

Fifth WMO AI Webinar

AI Weather Quest: Annual Highlights and Awards

Webinar | 24 September 2026 13h UTC

Agenda

- ✓ Quest framework and participation
- ✓ Evaluation and scientific insights
- ✓ CMAandFDU presentation and Q&A
- ✓ MicroEnsemble presentation and Q&A
- ✓ CLINT presentation and Q&A
- ✓ What comes next



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 ask questions during the Q&As.

Quest framework and participation

The **ECMWF AI Weather Quest** is a **global benchmark and learning framework** launched in August 2025 to compare ML models for sub-seasonal weather prediction. It is open to all (researchers, tech companies, forecasting institutions, etc.).

Participants currently produce global weekly forecasts for **weeks 3–4** ahead for the following variables:



Near-surface (2m)
temperature (tas)



Mean sea level
pressure (mslp)



Precipitation
(pr)



Forecasts are **evaluated weekly** against reference observational datasets, and results are published live:



The Quest has established a global community fostering evidence and exchange on ML-based S2S skill

Broad usability & uptake



Model summary questionnaire for model S2S skill

ECMWF AI Weather Quest Data Portal

Index of dataset directory

For more information, visit our official AI Weather Quest website.

Name	Scheduled date	Size	Id
README_AI_Weather_Quest.txt	01-10-2025 10:54	5334	107793360
L3CM2_AI_Weather_Quest.txt	01-10-2025 10:54	380	107793360
By_team.zip	20-09-2025 11:15	-	-
By_team.zip	20-09-2025 00:23	-	-
By_team.zip	20-09-2025 11:15	-	-
By_team.zip	20-09-2025 00:23	-	-



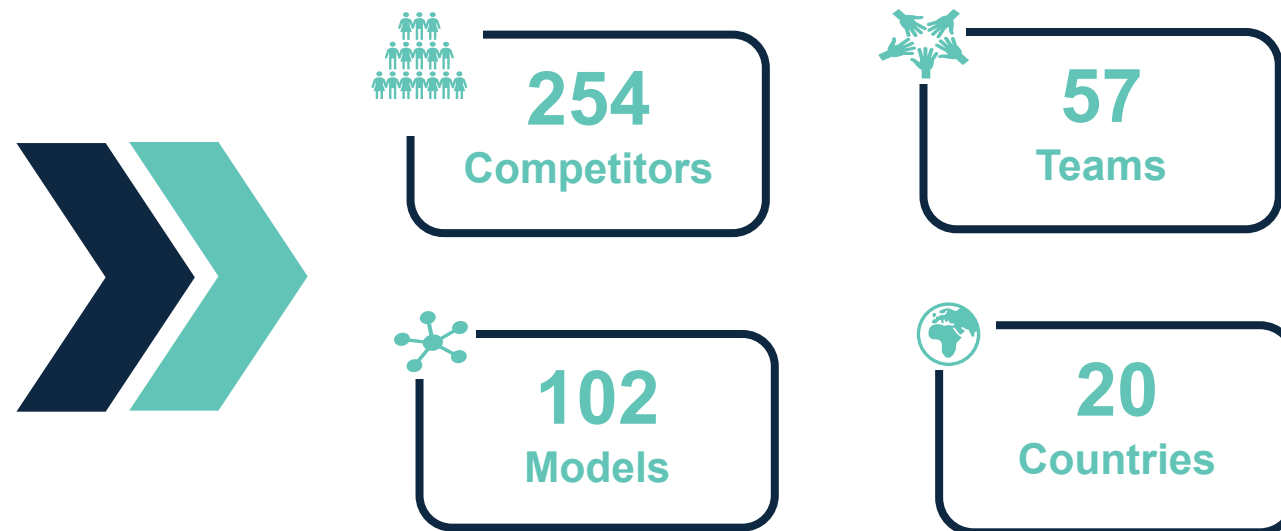
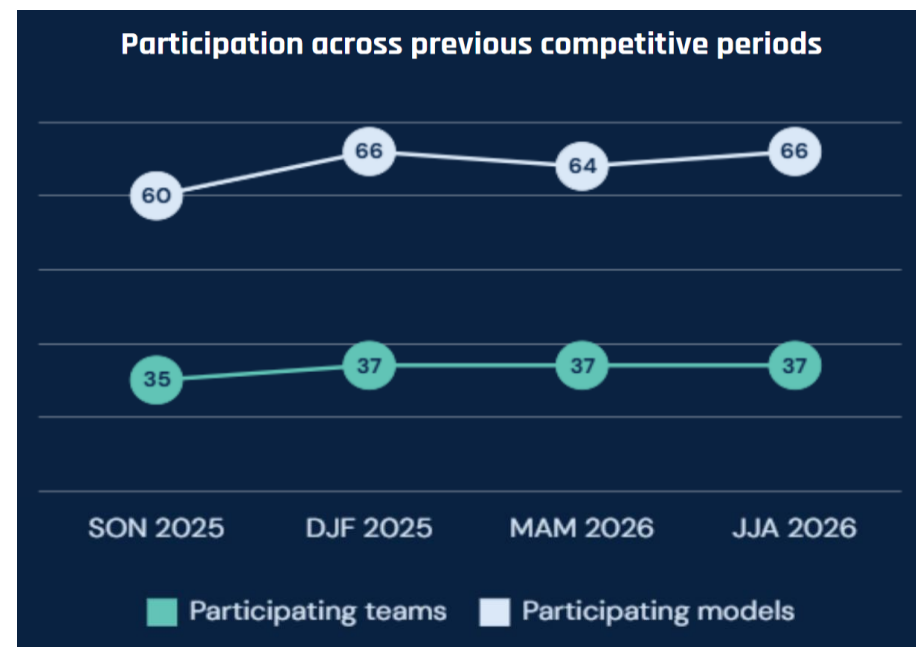
Strong participation

250+ Competitors

50+ Teams

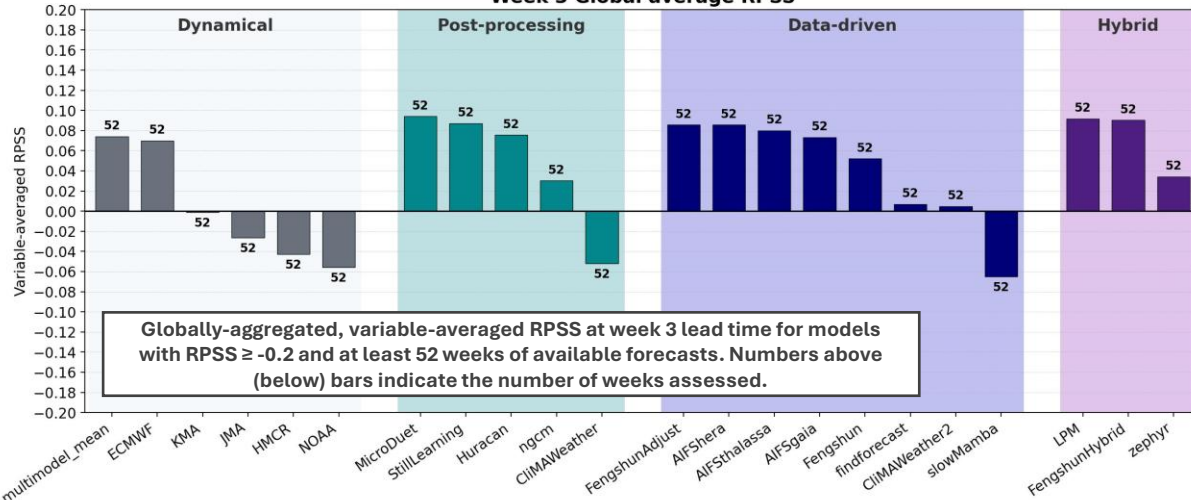
100+ Models

Participation overview

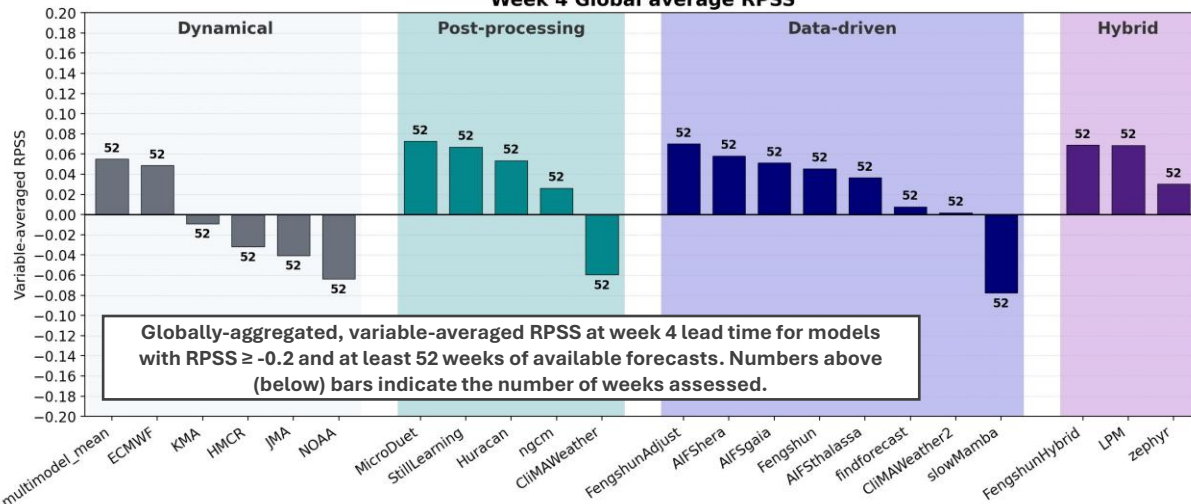


Evaluation insights : First competitive year overall scores

Week 3 Global average RPSS



Week 4 Global average RPSS



Dynamical predictions calibrated against a reforecast climatology.

The most skilful models use dynamical forecasts.

- (1) Top-performing approaches shifted from post-processing techniques (e.g. MicroEnsemble's **MicroDuet**) early in the forecasting year to hybrid models (e.g. CMA&FDU's **FengshunHybrid**).
- (2) Several data-driven (DD) systems outperform IFS including:

- AIFS **S2SHera**
- CMA&FDU **FengshunAdjust**

Highlight contributions from:

Academia and technological corporation collaboration

Meteorological forecasting centre



Given skill in **post-processing** and **hybrid** forecasting systems, we acknowledge efforts led by CMA, ECMWF and IRI to host the WWRP/WCRP S2S database.

Evaluation insights : First competitive year overall scores

The Quest has observed the development of innovative forecasting methods.

Team **CLINT** developed multiple skilful forecasting approaches:

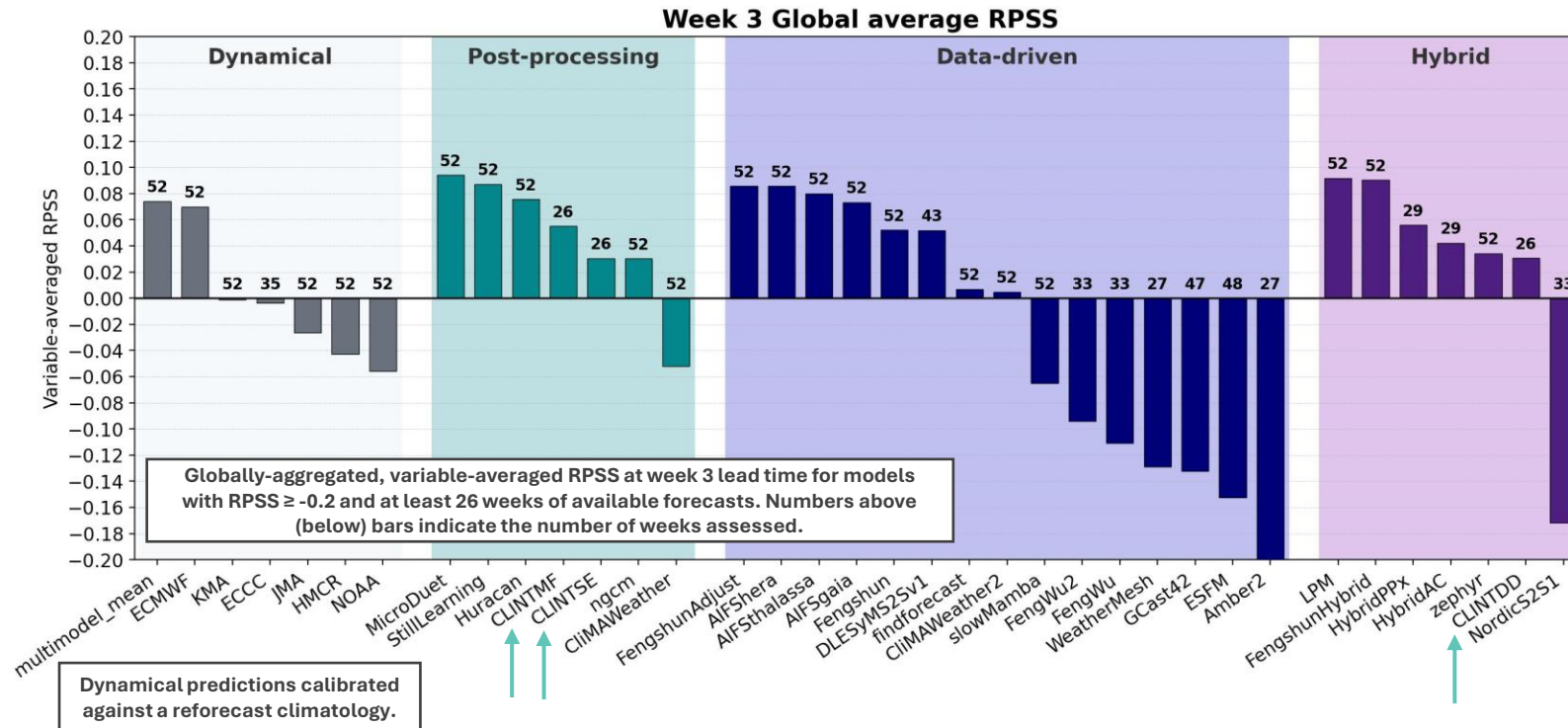
- Bias-adjusted, lagged ECMWF ensemble forecasts.
- A pre-trained U-Net model for probabilistic forecast calibration.



