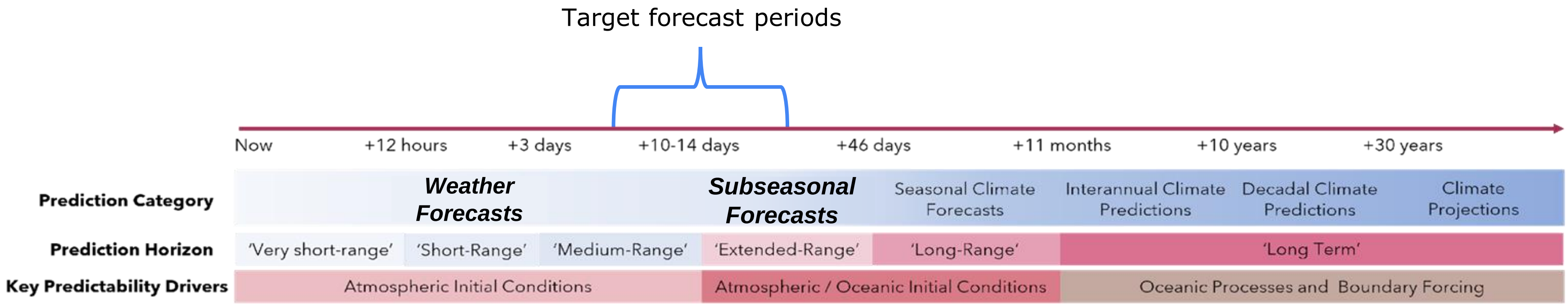
Training Program Forecast Targets: 1-4 Weeks Ahead



Planned Forecast Range for Training Program:

- **'Medium-Range' forecasts** (e.g. 10 days ahead): timeframe where AI-based methods have performed well
- **'Subseasonal' forecasts** (14 to 40+ days ahead): key decision-making window for agriculture and a known predictability challenge for all models. AI-based methods are improving over this window

Current model plan: cover **global models** needed for forecasts at medium-range and subseasonal timescales, including **deterministic and probabilistic outputs** (deterministic less useful after 2 weeks ahead)

Traditional Forecast Pipeline: Numerical Weather Prediction (NWP)

Approximate Current State of Earth System

Inputs vary, e.g. satellites, radar, ground sensors

Assimilate New Observations into Model

Additional observations can keep model more realistic

Post-process raw model outputs to correct biases

Machine learning statistics often already used to correct raw outputs

Generate Forecast

Correct Biases

Initial Approximation → *Updated Approximation*

Raw Forecast Output

Use Physical Equations to Model How Earth System Will Evolve

Standard governing equations to approximate motion and energy

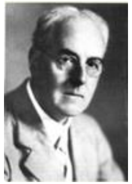
AIM FOR SCALE | WEATHER PACKAGE

Recent AI Model Developments Are Changing the Weather Forecasting Field

1st Revolution in Weather and Climate Prediction

2nd Revolution

1922
NWP Concept Tests



Lewis Richardson computes NWP forecast by hand
(6 days to calculate) - results fail in part due to initial condition limitations

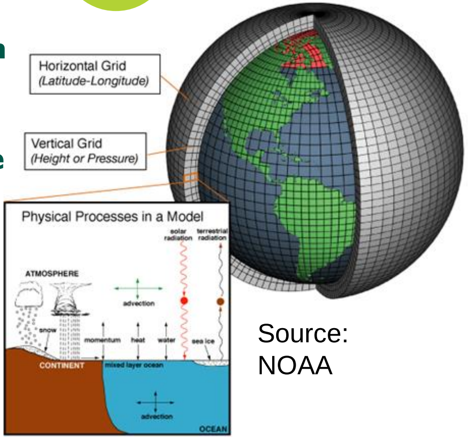
1950
Single layer computational model demonstrated

Single Variable Computational Forecast
(Charnay, Fjortoft and Neumann)

1980s
15-layer computational model run operationally at UK Met Office

Improvements in Supercomputers Advance NWP Forecast Pipeline

1990s
Supercomputers enable ensemble-based forecasting at multiple centers globally



