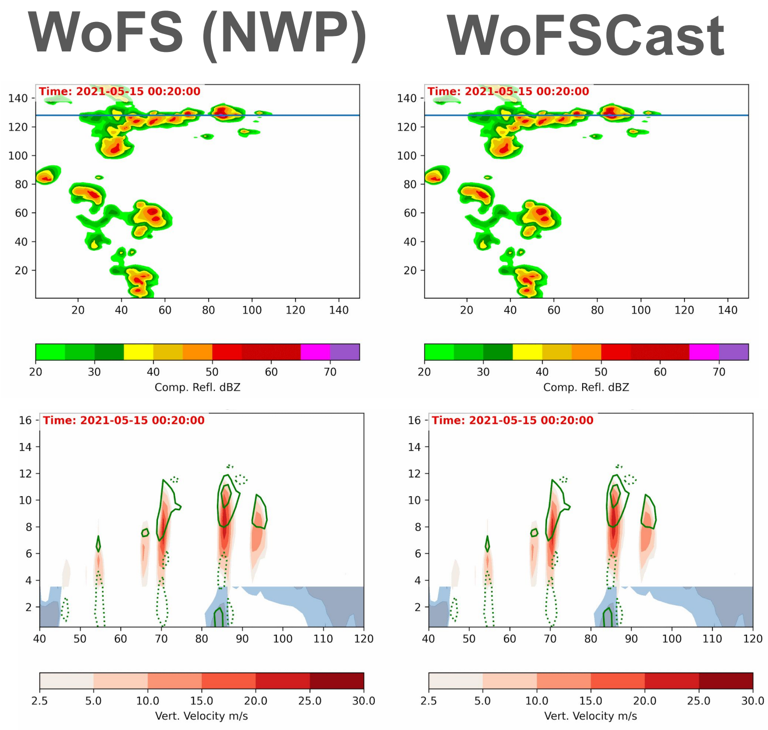



Slide courtesy of David Harrison

- GFS received the highest ratings on average, but GraphCast was statistically similar
- AI models generally rated higher in strongly-forced environments; **struggled with weak/subtle forcing and split-flow regimes**
- Derived soundings were physically realistic but very **smooth – difficult to get inversions**
- **AI models showed less run-to-run and day-to-day consistency** than the GFS – the “best” AI model varied considerably each day
- Participants impressed by current state of models but indicated there’s still work to do before they’re ready for operations

# NSSL: WoFSCast

- Modified GraphCast to work with regional domain
- Trained on archive of 3-km WoFS forecasts.
- High-fidelity emulation of WoFS, even without diffusion, owing to 10-min data interval

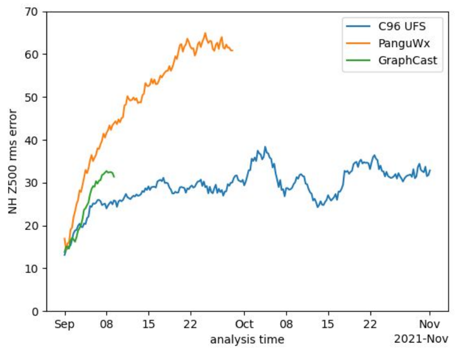


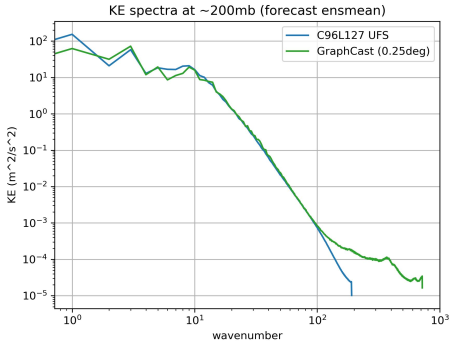

**Composite reflectivity**
**(horiz line → vertical cross-sections below)**

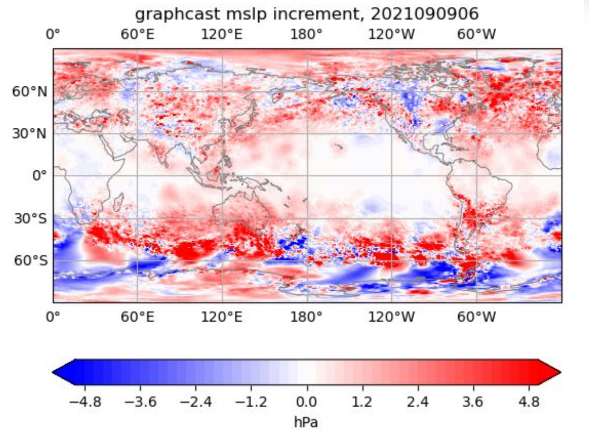
**Updraft speed, divergence, $\theta_v$ perts (cold pools)**