Highlight contribution from:
SME-Research center collaboration



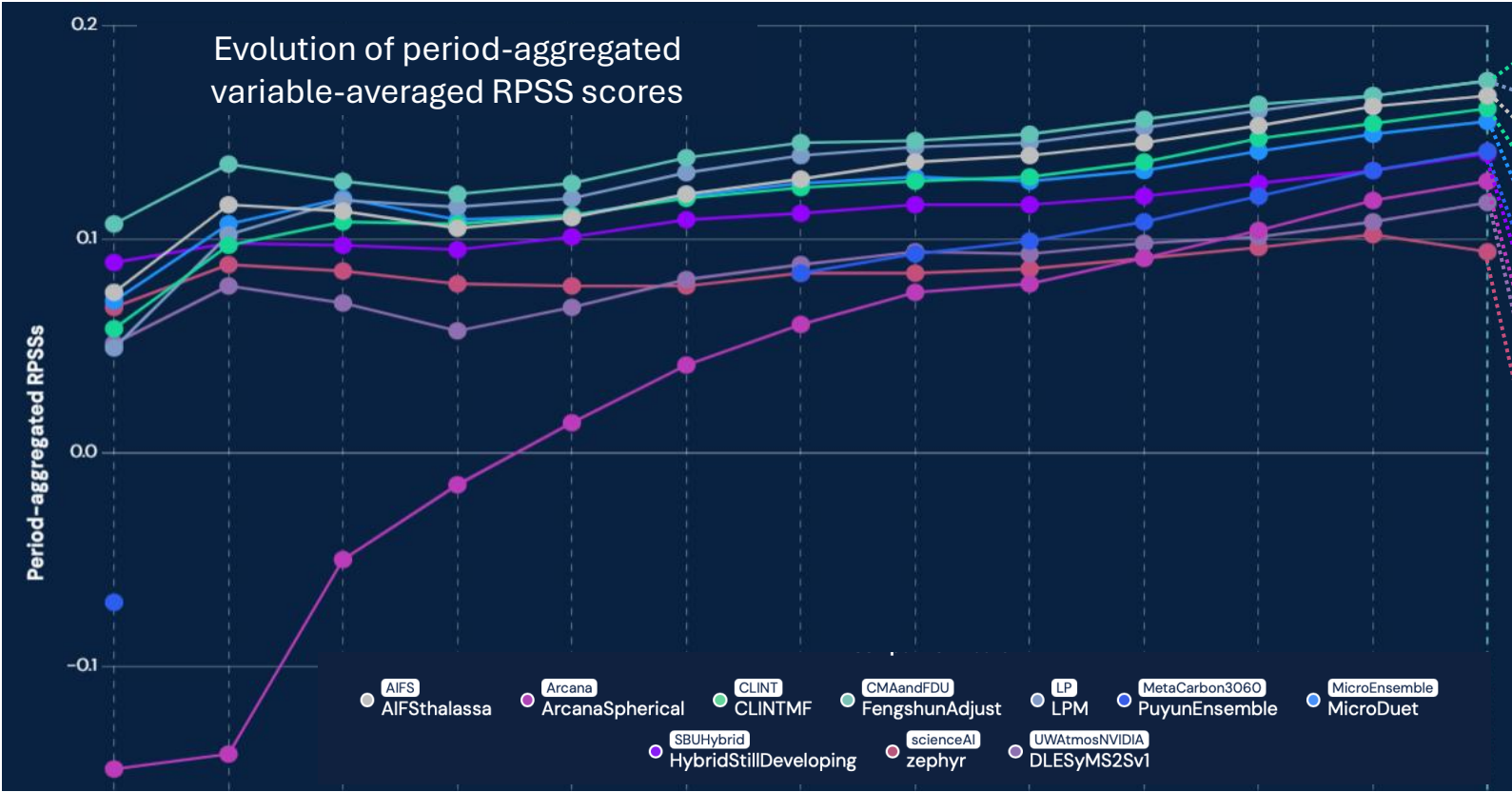
the
climate
data factory



Evaluation insights : JJA Top-performers overview for week 3 ahead

20 Teams eligible for variable-averaged, period-aggregated scores

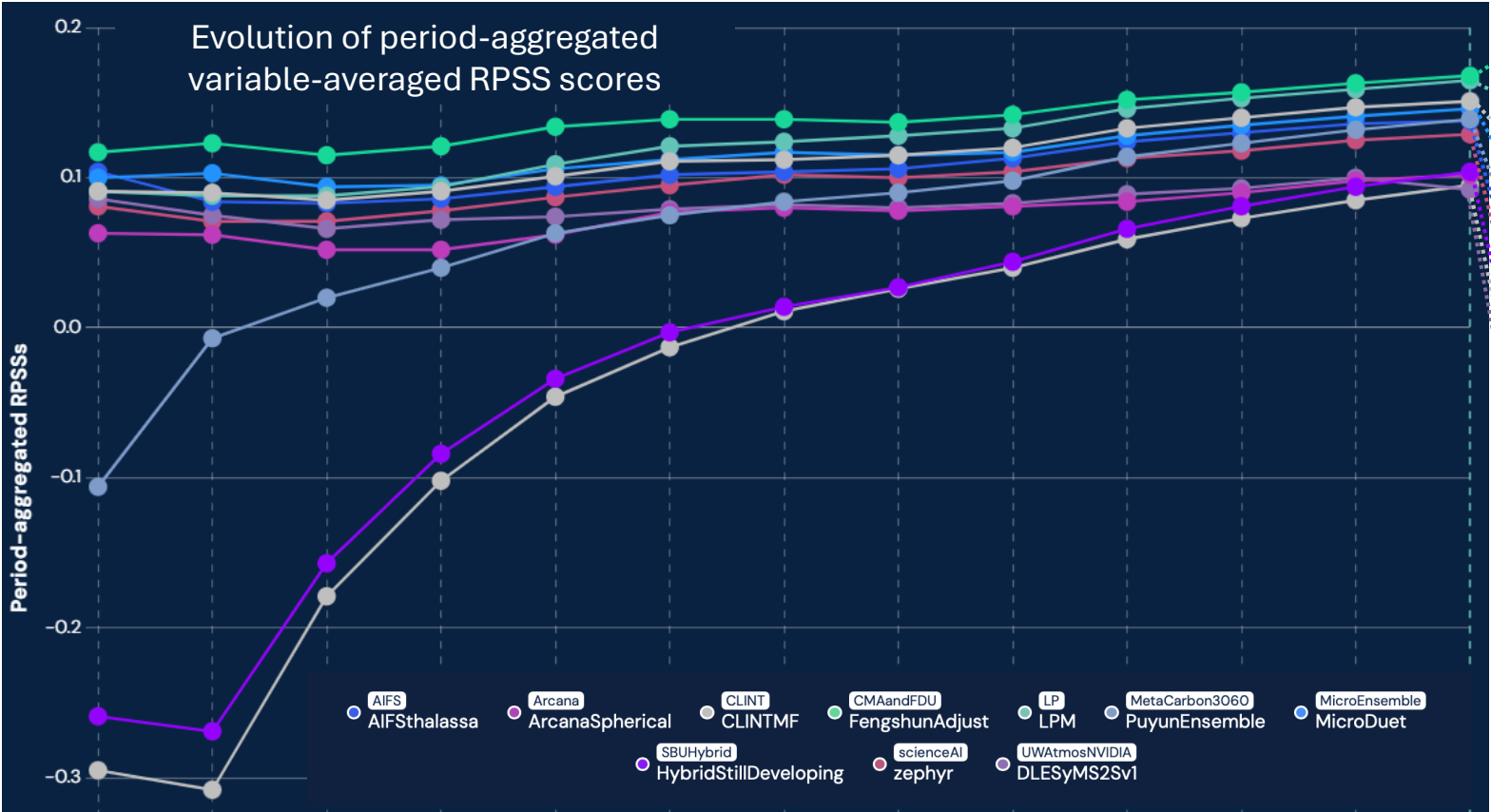
13 Teams above climatology



Evaluation insights : JJA Top-performers overview for week 4 ahead

20 Teams eligible for variable-averaged, period-aggregated scores

11 Teams above climatology





Presentation by team CMA and FDU





Fengshun Family by CMAandFDU

China Meteorological Administration | Fudan University

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**

Meteorologists

AI Experts

CMA



Bo Lu



Jie Wu



Chenguang Zhou



Jiahui Hu



Qifeng Qian



Chenpeng Wang



Zesheng Dou



Yang Zhao



Chunyan Zhao

FDU
University



Xiaohui Zhong



Hao Li



Lei Chen

Model portfolio

DATA-DRIVEN

- Fengshun
- FengshunAdjust

AI + NWP

- FengshunHybrid

FuXi-S2S
(Chen et al., 2024)

CRA40

Fengshun 1.0: First Data-Driven S2S Forecasting Operational System

nature communications

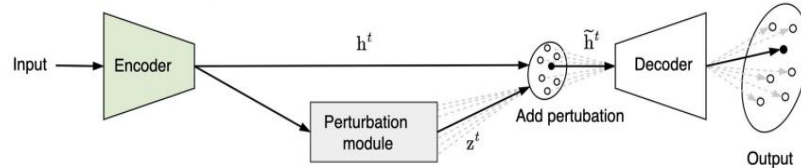


Article

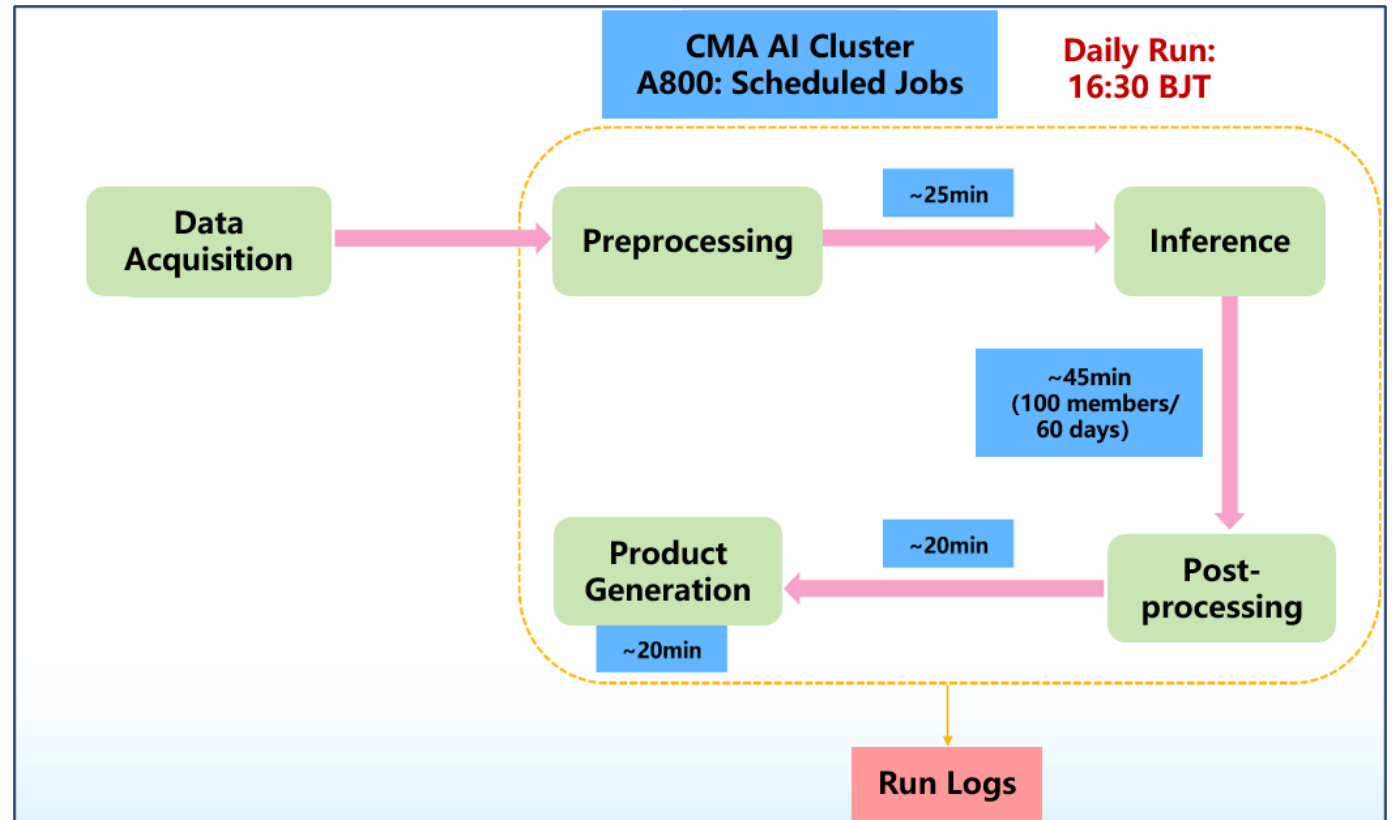
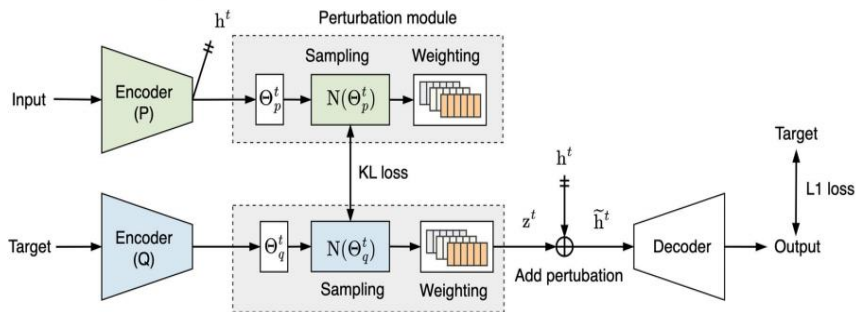
<https://doi.org/10.1038/s41467-024-50714-1>

A machine learning model that outperforms conventional global subseasonal forecast models

a FuXi-S2S inference stage



b FuXi-S2S training stage

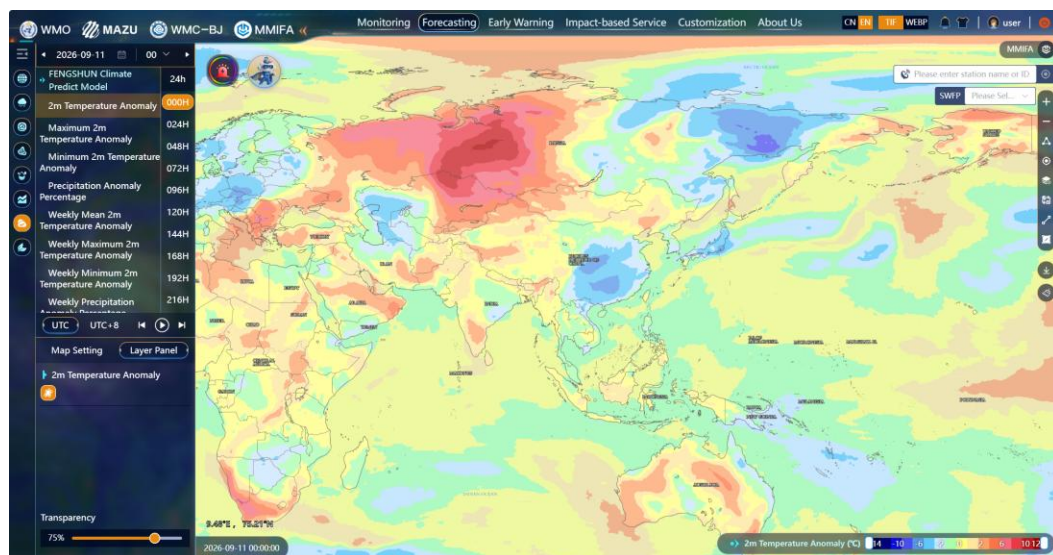


Became Operational ever since September 2024

Fengshun 1.0 : open source



Graphic Products at MAZU Platform



<http://ew4all.wmc-bj.net/EW4ALL/predictions>

Fengshun 1.0 has been open-sourced



<https://github.com/NCC-China/CMA-AIM-S2S-Fengshun>

A light-weighted version

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

Use Monday initial conditions only, due to ERA5 delay

Week number	Day of week						
	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

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

Model portfolio

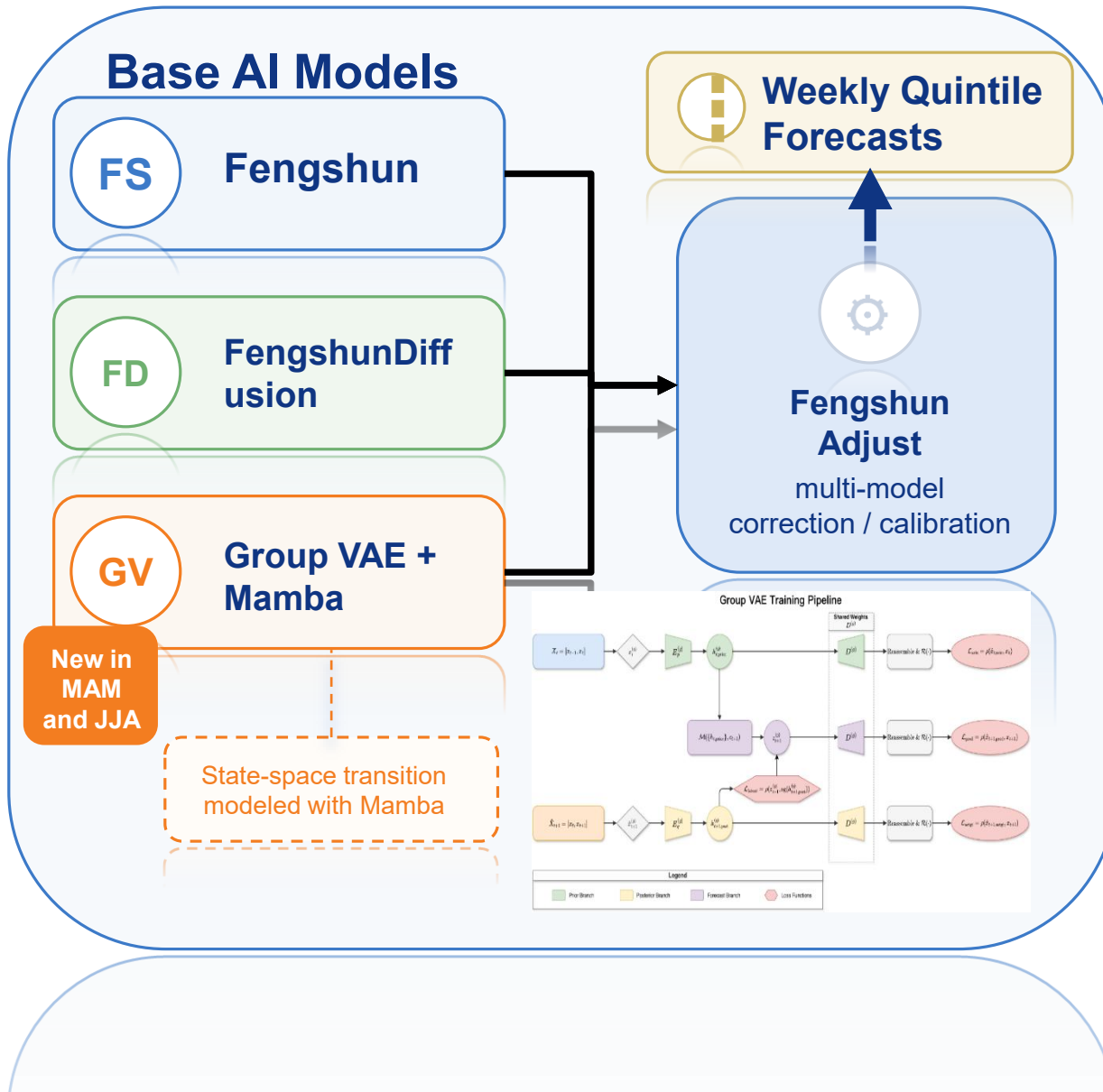
DATA-DRIVEN

- Fengshun
- FengshunAdjust

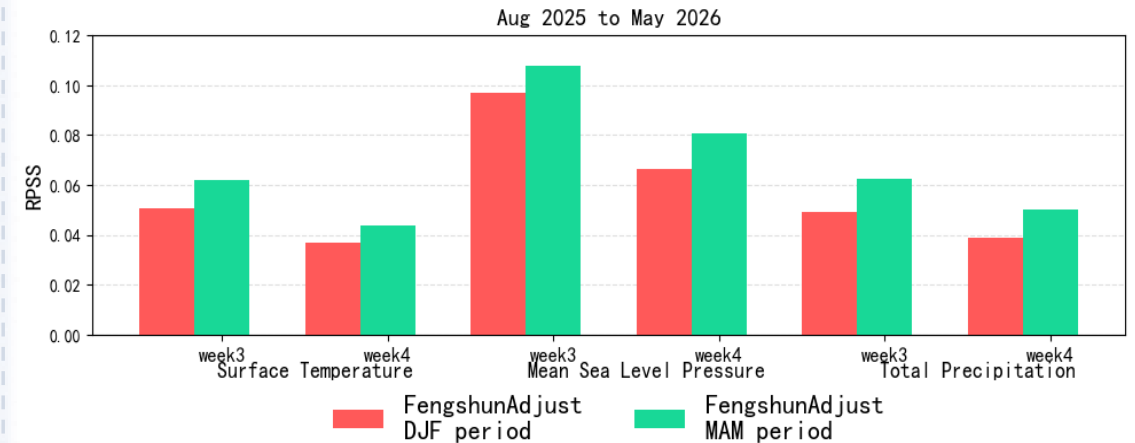
AI + NWP

- FengshunHybrid

FengshunAdjust: A data-driven ensemble model



- For the **MAM** and **JJA** seasons, we extended the original Fengshun and FengshunDiffusion framework by adding a **new model with Group VAE and Mamba for state-space transition modeling**



Key update in MAM:

- + **Group VAE base model**
- + **Mamba state-space transition**
- + **Better forecast skill**

FengshunAdjust: Case Study



ERA5

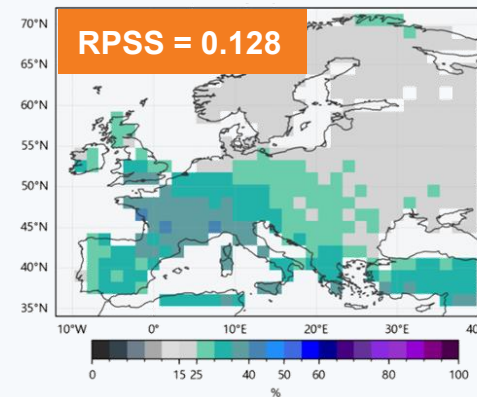
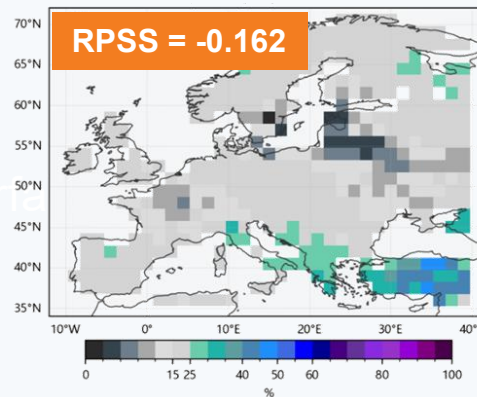
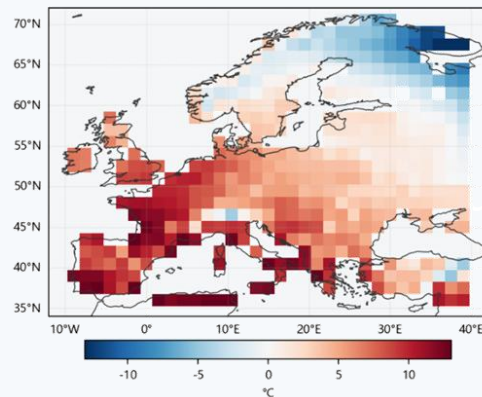
ECMWF

