AI Weather Quest ECMWF

Enhancing sub-seasonal predictions with AI/ML: A competition by ECMWF, endorsed by WMO

SON Awards Webinar

ECMWF

Destination Earth

WORLD METEOROLOGICAL ORGANIZATION

Agenda

- **Presentations**
  - ✓ Introduction by Matthew Chantry
  - ✓ SON Period participation overview
  - ✓ SON Period forecast evaluation
  - ✓ Presentation from first top-performing team MicroEnsemble
  - ✓ Presentation from second top-performing team CMAandFDU
  - ✓ Presentation from outstanding team Fahamu
  - ✓ Key milestones and actions

This session is being recorded.
The recording will be made available online after the webinar. If you do not wish to appear, please turn off your camera.

Please mute your microphone.
Please keep yourselves muted during presentations. You are welcome to take the floor during the Q&As or ask questions in the chat.

Introduction by Matthew Chantry

Strategic Lead for Machine Learning, ECMWF

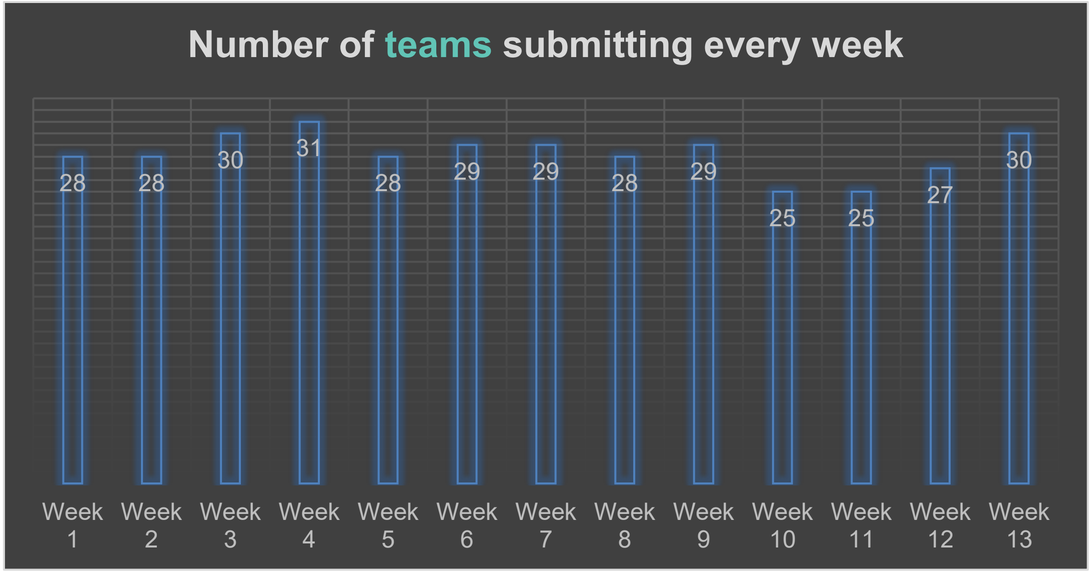
SON Period participation overview

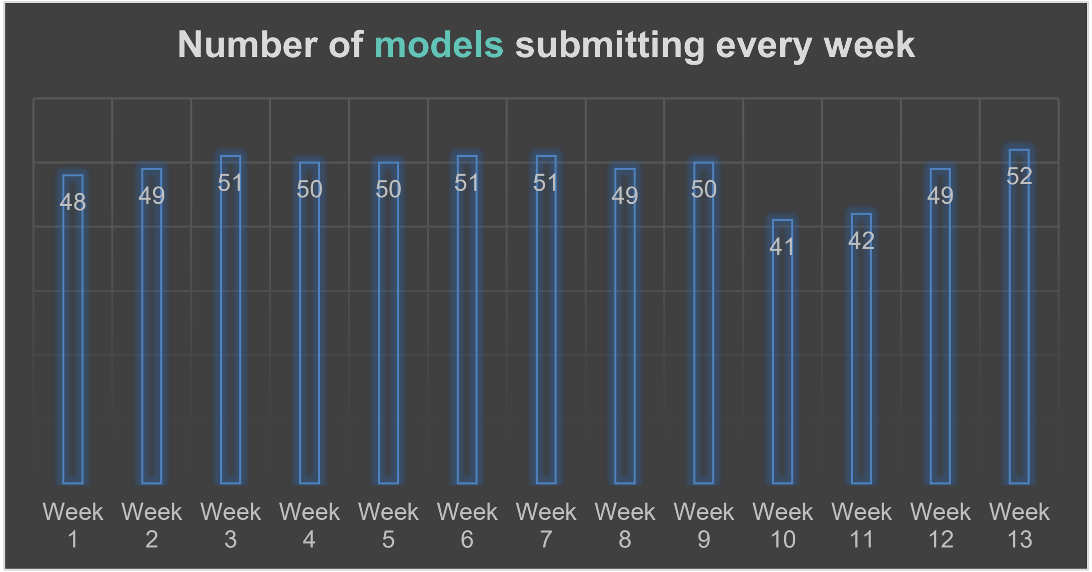
Accumulated participation over SON period