Slide courtesy of Corey Potvin, Monte Flora

# PSL: Cycling with emulators




- Assimilation of surface pressure only observations as in 20CrV3 reanalysis using EnSRF.
- All deterministic AI4NWP models grow instabilities and "blow-up".
- Path forward: train our own models that can be resilient to high-frequency noise in initial conditions, reduce non-physical noise in the analysis states.

Slide courtesy of Jeff Whitaker and Laura Slivinski

# Highlights of AI/ML Activities at NWS/NCEP Environmental Modeling Center

- Training GraphCast Weather model using GDAS data (MLGFS)
- ML-based Global Ensemble Forecast System (MLGEFS)
- ML-based Air Quality Model (AQcGAN)
- ML for bias corrections

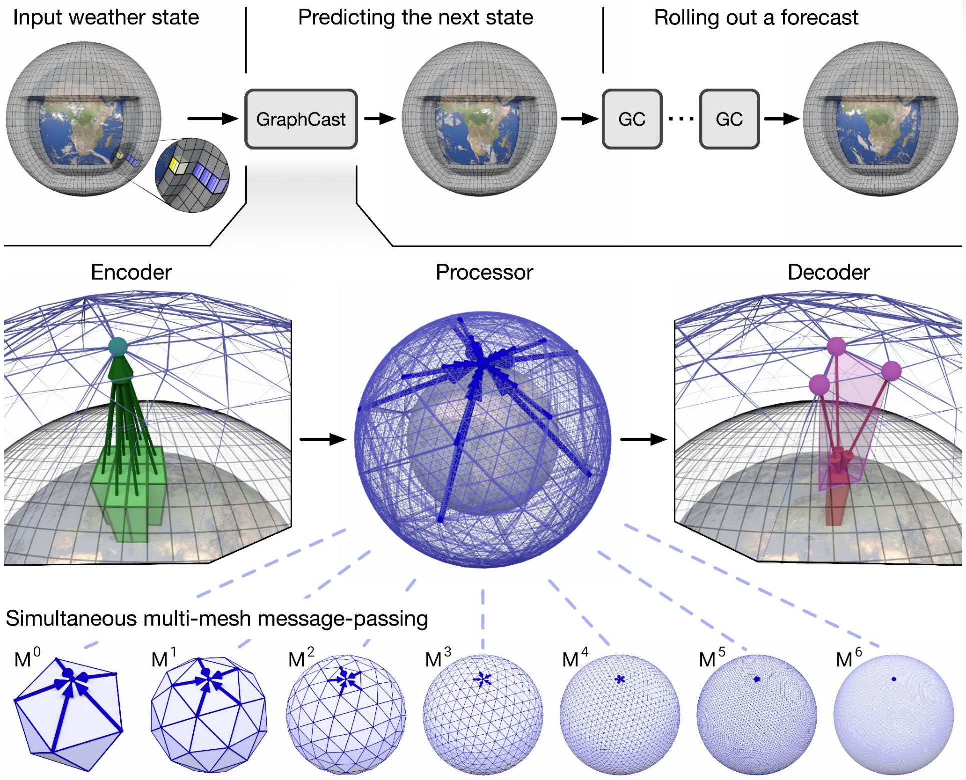
# GraphCast Model Implementation at NCEP/EMC

Resolution:
- 6-hourly temporal (up to 10 days lead time)
- 0.25-degree spatial

Num. Levels:
- **13** [50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, 1000]
- **37** [1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 925, 950, 975, 1000]

Surface Variables:
- 2m Temperature,
- 10m U&V Components of Wind,
- Mean Sea Level Pressure,
- Total Precipitation