FengshunAdjust

Forecast
period

near surface
temperature

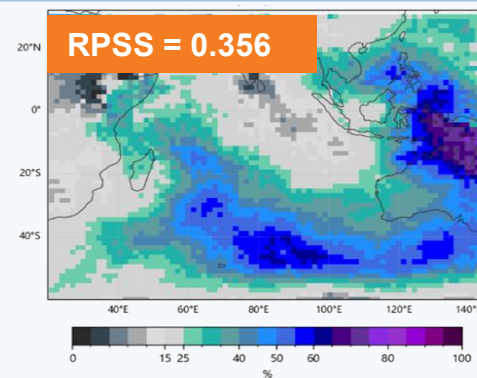
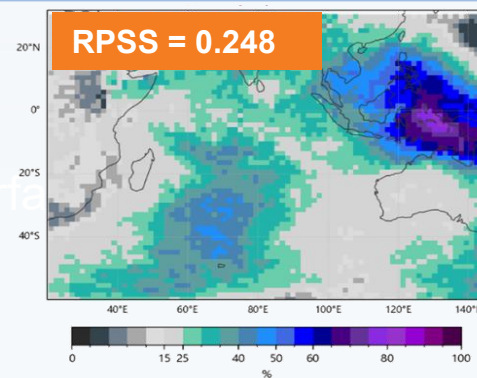
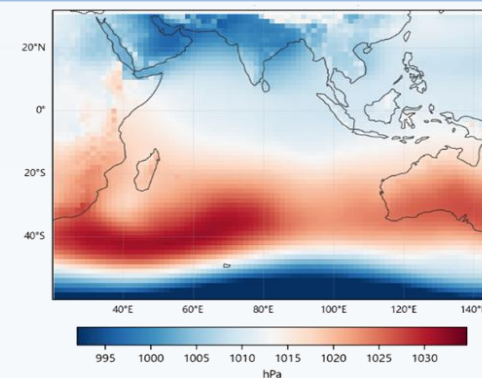
02–08 Mar, 2026



week-4
Q>0.8

mean sea level
pressure

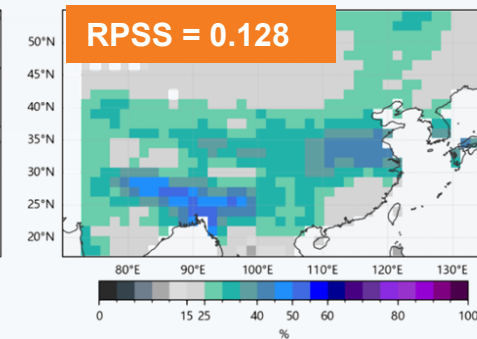
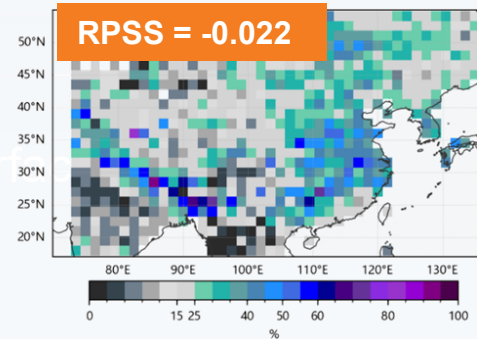
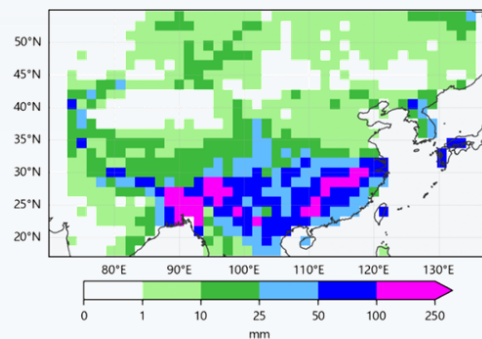
27 Jul–02 Aug, 2026



week-3
Q>0.8

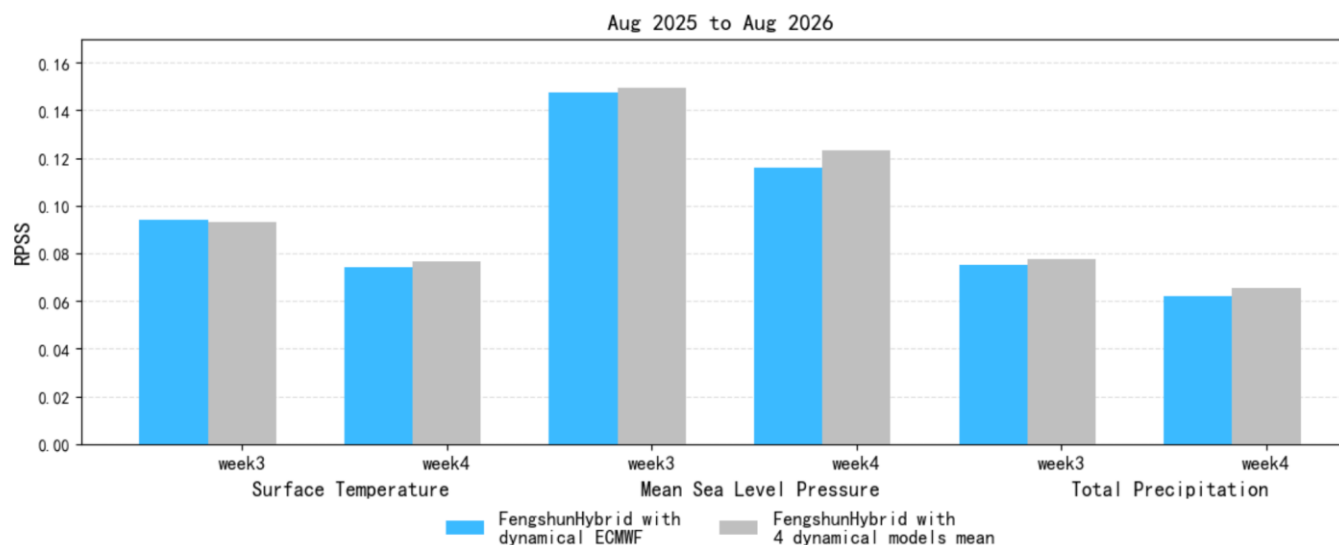
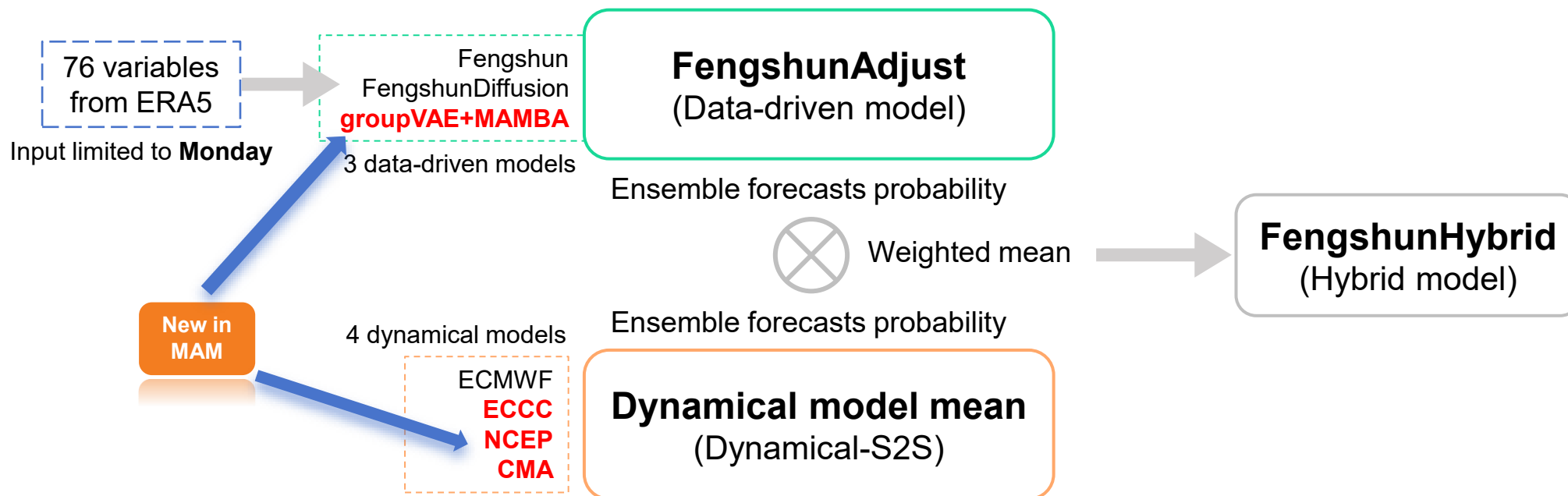
total precipitation

27 Apr–03 May, 2026



week-3
Q>0.8

FengshunHybrid: AI+NWP

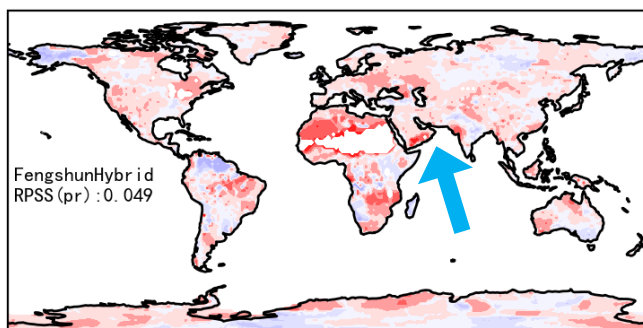
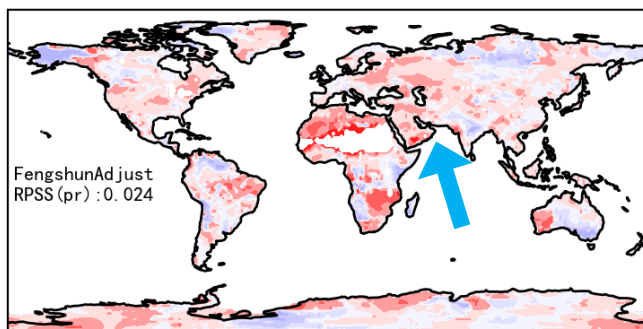
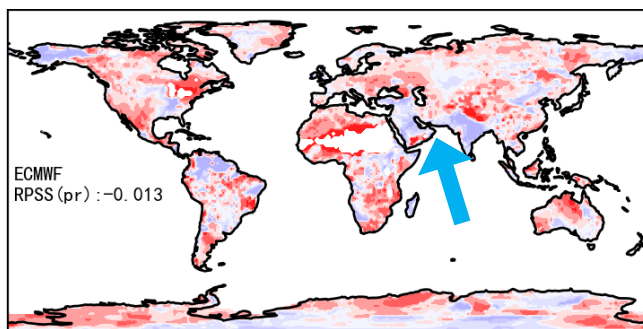
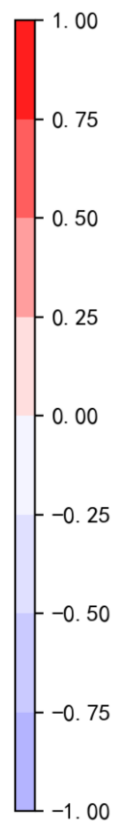


Variables:
Base time:
Forecast period:

total precipitation
26 February, 2026
week-3

surface temperature
09 April, 2026
week-3

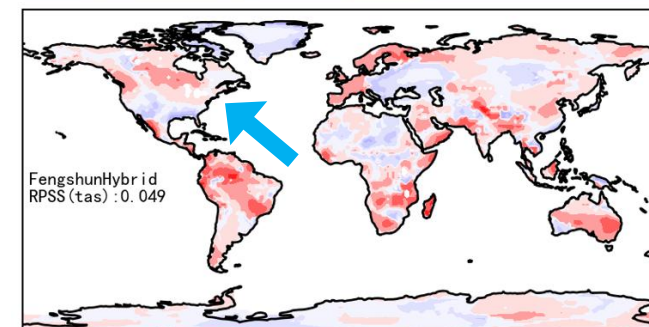
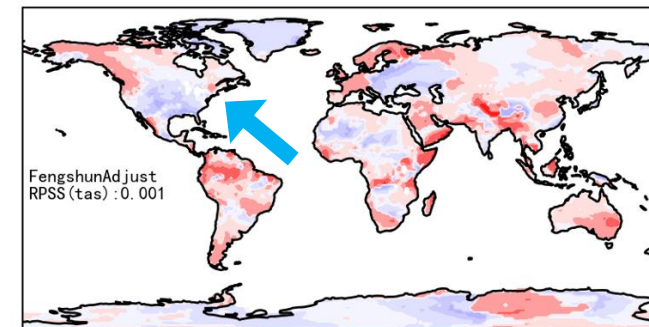
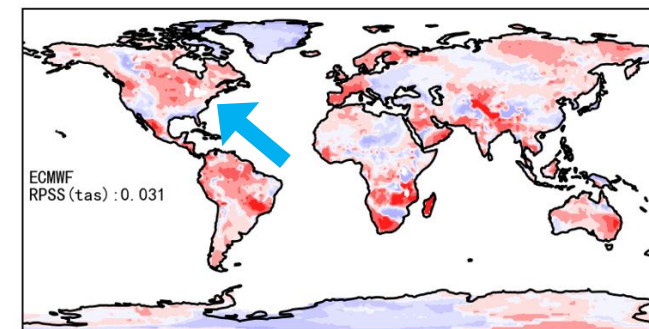
RPSS



ECMWF
(Dynamical-S2S)

FengshunAdjust
(Data-Driven)

FengshunHybrid
integrates the advantages of
both and balances their
respective weaknesses

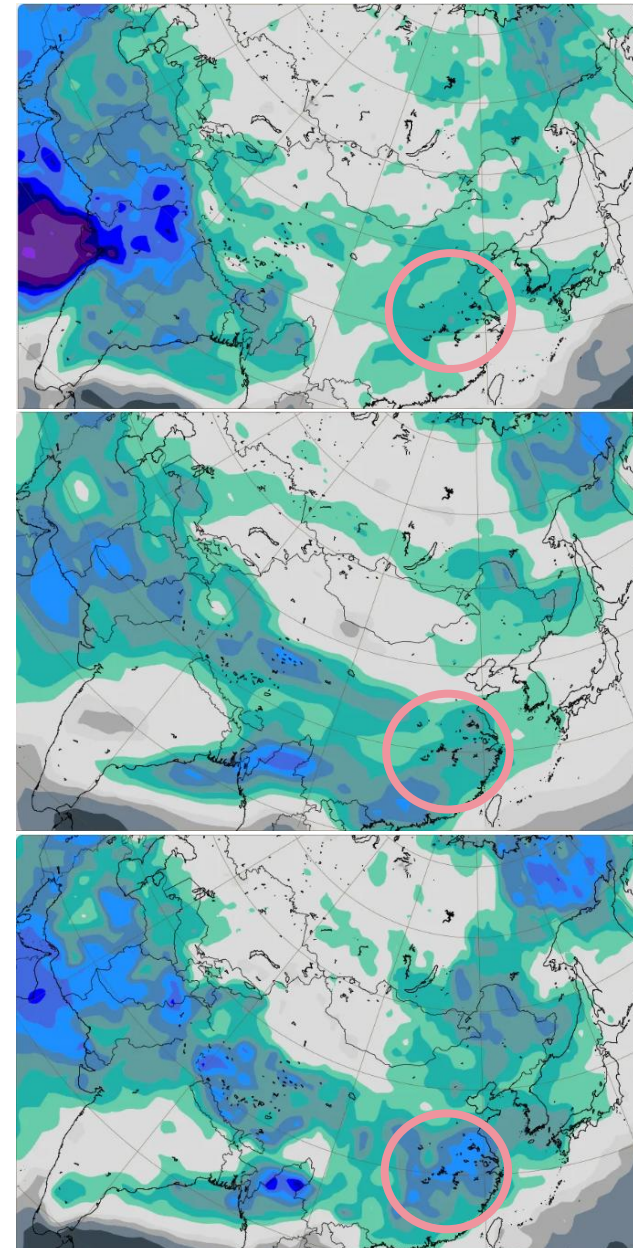
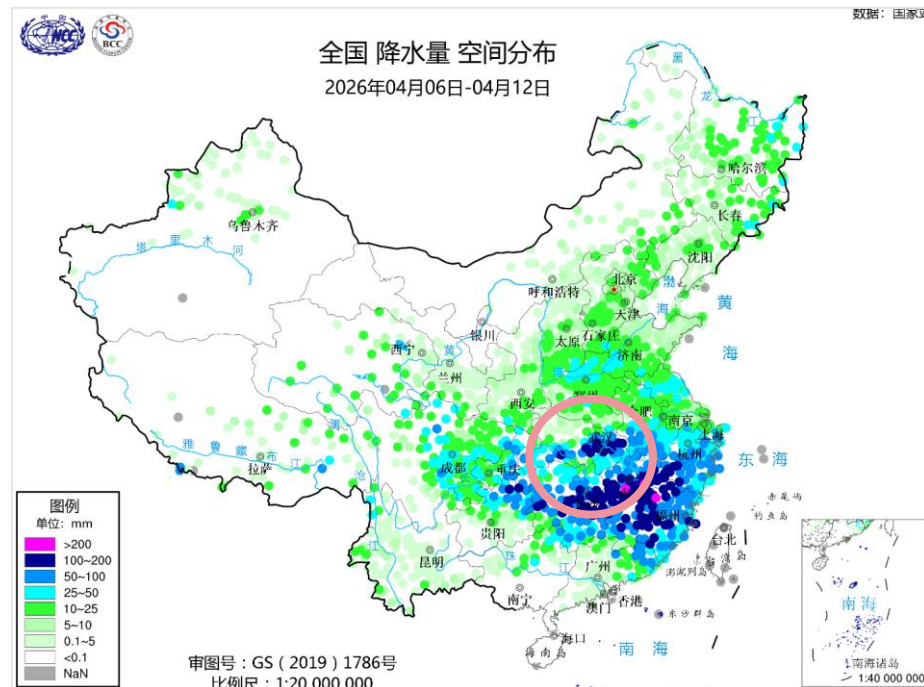


In case 1, the data-driven model (**FengshunAdjust**)
outperforms the dynamical_S2S model (**ECMWF**)

In case 2, the dynamical_S2S model (**ECMWF**)
outperforms the data-driven model (**FengshunAdjust**)

An abnormally northward and strong subtropical high caused heavy rainfall across the Jiangnan region in early April.