166
Number of competitors to the Quest

35
Number of teams

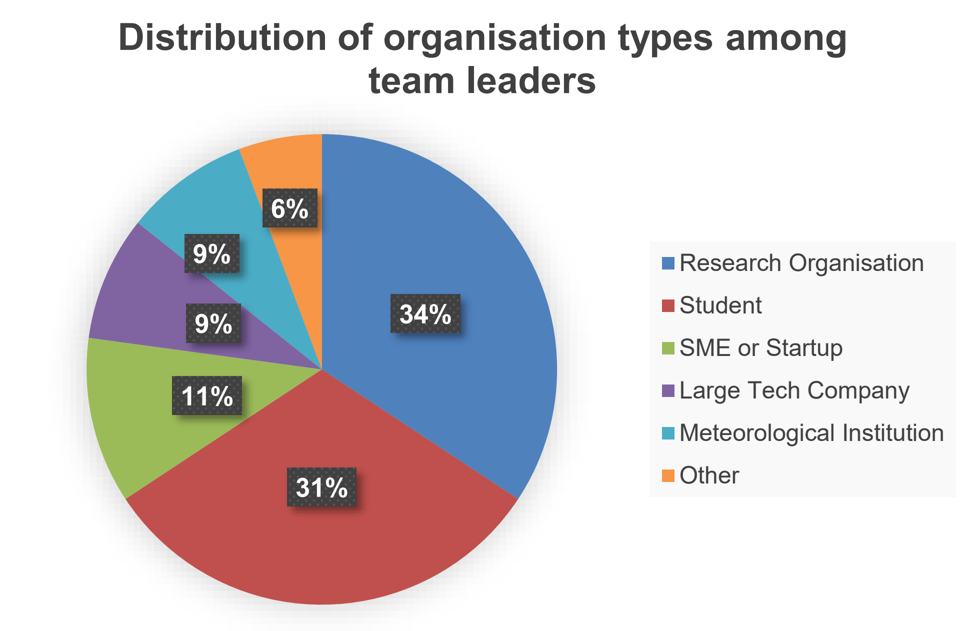
60
Number of models

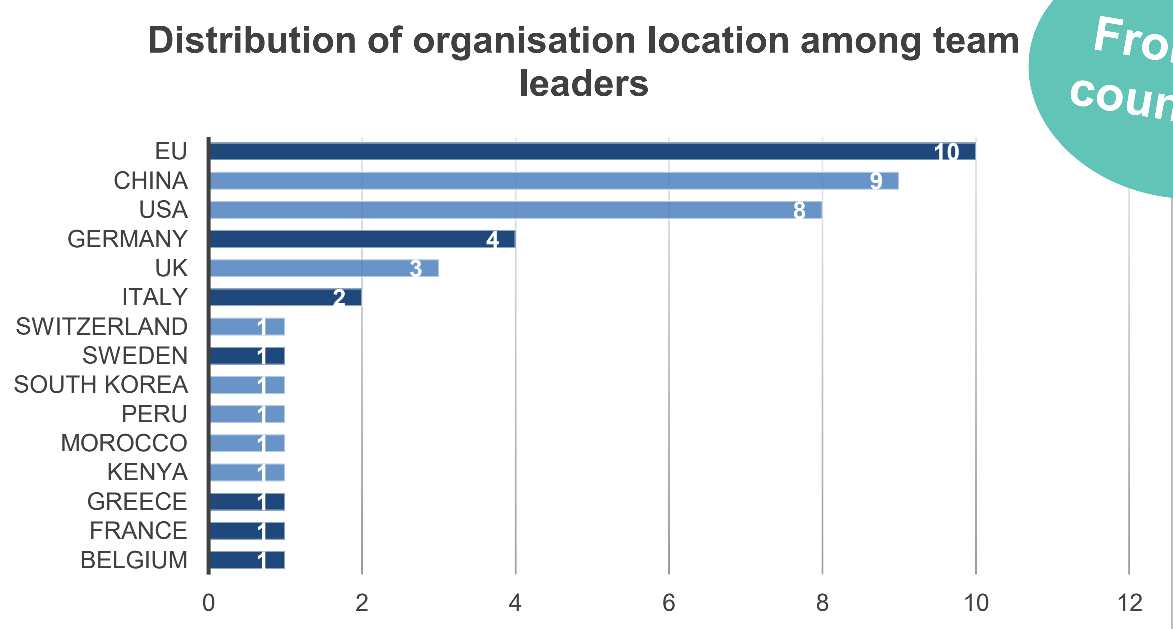




ECMWF Destination Earth WORLD METEOROLOGICAL ORGANIZATION

SON Period participation overview





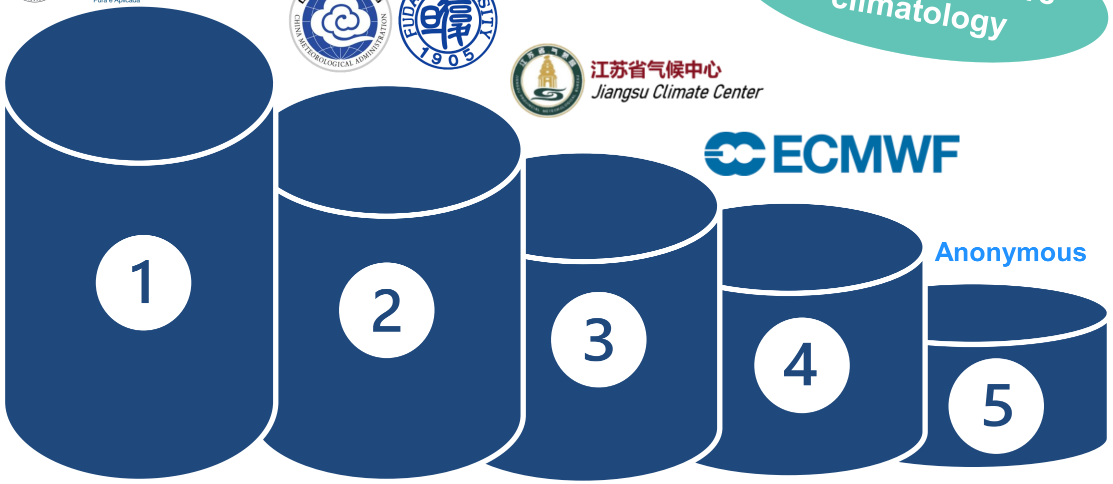
From 14 countries

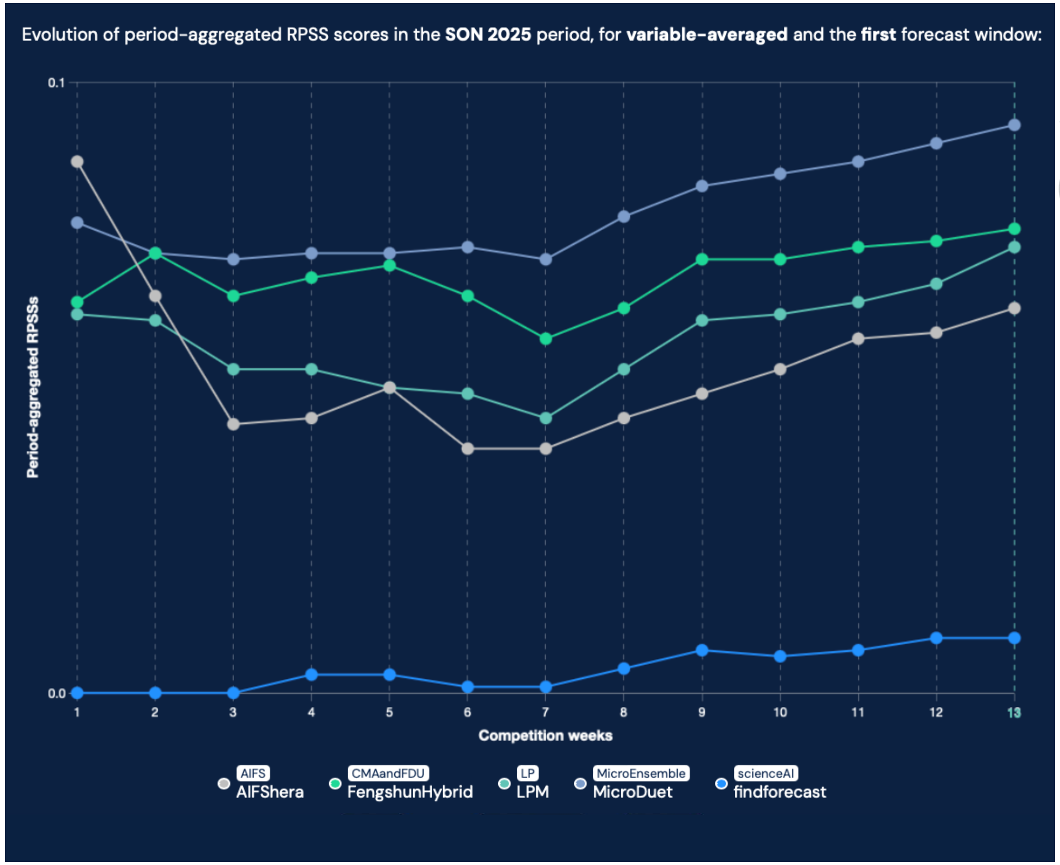
Affiliations of SON teams that accepted identities public display