Source: NOAA

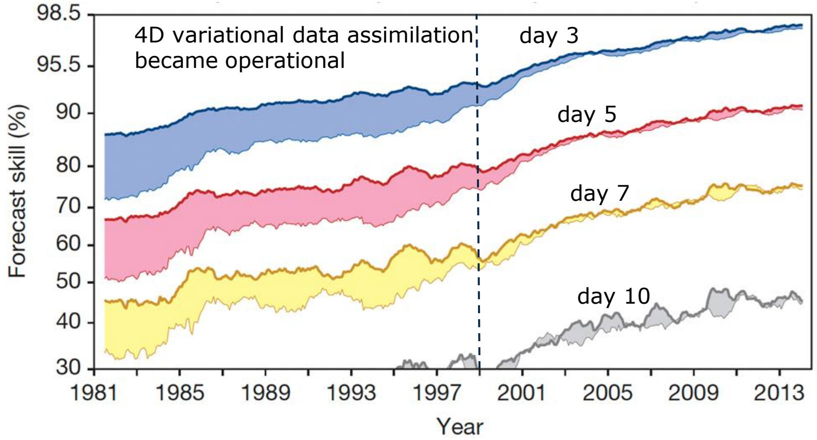
2023
Multiple groups show AI-based methods can compete with standard NWP approaches

Recent AI Advances Are Changing The Forecasting Paradigm

AIM FOR SCALE | WEATHER PACKAGE

Why have we shifted from incremental progress with computational power in NWP towards a new ‘revolution’ in forecasting with AI weather forecasts?

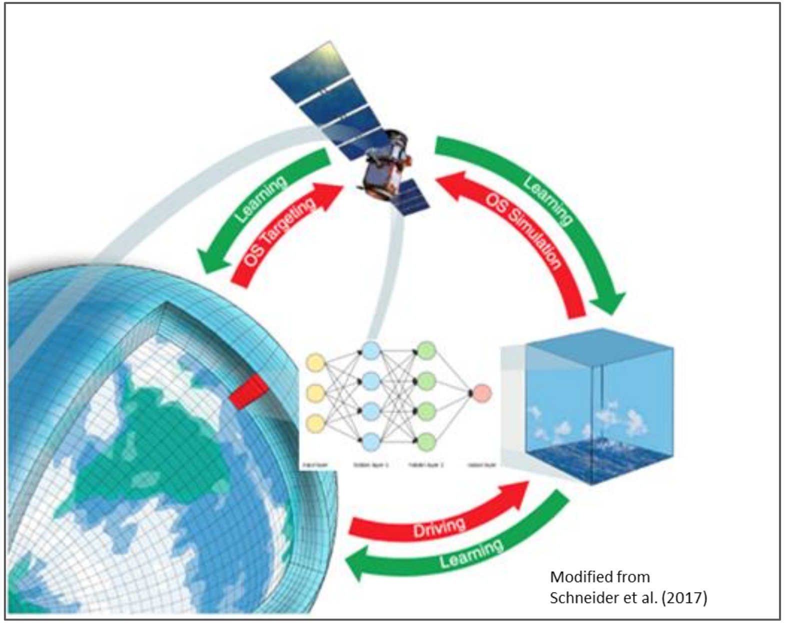
Progress in the forecast skill of European weather model



The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

Nature (2015)



Modified from
Schneider et al. (2017)

Recent AI Model Development Highlights

2023

2018

Scientific Tests of Modelling Possibilities

Dueben & Bauer (2018), Weyn et al. (2019, 2020, 2021)
Rasp & Theurey (2021), Clare et al. (2021)

2022

First Competitive Medium-Range Forecasts

NVDIA FourCastNet efficiencies vs. ECMWF IFS (Jan 2022)

Keisler GraphNN comparable to NOAA GFS (Jan 2022)

Huawei Pangu-Weather shows more accurate Tropical Cyclone tracks vs. IFS (Nov 2022)

Google Deepmind introduces GraphCast, comparable to IFS (Dec 2022)

2023

Operational Implementation

ECMWF introduces AIFS (2023) and releases experimental ensemble version (2024), running operationally alongside IFS as of 25 February 2025