Observed

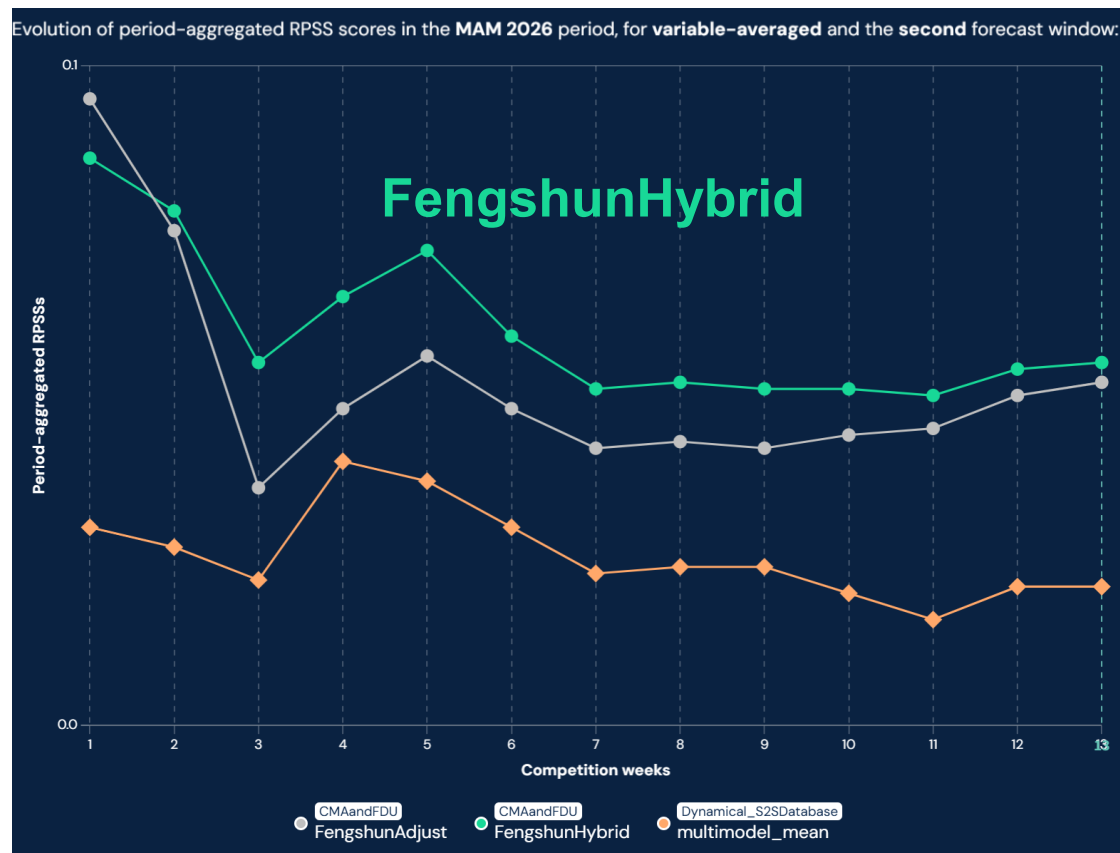


Dynamical model mean
Dynamical-S2S

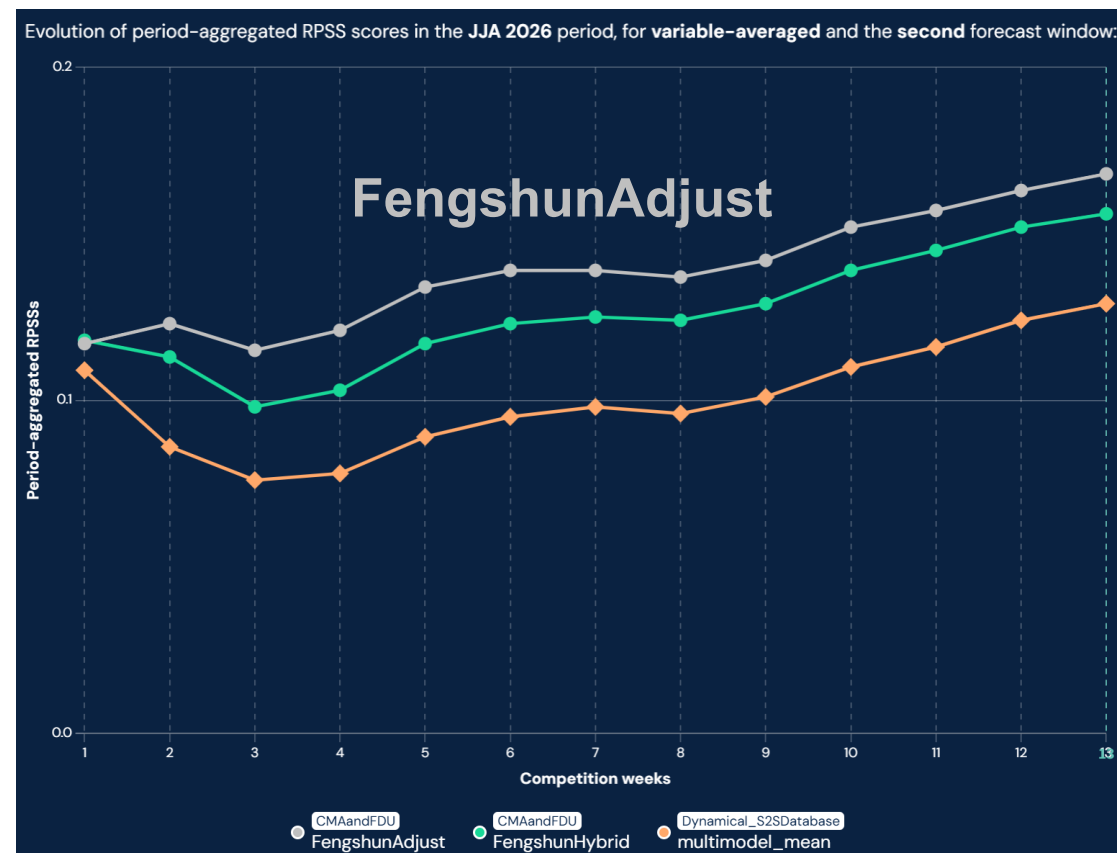
FengshunAdjust
Data-driven

FengshunHybrid
Hybrid-model

Performance in AI Weather Quest



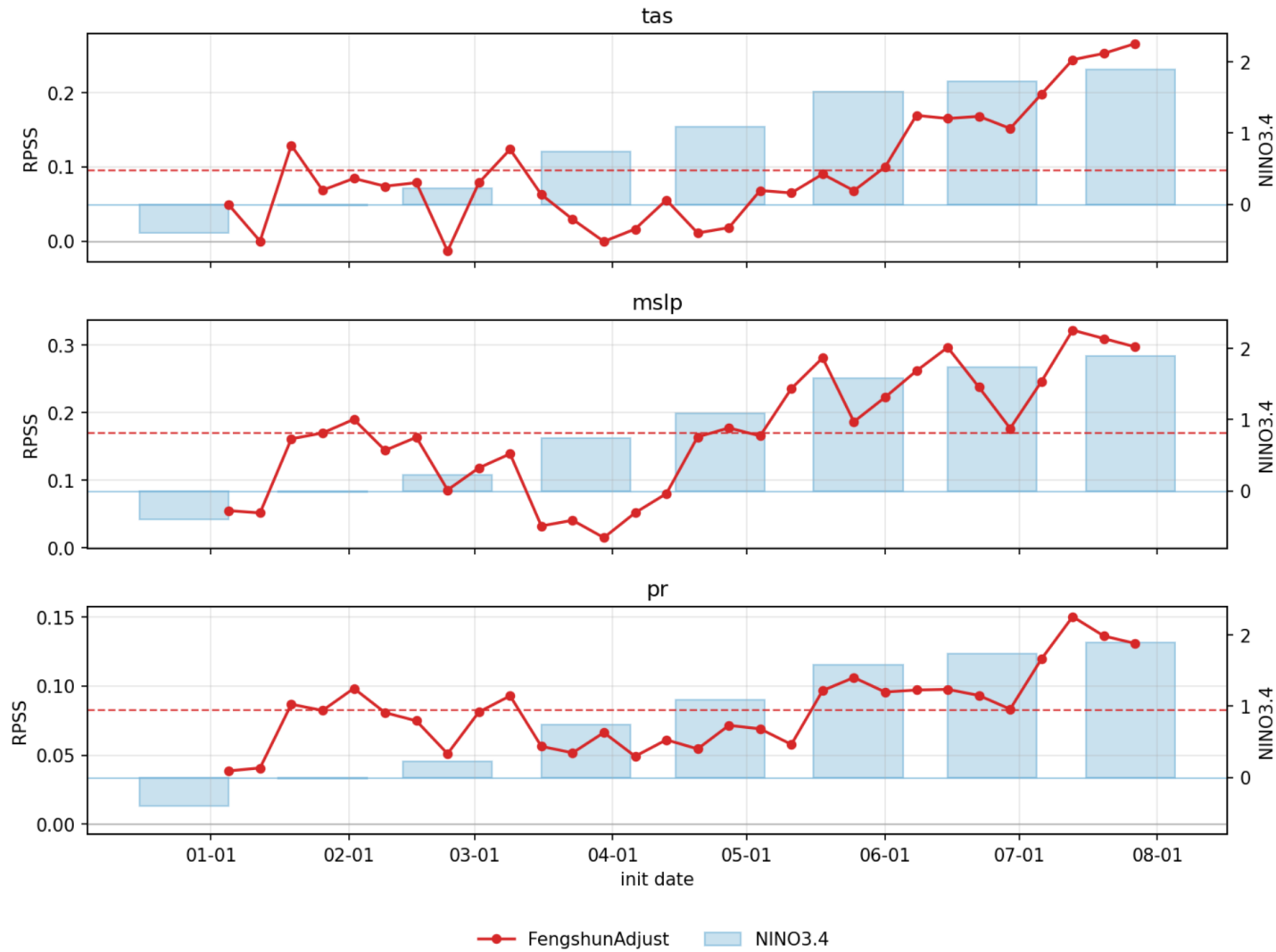
MAM Ranking (Week 4): Hybrid better



JJA Ranking (Week 4): AI-Ensemble better

- Ranking #1 for MAM (Week 4): FengshunHybrid
- Ranking #1 for JJA (Week 4): FengshunAdjust

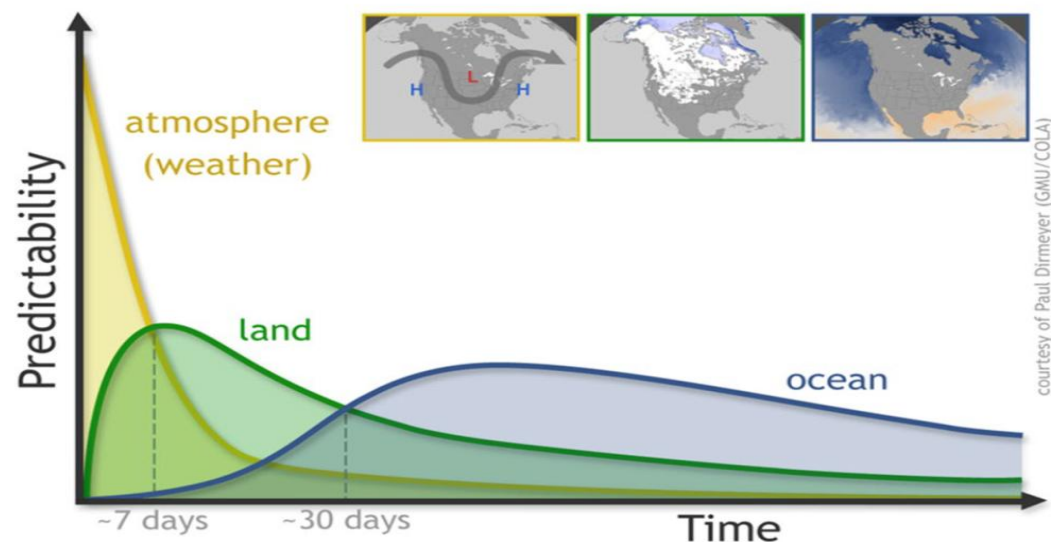
Higher prediction skill during El-Nino developing phase



ENSO also provide predictability for purely data-driven AI model

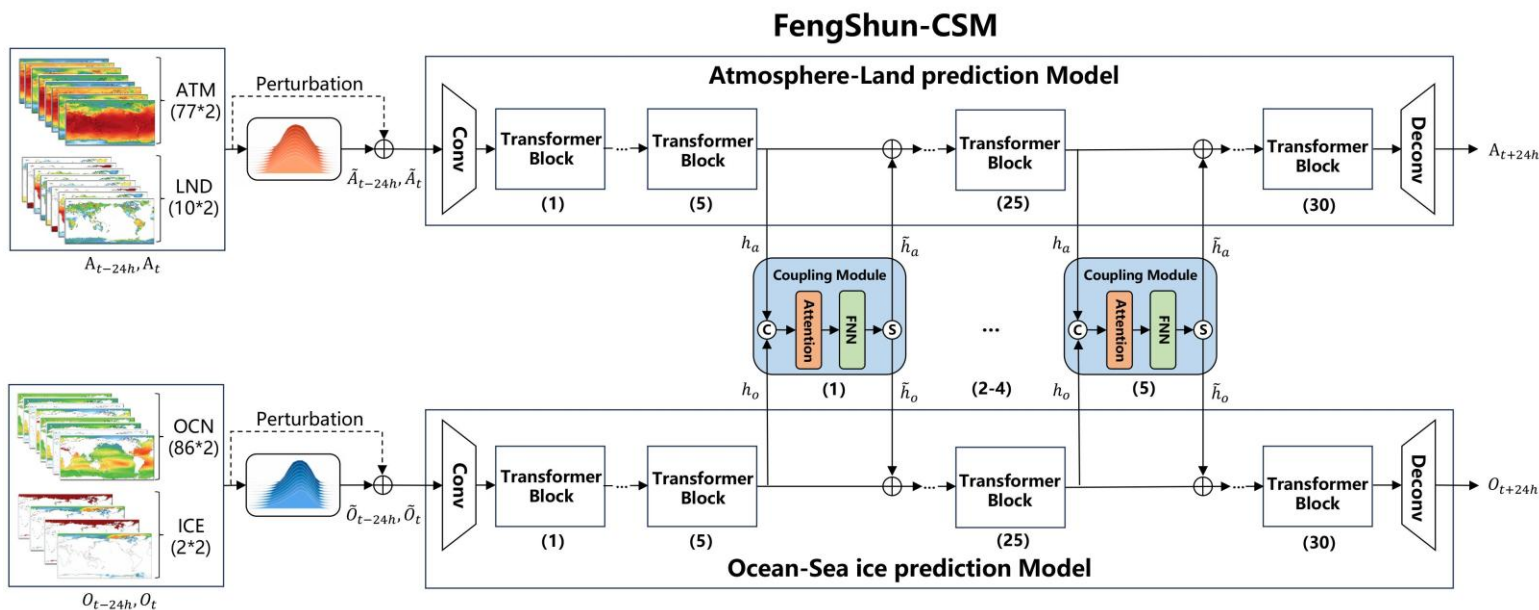
RPSS for FengshunAdjust in 2026

What's Next? Toward Climate System Model (CSM)



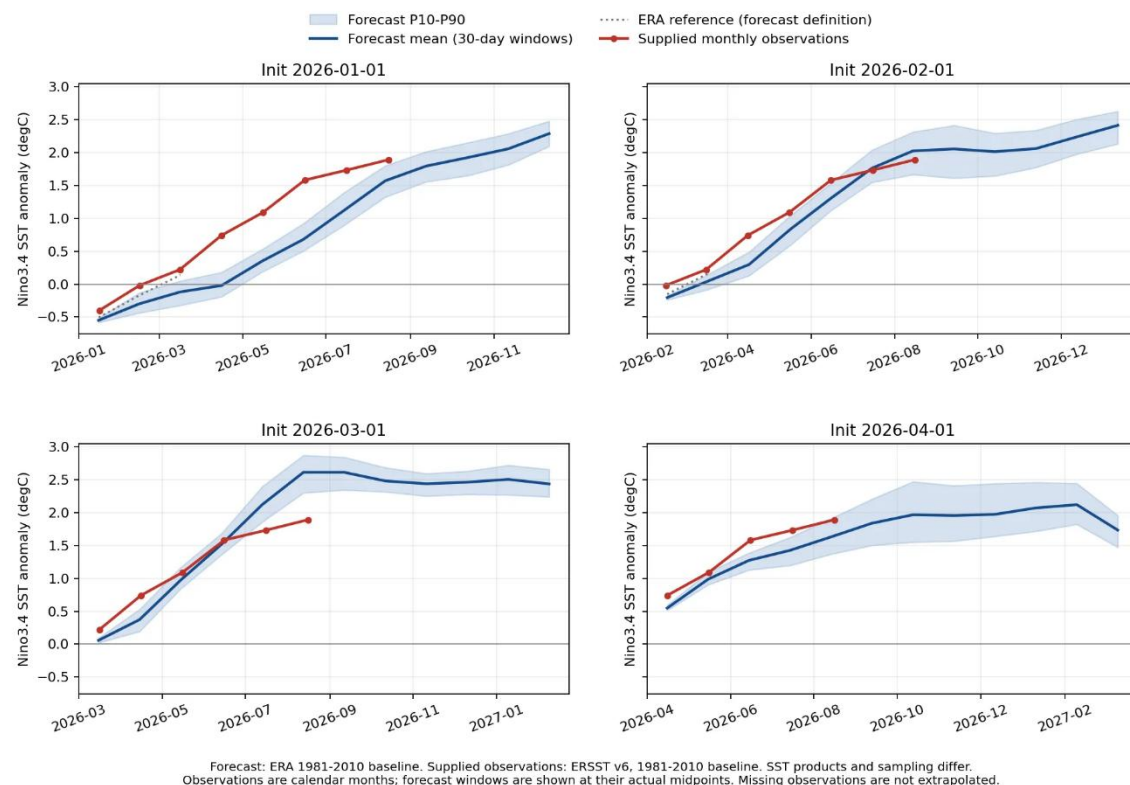
Our Team will participate AI Weather Quest in Year2
MJO, TC activity...
Beyond:

- ◆ Land & Ocean forcings provide predictability source for longer time range.
- ◆ Can Fengshun provide skillful prediction at seasonal or longer time range?