SON Period forecast evaluation: Top-performers overview

13 Teams eligible for variable-averaged, period-aggregated scores.

Five teams above climatology





Anonymous

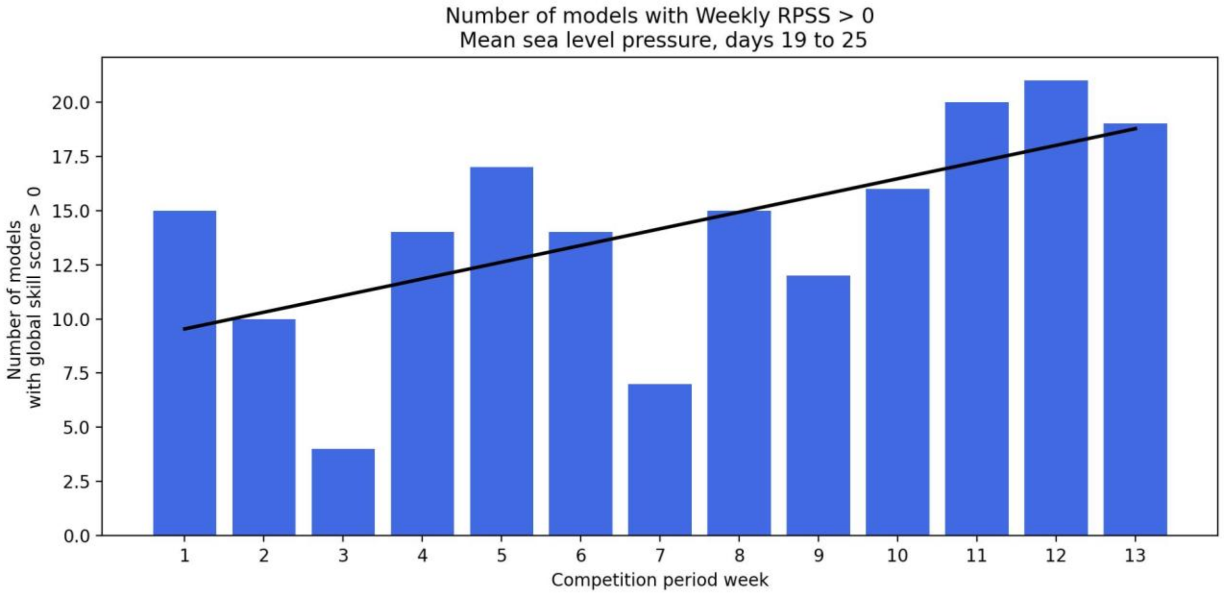
Global South participation

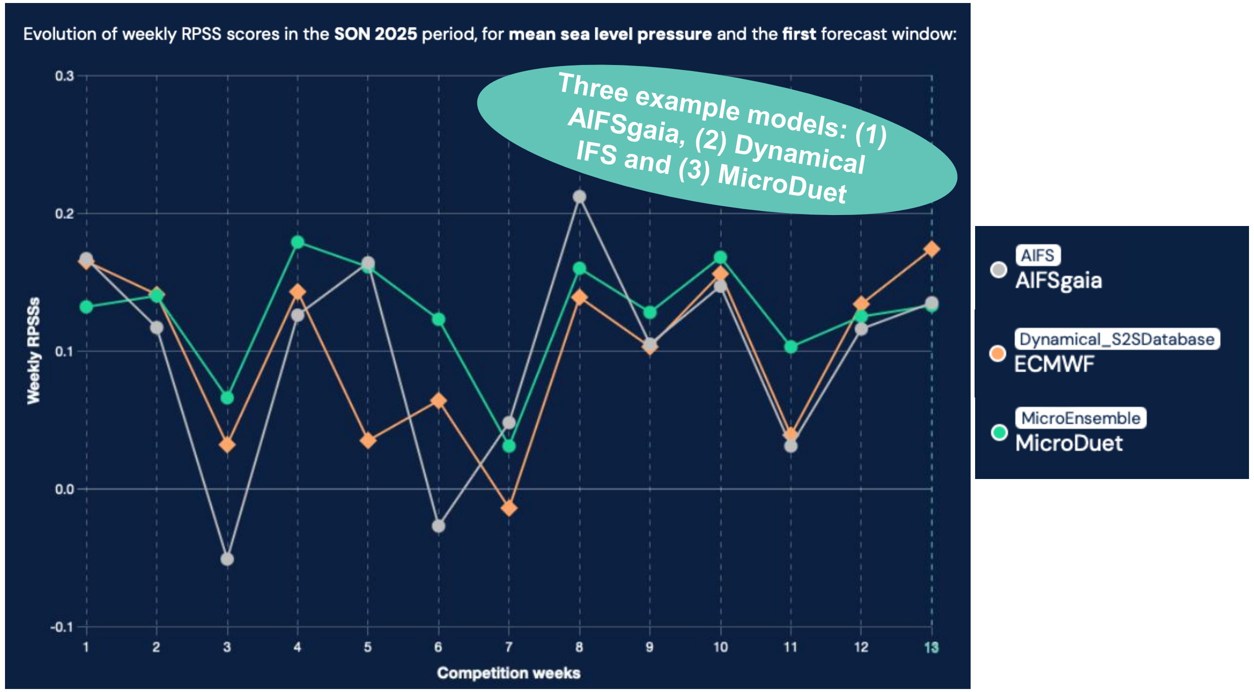
Universidad Nacional Mayor de SAN MARCOS Universidad del Perú. Decana de América + ICPAC

Two teams from Global South.

Team Fahamu (ICPAC, East African RCC) will be spotlighted due to their skilful temperature and mean sea level pressure predictions.

SON Period forecast evaluation: A platform for model development





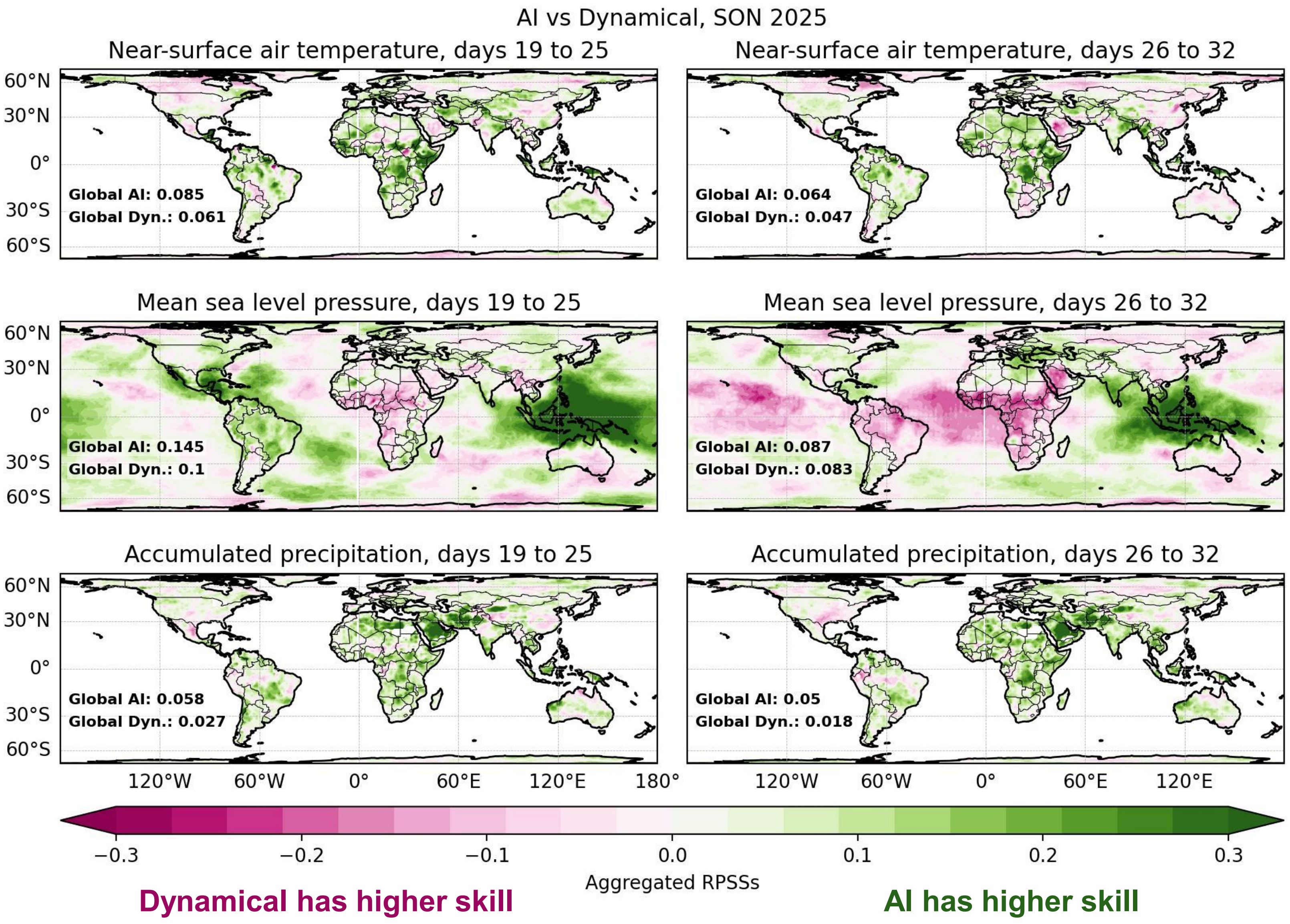
For all variables, the number of models performing better than climatology has increased throughout the SON period.

Skill scores for three consistent models remained stable throughout SON. The larger number of models now surpassing climatology is attributable to model development, not to any increase in the inherent predictability.

It is expected that DJF will be more competitive

+JR team (individuals)
+2 anonymous

SON Period forecast evaluation: Comparing AI & dynamical models



Differences in period-aggregated scores for all variables and lead times

+1 anonymous

AI model-mean *minus* Dynamical model-mean

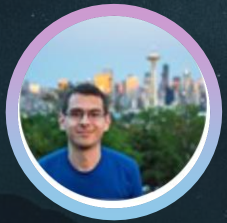
Initial conclusions

- For all variables and lead times, AI outperforms dynamical regarding global skill scores.
- Large improvements seen in tropical regions. In particular, pressure forecasts over Maritime Continent.
- We find a dipole in enhanced/degraded tropical skill for pressure predictions. MJO influence?

Presentation by MicroEnsemble

Best ranked-team of the SON Period for variable-averaged, period-aggregated scores, for both 1st and 2nd forecast windows