Atmospheric Variables:
- Geopotential Height,
- Temperature,
- Specific Humidity,
- U&V Components of Wind,
- Vertical Velocity

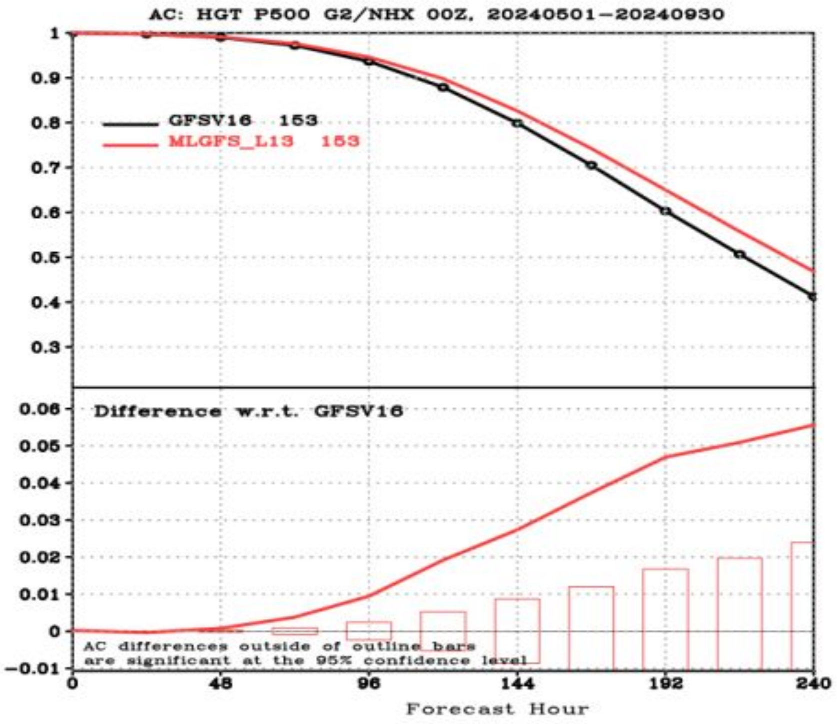


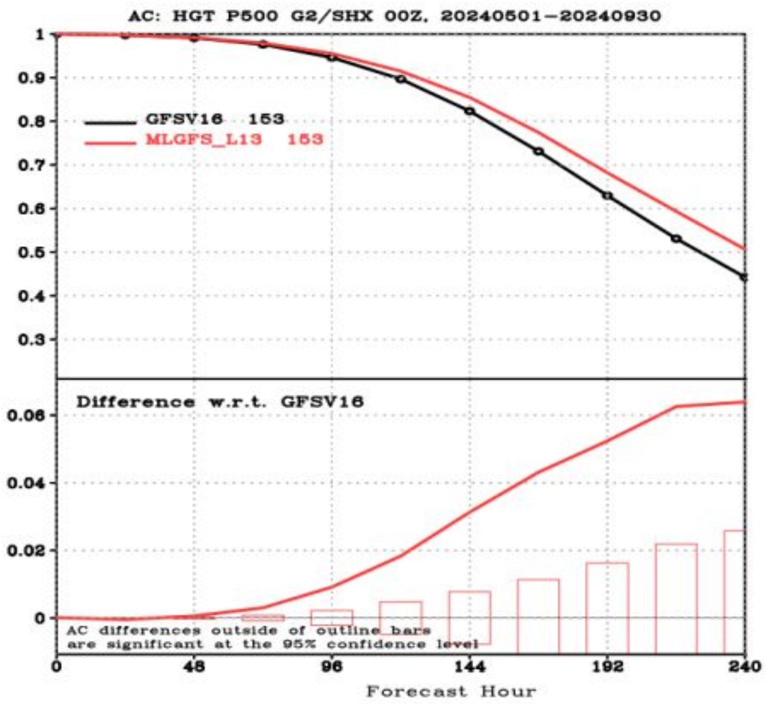
# Training GraphCast Weather model using GDAS data