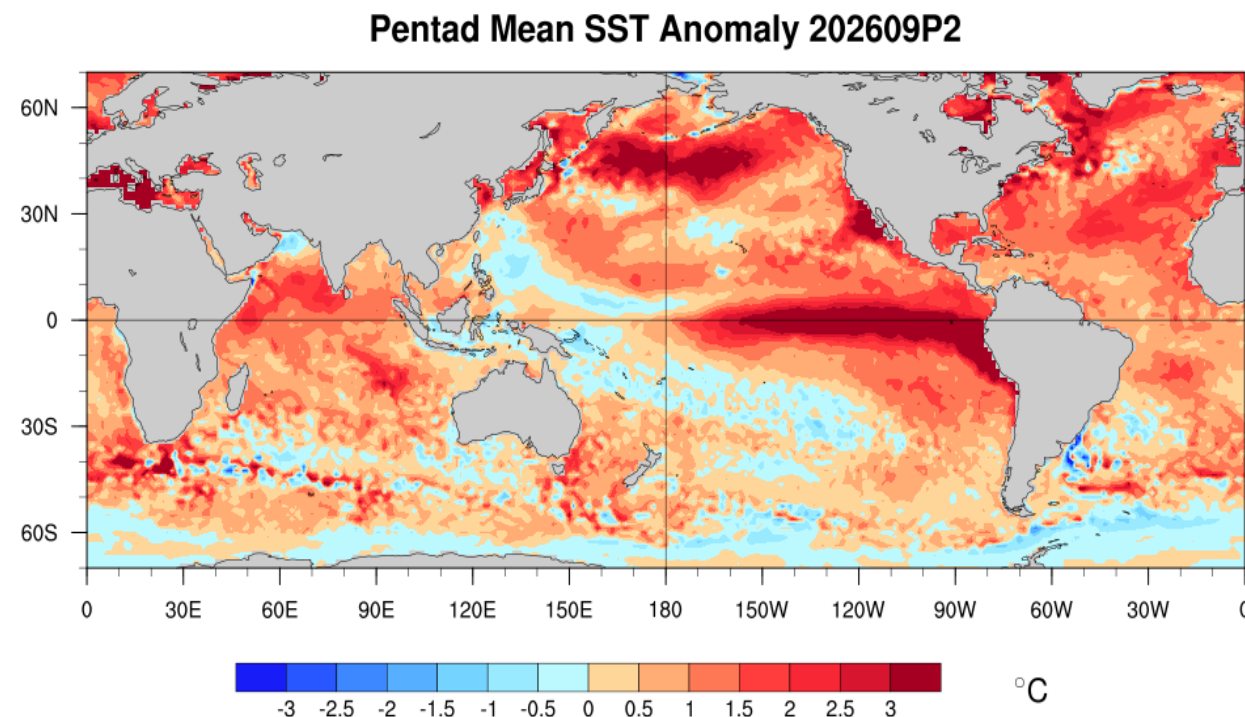


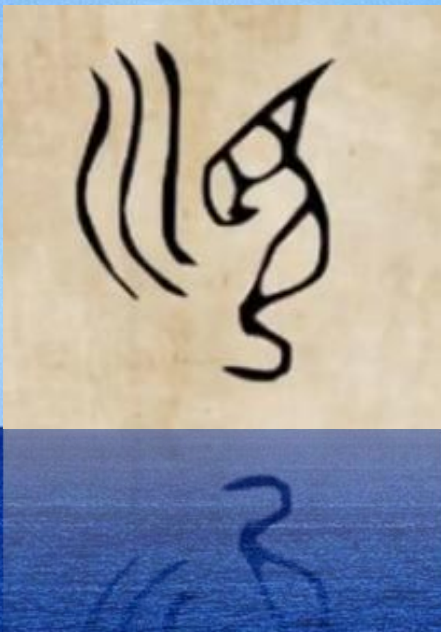
- How to train a model that captures variation at different timescale b/w mult-sphere?
- How to introduce physical constrains given the limited samples?

Nino3.4 predictions (blue curves)



Observed SST Anomaly





The END

THE END





Presentation by team MicroEnsemble



Microsoft





MicroEnsemble

StillLearning, Huracan, MicroDuet



Jonathan Weyn

Microsoft



Hannah Guan

Harvard University



Soukayna Mouatadid

University of Toronto



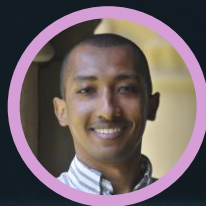
Paulo Orenstein

Instituto de Matemática Pura e Aplicada



Judah Cohen

MIT



Lester Mackey

Microsoft Research



Alex Lu

Microsoft Research



Genevieve Flaspohler

Rhiza Research



Zekun Ni

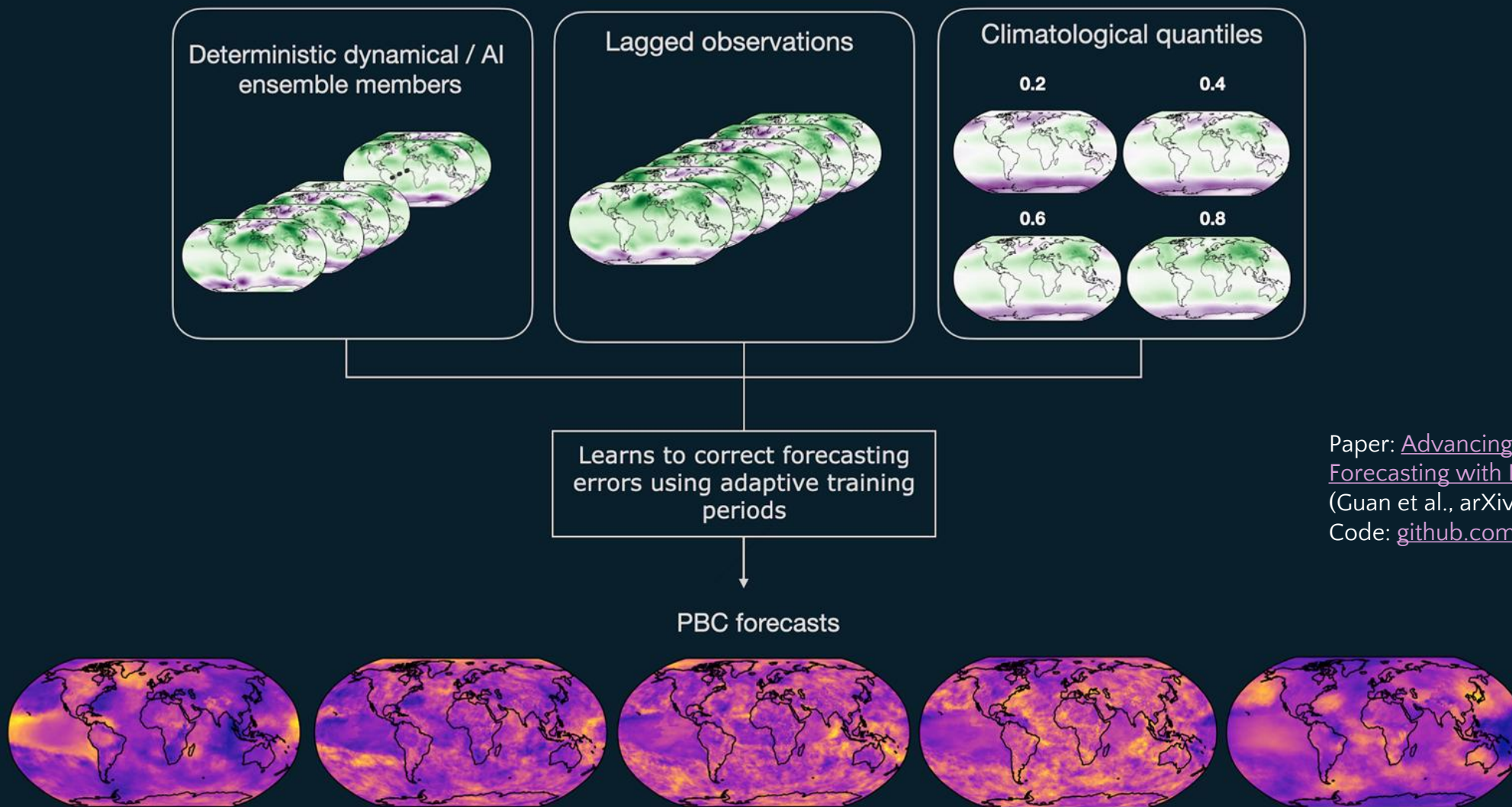
Microsoft



Haiyu Dong

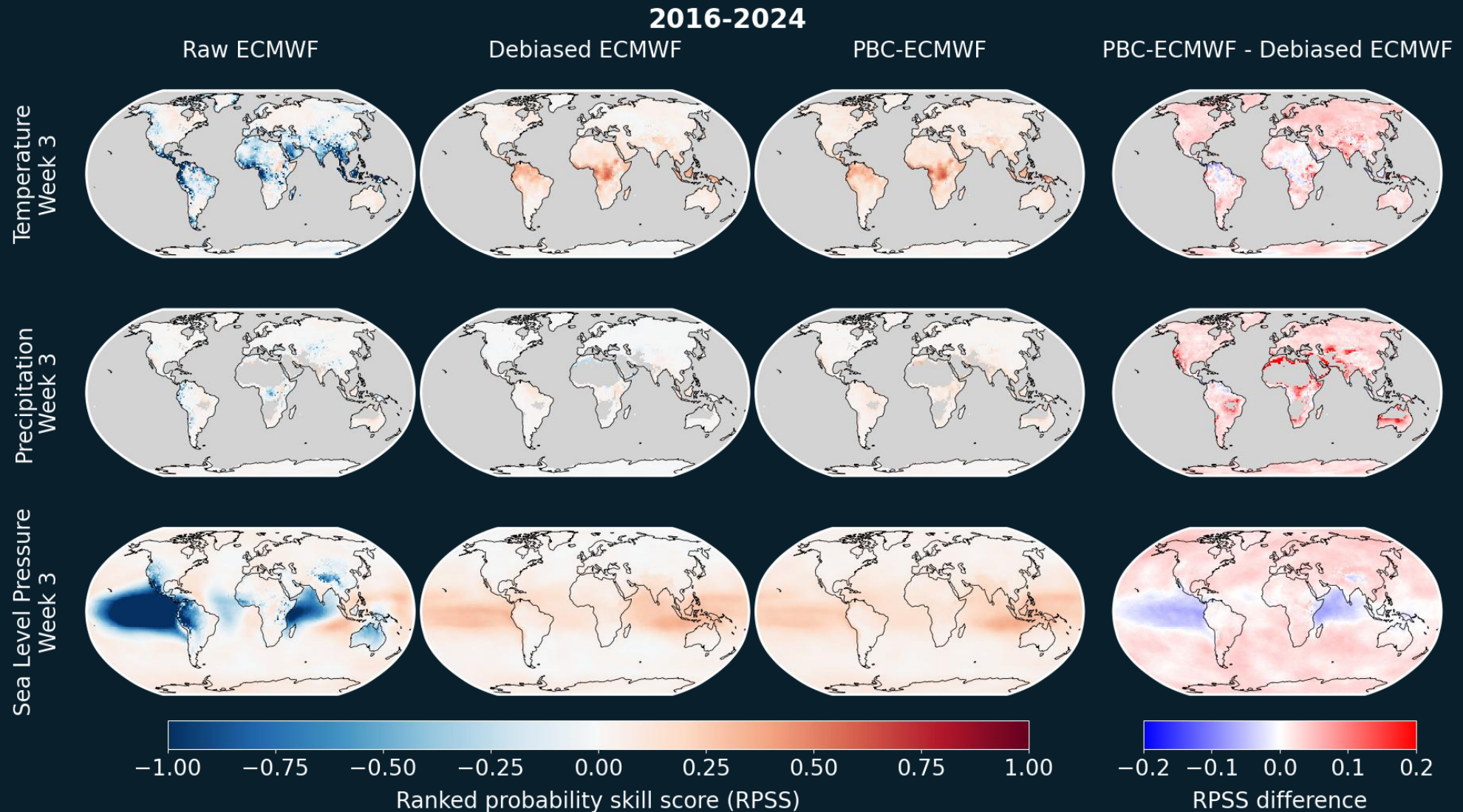
Microsoft

StillLearning = Probabilistic bias correction (PBC)

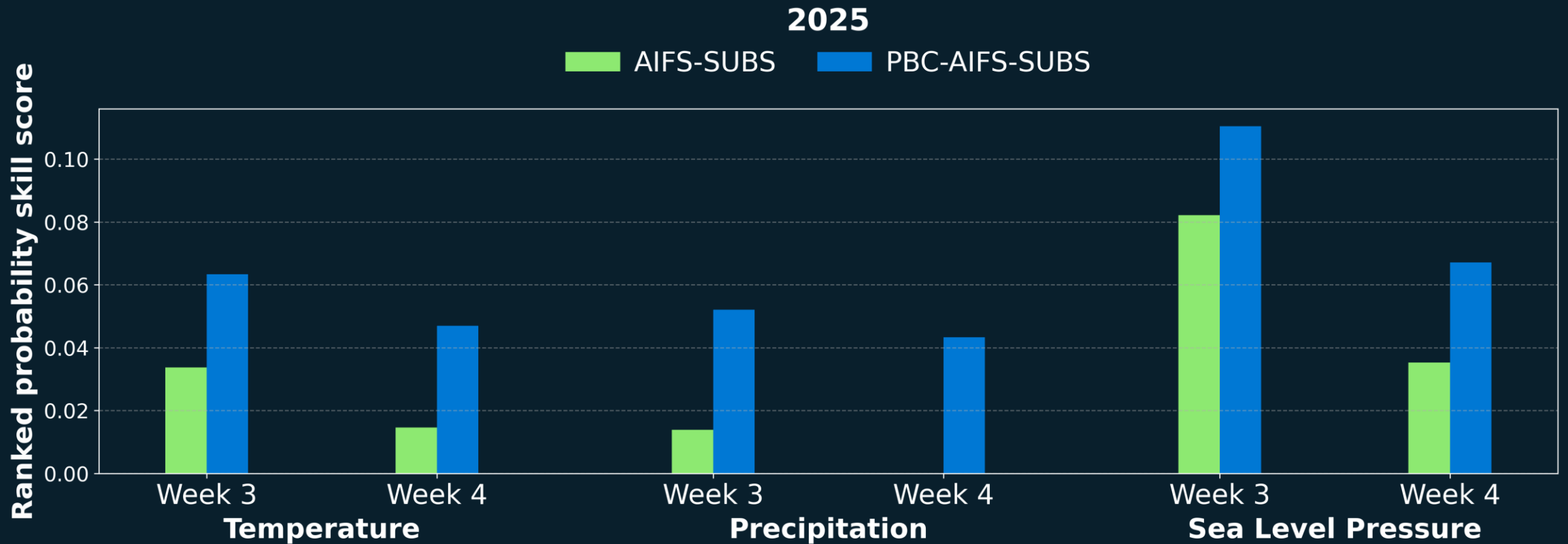


Paper: [Advancing Subseasonal Forecasting with Machine Learning](#)
(Guan et al., arXiv, 2026)
Code: github.com/mouatadid/pbc

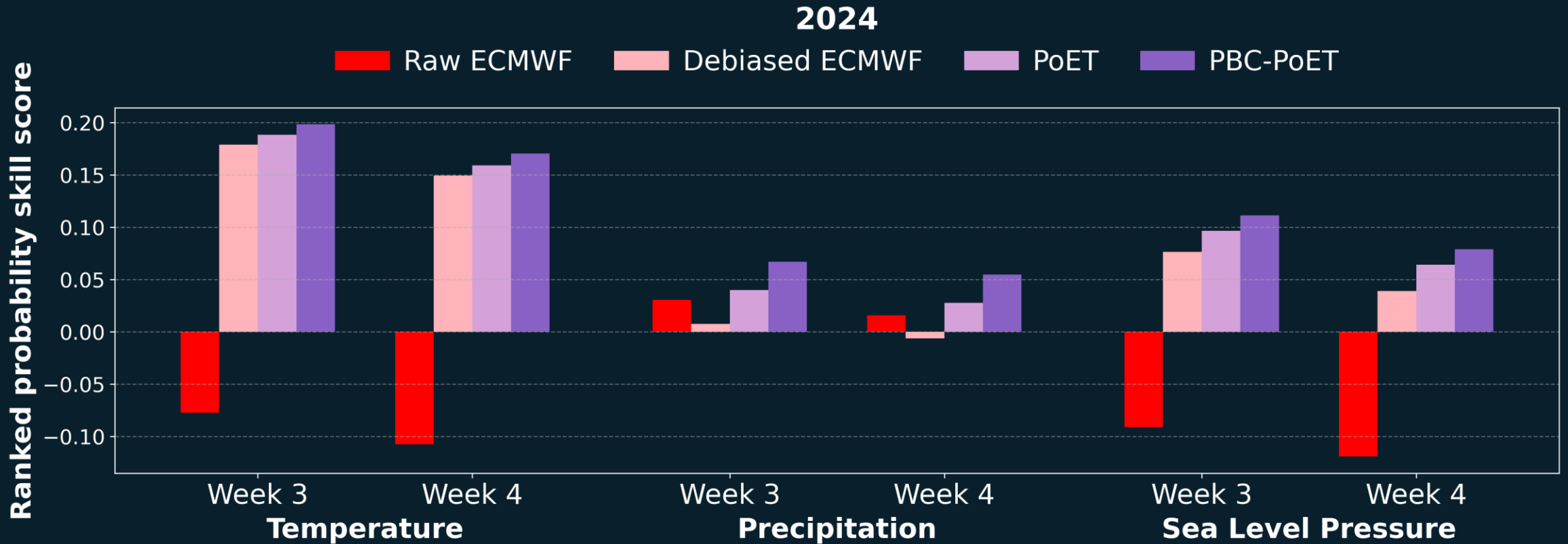
PBC broadly boosts the skill of raw ECMWF forecasts more effectively than operational debiasing



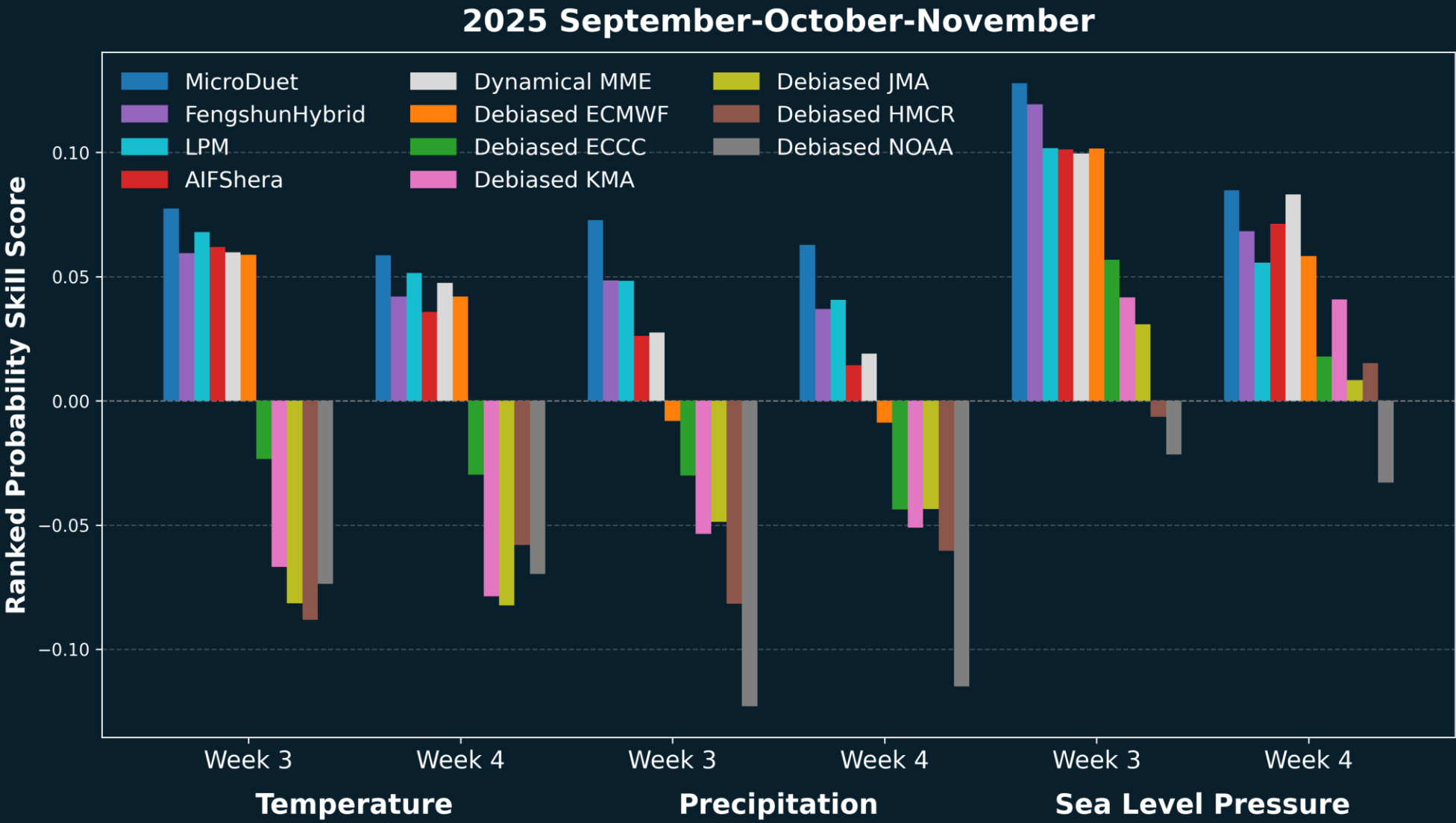
PBC enhances the subseasonal skill of ECMWF's AI Forecasting System (AIFS-SUBS)