MicroEnsemble



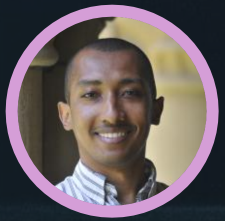
Jonathan Weyn
Microsoft

Hannah Guan
Harvard University

Soukayna Mouatadid
University of Toronto

Paulo Orenstein
Instituto de Matemática Pura e Aplicada

Judah Cohen
MIT



Lester Mackey
Microsoft Research

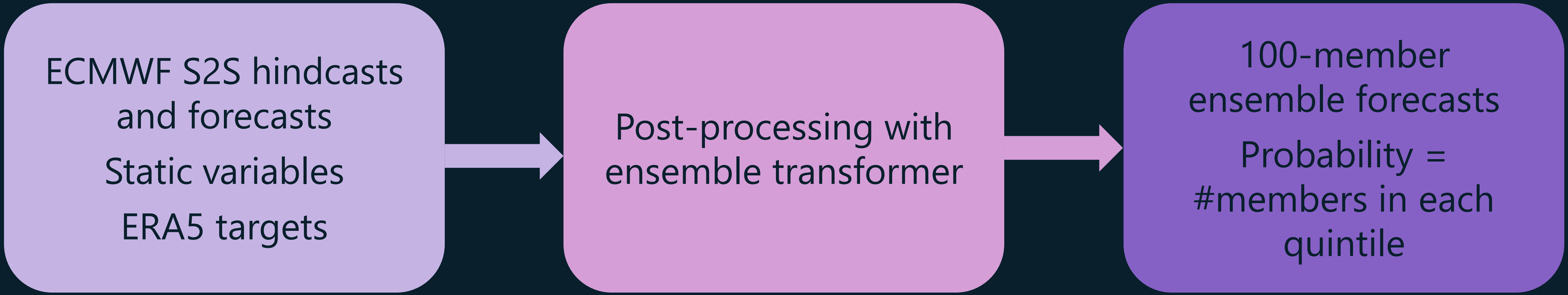
Alex Lu
Microsoft Research

Genevieve Flaspohler
Rhiza Research

Zekun Ni
Microsoft

Haiyu Dong
Microsoft

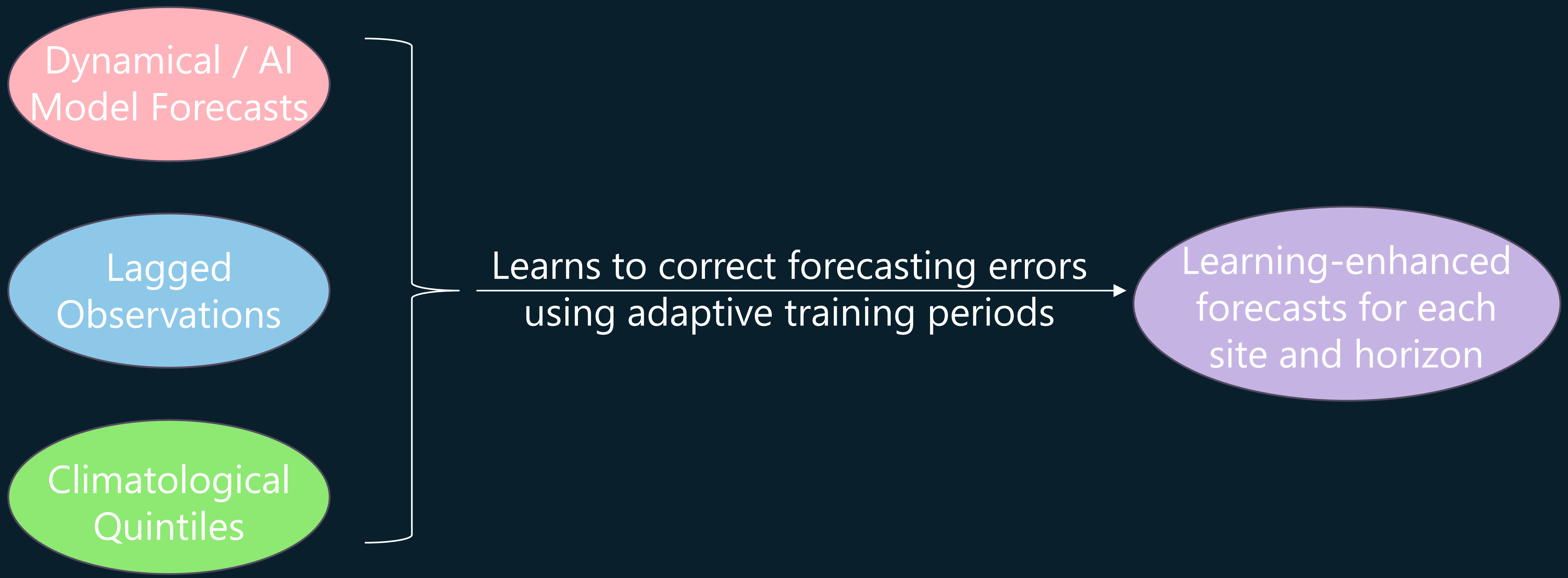
“Huracan” aka “PoET”



- Improving Medium-Range Ensemble Weather Forecasts with Hierarchical Ensemble Transformers (2024), a.k.a. “PoET”
- Inputs consist of all 85 parameters available in S2S plus 37 prescribed static variables
- 10-member perturbed hindcasts for training
- The model is agnostic to ensemble size so inference is done on the 100-member operational ensemble

RPSS (2024 test forecasts)	PoET	ECMWF (raw)
tas – day 19	0.1886	-0.0761
tas – day 26	0.1670	-0.0982
mslp – day 19	0.0979	-0.0899
mslp – day 26	0.0659	-0.1173
pr – day 19	0.0404*	0.0308
pr – day 26	0.0281*	0.0157

Probabilistic Bias Correction (PBC)



- Inspired by Adaptive bias correction for improved subseasonal forecasting (Mouatadid et al., Nature Communications, 2023)

StillLearning Seasonal RPSS (2016-2024)