Notes

- Model developments are advancing rapidly
- AI-based modelling methods vary (e.g. NeuralGCM, AIFS, GenCast) and perform differently at different timescales
- Operational centers are especially keen for feedback on where their models are working or not

AIM FOR SCALE | WEATHER PACKAGE

Data-driven prediction enables huge efficiency gains

FOURCASTNET: A GLOBAL DATA-DRIVEN HIGH-RESOLUTION WEATHER MODEL USING ADAPTIVE FOURIER NEURAL OPERATORS

A PREPRINT

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109

Ashesh Chattopadhyay
Rice University
Houston, TX 77005

Morteza Mardani
NVIDIA Corporation
Santa Clara, CA 95051

Thorsten Kurth
NVIDIA Corporation
Santa Clara, CA 95051

David Hall
NVIDIA Corporation
Santa Clara, CA 95051

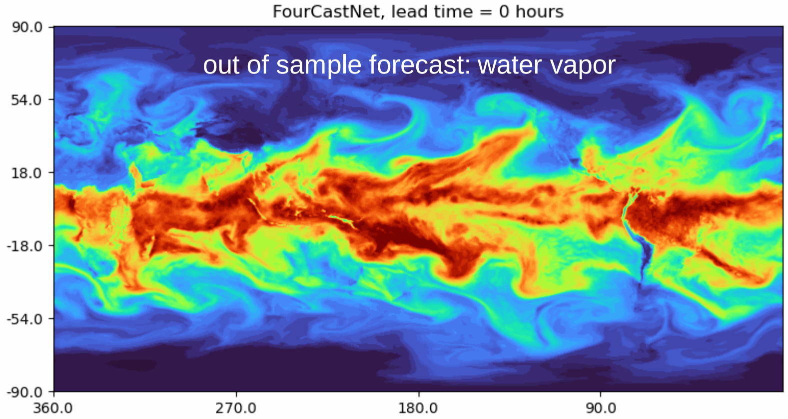
Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

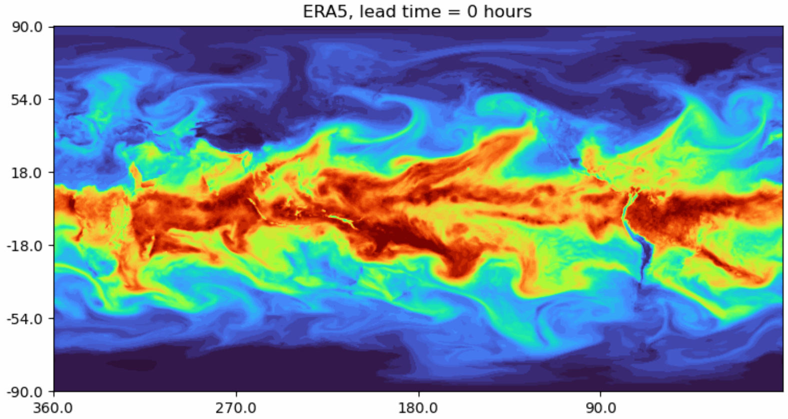
Kamyar Azizzadenesheli
Purdue University
West Lafayette, IN 47907

Pedram Hassanzadeh
Rice University
Houston, TX 77005

Karthik Kashinath
NVIDIA Corporation
Santa Clara, CA 95051

Animashree Anandkumar
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051





Trained only on
33 variables of 1979-2017 25km observation-derived data (ERA5)
O(10^5) x faster than the best numerical weather model (ECMWF's IFS)

Anomaly Correlation Coefficient (ACC) a common metric to measure model performance in AI weather forecasting community