- Compute Node:
  - PW AWS Cloud
  - 8 H100 80GB memory GPU cores
- Training period:
  - 2021-03-21 - 2022-09-01 (4 cycles/day)
- Validation period:
  - 2022-09-01 - 2023-01-01 (4 cycles/day)
- Verification period:
  - 2023-01-01 - 2024-01-01 (2 cycles/day: 0z and 12z)
- Training steps
  - Model 1: Fine tuning GC for 12, 13 and 14 Autoregressive steps using GDAS and ERA5 data
  - Model 2: Full Training GC with GDAS data

- Metrics:
  - RMSE,
  - ACC (ERA5 Climatology data were used to calculate ACC)
- Scenarios:
  - 1) Fine tuned GraphCast against ERA5
  - 2) Fine tuned GraphCast against GDAS
  - 3) Fully trained GraphCast against GDAS
  - 4) GFS forecast against GDAS
- Regions:
  - Global,
  - North America,
  - Northern Hemisphere,
  - Southern Hemisphere,
  - Tropics

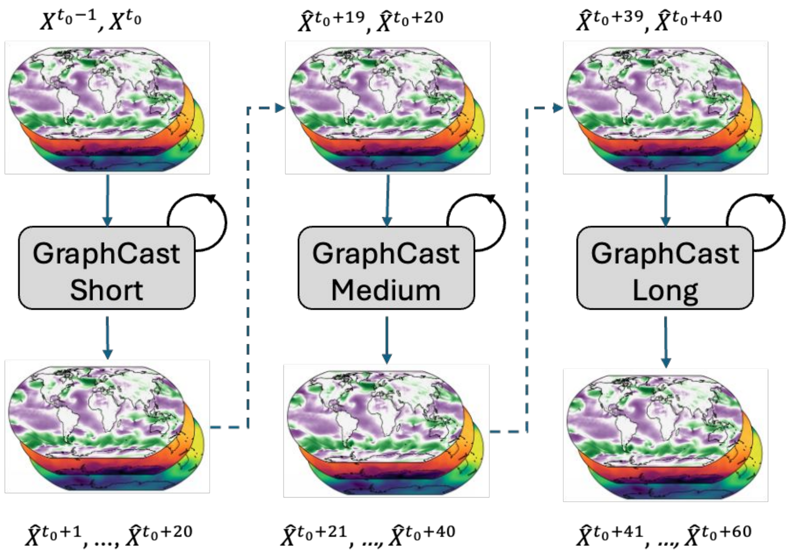
# 500-hPa HGT ACC, May-Sept 2024



# ML-based Global Ensemble Forecast System (MLGEFSv1.0)

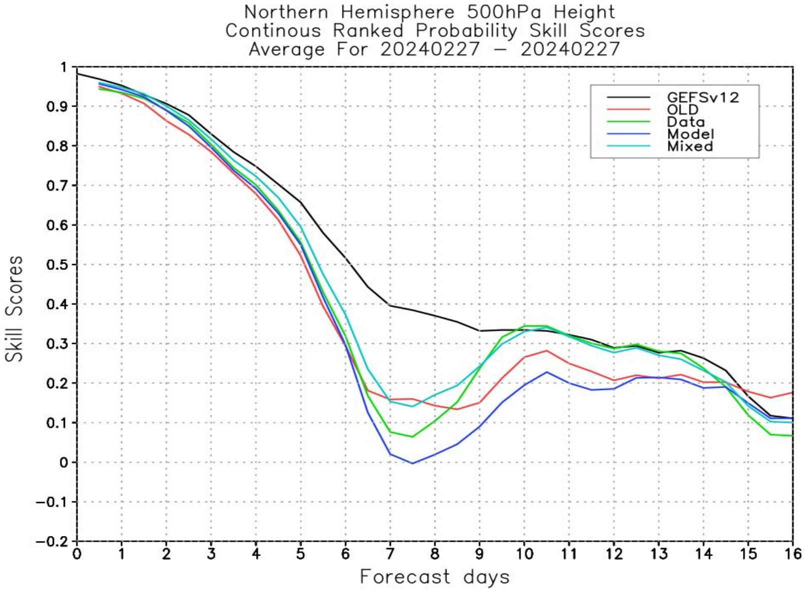
- Resolution:
  - 0.25 degree spatial resolution
  - 6-hourly temporal resolution up to 16 Days lead time
- Number of Levels: 13
- Core Architecture: GraphCastGFS
- Number of Ensemble Members: 31
- Data Uncertainty:
  - ICs: Perturbed GEFS Initial Conditions
    - 30 perturbed members
    - 1 control member