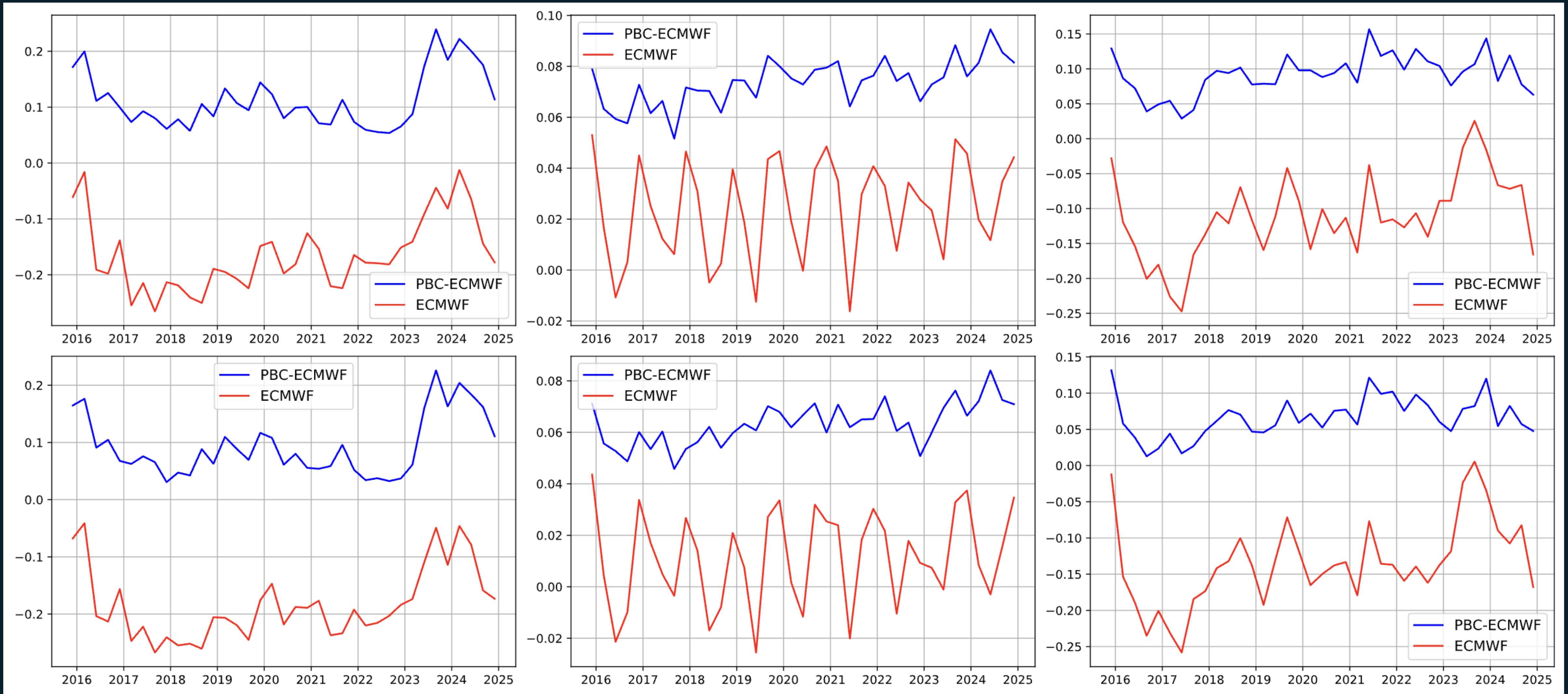
Temperature

Precipitation

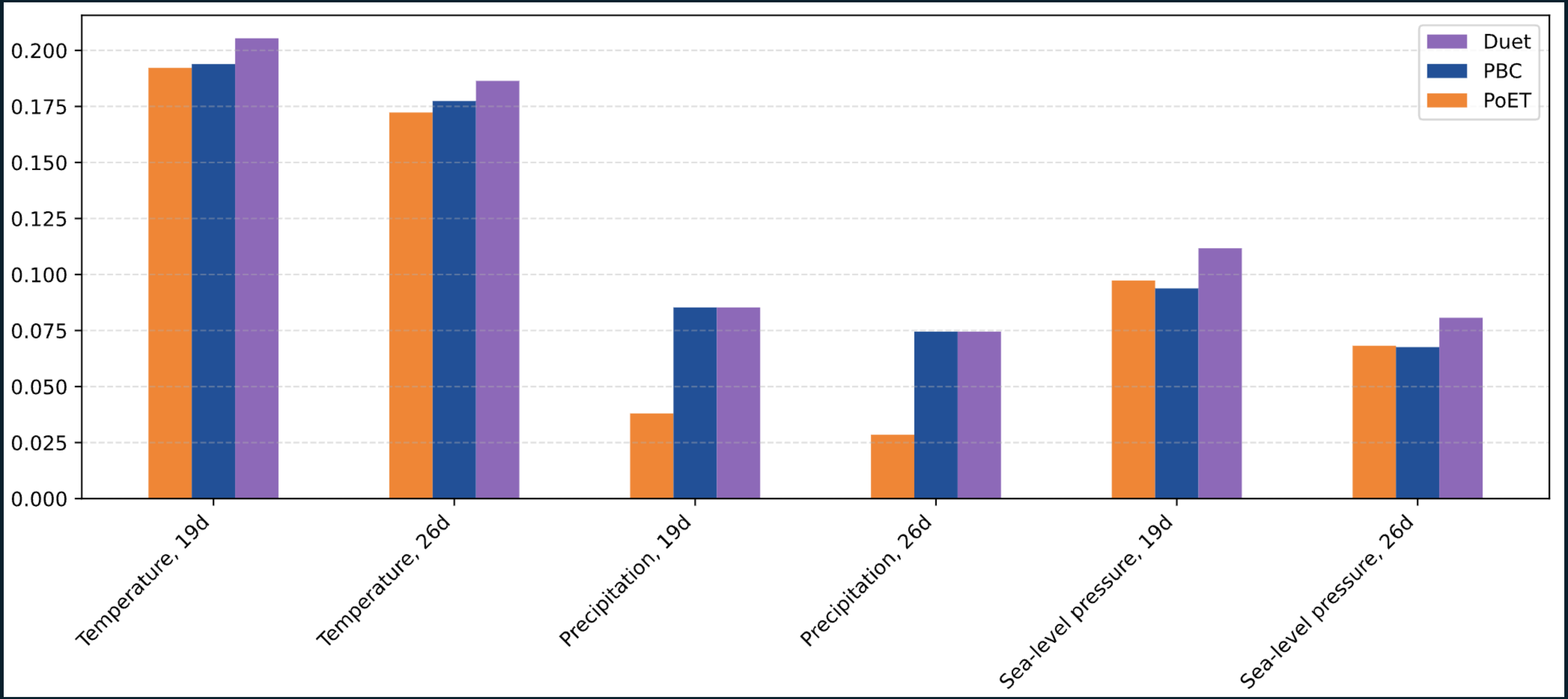
Sea-Level Pressure

Horizon 19

Horizon 26



Duet RPSS (2024)



Duet combines the strengths of PBC and PoET

Duet Extreme Temperature Forecast

WEATHER Cold Waves Add Topic +

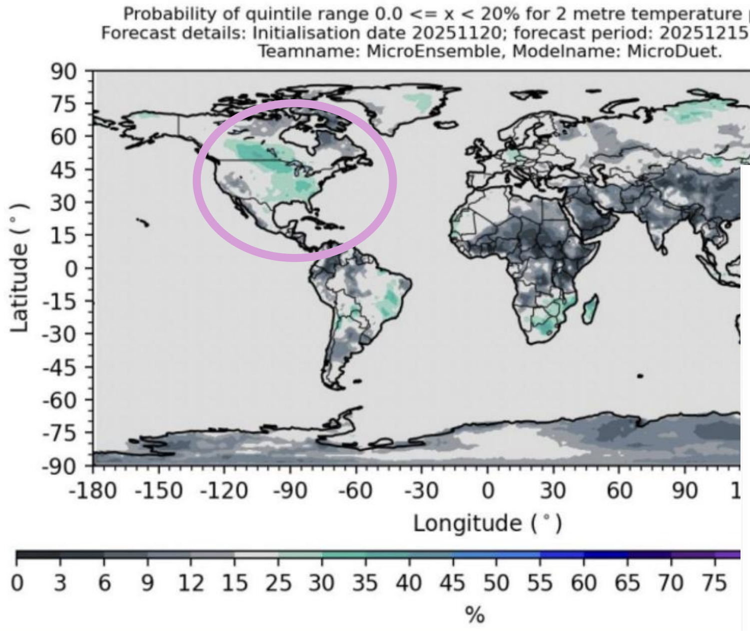
Some of Earth's most extreme cold may be headed for the US in December

Some forecasters say a complex dance involving the polar vortex could send some of Earth's most extreme cold toward the United States.

Doyle Rice
USA TODAY
Updated Nov. 27, 2025, 8:06 a.m. ET

Judah Cohen @judah47 · Nov 23

Lots of hype about the #PolarVortex but is it deserved? Our AI subseasonal model, which I can credibly claim as world's best is predicting that the most expansive region of most likely extreme #cold on earth stretches from the Canadian Plains to the US East Coast 3rd week of Dec.



Lester Mackey

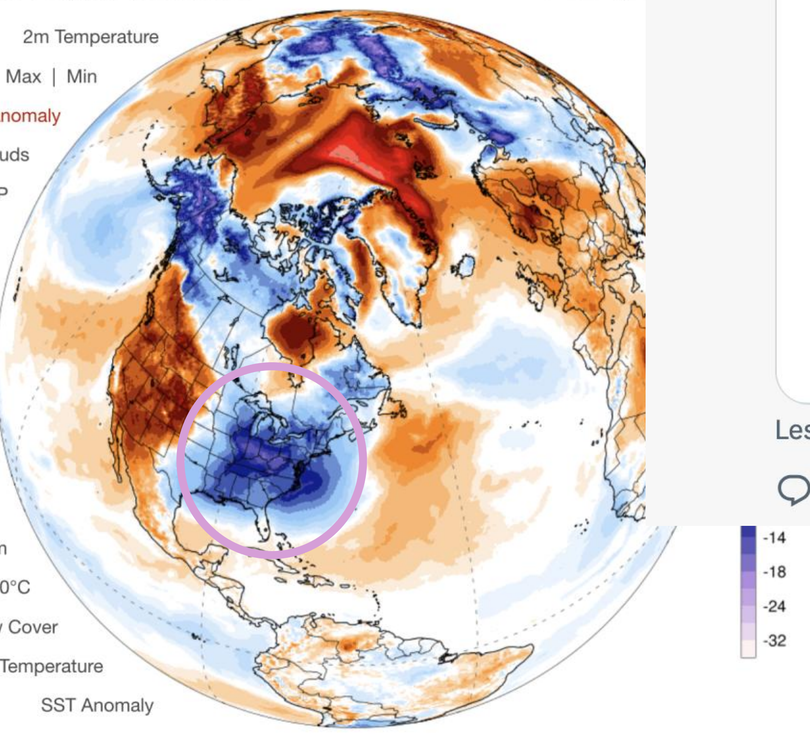
31 63 297 65K

About this page

GFS 2m T Anomaly (°C) [CFSR 1979-2000 baseline]
1-day Avg | Mon, Dec 15, 2025

ClimateR
Climate Change In

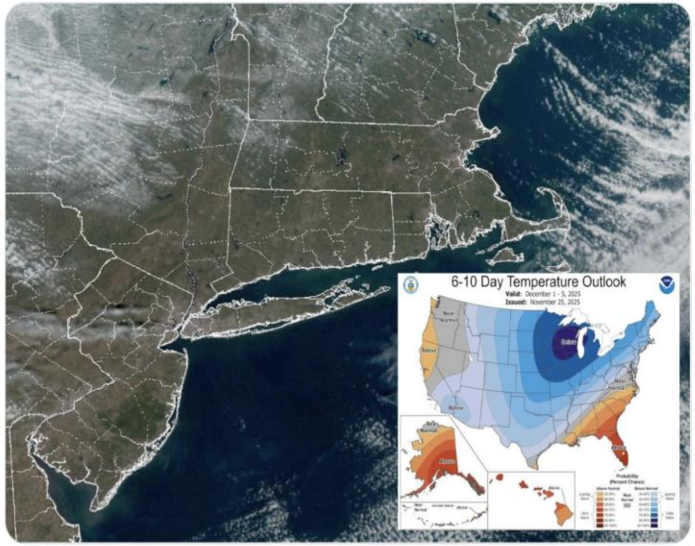
2m Temperature
Avg | Clim | Max | Min
2m Temp Anomaly
Precipitation / Clouds
Precipitation / MSLP
Precipitable Water
PWtr Sd Anomaly
10m Wind Speed
Mean SL Pressure
MSLP Sd Anomaly
500hPa Geopot. Hgt
500hPa GPH Anom
500hPa Jetstream
250hPa Jetstream
Snow Depth / T 0°C
Sea Ice / Snow Cover
Sea Surface Temperature
SST Anomaly