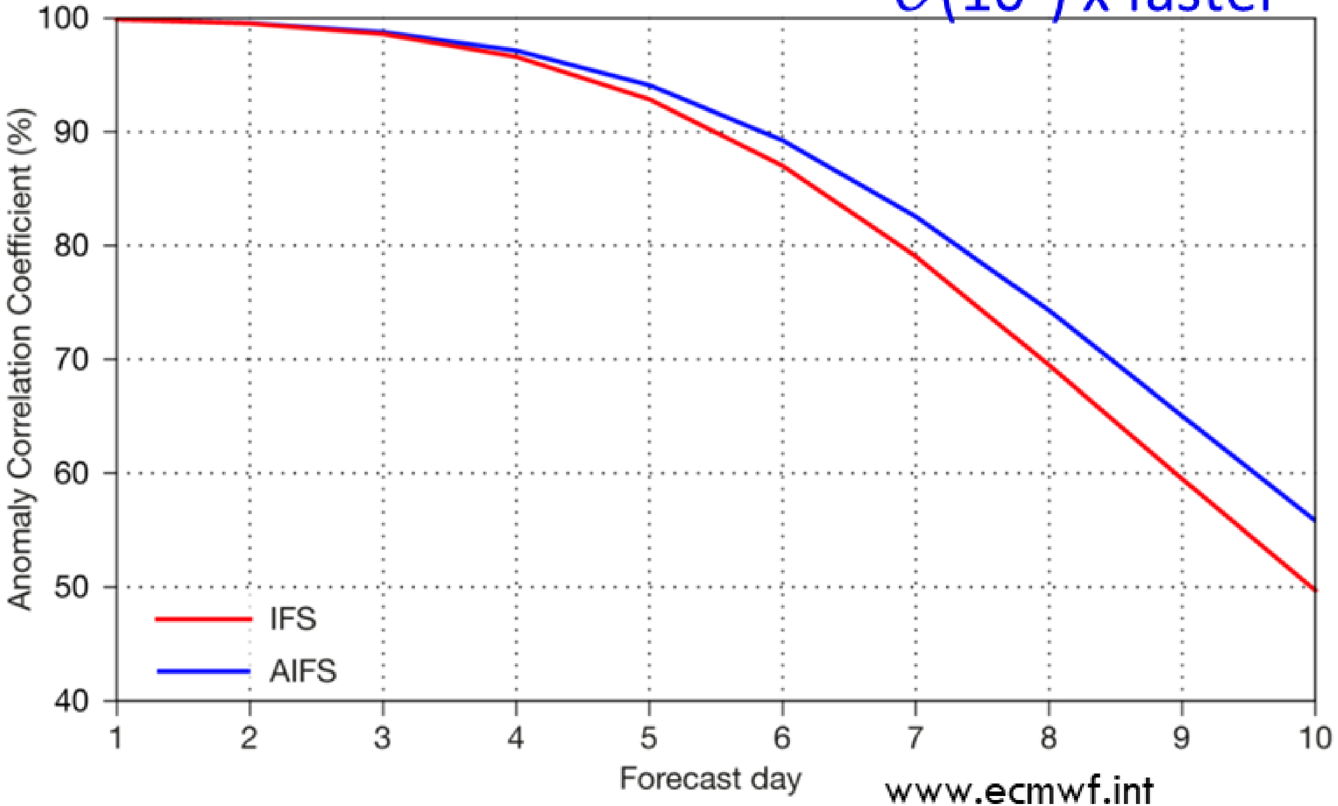
What are anomalies? departures from a long-term average (what we often call a climatology)

ACC measures how well the predicted anomalies of a forecast match these anomalies

Generally a good indicator for measuring **deterministic deviations from normal** - e.g. unusual temperature, and **directional change** - e.g. forecast anomalies vs observed anomalies