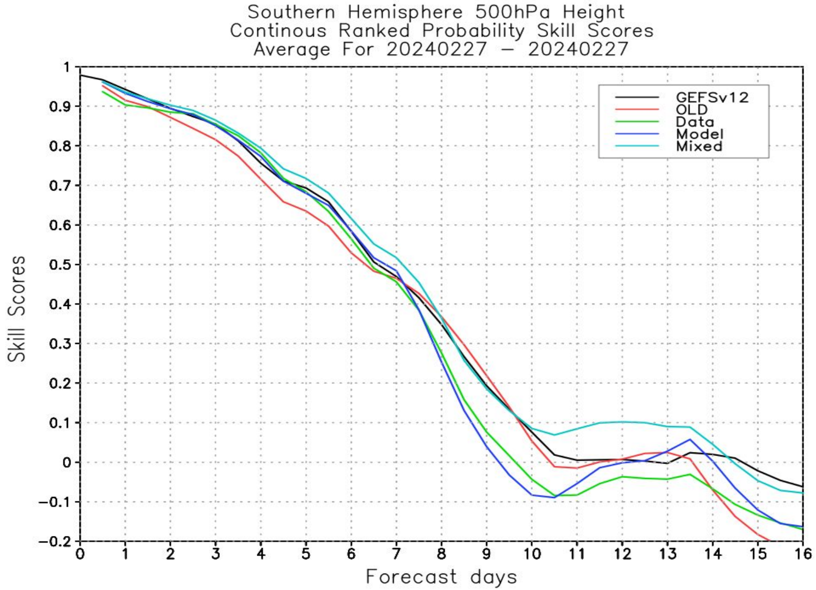



- Model Uncertainty:
  - Multiple configurations of GraphCastGFS (perturbing model weights) initialized with the control member
    - 30 members from configurations of GraphCastGFS
- Note: We are also considering to implement GenCast and SEEDS ensemble models from Google.

# ML-based Global Ensemble Forecast System (MLGEFSv1.0)

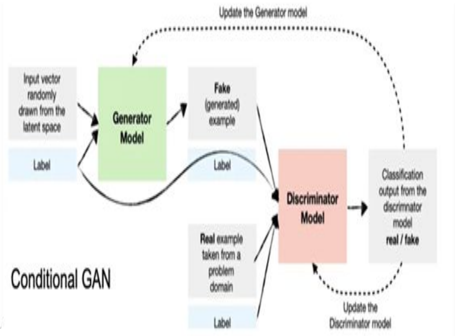
- MLGEFS global forecasts verification for 20240227 against GDAS analysis
- **MLGEFS CRPS score is higher than GEFSv12 in southern hemisphere, but is still lower than GEFSv12 in northern hemisphere**