U.S. Natural Gas Surges to 3-Year High as Deep Freeze Hits December

New York Post @nypost

Northeast set to be blasted by 'most extreme cold on Earth' before Christmas trib.al/skHrwSX



1:09 PM · Dec 1, 2025 · **10.5M** Views

Presentation by CMAandFDU

Second best ranked-team of the SON Period for variable-averaged, period-aggregated scores, for both 1st and 2nd forecast windows

A Brief Introduction to Fengshun Series Models by CMAandFDU

Team Leaders： Bo Lu & Hao Li

Team Members: Z. Dou, L. Chen, Y. Zhao, X. Zhong, J. Hu, Q. Qian, J. Wu, C. Zhao, C. Zhou, C. Wang, L. Du, Z. Shu, Y. Xin

2025/12/18

Diversity of Our Team

Meteorologist | AI Expert

CMA

University

Co-Team Leader:

- Bo Lu (Meteorologist at CMA)
- Hao Li (AI experts at FDU)

Three Submitting Models:

- Fengshun (AI-light weight)
- FengshunAdjust (AI models ensemble)
- FengshunHybrid (AI+NWP)

Fengshun

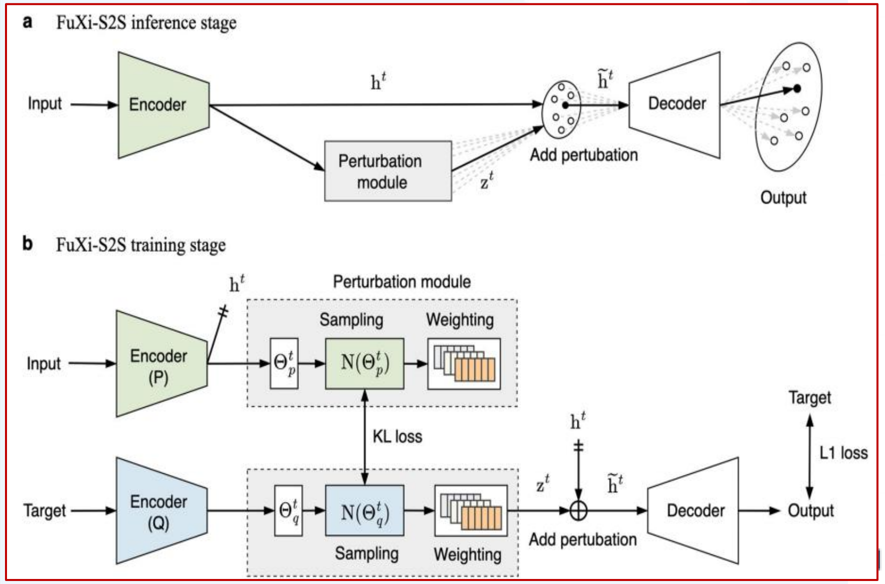
1. **Purely Data-Driven (ERA5);**
2. **Forecast weekly quintiles directly (tas, pr, slp);**
3. **Easy to update (no-need to calculate climatology)**

Chen et al., 2024——FuXi-S2S

nature communications

Article https://doi.org/10.1038/s41467-024-50714-1

A machine learning model that outperforms conventional global subseasonal forecast models



Only use initial condition each Monday (due to ERA5 delay)

Week number	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0				1	2	3	4
1	5	6	7	8	9	10	11
2	12	13	14	15	16	17	18
3	19	20	21	22	23	24	25
4	26	27	28	29	30	31	32
5	33	34	35	36	37	38	39
6	40	41	42	43	44	45	46

Day of week

Forecast submission window | First forecast period | Second forecast period | Publication of evaluation results

FengshunAdjust

Submitting results (weekly quintiles)

Calculate climatological quintile boundaries

ensemble forecasting

AI model base on SwinTransformer

Several base models

ERA5 daily data (from Monday)

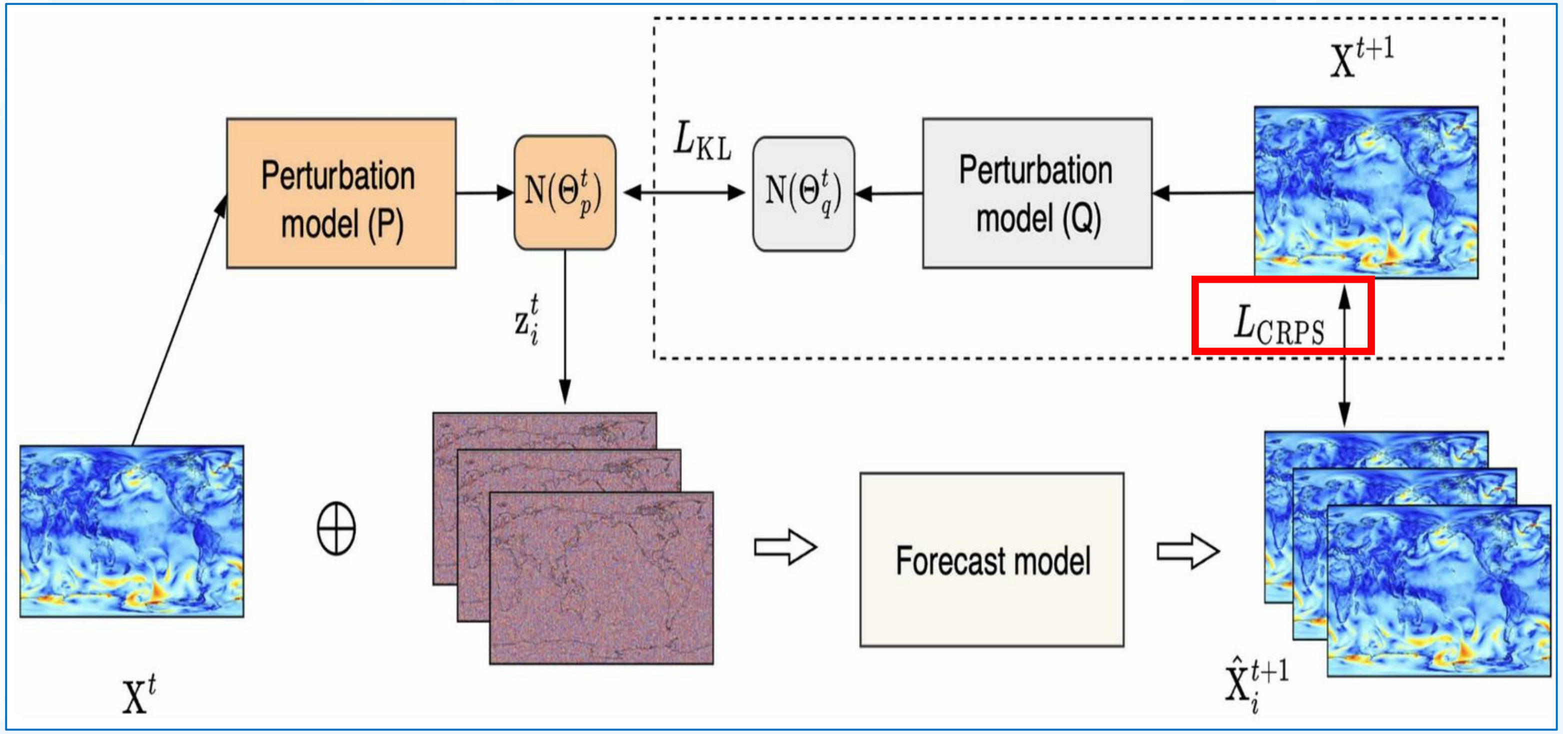
- A data-driven AI model trained for ensemble forecasting.
- Only ERA5 data (with 6days latency) been used for driving.