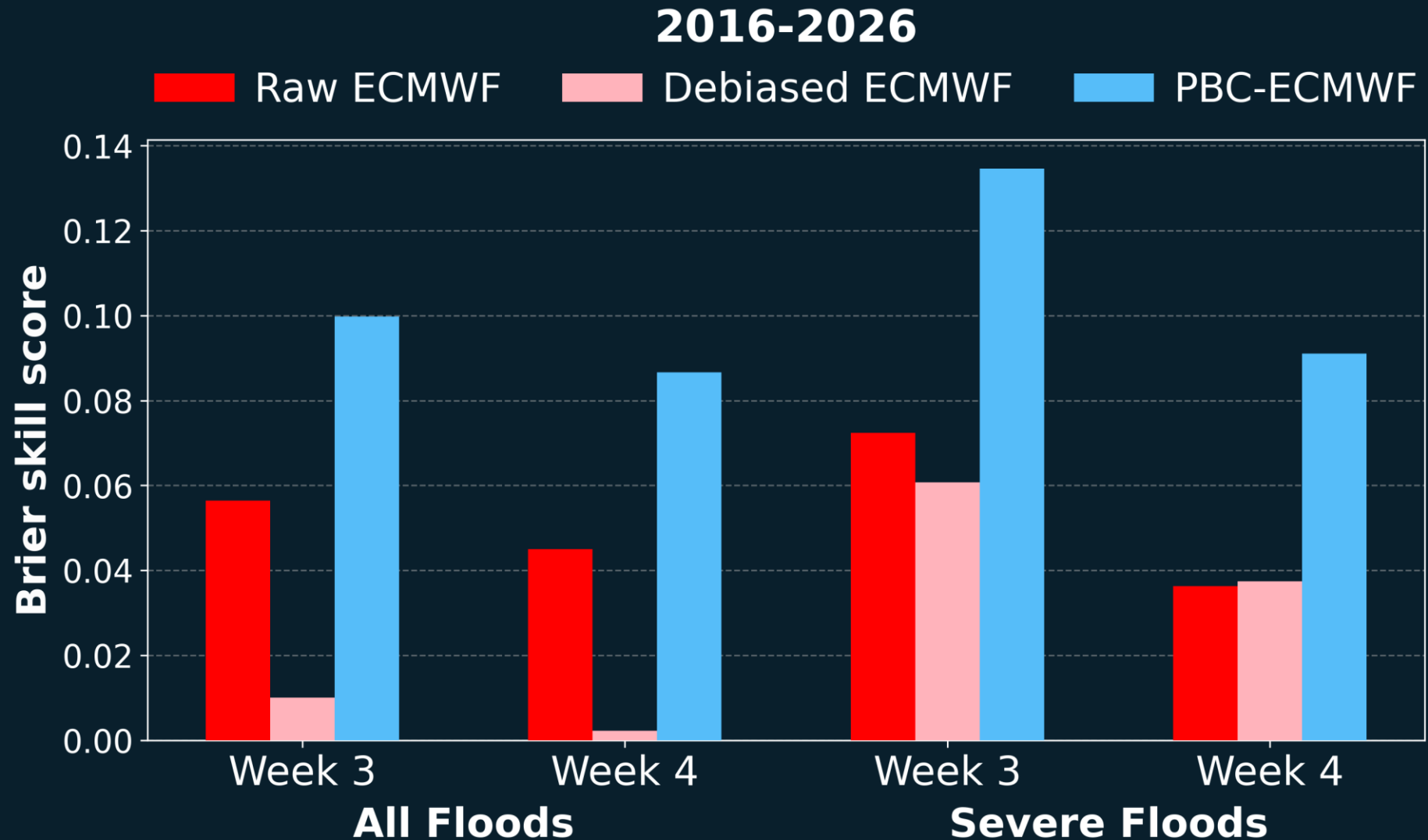
PBC further improves PoET (Huracan), a state-of-the-art AI post-processor



MicroDuet = PBC-ECMWF + PBC-Poet

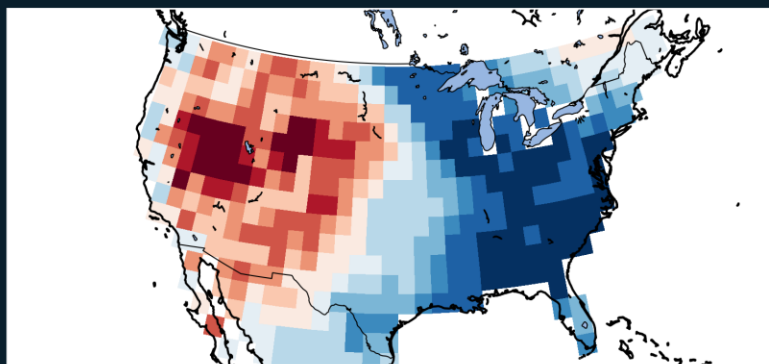


PBC displays higher skill in predicting floods



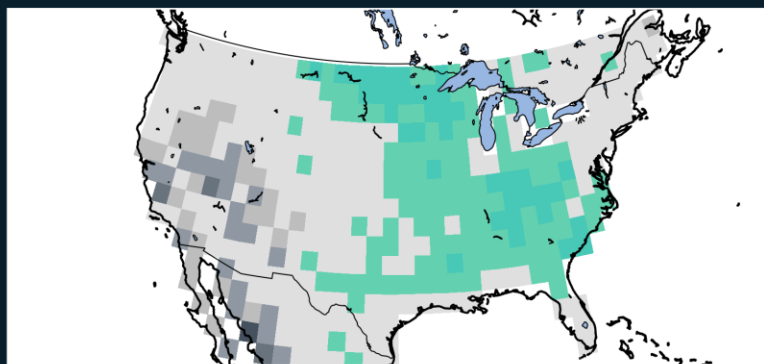
MicroDuet anticipated an extreme cold event missed by debiased ECMWF

Observed (ERA5)



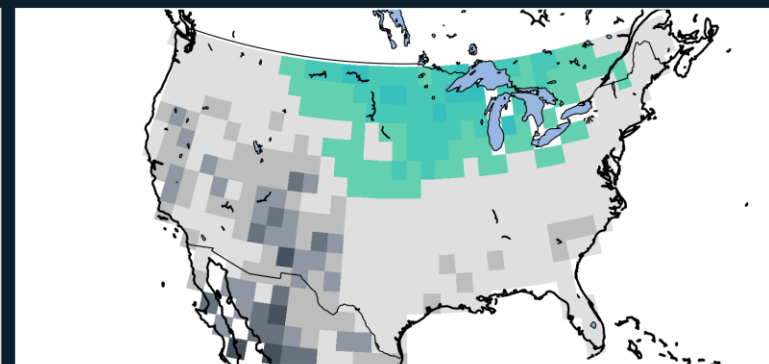
Temperature anomaly (°C): Dec 15-21, 2025

MicroDuet



Predicted probability of lowest temperature quintile: Dec 15-21, 2025

Debiased ECMWF



PBC models can offer agricultural support in African El Niño response

Following strong local benchmarking performance, the **Kenya Meteorological Services Authority** and the Climate Hazards Center (CHC) will pilot the PBC model during the record-breaking 2026 El Niño rainy season. Skillful subseasonal rainfall forecasts can drive decision-support for farmers in taking critical actions, including planting, fertilizing, and harvesting.



July 22nd 2026 field assessment photos from Western Kenya show the difference planting timing can make in maize maturation in two neighboring plots that planted at different times. Photos taken by Chris Shitote, CHC, just north of Kitale, in northern Trans Nzoia County.



Presentation by team CLINT



the
climate
data factory



cmcc



the
climate
data factory

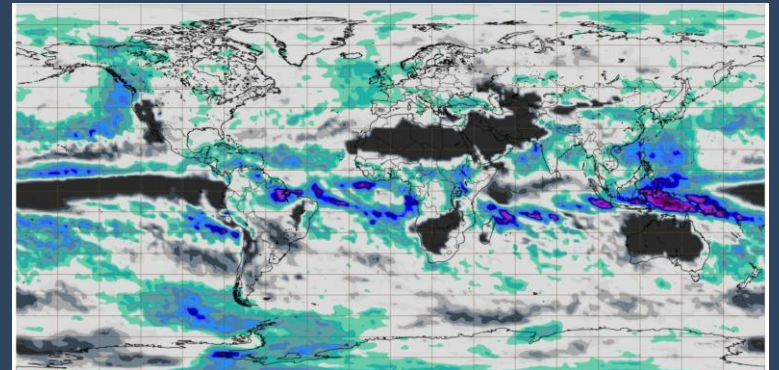
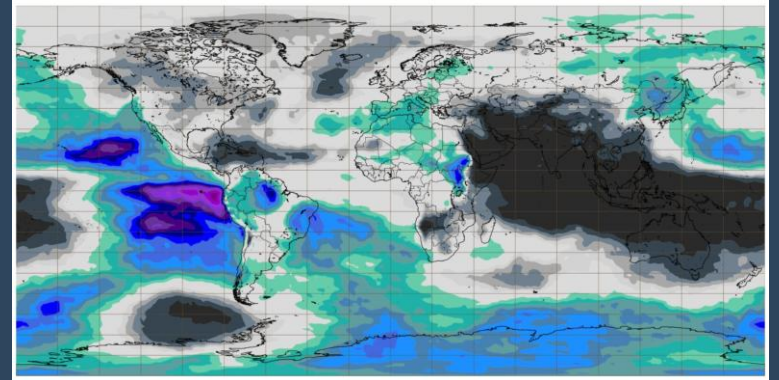
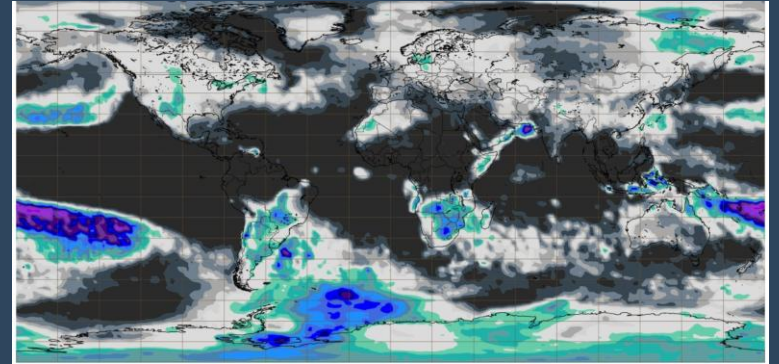


cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

AI Weather Quest: How Far Does Forecast Bias-adjustment Take You?

Harilaos Loukos, Thomas Noël, Fabio Merizzi,
Guilhem Vignal, Ronan McAdam.

WMO AI Webinars: AI Weather Quest
September 24th 2026



CLINT Team

The Climate Data Factory

Private provider of climate data & services

- Harilaos Loukos (founder)
- Thomas Noël (data engineer)
- Fabio Merizzi (ml engineer)
- Guilhem Vignal (ml engineer)

CMCC Foundation

(Euro-Mediterranean Center on Climate Change)

Research organisation

- Ronan McAdam (climate scientist)



What we do

Climate Data Products & Services

Climate projections

- Future hazard & weather variables

Climate forecasts

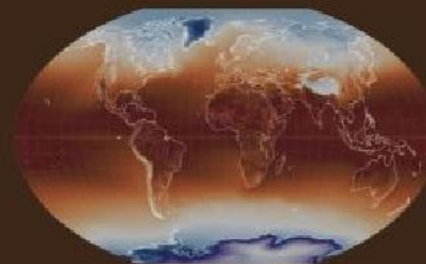
- 4 weeks and 6 months outlooks

Climate Storylines

- Observed weather under past and future climates

Climate service

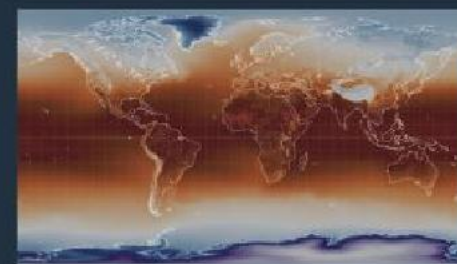
- On demand data processing



Climate projections

Our high-resolution and multi-model scenarios of climate hazards extend up to 2100. Our climate projections data are designed for direct use in climate change impact studies and physical risk assessments. They have a high spatial resolution, encompass multiple Earth System Models and emission scenarios, and cover a variety of climate hazards and variables. It is a coherent collection of climate information for risk analysis under future climate conditions.

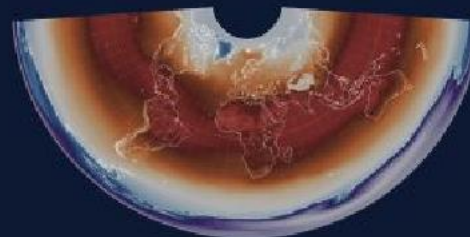
[explore projections >](#)



Climate forecasts

Our high-resolution multi-system sub-seasonal and seasonal forecasts provide valuable insights into climate fluctuations from four weeks up to six months ahead. These forecasts are based on extensive data from leading operational centers and go through advanced processing to ensure accuracy and reliability. Our climate forecasts are designed for direct use in decision-support applications for different business and economic sectors.

[explore forecasts >](#)



Climate service

Our climate service is dedicated to the custom management and processing of climate data according to your unique specifications. If you require specific climate projections or forecast data, or if you need to manage or process public or proprietary data, we can handle those for you with the same efficiency and quality standards applied to our own data. This to ensure reliable and actionable information to enhance the effectiveness of your climate-related projects.

[explore services >](#)



Climate storylines

Our Storyline simulations are reconstructions of observed weather under past and future climates. This approach enables professionals to explore how climate change is reshaping the magnitude of extreme events in an increasingly volatile climate. The ERA5 reanalysis is used as the proxy of observed weather and the storylines are contextualized version of ERA5 under preindustrial and future conditions at several global warming levels (+1.5°C, +2°C, +3°C, +4°C).

COMING SOON

Motivation to join the AI Weather Quest

- **Benchmark our bias-correction method of S2S forecast used in our services**
 - We focus on correcting daily data to have forecasts that can be used for impact models and daily metrics calculations
- **Test a data driven framework for S2S forecast developed during the EU funded CLINT (Climate Intelligence) collaborative research project**
climateintelligence.eu

AI Weather Quest CLINT models

- **CLINT-MF (Multi forecast)**

- Daily bias-correction and forecast pooling of ECMWF forecast (3 days/303 members)

=> Our top performer: Focus of this presentation (skill and where it comes from)

- **CLINT-DD (Data driven)**

- Deterministic data driven ML approach guiding ECMWF probabilistic forecast (no pooling)

=> Only the last 2 forecast weeks of JJA: Preliminary results (does it beat CLINT-MF?)

- **CLINT SE (Selection)**

- CLINT-MF with member selection (benchmark CLINT-DD guidance)

=> not performing: not shown

Methods

Bias correction, pooling, and the data-driven component.

Bias correction

Vrac, Loukos, Noël, Defrance 2025

Cumulative Distribution Function-transform (Quantile mapping method)

- Distribution-based bias correction
- Non parametric (empirical) and univariate
- Designed for non-stationary climates
- Applicable to both bias correction and statistical downscaling
- Applied by forecast week with all ensemble members pooled together

Article

Should We Use Quantile-Mapping-Based Methods in a Climate Change Context? A “Perfect Model” Experiment

Mathieu Vrac ^{1,*}, Harilaos Loukos ^{2,*}, Thomas Noël ^{2,*} and Dimitri Defrance ²

¹ Laboratoire des Sciences du Climat et de l’Environnement (LSCE-IPSL), CEA/CNRS/UVSQ, Université Paris-Saclay, Centre d’Études de Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France
² The Climate Data Factory, 75010 Paris, France; harilaos@theclimatedatafactory.com (H.L.); dimitri@theclimatedatafactory.com (D.D.)
* Correspondence: mathieu.vrac@lsce.ipsl.fr (M.V.); thomas@theclimatedatafactory.com (T.N.)

Abstract

This study assesses the use of Quantile-Mapping methods for bias correction and downscaling in climate change studies. A “Perfect Model Experiment” is conducted using high-resolution climate simulations as pseudo-references and coarser versions as biased data. The focus is on European daily temperature and precipitation under the RCP 8.5 scenario. Six methods are tested: an empirical Quantile-Mapping approach, the “Cumulative Distribution Function—transform” (CDF-t) method, and four CDF-t variants with different parameters. Their performance is evaluated based on univariate and multivariate properties over the calibration period (1981–2010) and a future period (2071–2100). The results show that while Quantile Mapping and CDF-t perform similarly during calibration, significant differences arise in future projections. Quantile Mapping exhibits biases in the means, standard deviations, and extremes, failing to capture the climate change signal. CDF-t and its variants show smaller biases, with one variant proving particularly robust. The choice of discretization parameter in CDF-t is crucial, as the low number of bins increases the biases. This study concludes that Quantile Mapping is not appropriate for adjustments in a climate change context, whereas CDF-t, especially a variant that stabilizes extremes, offers a more reliable alternative.

Keywords: bias correction; statistical downscaling; climate models; climate projections



Academic Editor: Yong Wang

Received: 11 April 2025

Revised: 20 June 2025

Accepted: 27 June 2025

Published: 1 July 2025

Citation: Vrac, M.; Loukos, H.; Noël, T.; Defrance, D. Should We Use Quantile-Mapping-Based Methods in a Climate Change Context? A “Perfect Model” Experiment. *Climate* **2025**, *13*, 137. <https://doi.org/10.3390/cli13070137>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last few decades, scientists have attempted to model, quantify, and understand the potential impacts of climate change. To investigate these impacts, various hydrological, agricultural, and, more generally, environmental “impact models” have been developed (e.g., [1,2]). Although such models can be different regarding their goals, underlying assumptions, and functioning, they all rely on the use of climate simulations of climate variables, such as temperature and precipitation, as inputs to simulate impacts (e.g., in terms of river flows, crop yields, etc.). Therefore, the realism and robustness of climate simulations are essential, both over historical and future periods and scenarios of interest, to obtain confidence in the modeled impacts (e.g., [3]).