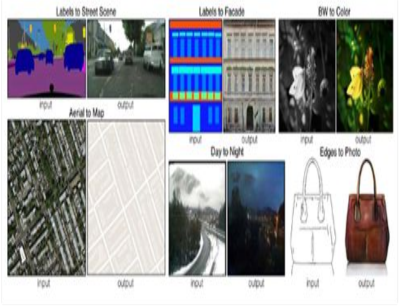


# Full Air-Quality Model Emulation DL Network - AQcGAN

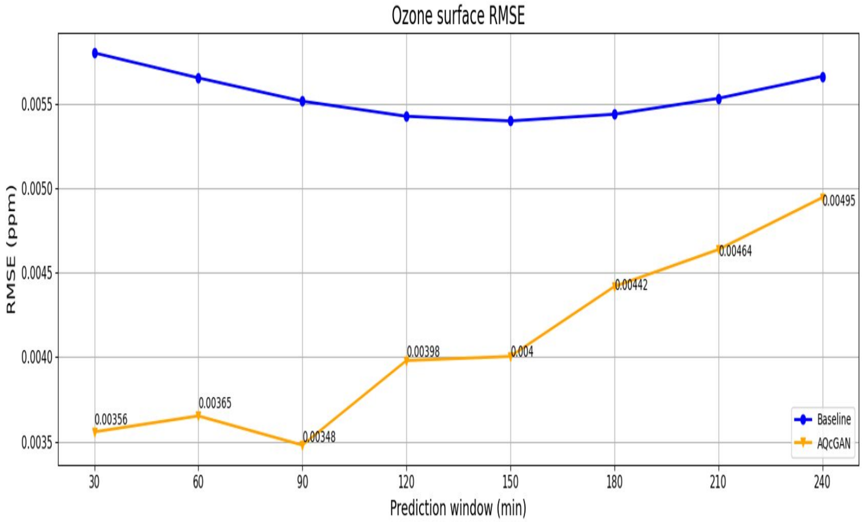
- Using a conditional GAN
  - Uses a sequence to sequence translation
  - Based on Pix2PIx architecture
- Based on Bihlo (2021)
- Adapting for Air Quality modeling
- **Forecasting “differences” through time**
- Introducing a 3$^{rd}$ dimension for vertical column



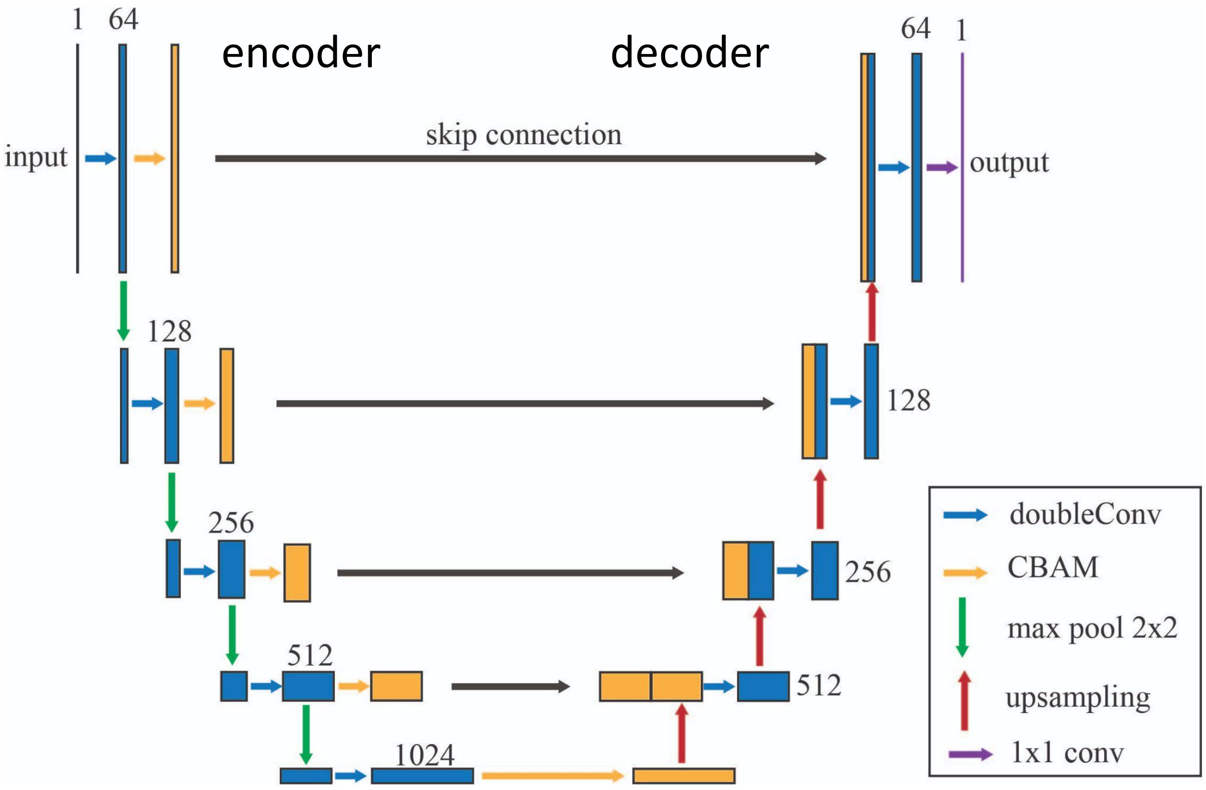

Based on Image-to-Image Translation (Isola 2017)