FengshunAdjust

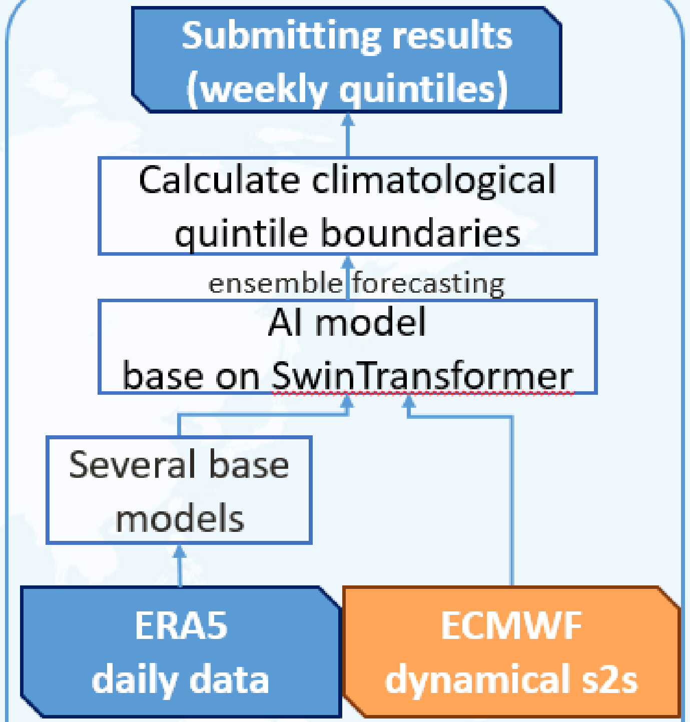
Multi-AI Models Ensemble (Purely ERA5-driven)

Original Version + CRPS-LOSS Version + Diffusion Version

Zhong et al., 2025 Science Advances



FengshunHybrid



- Similar to FengshunAdjust.
- Both ERA5 data and outputs from the ECMWF S2S operational model have been used for driving

FengshunHybrid

FengshunHybrid = FengshunAdjust + ECMWF dynamical model

#1 Forecast Window (FengshunHybrid better)

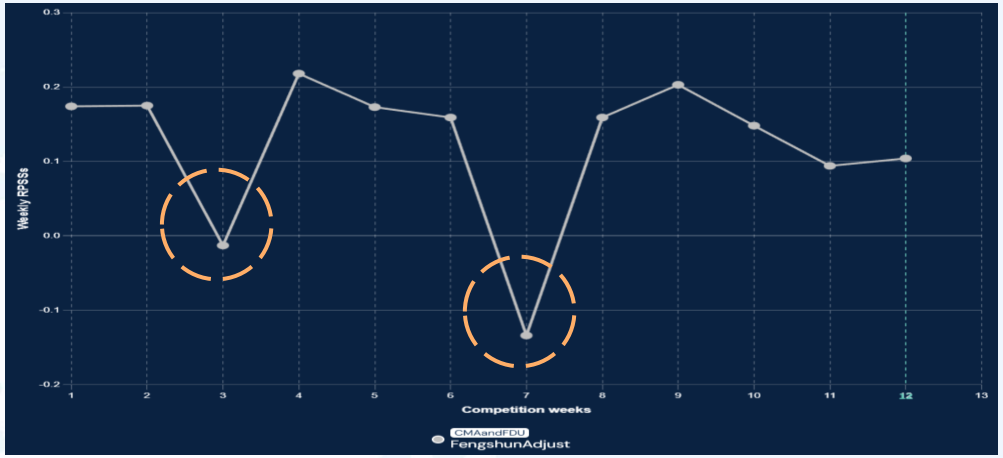
	CMAandFDU				
3	FengshunHybrid	0.055	0.121	0.046	0.074
4	FengshunAdjust	0.046	0.124	0.043	0.071
11	Fengshun	0.010	-0.008	0.023	0.008

#2 Forecast Window (FengshunAdjust better)

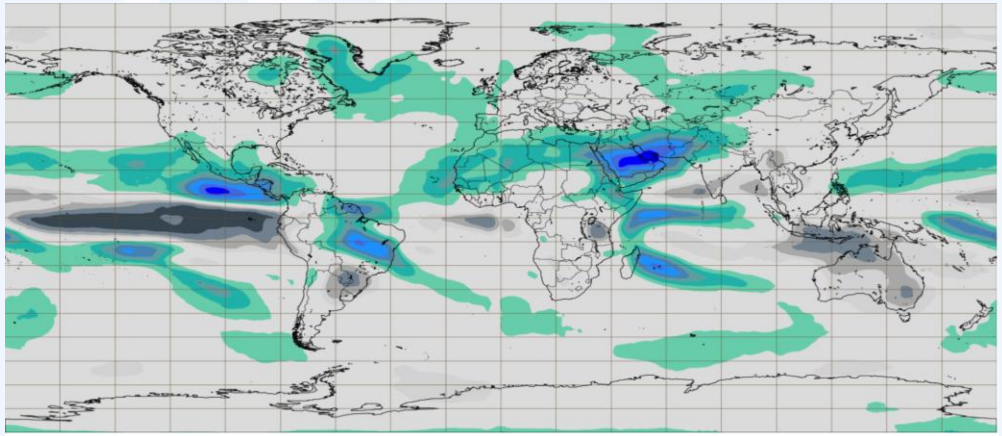
	CMAandFDU				
3	FengshunAdjust	0.041	0.075	0.033	0.050
4	FengshunHybrid	0.039	0.065	0.034	0.046
10	Fengshun	0.011	-0.015	0.023	0.006

What’s next

- **Real-time input**: ERA5 (delay) → IFS
- **Base model output**: Week3+4 → Week1+2+3+4
- **Fengshun (data-driven):** Easy to update, but relative limited score. Discover a more effective way to avoid long-time hindcasts.
- **FengshunAdjust (data-driven):** Unstable MSLP performance (two bad cases). Blurring issues in precipitation forecasting.
- **FengshunHybrid (AI+NWP):** How to balance the outputs b/w the AI model and dynamic model? More NWP models?



Mslp scores for FengshunAdjust dropped significantly on two cases



The precipitation faces blurring issue in s2s forecasting


Fengshun:
Wishing you smooth sailing ahead

Presentation by Fahamu

Spotlighted as an outstanding Africa-based team from the SON Period

ECMWF AIFS Sub-Seasonal Ensemble Forecasting with GPU Cloud Computing

Team Fahamu

Intergovernmental Authority on Development (IGAD)

Climate Prediction & Applications Centre(ICPAC), Kenya

Motivation and Background: FahamuAIFSv1 Model

- As part of CRAF'd funded project E4DRR and SEWAA initiatives at ICPAC(1)
- To Learn and build capacity on cloud computing GPU inference
- Utilizing open ECMWF AI Model AIFS ENS v1.0 for Sub- seasonal forecasting
- Enabled by Coiled notebook with Flash-Attention Anemoi inference Docker, LLM tools Claude Code
- 1. https://cgan.icpac.net/, https://icpac-igad.github.io/e4drr/

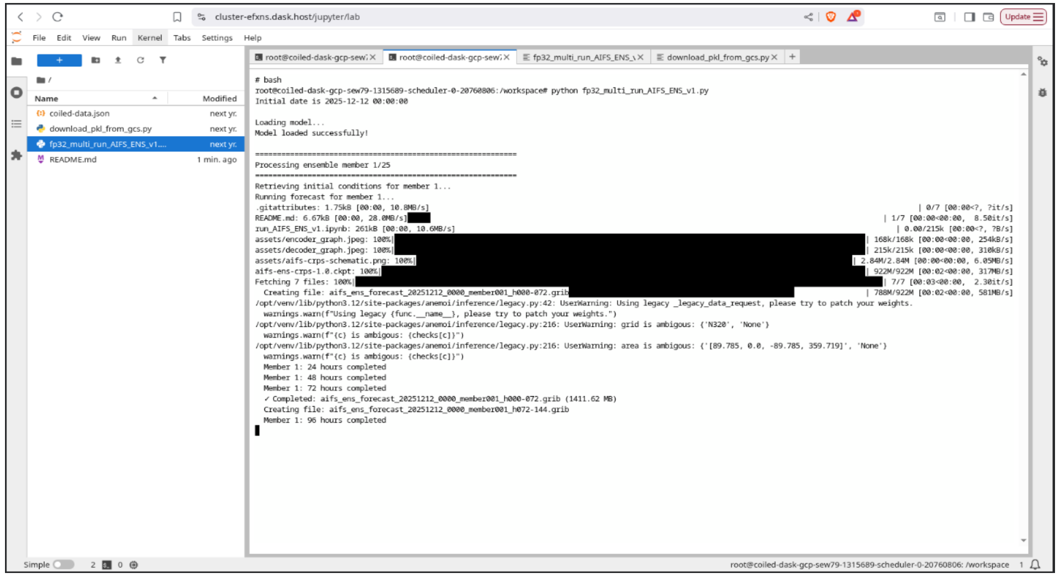
Week	FW1_tas	FW1_mslp	FW1_pr	Average	No. of ENS	Rank
Week 1	NA	NA	NA	NA	NA	NA
Week 2	-0.155	0.064	-1.222	-0.438	50	11
Week 3	0.004	-0.132	-1.318	-0.482	50	16
Week 4	-0.192	0.017	-1.452	-0.542	50	14
Week 5	-0.136	0.117	-1.353	-0.457	48	13
Week 6	-0.108	0.019	-1.34	-0.476	20	14
Week 7	0.013	-0.006	-1.132	-0.375	50	14
Week 8	-0.016	0.204	-1.002	-0.271	50	11
Week 9	0.078	0.14	-1.131	-0.304	50	14
Week 10	NA	NA	NA	NA	NA	NA
Week 11	-0.011	0.055	-1.192	-0.382	50	14
Week 12	0.038	0.068	-1.245	-0.38	50	16