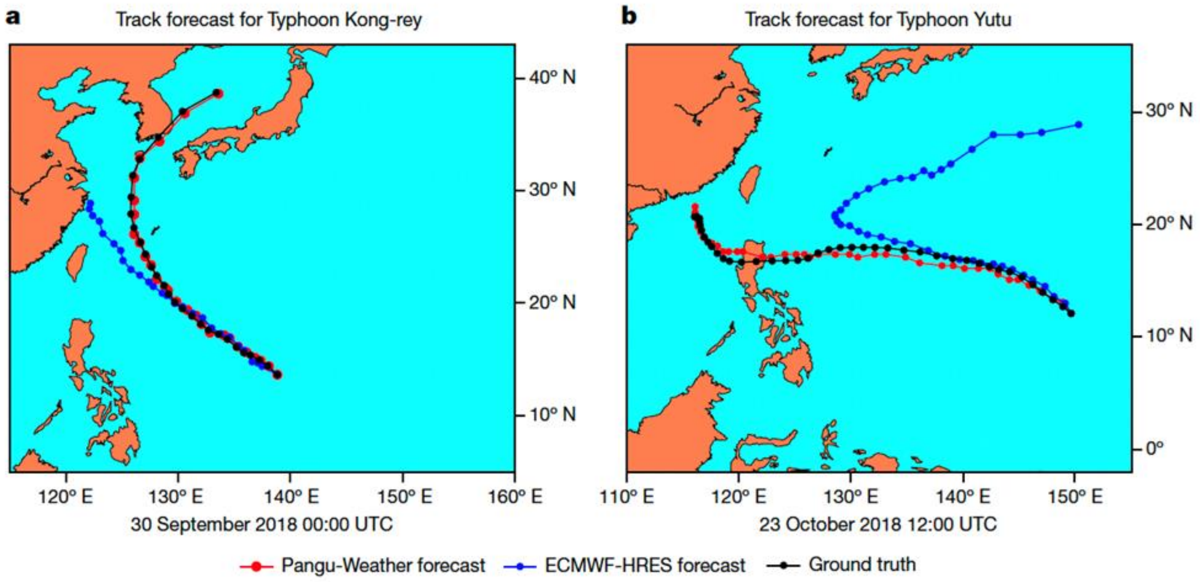
Efficiency and accuracy advances demonstrated with AIFS (ECMWF deterministic model) using ACC on Z500

$\mathcal{O}(10^5)$ x faster



www.ecmwf.int

Accuracy opportunities also demonstrated for extreme events like tropical cycle track predictions show potential



HUAWEI

Pangu Weather AI Forecasts: Bi et al. (2023) Nature - 200M parameters (transformer model) show Pangu captured 2018 Typhoons Kong-rey and Yutu tracks better than ECMWF HRES

Integration of AI into Forecast Pipeline Offers Several Opportunities, Including Efficiency, Accuracy, & Autonomy

Current Value Add from AI

Generate Forecast

Current Value Add from AI

Calibrate probabilities and downscaling

Initial Approximation/ Assimilation Steps (often ERA5 today)

Opportunities include:

- **Efficiency:** AI models can speed up the initial forecast runs significantly and prove more cost effective
- **Accuracy:** AI models can ingest multiple types of data and train on historical data to improve forecast accuracy
- **Autonomy:** Once an AI model is trained, it can be run locally on a desktop computer

Common machine learning benchmarking frameworks

Why benchmarking?

In the past 5 years, AI weather prediction models have **surpassed the skill** of leading numerical weather prediction (NWP) models using some scientific benchmarking standards.

These AI weather prediction models are **wildly** cheaper than NWP models.

AI model: 3 min on 1 A40 GPU - **$3K USD** ***(online GPU usage even cheaper)***

IFS 9km: 1 hour on 12500 CPUs of Cray XC40 supercomputer - **$125M USD**

Are these advances **useful** for operational decisions in an agricultural context?