## O3 Forecast, North American




*Courtesy of Jennifer Sleeman (JPL)*

# ML for Bias Correction - model architecture (UNet)



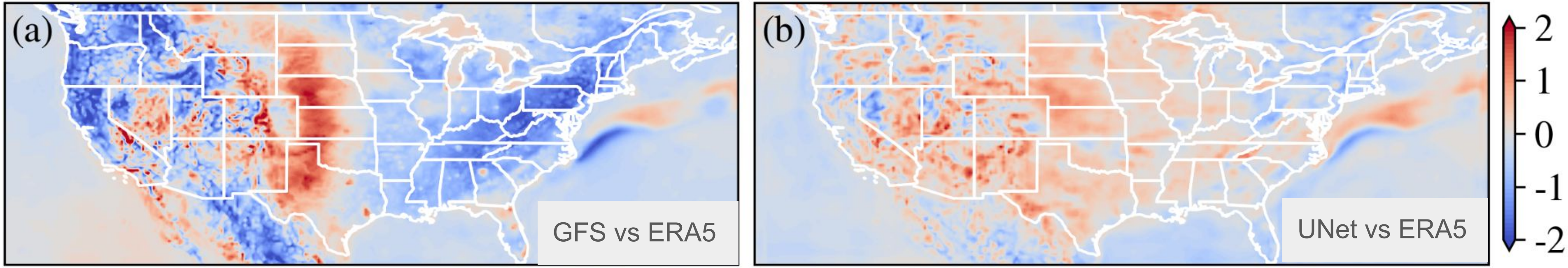

UNet with Convolutional Block Attention Module (CBAM):

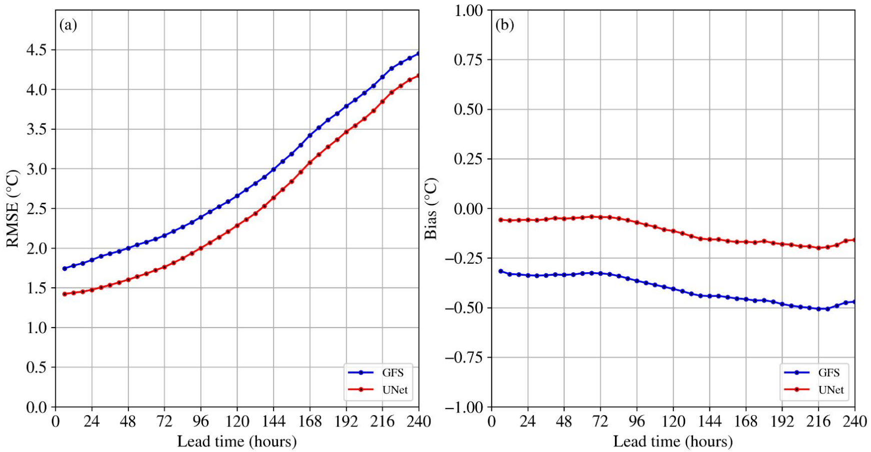
- doubleConv
  - 3X3 filter
  - Batch normalization
  - LeakyReLU
- Max pooling: 2X2

Attention Module:

- Identify important features across channels and spatial regions

# ML for Bias Correction - T2m

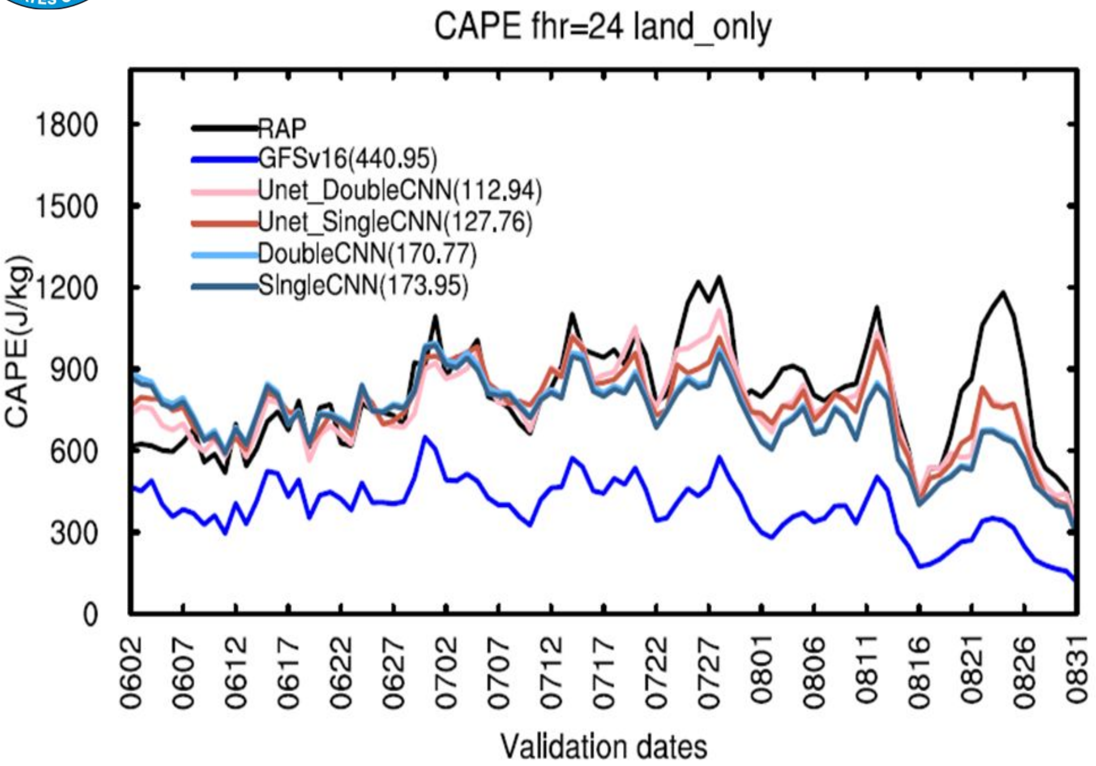


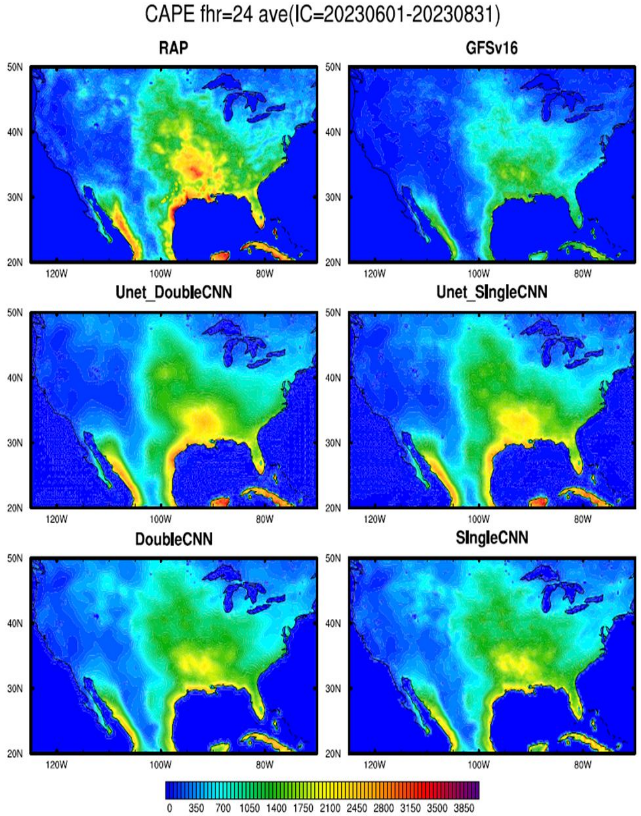

- Top: spatial distribution of five-month (January-May) mean bias of t2m for forecast hour 72, averaged over four cycles (00, 06, 12, 18).
- Bottom: domain-averaged RMSE and bias for forecast hour 6 to 240
- GFS shows cold biases over west and east CONUS, while warm biases over central CONUS.
- UNet effectively reduced both warm and cold biases.

Linlin Cui, Jun Wang

# ML for Bias Correction - CAPE




## Land: 2023 summer




Wei Li, Linlin Cui, Sadegh Tabas, Jun Wang, Fanglin Yang

# The End