However, because global climate simulations usually have a relatively low spatial resolution, a major drawback is that they are not designed to provide local information. Indeed, Global Climate Models (GCMs) have a mean resolution of approximately 150 × 150 km, whereas most impact models operate at a much more local scale, from a watershed or

Forecast pooling

Vitart, Takaya 2021

Lagged ensembles

- Combines forecasts initialized on previous days to increase ensemble size
- Tested with ECMWF S2S forecasts
- Longer lead times benefit most: 2 days for week 2, to 4 days for week 4.
- Improved probabilistic skill: if the daily ensemble has more than 20 members

Our experiments

- Never tested on bias corrected ensembles
- Use 3 and 4 days (303/404 members)

Received: 8 March 2021 | Revised: 8 July 2021 | Accepted: 9 July 2021 | Published on: 26 July 2021

DOI: 10.1002/qj.4125

Quarterly Journal of the
Royal Meteorological Society



RESEARCH ARTICLE

Lagged ensembles in sub-seasonal predictions

Frédéric Vitart¹ | Yuhei Takaya²

¹European Centre for Medium-Range
Weather Forecasts, Reading, UK

²Meteorological Research Institute, Japan
Meteorological Agency, Tokyo, Japan

Correspondence

F. Vitart, European Centre for
Medium-Range Weather Forecasts,
Shinfield Park, Reading, RG2 9AX, UK.
Email: frederic.vitart@ecmwf.int

Abstract

The benefits of using lagged ensembles for S2S predictions have been evaluated using the ECMWF sub-seasonal to seasonal (S2S) forecasts. The ECMWF S2S forecasts are currently produced twice a week from a 51-member ensemble. An experiment where the ECMWF forecasts are produced daily instead of twice weekly has been performed to identify the minimum lagged-ensemble configuration which will produce probabilistic forecasts at least as skilful as the current configuration. Results suggest that the extended-range prediction of the Madden–Julian Oscillation (MJO) is not deteriorated with the daily lagged ensemble approach, but there is a slight degradation during the first 10 days. In the northern Extratropics, the probabilistic skill scores are improved after week 2 when using lagged ensembles, but only if the daily ensemble size exceeds 20 members. The lagged ensemble approach is more beneficial in the Tropics than in the northern Extratropics, showing even benefits in forecast week 1 over the Tropics, despite the 24 hr difference between two consecutive forecasts. As expected, the daily lagged ensemble method is more beneficial for longer lead times and the optimal lagged ensemble window increases from 2 days for week 2 to 4 days for week 4. An idealized numerical simulation, based on a Monte Carlo approach, confirms the benefits of lagged ensembles. The results obtained are generally consistent with the ECMWF forecast experiment, suggesting that such a simulation could be used as a cheap alternative to address the benefit of lagged versus burst ensembles.

KEYWORDS

1. tools and methods, ensembles, forecasting (methods), sub-seasonal

1 | INTRODUCTION

The paradigm shift from a single weather forecast to an ensemble of forecasts started in 1992, with the implementation of global ensemble prediction systems in several operational centres (e.g. Tracton and Kalnay, 1993; Buizza and Palmer, 1995). Unlike single weather forecasts, ensemble forecasts provide an indication of the range of possible future states of the atmosphere. The multiple simulations are conducted to account for the two usual sources of uncertainty: the errors introduced by the

use of imperfect initial conditions and errors introduced because of imperfections in the model formulation, such as the approximate mathematical methods to solve the equations. Ideally, the verified future atmospheric state should fall within the predicted ensemble spread, and the amount of spread should be related to the forecast uncertainty. This approach is used to produce probabilistic forecasts.

There are a number of ways to generate initial conditions perturbations. The simplest way consists of starting the model integrations at slightly different times (e.g. a

Data driven framework

McAdam et al. 2025

Feature selection for data driven forecast

- Purely data-driven seasonal forecasting
- Optimally selects from a large pool of potential predictors
- Interpretable feature selection
- Pre-training on a paleoclimate simulation and fine tuning on ERA5
- Deterministic

Difficulties encountered for AI Weather Quest

- Needs training by forecast week
- Resource demanding



<https://doi.org/10.1038/s43247-025-02863-4>

Feature selection for data-driven seasonal forecasts of European heatwaves

Check for updates

Ronan McAdam¹✉, Jorge Pérez-Aracil^{2,3}, Antonello Squintu⁴, César Peláez-Rodríguez², Felicitas Hansen⁴, Verónica Torralba⁵, Harilaos Loukos⁶, Eduardo Zorita⁴, Matteo Giuliani^{1,7}, Leone Cavicchia⁸, Sancho Salcedo-Sanz² & Enrico Scoccimarro¹

The early warning of heatwaves using seasonal forecasting systems has the potential to mitigate economic losses and risk to life. Because of the limited reliability and computational expense of dynamical forecasting systems, efforts in recent years have turned to exploiting Machine Learning. Here, an inexpensive approach to forecasting summer heatwaves over Europe is developed, using an optimisation-based feature selection framework to detect a combination of variables, domains and time-lags used to skilfully predict heatwaves. The purely data-driven forecasts are shown to match, and in places outperform, the skill of the state-of-the-art dynamical multi-model products. Moreover, low skill over Scandinavia, a long-term issue common to most dynamical systems, is improved in our data-driven approach. This work also highlights that the greatest contribution to skill comes from predictors at 4–7 weeks time-lag (e.g. mid-March), and identifies predictors which can form the basis for future studies on mechanisms.

Heatwaves (HWs) are prolonged periods of extreme temperature, and lead to a wide range of impacts, including the collapse of agricultural yields¹, drastic increases in energy usage², impacts on human health, and increased mortality^{3,4}. Europe has experienced devastating heatwaves in the past decades, including, but not limited to, deadly events in 2003 and 2010⁵ and more recently in 2022^{6,7}. Climate projections suggest further intensification of HWs in the coming decades⁸, which will likely lead to an increase in deaths attributed to extreme heat unless mitigation measures are implemented⁹. Consequently, the ability to predict extreme summer heat several months in advance provides an opportunity for the agricultural industry and national health services to implement mitigation measures¹⁰.

Seasonal forecasting, the prediction of seasonal climate conditions several months in advance, has the potential to provide society with time and necessary information to take meaningful action prior to potentially damaging climate events. Such information already serves as the foundation for climate services in various sectors, such as the early warning of droughts for agriculture^{11,12} and snow cover for tourism¹³. The generation and maintenance of seasonal forecasts, however, is an enormous computational undertaking, with many centres around the world producing dozens of ensemble members of climate models, each of which couples several components of the climate system, at horizontal resolutions typically less than 1 geographical degree. Moreover, predicting heatwaves beyond the

deterministic limit (roughly 10–15 days) is challenging^{14–17}. The state-of-the-art operational seasonal forecasting systems from the Copernicus Climate Change Service (C3S) have demonstrated reliable forecast skill over large parts of Europe when predicting, up to 3 months in advance, seasonal heatwave indices^{18,19}. However, skill gaps remain; for example, over northern Europe, which is less influenced by more predictable mid-latitude variability²⁰.

The increasing use of Machine Learning (ML) for weather and climate science applications provides a means of reducing the resources required to make accurate weather forecasts without compromising on skill, as demonstrated in the field of weather forecasting²¹. Dynamical forecasting systems, named after their ability to solve dynamical and thermodynamic equations numerically, are now matched in skill or even outperformed by purely data-driven approaches using a range of machine or deep learning architectures^{22–26}. These data-driven approaches leverage techniques designed to identify relationships between multiple variables from large datasets of observations or model simulations. The current frameworks used are being adjusted for subseasonal²⁷ and seasonal timescales²⁸, and any data-driven approach to seasonal forecasting requires considerably more data than is available in the observational records²⁹.

Prior to the widespread use of ML, statistical seasonal forecasting showed that the input of known predictors, such as soil moisture for

¹CMCC Foundation - Euro-Mediterranean Center on Climate Change, Bologna, Italy. ²Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain. ³Programa de doctorado en Computación Avanzada, Energía y Plasmas, Universidad de Córdoba, Córdoba, Spain. ⁴Helmholtz-Zentrum Hereon, Geesthacht, Germany. ⁵Barcelona Supercomputing Center (BSC), Barcelona, Spain. ⁶The Climate Data Factory (TCDF), Paris, France. ⁷Department of Electronics, Information and Bioengineering, Politecnico di Milano (POLIMI), Milan, Italy. ✉e-mail: ronan.mcadam@cmcc.it

New ML approach

Merizi, Loukos, Noël submitted

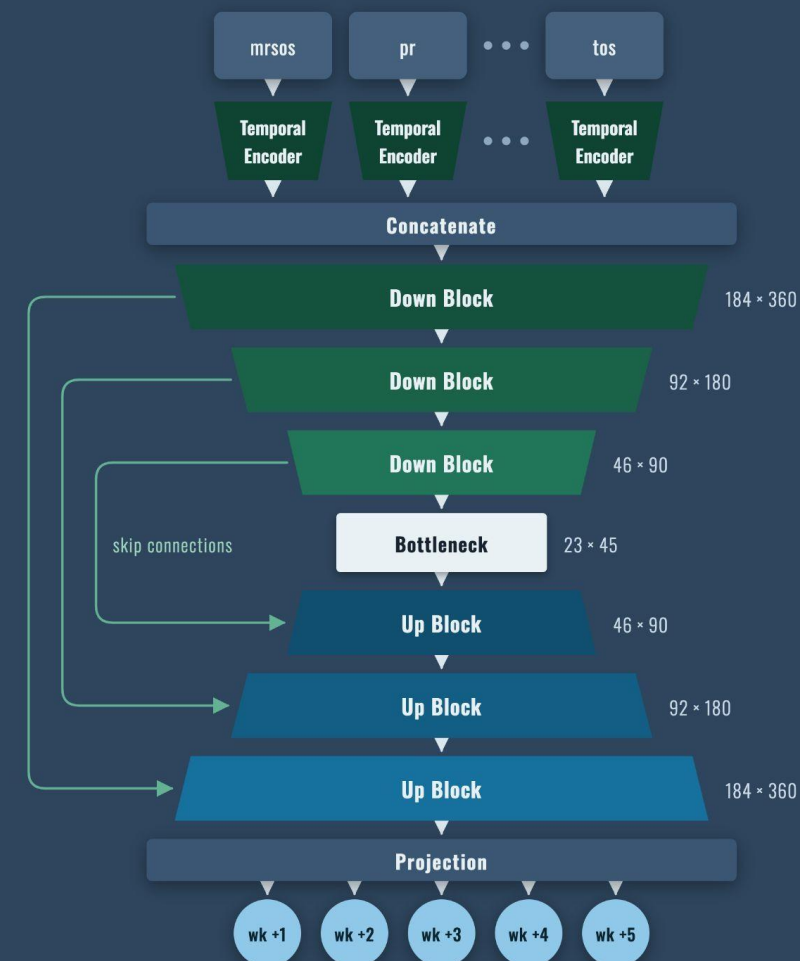
A multi-encoder U-Net trained twice

- 10 predictors × 12 weekly means
- One temporal encoder for each
- Residual U-Net, periodic in longitude
- 5 lead weeks in one forward pass, no rollout.
- Pre-trained on past2k (1850 years), fine tuned on ERA5 (85 years) with a weight anchor
- Guides the ECMWF ensemble (weighted distribution shift)

Advantage for operational use

- One training of all forecast weeks, fast inference.

10 predictors × 12 weekly means = 120 input channels
mrsos · pr · psl · rlut · tas · tasmax · tasmin · zg500 · siconc · tos



5 forecast weeks × 3 variables (tas, zg500, pr) produced in a single forward pass — not autoregressive



Hybrid data driven framework

McAdam et al. 2026

Blending deterministic data driven and ensemble dynamical forecast

- Hybrid data-driven/dynamical approach
- 120-member multi-model ensemble
- Member selection by similarity
- Two methods: grid point or over a domain
- Combines complementary skill patterns
- Improved forecast skill about 10%
- Improved physical representation (summer z500 atmospheric circulation)

S2S2D Conference 2026

Blending data-driven and dynamical seasonal forecasts
with ensemble sub-sampling:
10% increase in skill of heatwave predictions over Europe

S2S2D
2026

Selecting dynamical members based on similarity to data-driven forecasts

Dynamical and data-driven (McAdam et al., 2025) seasonal forecasts of heatwaves Europe demonstrate skill in different places - can selection bring out the best in both?

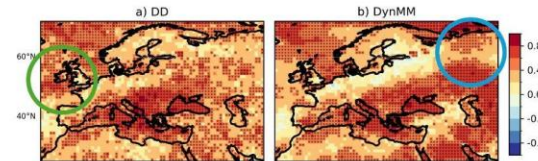


Figure 1: Correlation skill of number of summer heatwave days between seasonal forecasts (DD and DynMM) and ERA5 over 1993-2016.

McAdam, R., Pérez-Aracil, J., Squintu, A. et al.
Feature selection for data-driven seasonal forecasts of European heatwaves. Commun Earth Environ 6, 842 (2025).

Criteria: select the n% members with least difference to data-driven forecasts of heatwave index

Ensemble Selection Techniques

DynMM	120-member ensemble from 4 burst-mode dynamical systems (ECMWF5.1, CMCC-SPS3.5, DWD2.1, MF9)
DD	Deterministic Data-Driven forecast from McAdam et al., 2025
ESP-DD	Ensemble Selection by grid Point difference
ESD-DD	Ensemble Selection by average Domain difference

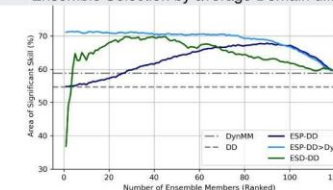


Figure 2: Area of significant skill over Europe as a function of ensemble subsample size.

Increase in heatwave skill

Subsample is informed by both data-driven and dynamical skill, emerging regions of skill (British Isles) and widening of existing zones (western-to-central Europe).

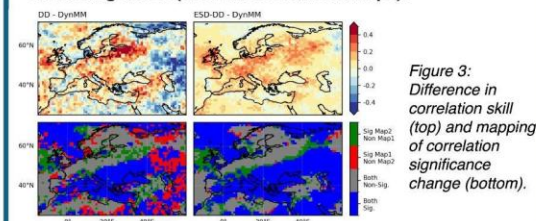


Figure 3: Difference in correlation skill (top) and mapping of correlation significance change (bottom).

Improved representation of heatwave drivers

Subsampling based on heatwave index also leads to increase in skill of atmospheric circulation.

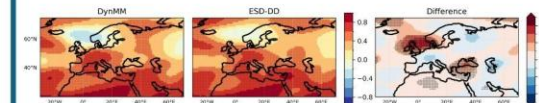


Figure 4: Correlation of summer z500 anomalies in DynMM and ESD-DD with ERA5. Stippling: significant difference w.r.t random sampling.

Ronan McAdam¹, Luca Famooss Paolini¹, Marianna Benassi¹,
Harilaos Loukos², Enrico Scoccimarro¹
¹CMCC Foundation – Euro-Mediterranean Centre on Climate Change,
Italy
²The Climate Data Factory, France