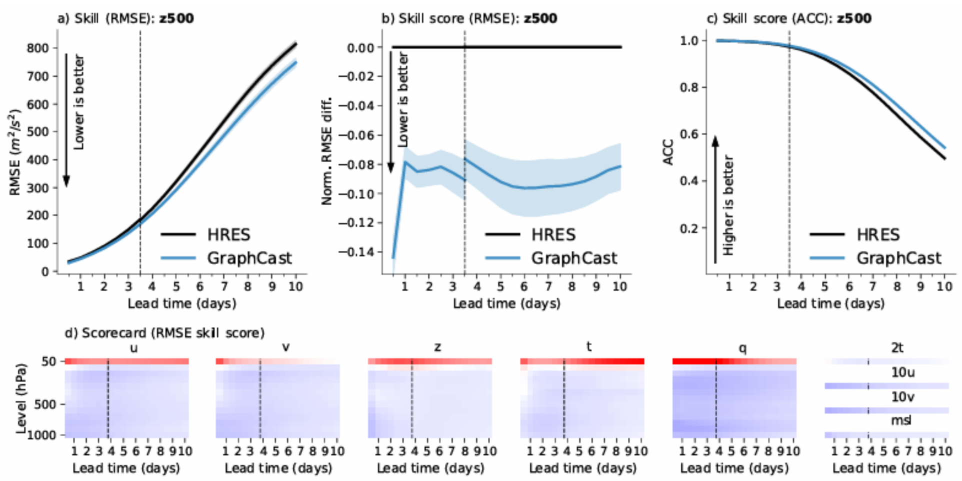
WeatherBench 2

		2m Temperature RMSE [K]					Precipitation 24h precipitation SEEPS				
	Lead time [days]	1	3	5	7	10	1	3	5	7	10
Physical models	IFS HRES	0.51	0.87	1.33	1.88	2.56	0.20	0.33	0.46	0.60	0.76
Physical models	IFS ENS (mean)	0.56	0.86	1.20	1.58	2.00	0.21	0.35	0.49	0.63	0.77
Physical models	ERA5-Forecasts	0.71	1.06	1.51	2.06	2.70					
ML / hybrid models	Pangu-Weather (oper.)	0.58	0.95	1.40	1.95	2.64					
ML / hybrid models	GraphCast (oper.)	0.44	0.75	1.18	1.73	2.43	0.21	0.32	0.43	0.56	0.72
ML / hybrid models	GenCast (oper.) (mean)	0.44	0.75	1.13	1.54	1.98	0.20	0.31	0.44	0.58	0.74
ML / hybrid models	Pangu-Weather	0.56	0.91	1.39	1.97	2.68					
ML / hybrid models	GraphCast	0.51	0.82	1.25	1.81	2.56	0.21	0.50	0.60	0.69	0.81
ML / hybrid models	FuXi	0.53	0.84	1.27	1.65	2.17	0.14	0.31	0.44	0.66	0.79
ML / hybrid models	GenCast (mean)	0.49	0.80	1.16	1.57	2.02	0.14	0.29	0.43	0.57	0.73
ML / hybrid models	Stormer ENS (mean)	0.66	0.93	1.30	1.74	2.23					
ML / hybrid models	Excarta (HEAL-ViT)	0.55	0.84	1.27	1.83	2.56					
ML / hybrid models	ArchesWeather-Mx4	0.51	0.83	1.27	1.77	2.32					
ML / hybrid models	ArchesWeatherGen (mean)	0.54	0.83	1.20	1.60	2.04					
	Climatology	2.61	2.61	2.61	2.61	2.61	0.95	0.95	0.95	0.95	0.95

−50 −20 −10 −5 −2 −1 1 2 5 10 20 50

Better ⟵ % difference in RMSE/SEEPS vs IFS HRES ⟶ Worse

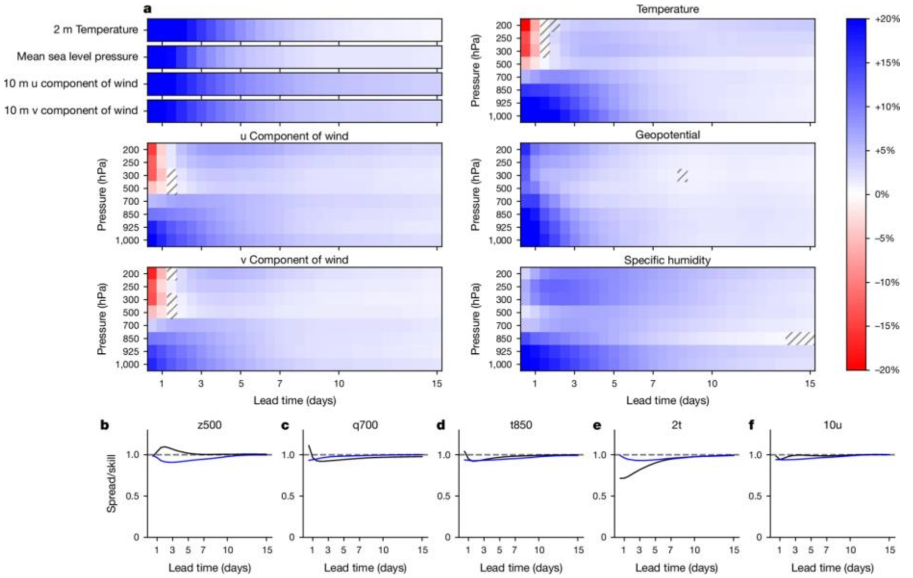
Graphcast demonstrates skill relative to ECMWF HRES