FahamuAIFSv1 model SON season weekly forecast performance summary

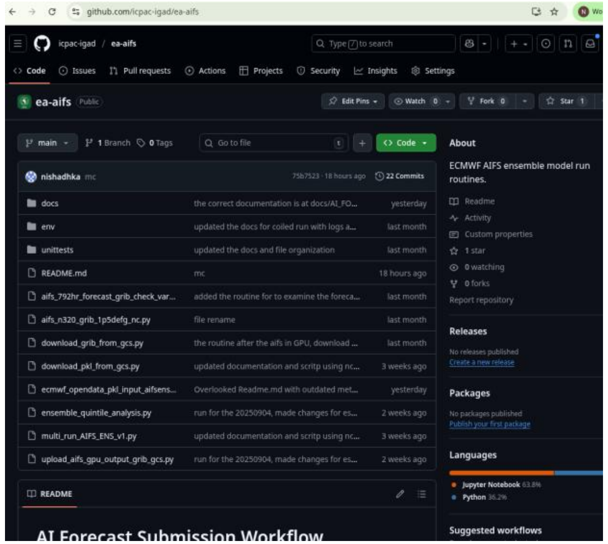
Coding for AIFS inference for Sub seasonal Forecast

Workflow Index

1. Initial Condition Preparation → `ecmwf_opendata_pkl_input_aifsens.py`
2. Transfer to GPU Environment → `download_pkl_from_gcs.py`
3. AI Model Execution → `multi_run_AIFS_ENS_v1.py`
4. GPU Output Upload → `upload_aifs_gpu_output_grib_gcs.py`
5. Forecast Download & Regrid → `aifs_n320_grib_1p5defg_nc.py`
6. Ensemble Analysis → `ensemble_quintile_analysis.py`
7. Forecast Submission → `forecast_submission_20250918.ipynb`



Coiled notebook for GPU inference



https://github.com/icpac-igad/ea-aifs

sewaa ICPAC COMPLEX RISK ANALYTICS Fund E4DRR NOR CAP

GPU Computing Infrastructure needed for AIFS ENS

Hardware Specifications

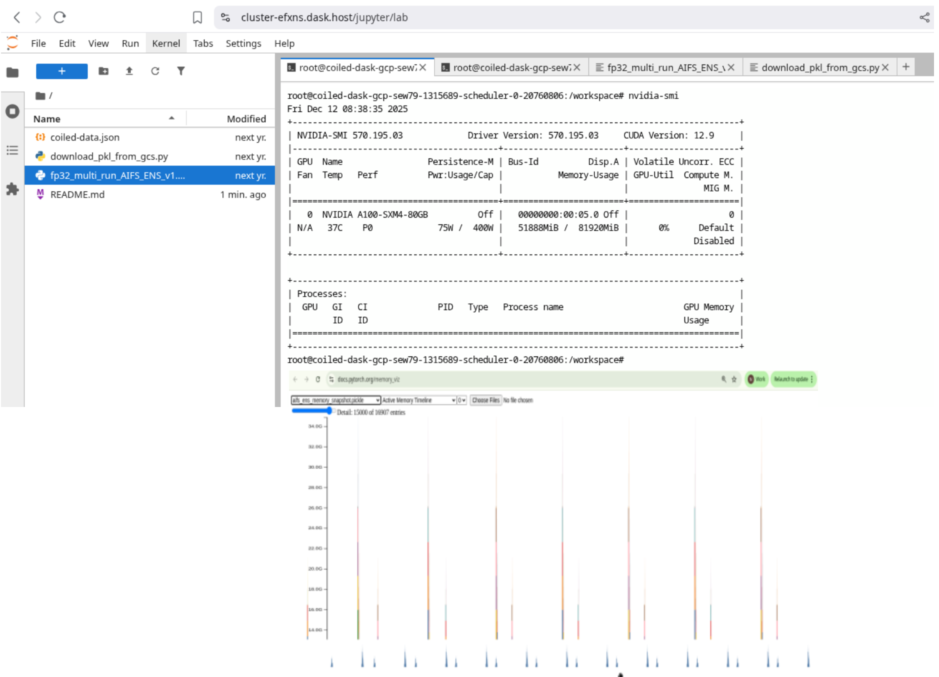
- GPU: NVIDIA A100 80GB via Google Cloud a2-ultragpu-1g, flash-attn requirement
- Storage: 369GB NVMe SSD for high-speed I/O, 72 hours forecast per member 1.5 GB in grib format

Performance Metrics

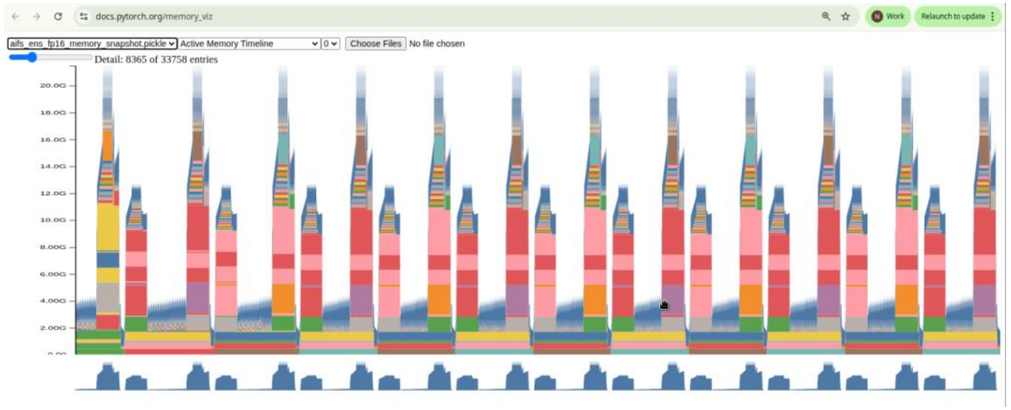
- Processing time: ~443 seconds per ensemble member, throughput: ~99 GRIB fields per second, zarr/icechunk for climatology analysis and 1.5° AIQuest forecast submission
- Memory efficiency: 52GB GPU memory utilization
- Total runtime:7 hours for 50 members, totaling 550 files and ~800 GB

Cost Optimization

- Per-member cost: ~$0.92 USD (~ 40$ for 50 members)
- Hourly rate: ~$5.08 USD for sustained GPU usage
- Regional efficiency: Same-region deployment reduces transfer costs by 60-80%
- The storage cost for ~800GB and Docker image, 15$ per week



AIFS ENS GPU profile for FP32 full precision inference, max 52GB



AIFS ENS GPU profile for FP16 half/reduced precision inference, max 21GB
https://huggingface.co/ecmwf/aifs-ens-1.0/discussions/17

Challenges and Way forward

- Current execution relies on manual, script-based steps; lacks full automation and orchestration for routine sub-seasonal forecasting.
- AIFS-ENS requires large memory GPUs (80 GB A100), and limited availability creates operational bottlenecks. Profiling indicates AIFS-ENS can run in reduced precision, enabling the use of widely available mid-tier GPUs (e.g., 24 GB T4/L4) and lowering costs. Containerized, event-driven pipelines (Cloud Run / microservices) to trigger inference, post-processing, and submission workflows.
- AIFS ENS inference data stored as GRIB format and in scale of ~800GB per week run for N320 grid, needed to format in Zarr/Icechunk for wider usage.
- Initialization relies on ECMWF Open Data with only a 4-day window, restricting reproducibility and re-forecast generation. Use AWS S3 Open Data and Virtual Datasets (Kerchunk/VirtualiZarr/Icechunk) to overcome the 4-day availability limit and enable multi-year initialization.
- Observation blending and surrogate AI refinement are still experimental. Continue experimentation with observation-blending and bias-correction using surrogate modeling to enhance FahamuAIFS forecasts.

Thank you

Key milestones and actions

12/02/26

Competition Phase

SON period	DJF period	MAM period	JJA period

SON Awards Webinar

DJF Awards Webinar

MAM Awards Webinar

JJA & Yearly Awards Event

19/03/26

Join the competition and start submitting your forecasts weekly **at any time**!

First DJF leaderboards

Access SON data
Access the complete SON forecast datasets and participating teams' models' summaries.

Register for the DJF Awards webinar
Top-competitors of the DJF Period will share their insights.

Submit to S2S2D
Submit an abstract to the International S2S2D Conference, indicating you are an AI Weather Quest participant

ECMWF · Destination Earth · WORLD METEOROLOGICAL ORGANIZATION

Thanks!

To everyone involved in the organisation of the AI Weather Quest and everyone involved in its journey!