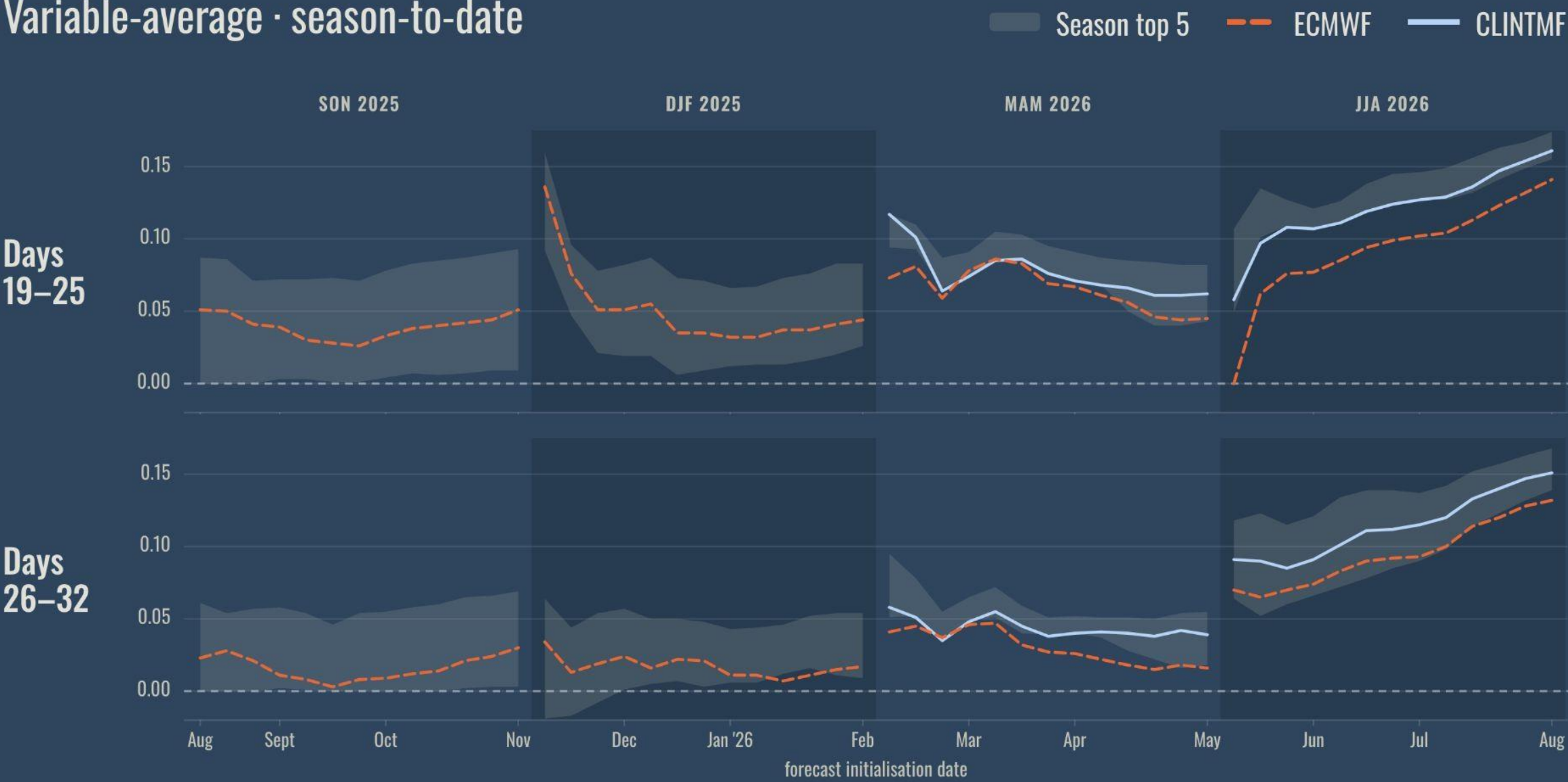


Results

Season-to-date skill, focus on CLINT-MF

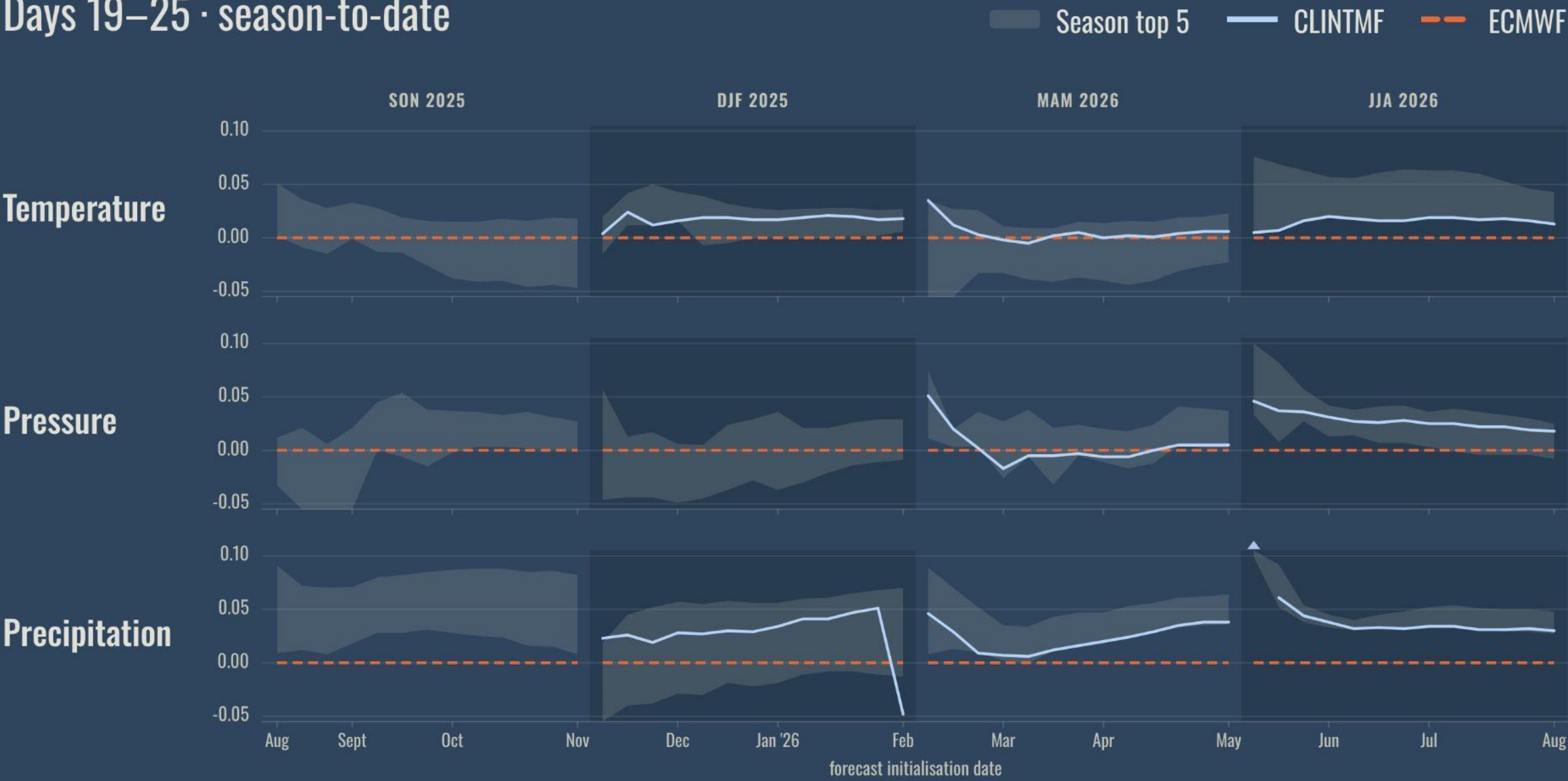
Season-to-date skill (RPSS)

Variable-average · season-to-date



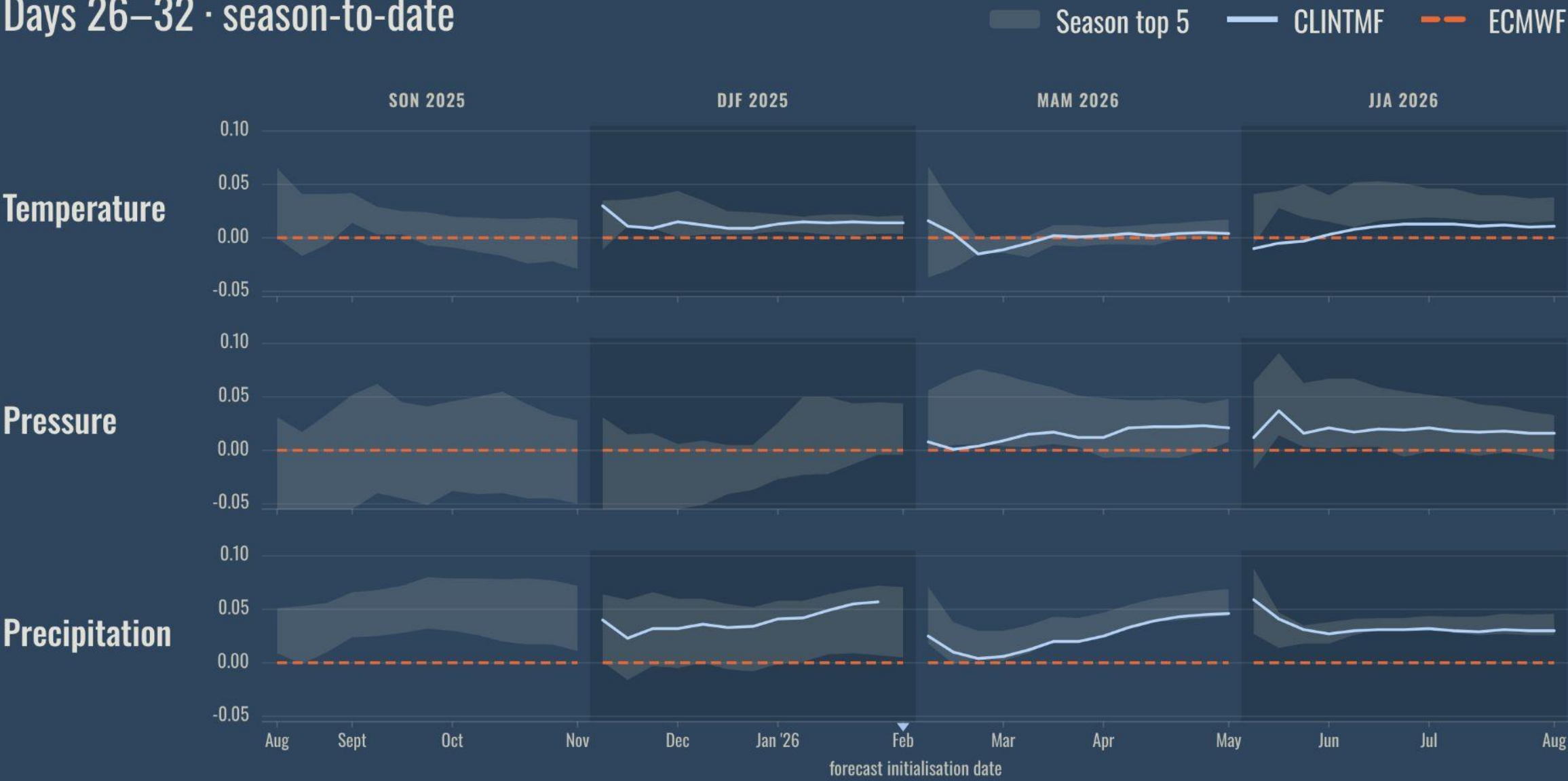
Season-to-date skill (RPSS): gain over ECMWF

Days 19–25 · season-to-date



Season-to-date skill (RPSS): gain over ECMWF

Days 26–32 · season-to-date



Where CLINTMF's added skill over ECMWF comes from (RPSS)

Mean of weekly gains

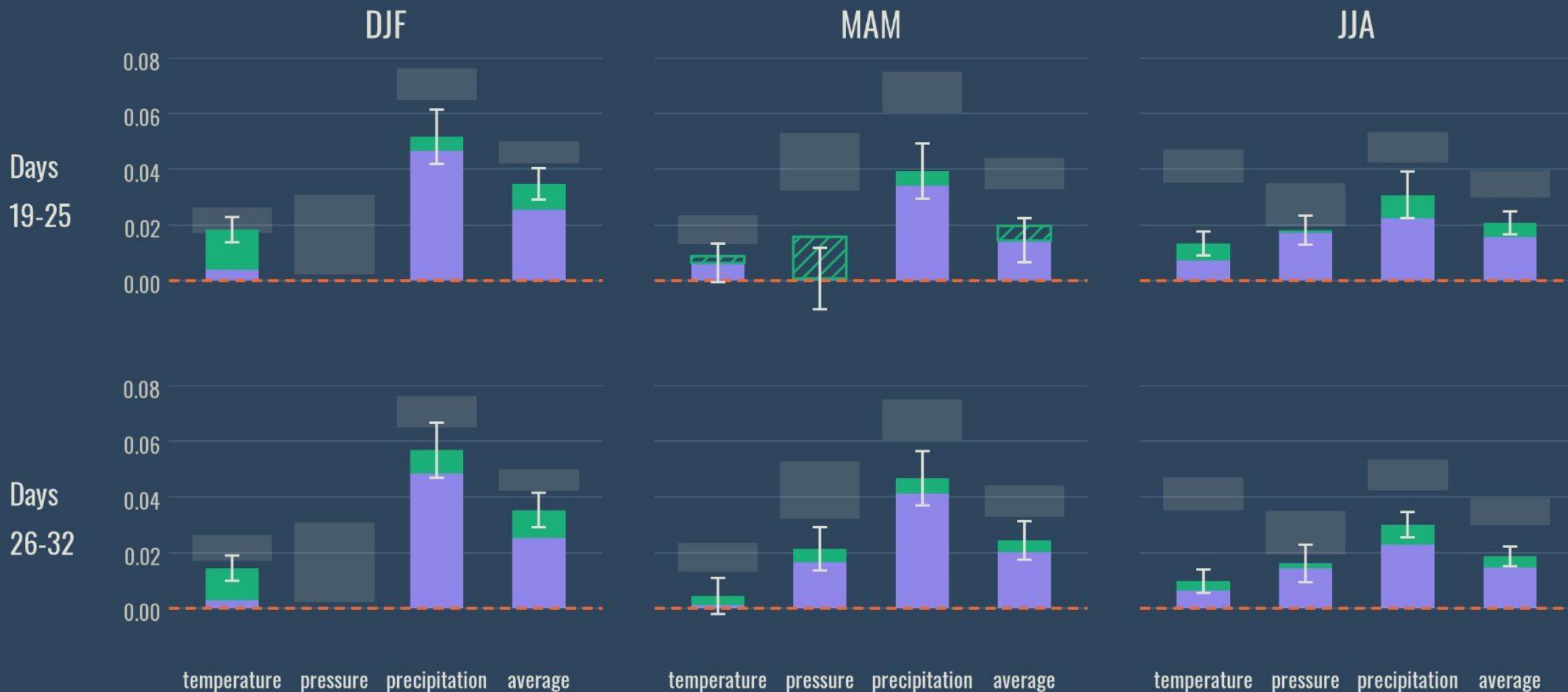
Season leader

ECMWF

Bias correction

Pooling

Pooling, negative

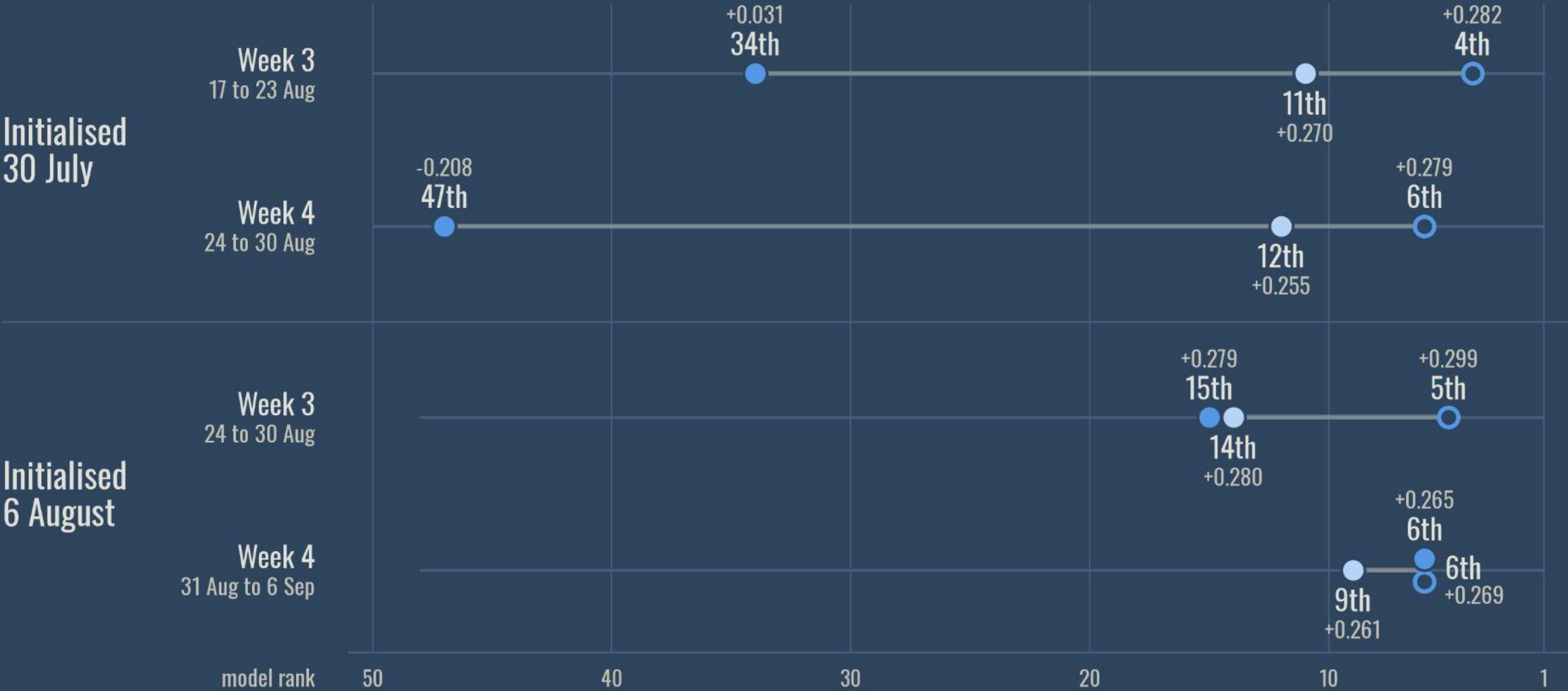


CLINT-DD
(last 2 weeks of JJA)

CLINT-DD on the weekly leaderboard: rank and skill (RPSS)

Temperature

● CLINT-DD ● CLINT-MF ○ CLINT-DD corrected



Next steps for the next AI Weather Quest season

- **CLINT-MF (Multi forecast)**
 - Use to the right ERA5 climatology (6-hourly as AI Weather Quest instead of hourly)
 - Addapt pooling by variable?
- **CLINT-DD (Data driven/hybrid)**
 - Retrain/optimize the hybrid data driven model on the right ERA5 climatology
 - Extend to pressure and precipitation
- ~~CLINT SE (Selection)~~ => **CLINT ED (Ensemble Diffusion/probabilistic)**
 - Add ensemble diffusion to CLINT-DD for a purely probabilistic data driven forecast
=> optimise RPSS skill by adjusting the ensemble variance

Controlling ensemble variance in diffusion models

Merizzi, Evangelista, Loukos 2025

Generative ensemble downscaling

- Denoising Diffusion Implicit Model (DDIM) for ERA5 to CERRA downscaling.
- Ensemble generation: multiple realizations for the same input.
- Variance control through diffusion steps
- Variance calibration against a reference ensemble
- Uncertainty representation
- Efficient high-resolution ensembles

Controlling ensemble variance in diffusion models: an application for reanalyses downscaling

Fabio Merizzi · Davide Evangelista · Harilaos Loukos

Received: 31 March 2025 / Accepted: 5 December 2025

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2026

Abstract

In recent years, diffusion models have emerged as powerful tools for generating ensemble members in meteorology. In this work, we demonstrate how a Denoising Diffusion Implicit Model (DDIM) can effectively control ensemble variance by varying the number of diffusion steps. Introducing a theoretical framework, we relate diffusion steps to the variance expressed by the reverse diffusion process. Focusing on reanalysis downscaling, we propose an ensemble diffusion model for the full ERA5-to-CERRA domain, generating variance-calibrated ensemble members for wind speed at full spatial and temporal resolution. Our method aligns global mean variance with a reference ensemble dataset and ensures spatial variance is distributed in accordance with observed meteorological variability. Additionally, we address the lack of ensemble information in the CARRA dataset, showcasing the utility of our approach for efficient, high-resolution ensemble generation.

Keywords Diffusion models · Ensemble generation · Reanalysis downscaling · Calibrating variance

1 Introduction

Neural approaches to downscaling meteorological reanalyses have proven effective, enhancing spatial and temporal resolutions while reducing computational costs compared to traditional numerical models.

The task is typically framed as an image-to-image prediction, transforming a low-resolution input into a high-resolution output. Widely adopted prediction-based models, such as U-Net [1] and ViT [2], operate deterministically and predict the most likely outcome for a given input. While this can produce high-quality predictions, these models lack the ability to quantify uncertainty or capture extreme events. To address this limitation, conditioned generative approaches, such as diffusion models [3], have gained traction in the scientific community. Unlike deterministic models, diffusion models sample outcomes from a learned distribution, producing diverse ensemble members even when conditioned on the same low-resolution input. These ensembles provide valuable insights into uncertainty estimation, the underlying physical system, and extreme events.

In this paper, we apply a Denoising Diffusion Implicit Model (DDIM) [4] to downscale meteorological reanalyses, focusing on wind speed. Specifically, we perform the transformation from a global reanalysis (ERA5 [5]) to a regional reanalysis (CERRA [6]) over the European domain. Building on prior work [7], we develop a diffusion model capable of operating across the full CERRA domain and we explore architectural modifications to



To conclude

- Bias-adjustment and pooling of ECMWF took us into the top five, every season, both lead times
- Closing that gap is now the job of the data-driven forecast

Thanks

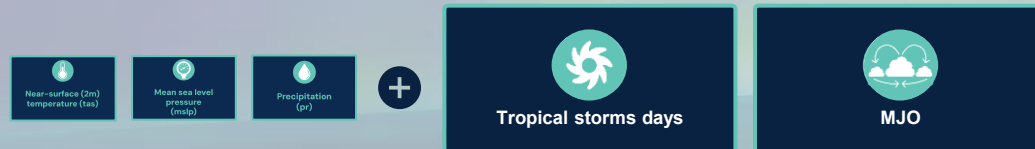
Harilaos Loukos - harilaos@theclimatedatafactory.com



What comes next

Building on the success of the first forecasting year, the Quest continues as a long-term benchmarking and learning infrastructure, with an expanded scientific framework and new improvements.

Participants will be able to produce weekly forecasts for weeks 3–4 for **two additional variables**:



Continental evaluation views and clear **information on model types** will be provided on the leaderboards:

	Team Rank	Model Rank	Team name Model name	Model type
	1		MicroEnsemble	
Region		1	Huracan	Post-processing
Global				
Africa				
Asia				
Australasia				
Europe				
Central & South America				
North America				

Strong participation

Where capacity exists, **participation by African institutions** will be supported through targeted outreach **aligned with ArcX-linked initiatives and WMO networks.**



Broad usability & uptake

Access to submitted data will be granted on day 5:

Week number	Day of week						
	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

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



Thanks!

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