Graphcast

Lam et al. 2023 (Science): *300M parameters (graph neural nets)*

Gencast shows skill relative to ECMWF ENS



Gencast

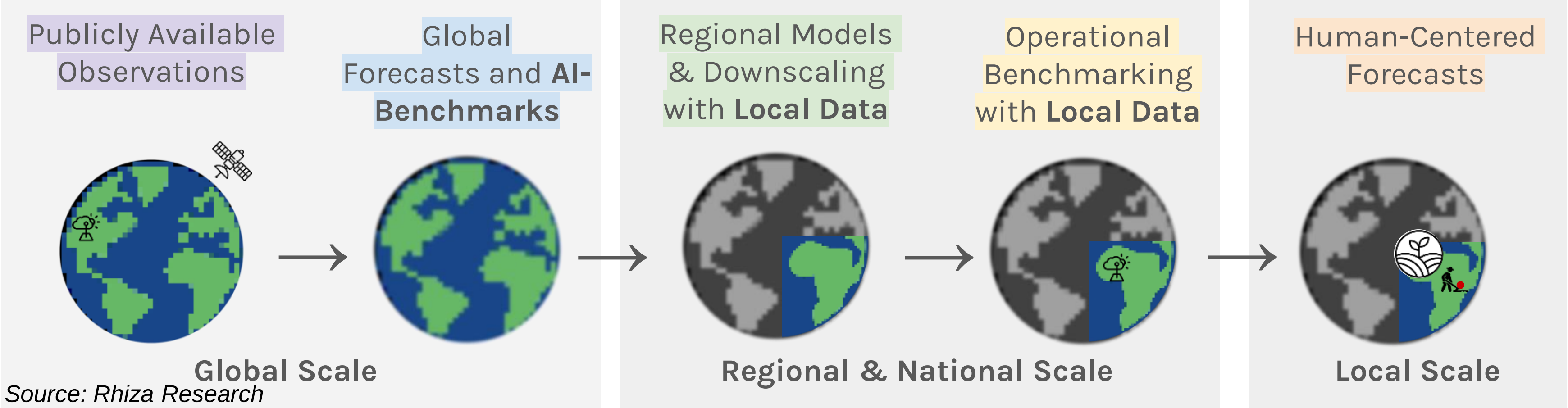
Price et al. 2024 (Nature): (conditional diffusion model)

Google DeepMind

Scientific benchmarking for data-driven models often prioritizes - average error and correlation metrics, global estimates, and 'high-impact' prediction tasks like TC tracks

Benchmarking has two theories of change

1. **AI benchmarking:** Using benchmarks to drive breakthroughs in AI-based forecasting skill on high-impact prediction tasks.
2. **Operational benchmarking / verification (our focus):** Helping NMHSs decide which novel forecasting models and outputs (if any) to operationalize.



Source: Rhiza Research

Key Need for Training Programs – Flexible Tool Agnostic Capacity

Awareness of the Latest Models and Methods

- The weather forecasting field is changing rapidly and training can help keep up with latest advances

Understand Limits of Conventional NWP and new AI Weather Forecasts

- AI-based models are data-driven – if there is limited data to train on, results will be impacted
- AI-based models do not currently provide a replacement for NWP models
- Extreme weather events need careful consideration for model applications

Evaluation is Critical For Effective Use

- Comparisons will have different results depending on the models, metrics, and ground-truth data used
- Changing climate can affect predictability, making frequent model evaluation even more critical

Integration of Weather Forecasts into Operational Decision-Making Remains a Significant Challenge

- Bringing people together across disciplines to further strategize how to operationalize weather information and make it effective for decision-making is a significant need

Building a community of practice

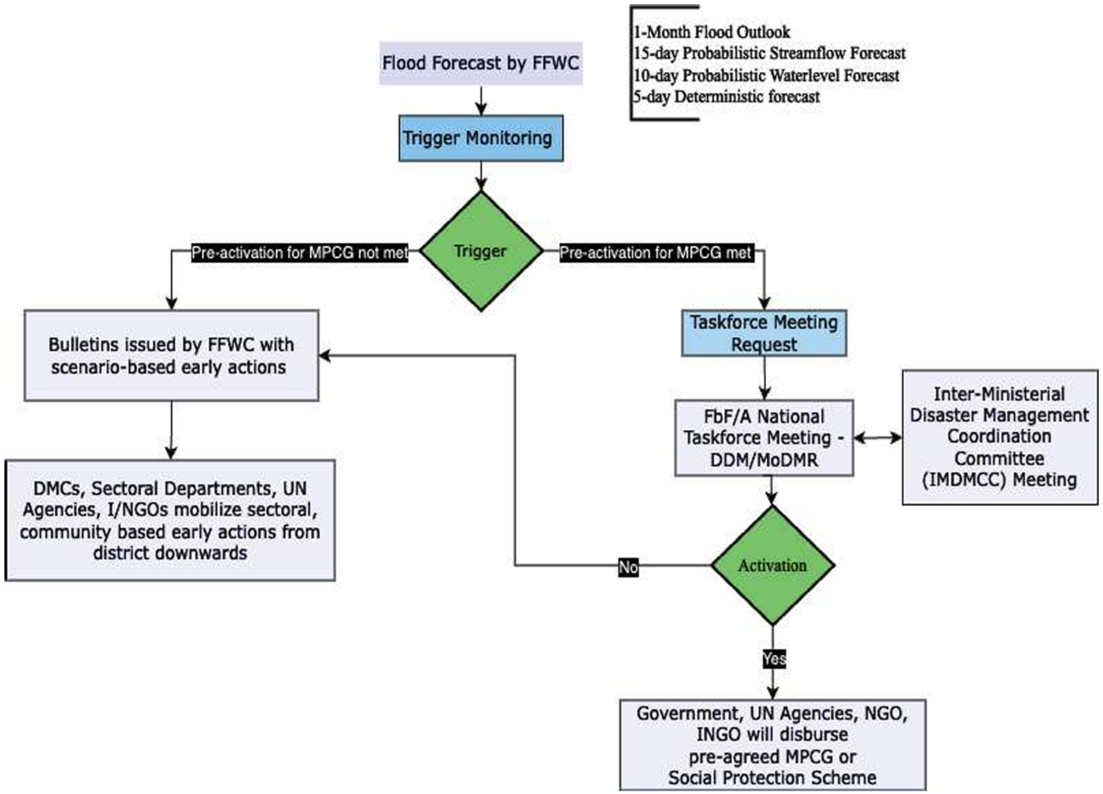
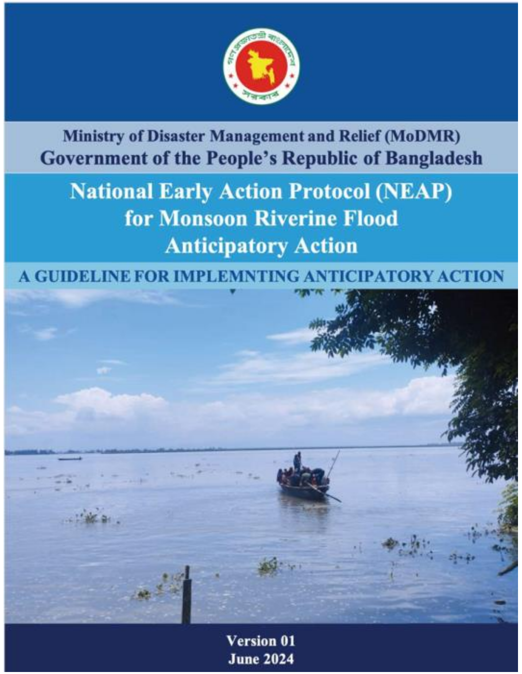


Figure 4: Trigger Activation Process



How can we learn from protocols in other countries and disciplines to improve action oriented forecasts for agriculture?